READINGS IN ENGLISH PHONETICS AND PHONOLOGY

Rafael Monroy-Casas,
Inmaculada Arboleda-Girao, eds.

1 monographs

IULMA monografías

UNIVERSITAT DE VALÈNCIA
INSTITUT UNIVERSITARI
DE LLENGÜES MODERNES APLICADES (IULMA)

INSTITUT UNIVERSITARI DE LLENGÜES MODERNES APLICADES DE LA COMUNITAT VALENCIANA (IULMA)

IULMA MONOGRAPHS

IULMA-UV Monograph collection

This collection is issued by the *Instituto Interuniversitario de Lenguas Modernas Aplicadas* (IULMA), an association which promotes research and disseminates publications dealing with key areas of applied linguistics. We publish leading empirical research linked to theoretical discussions on the following topics:

Translation and contrastive studies
Genres of specialised languages
The discourse of science and the professions
Pragmatic analysis of cybergenres
Corpus linguistics
Computational linguistics
Lexicology, lexicography and terminology
Information and communication technologies (ICT)
Critical discourse analysis
Discourse in the media

Proposals should be sent by email to the General Editor or to the Executive Secretary:

Dr. Miguel Fuster Márquez (miguel.fuster@uv.es)
Dr. Jesús Fernández Domínguez (jesus.fernandez-dominguez@uv.es)

Submissions are accepted in the following languages: Spanish, Catalan, English, French, German, and Italian.

The monographs in this collection undergo an external blind-review evaluation by international specialists.

Monographs are published biannually. However, the scientific board reserves the right to release additional issues if there are sufficient submissions of outstanding scientific quality.

Prospective contributors to IULMA monographs should go to the following address: (http://www.iulma.es/noticia.asp?idnoticia=2306).

READINGS IN ENGLISH PHONETICS AND PHONOLOGY

Rafael Monroy-Casas

Inmaculada Arboleda-Guirao (eds.)

UNIVERSITAT DE VALÈNCIA

INSTITUT UNIVERSITARI DE LLENGÜES MODERNES APLICADES

(IULMA)

Maquetación: los autores
Diseño de la cubierta: Celso Hernández de la Figuera

ISBN: 978-84-370-9456-4
Edición digital

CONTENTS

CONTRIBUTORS

Inmaculada de J. Arboleda Guirao holds a BA and MA in English Studies as well as a MA in Legal Translation English-Spanish. She obtained the First Rank Outstanding Student Achievement Award at the Spanish National Level (2009). After holding a PhD Research Scholarship, she joined the Univ. of Murcia and the National University of Distance Education as a lecturer (2011). She has co-authored *Systems for the Phonetic Transcription of English: Theory and Texts* (Peter Lang, 2011) and done research with Prof. Clive Upton (Univ. of Leeds). She is currently completing her PhD thesis on phonostylistics.

Silvia Barreiro Bilbao is a Senior Lecturer in the Department of Modern Languages and Foreign Languages in UNED (Madrid, Spain). She has a degree and a PhD in English Philology (University of León, Spain), and an MSc in Speech and Hearing Sciences (University College London). Since 1994 she has been teaching English Phonetics, first in the University of León and then in UNED. Her teaching and research interests have always been related to Phonetics, mainly on the acoustic sides of speech, with important publications in national and international journals. She is also the author of different phonetic materials published by the UNED.

Jean-Louis Duchet is Professor emeritus of English Phonetics and Linguistics and Dean emeritus of the Faculty of arts & languages at the University of Poitiers (France). He is the author of several books about phonology and the phonetics of English (*La Phonologie,* Paris: Presses universitaires de France; *Code de l'anglais oral,* Paris: Ophrys. Italian translation by Fiona Aglietti, *Codice della Pronuncia Inglese,* Firenze: Festina Lente; and co-author with Alain Deschamps, Michael O'Neil & Jean-Michel Fournier of *English Phonology and Graphophonemics,* Paris:

Ophrys). His research interests concentrate on the historical development of English pronunciation and stress-patterns since 1700, as evidenced by orthoepic treatises and pronouncing dictionaries. He also supervises research on large multilingual corpora of English for legal purposes.

Eva Estebas Vilaplana currently lectures English Phonetics at the UNED (Spain). She obtained her *PhD in Phonetics* at University College London (UCL) where she taught phonetics and phonology and worked as a research assistant for more than six years. She holds an *MA in Phonetics* (UCL) and an *AM in English Philology* (UAB). She has a degree in *English Philology* and another one in *Spanish Philology* (UAB). Her research, mainly on prosody, focus and intonation, has been published in a variety of national and international journals such as *Journal of Phonetics*, *Onomázein* or *RESLA*. She is the author of the book and self-learning materials *Teach Yourself English Pronunciation.*

Robert Fuchs is a PhD student at the University of Münster (Germany). In his PhD thesis, submitted in 2013, he compares the speech rhythm of educated Indian English and British English. He studied English linguistics, computer science and law at the Universities of Dresden (Germany), Hyderabad (India) und Cracow (Poland), and has published on various post-colonial varieties of English.

Luisa Granato is doctor in Linguistics and consultant academic at the University of La Plata, Argentina. Her teaching activities both at graduatre and post-graduate levels, as well as her research interests have always been centred around the field of spoken discourse, both in English and Spanish. She was senior lecturer in English Phonetics, Diction and Language, and is at the moment the head of a research program on casual conversation in Spanish. The result of her work has been published in a book, in chapters of books and national and international journals. She has been invited to lectures and give plenary presentations at different universities in Argentina, México, Venezuela, Uruguay, Chile, Brasil, Sweden and Spain and has participated with lectures in over sixty national and international conferences.

Sylvie Hanote is a professor at the University of Poitiers where she teaches English linguistics. Her fields of research are English phonetics, morpho-phonology and oral discourse analysis. She is particularly interested in the questions of prominence and reported speech in an oral corpus. She recently co-edited a book in French entitled *Focalisation(s). Saillance dans les langues: lexique, syntaxe, prosodie* published by the Presses Universitaires de Rennes (2013).

Erika Larsen has an MA degree in Interpretation and Translation and also a European Master's degree in Conference Interpreting, both awarded by Copenhagen Business School. She has worked for a Scandinavian translation company in Southampton, UK, and is currently a translator in the Danish Language Department of the European Commission in Luxembourg.

David Levey lectures at the University of Cádiz (Spain) on Phonetics and Phonology and Sociolinguistics. His main research interests include accents and language variation, bilingualism and second language acquisition. In addition to publications on various aspects of pronunciation and language teaching, he has written extensively on the sociolinguistics and language situation of Gibraltar.

Inger M. Mees was educated at the universities of Leiden and Edinburgh. She then moved to Copenhagen Business School, where she is now Associate Professor teaching English phonetics. She has written on contrastive phonetics and pronunciation training for Danish and Dutch students, and has also published on sociolinguistics, historiography of phonetics, English-medium instruction, and translation processes. Together with Beverley Collins, she has co-authored several books and numerous articles on English pronunciation, accent varieties, and historiography of phonetics. Their most recent publication is *Practical Phonetics and Phonology*, now in its third edition (Routledge, 2013).

Jose Antonio Mompeán obtained his B.A. and Ph.D in English studies from the University of Murcia in 1997 and 2002 respectively. He also obtained an MA in Phonetics from University College London in 2006. He has published extensively on English phonetics and phonology, with special emphasis on phonology from a Cognitive Linguistics perspective, variation in English pronunciation, and pronunciation teaching.

Brian Mott teaches Phonetics and Phonology, Semantics and Translation in the English Department of the University of Barcelona, where he was Linguistics Coordinator from 1984 to 1990 and from 2007 to 2010. He was also a teacher in the "Escola d'Idiomes Moderns" at UB for thirty years (English coordinator 1976-1980). From 2004 to 2010 he tutored on the Summer Course in English Phonetics at University College London directed by Professor John Wells. He has an MA in Spanish Studies (Aberdeen, 1969) and a PhD in Aragonese Dialectology (Barcelona, 1978). He has also studied the Mirandese dialect spoken in North East Portugal, and has published some twenty books.

María Riera is a Ph.D. candidate in the Department of English and German Studies at the Universitat Rovira i Virgili, Tarragona, Spain, where she also teaches English language and phonetics/phonology courses. Her research interests are acoustic phonetics, speech production and perception, the phonetics-phonology interface, sound change and the teaching of English pronunciation to speakers of Spanish/Catalan.

Luis Fernando Rodríguez Romero (Sevilla, 1977) has a degree in English Philology from the University of Seville. In 2005 he was awarded an M.Phil and is currently working on his Ph.D. dissertation on Pronunciation Teaching. He has lectured on English Phonetics and Phonology at the Universities of Seville and Greenwich (London), and has also presented several papers at both national and international conferences held in Spain for the last twelve years. His fields of interest include phonotactics, English & Spanish contrastive phonetics, prosody and pronunciation teaching.

Joaquín Romero. Master of Arts (1992) and Ph.D. (1995) in Linguistics, University of Connecticut. Associate Professor at Universitat Rovira i Virgili, Tarragona, Spain (1996 to present). Research interests include articulatory phonetics and phonology, the phonetics-phonology interface, sound change and English pronunciation teaching methodology.

Paul Tench is an Associate Researcher at Cardiff University, Wales, UK. He has now retired from full time academic work after working for 40 years at Cardiff and in Nigeria. His main teaching responsibilities and research interests focussed on phonetics and phonology, systemic-functional linguistics, linguistics applied to language teaching and learning. His most well known works include *Pronunciation Skills* (Macmillan, 1981), *The Roles of Intonation in English Discourse* (Peter Lang, 1990), *Intonation Systems of English* (Cassell, 1996) and *Transcribing the Sound of English* (Cambridge UP, 2011).

INTRODUCTION

Rafael Monroy-Casas
Murcia University

Introduction

The teaching of the phonetics and phonology of English to a Spanish university audience –I wonder whether this may be extrapolated to other contexts– has more often than not characterized itself by a bare, uncritical presentation of facts as they are usually portrayed in classic publications such as Gimson-Cruttenden, Collins & Mees, Roach, etc. These textbooks, while being systematic and even comprehensive in their coverage, present a 'canonical' view of language in the sense that they do not usually take account of contrasting views of the phonetics and phonology of English. On the other hand, it is not always easy for learners to find relevant research where well-established views may be called into question; as a result, students end up having a quasi-dogmatic opinion of the phonology of English, hardly being aware of other views and approaches.

With this in mind, we decided to contact a group of experts in English phonetics and phonology and invite them to participate in this project. We made it clear to them that our intention was to produce a book of readings with a balanced combination of theory and praxis, leaving out everything related to models of English, theories of pronunciation, pronunciation teaching or teaching in the classroom. The theoretical component from them was to be a state-of-the-art survey of a given topic, while the gist of the contribution was to be empirical research on a specific aspect of it within the broad thematic range of English phonology. For methodological reasons as well as for reasons of space, we decided to stick to the most fully described variety of British English: RP. The result is this book of readings, which contains a collection of contributions with a fundamentally pedagogical aim and which, hopefully, will provide students with reliable and up-to-date information on key issues of English. There are indeed texts on the market that cover different aspects of the phonetics and phonology of English, but they are either too broad in scope for our purposes, such as Ramsaran's (ed. 1990), *Studies in the Pronunciation of English, a Commemorative Volume in Honour of A.C. Gimson,* Windsow Lewis' (ed. 1995) *Studies in General and English Phonetics. Essays in Honour of Professor J.D. O'Connor*, or Dziubalska & Przedlacka (eds. 2005) *English Pronunciation Models: A Changing Scene*, or they have a specific, narrower orientation, which is the case of Romero–Trillo's (ed., 2012) *Pragmatics and Prosody in English Language Teaching*, where the focus is on the interaction of pragmatic~~s~~ meaning and prosodic features.

Those who have contributed to this volume are all experts in the English language. Their knowledge of and sensitivity to the problems of English pronunciation make their contributions important, since they not only offer an up-to-date view of specific phonological topics, but provide original inroads into the topic under analysis.

The volume is organized into three distinct parts: papers that focus on the phonetics and phonology of English segmentals, suprasegmental aspects of English with special emphasis on intonation and, finally, new developments in English, where some of the changes observed in current British English are reported on and assessed.

The first part consists of five chapters. In the opening chapter (Chapter 1), Luis Fernando Rodríguez studies the status of yod in pre-syllabic position. As the title suggests, this is a study of the behaviour of the semi-vowel /j/ in English, a controversial topic among phonologists (Giegerich, 1992) because, although it can act as a single head in a syllable followed by any vowel, this phoneme only clusters with preceding sonorants. This peculiar behaviour of /j/ leads the author to consider that there is a unique relationship between the onset and the syllabic nucleus, so that one could say that /j/ is part of both, the onset and the syllabic peak. The author argues that, if /j/ belonged only to the peak, there would not be any problem with a word like *crew* being pronounced as */krju:/, another consonant being eligible as part of the onset (it is well known that English phonotactics supports up to three consonants in syllable-initial position, initial /s/ being mandatory). The fact that /j/ cannot be inserted into the above-mentioned word and other phonotactically similar forms, favours the view that the onset consists of only two consonants, a third one being disallowed unless the first one is /s/.

In her contribution, 'Syllabic consonants vs. Schwa in English. A descriptive study' (Chapter 2), Inmaculada Arboleda raises the problem of the occurrence of syllabic consonants in English and their alternation with schwa plus /l/. This is a descriptive study which has as its starting-point the disagreement between British phoneticians as regards the use of one variant or the other. Samples of the speech of 80 informants, 40 men and 40 women –all of them speakers of a non-rhotic variety of English (RP)– were taken from the BBC Learning English website in order to see how natives (three people acted as referees) perceived potentially syllabic consonants, and the degree of agreement/disagreement among them. Monroy's pedagogical rules (2008) were used as a point of reference. These are context-dependent rules which predict whether in a given context we can have schwa or a syllabic consonant or both realizations. The author concludes that while English syllabic consonants appear to follow a pattern in line with Monroy's proposal, their behaviour is not as obvious as one might infer from these rules.

In 'Fricatives Revisited' (Chapter 3), Silvia Barreiro carries out a meticulous analysis of English fricatives from a perceptual standpoint. She focuses on the acoustic cues listeners use to perceive specific traits of a speech sound. After a thorough review of those studies that have centered on the acoustic characterization of English fricatives in an oral context, she pays special attention to the perceptual weight such acoustic cues play in fricative contrasts. Thus, categories such as place of articulation, manner of articulation and voicing are explained regardless of changes they may undergo due to speaker variation or in certain phonetic contexts. Specific acoustic features are also interpreted as a function of their articulatory correlates.

In ‘English plosives: beyond the [±voice] distinction’ (Chapter 4), Joaquin Romero and Maria Riera deal with the voicing parameter in English plosives. They consider that the voiced/voiceless opposition is not convincingly dealt with in current textbooks on the phonetics and phonology of English, since sounds are ascribed to either one category or the other. Although they deem this classification satisfactory at an introductory level, it is not considered adequate at a higher level, as it does not take into account the degree of variation sounds undergo according to context. They contend that other parameters such as voice onset time (VOT) allow a more rigorous description of the true nature of plosives. Factors such as aspiration or devoicing are better explained when related to VOT. Further characteristics of English plosives, such as glottalization, lateralization, flapping or deletion, are analysed on account of the fact that the study of such phenomena serves to give a more complete and realistic view than the simple voiceless/voiced dichotomy. They acknowledge that this may mean introducing greater phonetic complexity, but it is their conviction that this should not be a serious stumbling block within a university context.

The work of Jean-Louis Duchet on English lexical stress opens the second section of the book. In Chapter 5, the author examines both the English accentual phonological hierarchy and the role played by stress placement rules, highlighting the nature of English as a stress-timed language. Taking as a starting point the fact that English primary stress is a distinctive marker of any lexical unit, he goes on to explain its behaviour in simple as well as compound words taking into account the role played by prefixes and suffixes. In his analysis of stress placement, Duchet follows the English linguistic tradition of Kingdon (1958) and Fudge (1984), but also the French contribution of Guierre in particular (1965, 1969). Questions raised by Chomsky and Halle's theory (1968) and some conflicting rules are also addressed. As regards secondary stress and stress shift at phrase level, the author holds the view that both are determined by rules based on the alternation of strong and weak syllables. The chapter ends with a succint summary of the main accentual differences between British and American English.

In Chapter 6, José Antonio Mompeán discusses English accentual shifting. It is well known that the use of both lexical and sentence accent is language specific, there existing striking differences among languages. In English, for example, stress placement, although variable, can be influenced by phonology or morphology (avoidance of stress clash and the effect of affixes, respectively). This is an issue on which there is plenty of research at both theoretical and empirical level –suffice it to mention the contributions of Generative Phonology (Chomsky & Halle, 1968) and Metrical Phonology (Giegerich, 1985; Gussenhoven, 1991) or recent developments in Optimality Theory (Cho, 2002). In order to investigate the degree of stress shift in English, Mompeán discusses this phenomenon by using for this purpose the variation found in the pronunciation of the *-teen* numbers both in compounds and in an attributive function. A number of expressions from a corpus of news bulletins ranging from 1999 to 2009 –taken from the BBC World Service website– were analysed for the occurrence of stress shift. The conclusion reached by the author is that such shift seems to be the norm rather than the exception in the above-mentioned contexts.

Chapter 7 is devoted to the study of rhythm in English. Robert Fuchs addresses in 'You got the beat: Rhythm and timing in SSBE' different suprasegmental aspects of this phenomenon. After criticizing the simplistic view of rhythm that categorizes languages either as syllable-timed or stress-timed, the former having syllables of equal length, whereas in the latter feet would show a tendency to being isochronous, he analyses rhythm by focusing on differences in the length of adjacent syllables, a much more promising approach. He addresses the concept of rhythm as a set of metrical units with a periodic recurrence, discussing some popular metrics that he applies to data taken from the database of SSBE (Nolan et al, 2009). The chapter ends with a brief consideration of some of the factors contributing to the phenomenon of rhythmic variation in a given speech community.

In ‘Studies on the intonation of English: a critical review’ (Chapter 8), Luisa Granato assesses the main theoretical and analytical approaches to English intonation throughout the 20th century and early twenty-first century. Her purpose is to present and discusss their contributions in the light of new linguistic theories, and to see how their principles are applicable to the analysis of texts. The author emphasizes how crucial these contributions have proved to be, especially in an EFL context. She describes thoroughly the School of London concern with both the grammatical and attitudinal function of intonation; the contribution of Systemic Functional Linguistics in the definition of linguistic and non-linguistic meanings of intonation, and the impact of a discourse pragmatic approach. She completes her review by referring not only to the work done on intonation in the United States and in England, but also to other academic circles in other parts of the world who have an interest in the interface between phonology and the semantic or the lexico-grammatical systems of language.

In Chapter 9, Eva Estebas Vilaplana deals with ‘Phonological models of intonational description in English’. As the title shows, this is an introduction to the studies on English intonation focusing on the British and American contributions. Regarding the British school, she discusses the fundamental tenets on which it is based: the distinction between tonality, tonicity and tone (Halliday, 1967), and their idea of intonation as a configurational analysis of pitch contours, each of which consists of a nucleus as a mandatory element along with other optional components. As for the American School, the author reviews the evolution of intonation studies from their inception, with seminal contributions from Pike (1945), Trager and Smith (1951) and others, until now. In these pioneering studies, intonation was framed around four level tones which would account for the pitch movements at the end of an intonation phrase. More recent proposals, such as Autosegmental-Metrical phonology (Pierrehumbert, 1980), or the ToBI system of prosodic notation (Beckman & Hirshberg, 1994), have reduced intonational analysis to two tonal levels associated with stressed syllables and the edges of prosodic domains. The chapter finishes with a comparison on how different approaches can describe some basic intonation patterns in declarative, interrogative and imperative sentences.

In Chapter 10, ‘A systemic functional model of the intonation of clauses in English’, Paul Tench has carried out a detailed description of all the possibilities inherent in sentence intonation in spoken English. Using a systemic functional model, based on the work of Halliday (1967, 1970), Wells (2006) and himself (Tench, 1990, 1996, 2005), he presents the three basic systems that underlie the English intonation system: tonality, tonicity and tone. The author illustrates each of these systems comprehensively with examples in which all possible intonation options at clause level are represented. This is a unique illustration of all the intonation systems at work.

‘Connected speech: pronunciation of words in context’ is the title of Chapter 11, in which the author, Sylvie Hanote, canvasses the phonetic phenomena observed in the phonic substance of lexical units at speech level. Using the work of linguists like Jones (1957), Cruttenden (2001) and Roach (2009) as a backdrop, she deals with the phonetic substance of such units, discussing phonotactic processes such as rhythm, assimilation, elision, linking and prominence. In the first, she addresses the phenomenon of stress shift and vowel reduction as well as the notion of prosodic contour. As for assimilation, she examines how progressive or regressive assimilation can affect point, mode of articulation and voicing. Elision is analysed contextually to show how it may affect consonant clusters. She discusses how linking is achieved either through a final consonant or by means of intruding elements such as [j] , [w] or [r]. Finally, she addresses the question of prominence by taking several of its acoustic correlates into account (variation of intensity and fundamental frequency, vowel duration, change of rhythm, or the presence of a glottal stop). The phonetic description is illustrated by means of spectrographic representations of different sound sequences taken from the radio.

The third section, dedicated to New Developments in English, consists of three chapters. In the first (Chapter 12), Erika Larsen & Inger Mees study ‘four phonological variables in Cliff Richard’s songs and speech over a period of 50 years’. Based on the fact that British pop singers have shown from the 50s a tendency to use American English forms, motivated no doubt by the impact of the American popular music of the time, the authors have studied the music and the speech of Cliff Richard in a sample of 25 recordings taken from his musical repertoire and from interviews with the singer. The aim was to investigate three fundamental issues: to see whether there is any variation in pronunciation when Cliff Richard speaks or sings; to analyse whether the musical genre has exerted any kind of influence on his way of singing or speaking, and finally, to see to what extent the use of American speech habits have evolved over time in his pronunciation.

'Recent changes in English phonetics and phonology and their representation in phonetic notation’ is the title of Chapter 13, in which Brian Mott describes and analyses those changes that have taken place over the last 50 years in the phonetics and phonology of British English. The author reviews different approaches followed by British scholars to represent such changes in the leading pronunciation dictionaries (*The Longman Pronunciation Dictionary, The Cambridge English Pronouncing Dictionary, and The Oxford Dictionary of Pronunciation for Current English)*. As regards the notation to use, the author opts to follow the established usage as it appears in such dictionaries, making only minor concessions when it is understood that the change represents a significant advantage in terms of higher phonetic accuracy with respect to traditional practice. At the phonological level, he discusses some recent changes such as those affecting the short vowels, the merging of some of them in certain contexts, the widespread use of schwa, the smoothing of some diphthongs, *happY* tensing, etc. As for the consonants, he deals with current phenomena such as T-tapping and glottalling, L-vocalization, Yod- coalescence, and intrusive [r], etc.

The last chapter (Chapter 14) deals with ‘Some recent changes and developments in British English’. In it, David Levey delves into the changes that are occurring in contemporary British English. He assumes that both the RP of Daniel Jones and the English used by the BBC in its early days are far from the English one can hear today. New attitudes, education, changes in social structure, mobility and, of course, the impact of the popular media have helped shape today’s English. Based on recent research, like Brian Mott in the previous chapter, the author reviews the phonetic changes that are taking place in British English. He considers their place and status in the standard pronunciation models and analyzes to what extent one can bridge the gap between the type of pronunciation presented in textbooks and the English spoken today. Phenomena such as vowel shortening, H-dropping, TH-fronting, L-vocalization and other forms stigmatized until recently are now increasingly accepted. As perception and attitudes change over time, sometimes becoming the standard, Levey understands that the student has to be aware and recognize these developments, regardless of whether (s)he will finally adopt them or not.

In summary, this book of readings deals with the phonetics and phonology of British English providing an up-to-date and critical view of fundamental aspects of English that the student cannot ignore. We hope that the contributions collected in this volume will serve to give the student a less monolithic, less simplistic and more realistic view of the complex phenomena underlying the English language.

PART I

English Segmental Phonetics and Phonology

CHAPTER 1. THE STATUS OF YOD IN ‘PRECENTRAL’ (GIMSON) POSITION

Luis Fernando Rodríguez Romero
University of Seville

1. A brief phonetic description of /j/

The semivowel “yod” /j/ belongs to the group of the so-called “approximant” sounds, which, according to Knight, are formed “when the articulators are positioned in wide approximation, so that the gap between them is not narrow enough to cause friction” (2012: 40). Because of its palatal articulation, it is made with the front of the tongue approaching the hard palate. Semivowels are also called “glides” in that they consist of vowel-like movements, very much like diphthongs: “a semi-vowel is a rapid vocalic glide onto a syllabic sound of greater steady duration” (Gimson, 2001: 210).

Table 1. Sonority Scale.

	Vowels
Most sonorous (weakest consonantality)	Approximants
	Nasals
	Fricatives
Least sonorous (strongest consonantality)	Affricates
	Plosives

In the sonority scale, semivowels are, of all consonant sounds, the closest to pure vowels, even though they have “a reduced low-frequency spectrum amplitude, an additional decrease in amplitude at higher frequencies, and reduced prominence of the second or third formant peak" (Clements, 2009: 167-168).

2. Phonological behaviour of /j/ in single onsets

In spite of all those similarities, semivowels are not treated as vowels. To begin with, they lack the degree of prominence of pure vowels, as exemplified by such pairs of words as "**y**ou" and "b**e**". To consider the initial sound in "you" to be a vowel by itself would imply that the combination of vocalic sounds forms a hiatus and that, therefore, the monosyllabic word should, actually, be a di-syllabic word, which is obviously not the case there. The other "temptation" would be to consider the /ju:/ combination as a "diphthong". However, all eight diphthongs in Standard British English, as exemplified by the following words: "h**ere**", "p**oor**", "m**ai**n", "l**ow**", "b**oy**", "**air**", "p**ie**" and "m**ou**th"... are characterized by having a more prominent "onglide" followed by a less prominent "offglide". In other words, English diphthongs are said to be "falling diphthongs" in that the first element has more prominence than the other. However, in the sequence /ju:/ it is rather the other way round: it is the second element that is more prominent than the first. Because of that, rising combinations in English are treated as sequences of approximant plus vowel rather than diphthongs.

There is yet another reason why the semivowel in "you" could not be considered as part of a syllabic nucleus. Let us consider the following examples:

a pear

*a**n** apple*

According to the rule, the indefinite article morpheme could be realized either by "a" or "an" depending on whether the following word begins with a consonant or a vowel. In the first example, the "a" allomorph is chosen because the next word begins with a consonant, while in the second case, "an" is selected due to phonological conditioning. The question that needs to be addressed now is why is the "a" allomorph also used when the next word begins with a semivowel as in, for instance, "uniform"?

In the case of the word “pear”, the word begins with a consonant, which means that there is a consonant sound in the “onset” position of the syllable, followed by a diphthong in the nucleus position. In the case of “apple”, however, the “onset” position of the syllable, occupied by optional consonant sounds preceding the nucleus, is empty. This might justify the inclusion of a nasal consonant between the two vowels in what could be considered as a special kind of linking, as in “an apple”. The great difference between “pear” and “uniform” is that the first sound in these two words sounds quite different from each other. The initial plosive in “pear” is clearly a consonant sound, both phonetically and phonologically. It sounds like a consonant, and behaves as a consonant as well. As a plosive, it is the least sonorous of all consonants, so much so that it could never occupy the nucleus position, unlike some approximant sounds such as the lateral or the rhotic.

On the other hand, although the initial sound in “uniform” might sound like a vowel, as explained above, however it does not behave as such from a phonological point of view. If the semivowel were part of the nucleus as a hiatus or diphthong, the onset position would, then, be left empty as well, in which case the use of the “an” allomorph would be asked for as in “pear”. The fact that the other allomorph is used indicates that the onset position in “uniform” is, indeed, occupied by a sound. Therefore, in spite of its phonetic resemblance to vowels, it seems clear that the semivowel phonologically behaves like a consonant, because it occupies a purely consonantal and marginal position inside the syllable rather than a central position.

It seems clear, therefore, that the traditional distinction between consonants and vowels is not enough to account for all these intricacies, as “sonorant” consonants seem to form a somehow intermediate category between “obstruent” consonants and vowels, thus blurring the differences between them. To overcome these problems, Pike (1943) proposed considering two different distinctions, one strictly phonetic (“vocoid” vs. “contoid”), and the other based on the phonological functioning of the sound inside the syllable (“vowel” vs. “consonant”). Thus, the yod semivowel may fit the phonetic definition of a vowel, since it is produced with no audible noise due to a constriction in the vocal tract. However, it does not function as a vowel, because it is never syllabic, that is, it can never occupy the nucleus position of a syllable. In other words, it is a non-syllabic vocoid sound, acting as “very brief transitional onsets to the vocoid at the nucleus of the syllable” (Laver, 1994: 297). The first syllable of “uniform”, therefore, would be analyzed as follows:

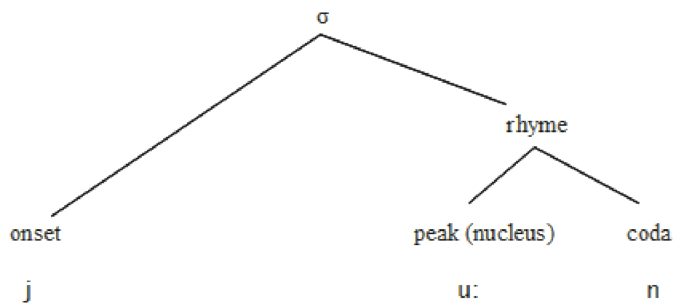


Figure 1. Syllabic structure of “un-”.

3. Phonological behaviour of /j/ in onset clusters

So far evidence has been given to support the claim that the semivowel does not belong to the nucleus of the syllable together with the following vowel, but rather to the onset of that syllable. However, this analysis poses some intriguing problems as well whenever the semivowel occurs in onset clusters. But before proceeding any further, it is necessary to provide a brief review of the onset phonotactics in English syllables.

The maximum number of elements allowed in the English onset is two. Each one of these two consonants will occupy one X-position in the onset. Giegerich defines the term “X-position” as a “timing unit” (1992: 142-143). It is important to note that onset and peak must constitute an upward slope in sonority, which means that the first one of these two consonants allowed in onset clusters must be less sonorous than the next one. In other words, as the elements in the onset get closer to the nucleus or peak of sonority, they become more and more sonorous.

The only exception to this generalization involves onset clusters of three X-positions, the first one of which is occupied by /s/, as in “spring”. Since the fricative sound is more sonorous than the plosive, there is a clear violation of the Sonority Sequencing Principle, according to which “segments are syllabified in such a way that sonority increases from the margin to the peak” (Clements, 2009: 169). Therefore, and since there are two peaks of prominence (the initial fricative sound and the vowel), there should be two syllables instead of one. However, an alternative solution would be to consider that “offending” coronal sound to be a special kind of segment that is appended to the core syllable:

> It is only /s/ that can violate the sonority generalisation of the onset, and whenever an onset contains three X-positions then the first one of these will be an /s/ that also violates the sonority generalisation. We can maintain our generalisation that core onsets contain two X-positions if increasing sonority by simply treating the occurrences of /s/ (…) as appendices: a core onset may be preceded by an appendix /s/ (1992: 150).

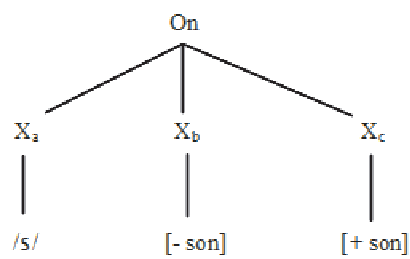


Figure 2. English onset template.

The reasons to consider /s/ as an external appendix rather than as part of the core syllable are twofold: on the one hand, its occurrence implies a violation of familiar principles of syllabification based on sonority such as the Sonority-Syllabicity Alignment Principle, which states that “sonority peaks correspond to syllable peaks and viceversa” (Clements, 2009: 169); on the other, onset appendices can only be /s/: all words with three consonants in the onset include the fricative sound as the very first segment in the cluster: “spring”, “strike”, “scream”, “split”, etc. In all of these words, the first element in the cluster is more sonorous than the second, so it should be treated as a somehow external element that is appended to the core syllable (“s + pring”, “s + trike”, “s + cream”, “s + plit”).

Taking all this into consideration, now it is time to reconsider the casuistry raised by the behaviour of the palatal semivowel in onset clusters. It is relevant to notice, first of all, how freely this sound clusters with nonsonorant sounds. In fact, there is no other sonorant sound with so many possible combinations with obstruents in onset clusters. It clusters with all oral plosives, both voiced and voiceless, as well as with most fricatives as well. Furthermore, some of these nonsonorant sounds do not occur in clusters at all, unless the following sound is /j/.

Certainly, it does not cluster with either affricate sounds or palato-alveolar fricative sounds, although it is important to realize that unaccented sequences of /tj, dj, sj, zj/ had already coalesced, in earlier stages in the evolution of the language, into /ʧ, ʤ, ʃ, ʒ/, as in “statue”, “educate”, “usual” or “visual”... which means that it would be redundant for a sound like /ʧ/ to cluster with /j/ in words such as these. Furthermore, even if they function, in phonotactic terms, as single units (otherwise there would be two nonsonorant sounds in onset position, thus violating the aforementioned sonority requirements), affricate sounds are stop-plus-fricative sequences, which explains why they do not cluster with any other sound in syllable onsets in English.

On top of that, it is striking to notice that the palatal semivowel can also occur in onset clusters after sonorant sounds, as in “mule”, “new” or “lewd”. There is yet another peculiarity that should call our attention. In single onsets, /j/ can be followed by any vowel in the peak: “yield” (/iː/), “yip” (/ɪ/), “you” (/uː/), “**<u>Yuc</u>**atan” (/ʊ/), “yes” (/e/), “yearn” (/ɜː/), “yaw” (/ɔː/), “yacht” (/ɒ/), “yarn” (/ɑː/), “yak” (/æ/), “young” (/ʌ/), “**<u>Ya</u>**kutsk” (/ə/), “yoke” (/əʊ/), “year” (/ɪə/)… however, what do all the words included in the table below (“pew”,

Table 2. Possible consonant clusters in British English onsets.

	m	n	ŋ	l	r	w	j	Examples
p				+	+		+	play, prey, pew
b				+	+		+	blue, brew, beauty
f				+	+		+	fly, fry, few
v[1]				+			+	view, vlog
t					+	+	+	try, twig, tune
d					+	+	+	dry, dwell, duty
θ					+	+	+	throw, thwart, thew
ð								
s	+	+		+		+	+	smear, sneer, slow, swing, sue
z							+	Zeus
ʃ					+			shrew
ʒ								
ʧ								
ʤ								
k				+	+	+	+	clue, crew, queen, queue
g				+	+	+	+	glue, grey, Gwen, gules
h								huge

"beauty", "few", "view", "tune", "duty", "thew", "sue", "Zeus", "queue", "gules" and "huge") have in common? All of them have the same vowel in peak position after /j/, that is, /uː/. There is, therefore, enough evidence to support the claim that cluster-/j/ tends to occur in syllables where the peak is /uː/ (/ʊə/ and /ɔː/ could be also found in such words as "cure"). It is important to realize that this constraint is only given with consonant-plus-/j/ clusters. There is no other possible onset cluster combination in English demanding any specific vowel as the peak of that syllable.

The peak and the coda are usually grouped together in a single phonological unit called "rhyme". There is a very good reason why these two segments of

1. The frequency of words such as "vlog" or "Vlach" might suggest that /vl/ has already become a licit initial cluster in present-day English.

the syllable should be regarded as part of another phonological unit, apart from the role played in the rhyming conventions of poetry. Peak and coda function together, so much so that, as Giegerich (1992) points out, “it is the number of X-positions in the rhyme (rather than the number of X-positions in peak and coda counted separately) that determines whether or not a syllable is well formed” (143). The maximum number of X-positions allowed in the rhyme of an English syllable is three. If the peak contains two X-positions, there will be just one X-position left in the coda, or vice versa. In other words, “it is the sum of X-positions in peak plus coda that counts, not the way in which these X-positions are distributed” (144). This would be the reason why */siːlp/ is not a plausible syllable in English[2].

The next logical question would be: does the peak of the syllable function together with the onset as well as with the coda? The answer seems to be negative: peak and onset do not interact in the calculation of maximal syllables, since both function separately: “there is no evidence for onset and peak being tied up together in a single unit in the way peak and coda are in the rhyme” (144). And yet, since cluster-/j/ and peak seem to have such a close relationship, there must be another reason to account for the fact that this cluster-/j/ does only occur before /uː/. According to Giegerich, this is because the palatal semivowel, when part of an onset cluster, is part of the peak as well:

> (...) if /j/ is part of the peak then we cannot expect it to be constrained by conditions that hold within onsets (…) We now have an explanation for the fact that /j/, itself [+sonorant], can occur after [+sonorant] phonemes as well as after [-sonorant] phonemes that otherwise fail to cluster altogether, such as /z, v/ etc. (1992: 157-158).

So far there seems to be enough evidence to accept and support this claim that cluster-/j/ must be part of the peak. However, and since the relationship between the semivowel and the peak is so close... is there any reason why cluster-/j/ should be also part of the onset? Let us consider the following examples: “crew”, “flew”, “shrew”[3].

---

2. However, it does not account for the correctness of words such as “quaint”, having a rhyme with four X-positions (/eɪnt/).

3. It's interesting to note that these words are usually mispronounced by Spanish students of English as [krjuː], *[fljuː], *[ʃrjuː]. They might be drawing an analogy with other similar words

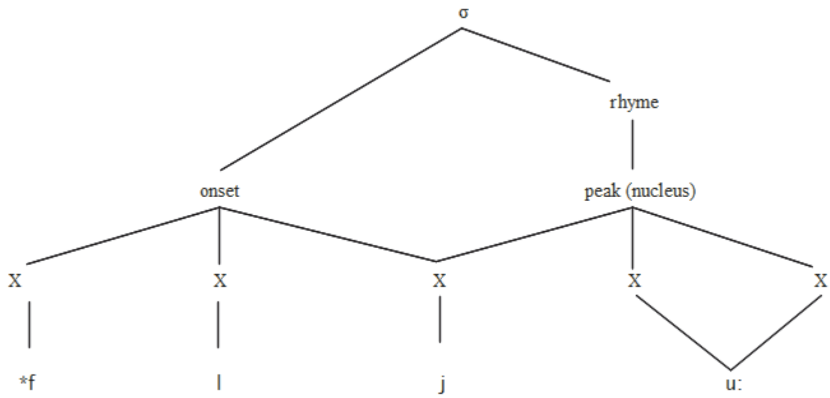


Figure 3. Wrong analysis of “flew”.

If the palatal semivowel were only part of the peak together with /uː/, then *[krjuː] would be an acceptable pronunciation of “crew”, since there would be two X-positions in the onset and another two X-positions in the peak.
However, the fact that these words are, actually, pronounced as [kruː], [fluː], [ʃruː]... proves that cluster-/j/ is not entirely part of the nucleus. Otherwise, there would be no reason why “few” and “flew” should be pronounced so differently. While in “few” there are two consonants and, therefore, two X-positions in the onset, in “flew” there would be three X-positions in the onset,

containing the digraph “-ew” in their written form, such as “view”, “few” or “dew”. Alternatively, non-native disyllabic realisations should be also considered in these examples.

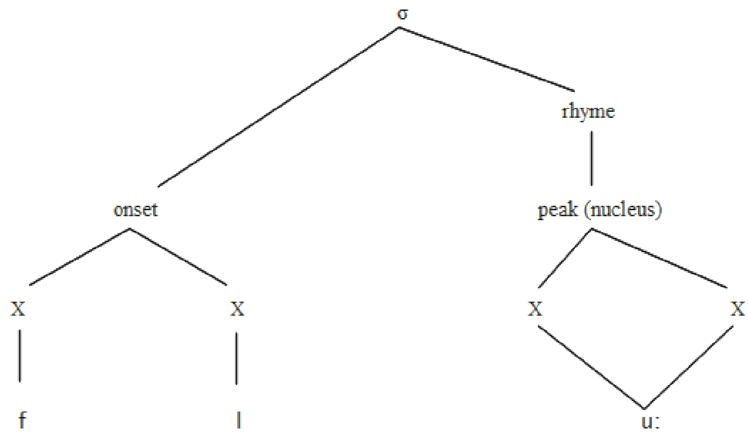


Figure 4. Correct analysis of “flew”.

in “flew” there would be three X-positions in the onset if the semivowel were actually pronounced. As explained above, the maximum number of X-positions allowed in the onset of a well-formed English syllable is two, with an optional third segment (an appendix). This is, therefore, the evidence needed to justify why cluster-/j/ can't be exclusively part of the peak, otherwise */fljuː/ would not be unacceptable in English: /fl/ would be part of the onset, while /juː/ would be part of the peak. It is because cluster-/j/ will always occupy one of the X-positions provided by the onset template that it can be only preceded by a single consonant or an appendix plus just one more consonant. And since the only appendix allowed in onset position is /s/, this explains why onset combinations such as /stjuː/ (“Stu”) are possible in English, while others like */trjuː/ (“trew”) or */fljuː/ (“flew”) are not.

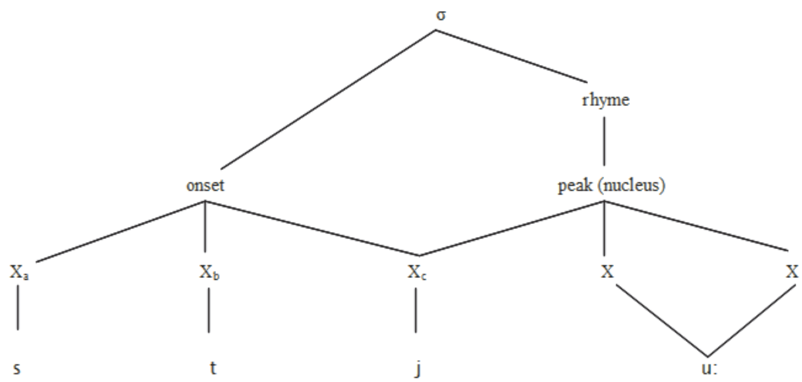


Figure 5. Syllabic structure of “Stu”.

This analysis poses one intriguing problem. Consider the following rhyme template for English monosyllabic words:

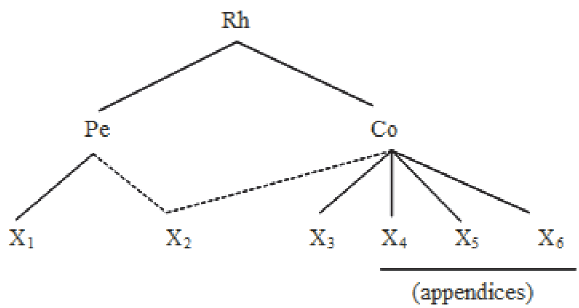


Figure 6. English rhyme template.

If cluster-/j/ is also part of the peak, thus occupying an X-position shared by both the onset and the peak, and if the maximum number of segments (and therefore X-positions) allowed in the rhyme is three, plus three optional coronal segments as appendices, how should a word like “stupe” /stju:p/ be analyzed? Is the semivowel occupying the X_1 and X_c positions simultaneously? Is the X_3 position in the peak as well instead of the coda? Is it possible to have three X-positions in the peak alone? And what about the final consonant? The stop /p/ is clearly not a coronal sound, and is not among the possible syllable-final appendices as listed by Giegerich, so it could not occupy positions X_4 to X_6. All these issues should be thoroughly addressed in future studies on this topic.

Needless to say, this brief outline concerns those accents that are not characterized by “yod-dropping”, whereby /j/ is lost after coronal alveolars /t, d, n, l, s, z/, as in “tune” (/tuːn/) or “new” (/nuː/). It is important to notice, nevertheless, that this phenomenon is not restricted to American accents such as G.A. As Glain (2012) and Cruttenden (2008) have already pointed out, in some British accents, including R.P. English, both /uː/ and /juː/ seem to coexist nowadays in such words as “lute”, “assume”, “supermarket” or “suitable”, the former being more and more common after /l/ and /s/ in accented syllables, while the latter remains predominant after /θ/ and /z/.

The loss of /j/ in onset clusters might be also due to “yod coalescence”, very specially in /tju/ and /dju/ sequences. Recent research carried out by Glain has shown that the loss of the semivowel is not restricted to unstressed syllables, thus developping in stressed syllables as well, as a consequence of either dropping or palatalisation. Furthermore, he adds that “this is the continuation of historic processes that have invariably led to the loss of /j/ from /Cju/ sequences since the beginning of modern English” (2012: 21).

4. The nature of post-central elements in dipthongal glides

So far /j/ and /iː/ have been treated as different phonemes, in spite of their phonetic similarities, because they behave in very different ways: while the former has an almost exclusively marginal position in the syllable, occupying the onset position, the latter has a syllabic function, occupying the peak position. The semivowel can be part of the peak together with /uː/ as well, but only when it is part of an onset cluster. As far as pre-central single-/j/ is concerned, therefore, enough evidence has been already provided to justify the reasons why it should be treated as belonging to the onset instead of the peak as part of a possible rising diphthong (e.g. /je/ in a word like “yes”). What about post-central position? How different is the [j] sound in “yes” from the [ɪ] sound (offglide) in the diphthong /eɪ/ (as in “make”) so as to justify the use of different phonetic symbols? Would it be possible to consider the occurrence of semivowels in post-central position in the rhyme of the syllable? In other words, is /ej/ a plausible representation of the vowel glide in such words as “pain”?

Gimson (2001) rejects the validity of that analysis in his treatment of RP post-central [ɪ, ʊ] elements because of the following reasons:

> (i) they do not have a distribution after all vocalic elements as general as that which we find in the case of pre-central /j, w/;
> (ii) they are in RP very weakly articulated (compared with pre-central /j, w/) and may correspond to monophthongal pronunciations in many other accents, cf. [eɪ] and [eː] or [ɛː];
> (iii) they have none of the fricative (phonetically consonantal) allophones characteristic of pre-central /j, w/ following /p, t, k/ (2000: 94)

While Gimson treats consonantal [j] and vocalic [ɪ] as separate segments from a phonemic point of view, Giegerich (1992), on the other hand, treats them as different realizations, in complementary distribution, of the same phoneme. [j] is always associated with onset positions (except when part of a cluster), while [ɪ] will always occur, exclusively, in peaks (never in onsets), either in the form of a lax monophthong or as the offglide of a closing diphthong. The only position where this phoneme can not be found is in the coda. This is why “if the X_2 position is [+consonantal] then it is a coda consonant and if it is [-consonantal] then it is a vowel, part of the peak” (166). Semivowels and vowels, as vocoids, are all [-consonantal] sounds. And only [-consonantal] sounds can occur as peaks in stressed monosyllables.

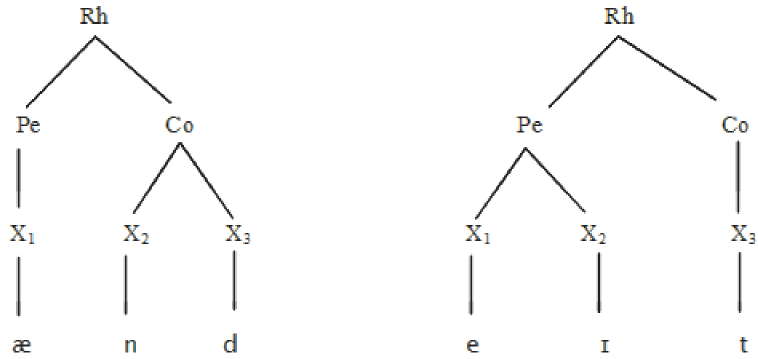


Figure 7. X_2 position: [+consonantal] and [-consonantal].

In a word like “and”, the X_2 position is clearly [+consonantal], since the sound following the peak is a nasal. Therefore, this nasal sound will be part of the coda. On the other hand, in the example “eight”, the X_2 position (since long vowels and diphthongs are associated with two X-positions) is considered [-consonantal], and therefore part of the peak.

To sum up, whether the semivowel is to be treated as a separate phoneme or as an allophone of the same phoneme, its occurrence in the peak will be restricted, in the analysis of Standard British English, to a pre-central position as part of an onset cluster (as in, for instance, “crew”). As far as diphthongal glides are concerned, /ɪ/ will be used in post-central position. While the onset accepts [+consonantal] (e.g. plosives for instance) as well as [-consonantal] (in this case, semivowels) segments, the peak only accepts [-consonantal] (vowels or, exceptionally, semivowels) segments and the coda only accepts [+consonantal] segments, as shown in the figure below.

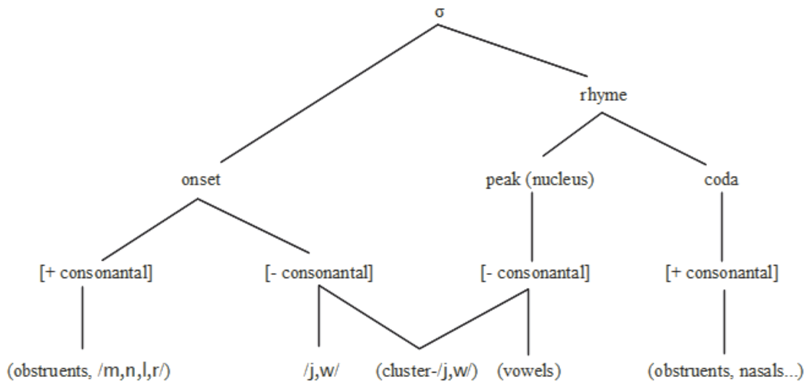


Figure 8. English syllable structure.

5. Conclusions

In spite of its phonetic resemblance to vowels (hence "vocoid"), its non-syllabic nature justifies the inclusion of yod in the group of consonants according to Pike's twofold distinction. In fact, according to Giegerich, the semivowel can be only part of the peak when it is, at the same time, part of an onset cluster, due to a series of constraints related to the nuclear vowel and the allowed number of consonants preceding the yod. Finally, although it clusters more freely with nonsonorant sounds than any other sonorant, it might nonetheless end up disappearing in /tju, dju, sju, zju/ sequences due to ongoing processes of either dropping or coalescence (palatalisation).

REFERENCES

Clements, George Nick (2009): Does Sonority Have a Phonetic Basis? *Contemporary Views on Architecture and Representations in Phonological Theory*, 165-175.

Cruttenden, Alan (2001): *Gimson's Pronunciation of English*. 6th edition. London: Arnold Education.

Cruttenden, Alan (2008): *Gimson's Pronunciation of English*. 7th edition. London: Hodder Education.

Giegerich, Heinz (1992): *English Phonology*. Cambridge: C.U.P.

Glain, Olivier (2012): The Yod /j/: Palatalise it or Drop it! How Traditional Yod Forms are Disappearing from Contemporary English. *Cercles*, 22, 4-24.

Knight, Rachael-Anne (2012): *Phonetics: A Coursebook*. Cambridge: C.U.P.

Kreidler, Charles (1997): *Describing Spoken English*. London: Routledge.

Laver, John (1994): *Principles of Phonetics*. Cambridge: C.U.P.

Pike, Kenneth (1943): *Phonetics, a Critical Analysis of Phonetic Theory and a Technique for the Practical Description of Sounds*. Ann Arbor: University of Michigan Press.

CHAPTER 2. SYLLABIC CONSONANTS VS. SCHWA IN ENGLISH. A DESCRIPTIVE STUDY

Inmaculada de Jesús Arboleda-Guirao
Murcia University

1. Introduction

The schwa and syllabic consonants are very common in English, and they are often found in everyday speech (Trask, 1996). Nevertheless, their usage has been a source of controversy in phonology. Many phoneticians, including Roach (2000) and Monroy (1980, 2008-2009), have formulated theories and rules about when the schwa and syllabic consonants are produced. These phoneticians tend to hold differing views about the production of these two phonemes, and they occasionally take a normative approach, recommending for certain words either a schwa or a syllabic consonant. As pointed out by Roach (2002, p. 75), "this area needs attention".

It cannot be denied that studying words and proposing pronunciations, the approach taken by the above phoneticians, is a highly tempting as well as necessary one in English phonetics due to the unpredictable nature of the English language, in which mismatches between pronunciation and spelling are exceedingly common. This approach is not challenged in this paper, as words standing alone and, in particular, the syllables in which either the schwa or syllabic consonant is found are the focus in this study. However, this is a narrow view and it is necessary to cast some light on the relationship between discourse and English syllabic consonants because in real life words are found in connected speech. Although little research has been undertaken into this issue by means of acoustic analysis, even less has been carried out in terms of perception. When do listeners perceive the same phoneme? Do these referees follow the same criteria when assessing syllabic consonant vs. schwa in a word? This study also intends to cast some light on this issue.

In this research project a real-speech corpus, in particular, a corpus whose source is the BBC news, will be explored in order to investigate how English syllabic consonants and the schwa in final position are produced and perceived. All this will be done insofar as the listeners' answers make it possible, since the study will be perceptual rather than acoustic.

2. Literature review

Research has been carried out into syllabic consonants and the schwa in languages in which these phonemes occur (Booij, 1983; Kager, 1989; Becker, 1998; Wiese, 2000; Windsor Lewis, 1990; van Oostendorp, 1995, 1998, 2000).

2.1. The production of syllabic consonants vs. schwa in English

The controversy around when the schwa or a syllabic consonant is produced is a mere shadow of an existing controversy involving the phonemic status of syllabic consonants. As stated by Álvarez (1980, p. 41), there is disagreement as to "whether syllabic consonants should be given phonemic status or should be interpreted as consisting of schwa plus consonant". Cohen (1957), Wells (1965) and Gimson (1970) tackle this issue. Some phoneticians assign "a different phonological function, depending on whether they are syllabic or not" (Monroy, 1980, p. 172, our own translation), as opposed to other authors, such as Jones (1992). Actually, according to Wells (1965, as cited in Monroy, 1980, p. 172–173),

> There are some types of speech in English RP which do not make a difference between syllabic and full consonants. Nevertheless, there are others which make systematic differences (our own translation).

As a result of these conflicting opinions, it does not seem odd that phoneticians differ from each other in the production of a syllabic consonant or a schwa.

2.1.1. Theories and rules

Theories and rules have been formulated by phoneticians about the occasions on which syllabic consonants and the schwa are pronounced. In order to propose them, these scholars explore only the phonemes within the word in question. Jones (1909/1992), Monroy (1980), Wells (1995), Roach (2000), Mora (2003) and Cruttenden (2001) provide comprehensive accounts of British English syllabic consonants. The focus in the present study will be on this variety because the overwhelming majority of the speakers selected use RP. Nevertheless, Kenyon (1956) and Abel (1962) and Gick (2002) are worth a mention for their theories in American English. British phoneticians deal with the schwa and syllabic consonants by distinguishing amongst their varying positions within the word: initial, mid and final. It is the latter which concerns our research. All the phoneticians mentioned above study three common consonants likely to precede a schwa or becoming syllabic consonants: /n/, /m/ and /l/. Two of these phoneticians, namely Jones (1909/1992) and Monroy (1980) also investigate /r/ but in the present research only non-rhotic accents will be studied. Disagreement amongst scholars is found once again, in this case regarding the use of a syllabic consonant or a schwa. For instance, in reference to /n/ as a final consonant, unlike Monroy (1980), Roach (2000) decides on a syllabic consonant when it is preceded by /f/ or /v/, e.g. seven /n̩/, except in word-initial positions. Even an established phonetician –in this case, Roach– changes his opinion over time. Thus, Roach (1986) holds the view that a syllabic consonant is to be used when it is preceded by more than a single consonant, including when the first consonant is /s/. In contrast, Roach (2000) states that in that case it is more usual to find a schwa, e.g. Wilton /ə/ or Boston /ə/.

In this paper, we discuss some of the rules for syllabic vs. non-syllabic consonants devised by Monroy (2008-2009). In particular we focus on four pedagogically-oriented rules operating in word-final position:

1ST RULE: Syllabic consonant
V (stressed) + C (except for r, l, m, b, g) + V (weak) + C (m, n, ŋ, l, r)
Examples: *lesson, people*.

2ND RULE: Schwa
V (stressed) + C (r, l, m, b, g)+ V (weak) + C (m, n, ŋ, l, r)
Examples: *melon, lemon, organ, ribbon*.

3RD RULE: Schwa
V (stressed) + N.[1]/NN+ V (weak) + N/ N+ Hom[2]
OR
V (stressed) + N+ Hom/Hom+ N/Plos.+N/N+Plos. +V (weak) + N (+ Hom)
Examples: *London, Clinton, Camden, cannon, human, diamond*.

4TH RULE: Syllabic consonant
V (stressed) + C + V (weak) +C (m, n, ŋ, l, r) C OR
V (stressed) + CC + V (weak) + C (m, n, ŋ, l, r)
Examples: *symbol, present, patient*.

As can be seen, Monroy (2008-2009) devises simple rules where alternative pronunciations are not envisaged (by alternative pronunciations we mean that both the schwa and a syllabic consonant are possible). He assigns one pronunciation to each rule. Rules 1, 2, and 4 apply equally to all the consonants liable to be syllabic or preceded by a schwa: /m/, /n/, /ŋ/, /l/, /r/. In contrast, Rule 3 applies only to nasals: /m/, /n/, /ŋ/, as seen on the preceding table.

In general terms, syllabic consonants are often said to be more acceptable (MacCarthy, 1965) and obligatory in a semi-formal register (Jones, 1918/1976). Words such as *obligatory*, *must* or *mispronunciation* are

[1] N. stands for Nasal

[2] Hom. stands for Homorganic. Trask (1996: 173) labels those “adjacent consonants” which share their point of articulation as *homorganic*; for example, *lamp* has two adjacent bilabial consonants, /m/ and /p/.

commonly found in the theories proposed such as those devised by Monroy (2008-2009). However, we think these words are too categorical, bearing in mind the disagreement amongst theories, rules and empirical studies. Giegerich (1992) claims that the schwa is usually elided when it precedes sonorant consonants and that in this context, even in the case of citation forms, the syllabic consonant may be used. We hold the view this claim is over-vague since, if we talk about sonorant sounds, we are thinking about /l/, /n/, /m/, /ɹ/ and /ŋ/; that is, all the consonants that come to mind when dealing with the choice between the schwa and syllabic consonants; but he does not specify when these phonemes are employed. The word *may* in Giegerich's account (1992) also contributes to vagueness*.* As can be seen, these theories are either too strict (speech can never be defined in black and white terms) or too vague.

2.2 The perception of syllabic consonants and schwas

The production of syllabic consonants and the schwa can also be studied by means of perceptual analysis, not just acoustically as was done by Lehiste (1964) or Álvarez (1980), among others.

A perceptual analysis is suitable not only as a means to explore speech production, but also as an end in itself, because it is listeners who are to perceive sounds in real-life speech. There are some research projects which focus on perception in other languages, for example, Bhatia (1973) in Hindi, and van Bergem (1995, in Dutch, in which he aims to know whether listeners are able to unambiguously distinguish between full vowels and the schwa). If focusing on English, there is a study on perception between lateral and nasal syllabic resonants, García (2006), but it is laboratory work. Mora and Mott (2003) also focus on the perception of consonant syllabicity by examining the role of length. Most of the perceptual studies in English which focus on language in use deal with phonetic aspects other than syllabic consonants and the schwa, e.g. Schaeffer and Eichorn (2001), in which a factor, vowel duration, is found to exert an effect on the listeners' perception of naturalness in sentences; this is achieved by exploring the listeners' level of agreement and disagreement. Only Arboleda and Monroy (2010) and

Arboleda (2012) deal with the perception of English syllabic consonants vs. schwa but in view of other factors. In our study, production, which is going to be dealt with by means of the listeners' answers, will be explored but attention will also be paid to perception as an end in itself. The point of whether the listeners share their assessment criteria, which needs research, will be covered as well and an attempt will be made to find out about agreement and disagreement in terms of Monroy's rules (2008-2009).

The perception of English syllabic consonants and schwa has not been approached in view of production rules. Nonetheless, there are studies which deal with the perception of surrounding phonemes in the word in question, but not addressing the difference between syllabic consonants and the schwa. For example, Eilers (1977, cited in Reese & Lipsitt, 1979) investigates the distinction of English fricatives in children. This researcher finds out that [sa] and [sha] distinctions are perceived by listeners before they are three years old, but that [sa] and [za] come three months later. The distinctiveness of this study, which makes it different from ours, is that in this case the listeners' answers are not geared to knowing about the difficulty/easiness experienced by listeners in general terms as for the phonemes in question but to distinguishing the difference between ages.

2.3 Corpus-based studies

As for studies dealing with corpus-based studies in newsreader English, Auran and Bouzon (2003) cover elision rules for the phonetisation of the MARSEC corpus written in French, summarised in English in Auran, Bouzon and Hirst (2004). For BBC news in particular, Mompeán and Mompeán (2006, 2007) and Deterding (2006) have carried out research on this corpus, concentrating on linking /r/ and the deletion of final /t/ and /d/, respectively (the latter is perceptual). We have made use of the same corpus (devoted to learners of English) as Mompeán and Mompeán (2006, 2007). No studies have been seen on news corpora which explore either the production or the perception of syllabic consonants and schwa.

3. Objectives

Our main aim in this research is to study the production and perception of English syllabic consonants and the schwa plus lateral/nasal in word final position at discourse level on a perceptual basis in view of the rules devised by Monroy (2008-2009). In particular, two main objectives followed by several research questions have been addressed:

1. Production of English syllabic consonants vs. schwa
   a) Does the listeners' level of agreement on syllabic consonants and schwa match Monroy's rules (2008–2009)?
2. Perception of English syllabic consonants vs. schwa
   a) Is there a relationship between Monroy's rules (2008-2009) and the listeners' level of agreement and disagreement?
   b) Do the listeners use the same assessment criteria in view of the rules above?

4. Methodology

The informants were 80 newsreaders (40 male and 40 female) from the BBC Learning English website (2009) who spoke non-rhotic native English (RP and other accents) with a normal or careful speaking rhythm. The instruments were a corpus, two questionnaires as well as an editing and recording audio program. The corpus employed was the News Archives of the freely available *BBC Learning English website* from the years 1999–2008. It consisted of transcribed news texts and their corresponding audio files. The questionnaires used were grouped into two types: *pre-task* and *task* questionnaires, depending on when they were filled in. The pre-task questionnaire consisted of four questions, some open-ended and others closed-ended, and it took about five minutes to complete. It was intended to obtain background information about the potential listeners, and the answers elicited by it served as a basis for our choice of three of them. All the possible listeners were to answer the same questions. The task questionnaire consisted of a closed-ended question in reference to ten words. The listeners answered one task-questionnaire per newsreader, and the time estimated to complete each questionnaire was about 10 minutes (see Appendix 1). Some software was also made use of. A recording, editing and mixing audio program (Audacity, [http://audacity.sourceforge.net/]), was employed to record the material that the listeners were to hear.

The study was conducted during a period of four months. From Mompeán's data collection of the corpus, 80 speakers were selected from those who met the conditions previously mentioned (quota non-probability sampling strategy). After checking text and audio, 800 words (10 per newsreader), representative of Monroy's rules, were chosen by focusing on the last syllable. Each word and its context were written and recorded (by means of Audacity). By considering the answers to the pre-task questionnaire, in order to avoid bias (as we knew the rules), three listeners not knowledgeable about the rules were chosen (triangulation contributes to validity). They had homogeneous characteristics: female, young, British, educated and with a good ear. They differed in their knowledge of phonetics and, to an extent, in their accent. They were emailed the task questionnaire for each speaker and the audio files (including words and context). Then all the information was transcribed into an Excel document.

5. Data analysis

Once the data had been collected, we proceeded to their treatment and analysis. The information from the variables (see Appendix 2 for a detailed account of the variables) was introduced in the data matrices from the Excel document in order to be captured later by the statistical package SPSS version 19.0.0 for Windows (SPSS, Inc., an IBM Company, 1989, 2010). Despite there being a more recent version, the University of Murcia holds a licence of this edition.

After recoding certain data, the information was subjected to statistical analysis, for which the procedures employed were as follows:

-Univariate analysis using techniques such as frequency and percentage counts; techniques of graphical nature, in particular, bar and pie charts.

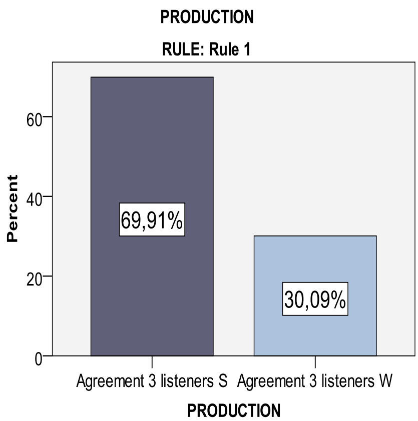
-Bivariate analysis techniques: contingency tables were drawn up.

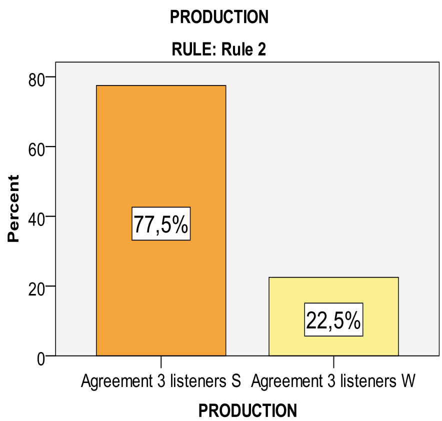
-Multivariate analysis techniques: Kendall's Coefficient of Concordance (W) was used. It measured the agreement amongst listeners. A null hypothesis needed to be tested: the referees were independent and they employed different assessment criteria. Rejecting the null hypothesis implied accepting some link amongst the listeners' assessment. A 95% confidence interval was assumed and the margin of error was of 5%, in other words, $\alpha = 0.05$ (typically used in Social and Human Sciences). Due to the low frequency of occurrence in certain categories of the variables, some of them were grouped on the basis of similarities of category. In some relevant cases in which grouping w as not possible, some visual inspection was undertaken.

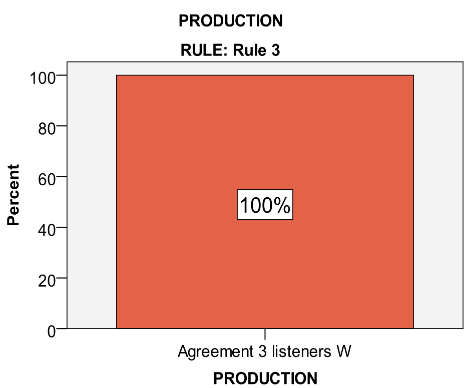
6. Results and discussion

The production found in our selection of the corpus will be explored in view of those rules formulated by Monroy (2008-2009). The aim is to find out more about the extent to which these theoretical rules are *applied* in real-life speech, that is, whether what is said by the rules constitutes what is pronounced by the informants in the corpus (see Figure 1). The aforementioned phonetician holds the view that in Rules 1 and 4 a syllabic consonant is produced. Our results are consistent with Monroy's formulation regarding these two rules. In fact, there is a high level of agreement relating to the syllabic consonant – 69.9% and 58.9%, respectively, but regarding the schwa to the listeners' ears there is also a remarkable percentage: around 30% and 40%. As for Rules 2 and 3, it is clear that the production in the latter follows Monroy's claim of there being a schwa (100% agreement in our results). It seems that the listeners all agree only when they think the schwa is produced, but they are uncertain as to the existence of a syllabic consonant. In contrast, the production in Rule 2 does not match what the phonetician sustains. Actually, while this scholar maintains that the schwa is to be pronounced here, the results show there is a high rate of syllabic consonants (almost 8 out of 10 cases) –even higher than that in Rules 1 and

4. In any case, we need to be cautious here, if we bear Rule 5 (a conflation of Rules 2 and 3) in mind. The percentage of syllabic consonants in Rule 2 is, in any case, surprisingly large (68.132%, if Rule 5 is added to Rule 2).







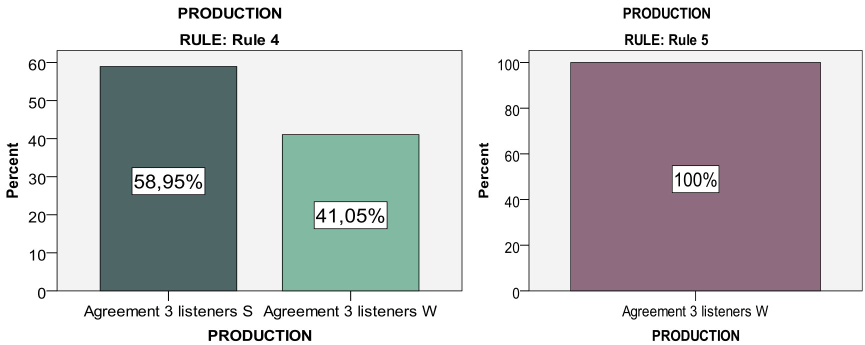
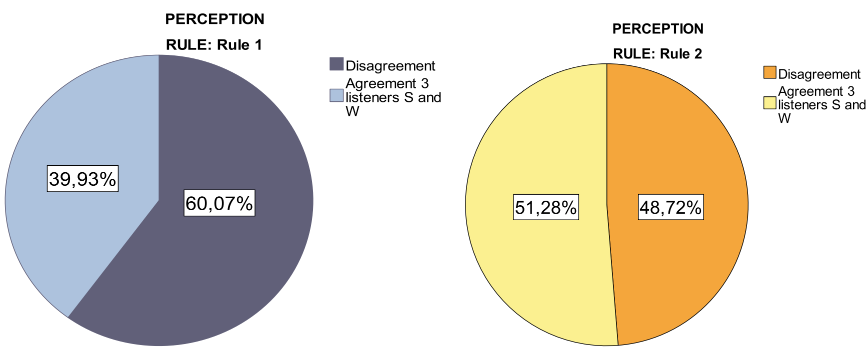


Figure 1. Agreement on S and W amongst the listeners in terms of rule (%)

Now the listeners' level of agreement (on schwa or syllabic consonant) and disagreement will be studied in terms of the rules above. It is observed that there is less agreement among the listeners in terms of Rules 1 and 4 (almost 40% of agreement) than Rules 2 and 3 (more than half of the cases are of accordance), as shown in Figure 2. This difference may be due to the fact that Rules 1 and 4 cover a wider spectrum of sequences in which larger discrepancy may arise. The case of Rule 5 is more striking, as there is a clear tendency towards discrepancy; the level of agreement here is very low (15.5%), as opposed to Rule 3, from which we would have expected a similar behaviour in terms of agreement, because there are nasals in between, the only difference being that in Rule 5 the nasal preceding is /m/, not /n/ as in Rule 3.



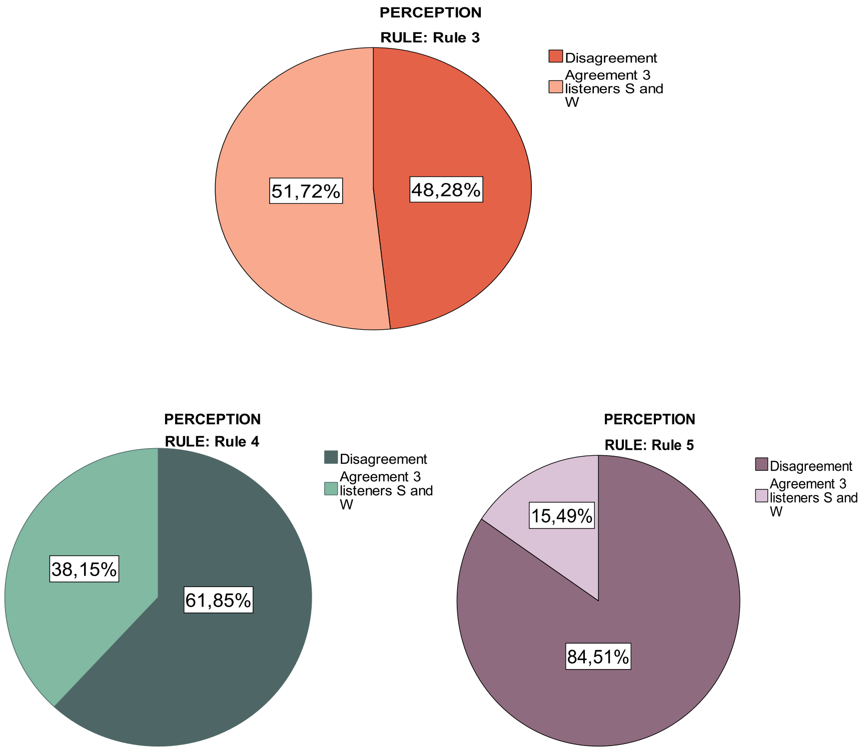


Figure 2. Disagreement and agreement amongst the listeners in terms of rule (%)

If trying to ascertain whether the listeners are independent in their assessment or they assess on the same basis in view of the rules at issue, Kendall's Coefficient of Concordance (W) indicates that for the five categories corresponding to the five rules, the null hypothesis of assessment independence among the listeners is rejected, and it can be claimed that there is statistically significant agreement amongst the listeners in this variable. We have W (df = 2) = 0.159 for Rule 1, W (df = 2) = 0.071 in the case of Rule 2, W (df = 2) = 0.207 for Rule 3, W (df = 2) = 0.192 in terms of Rule 4 and W (df = 2) = 0.476 for Rule 5, p being < 0.01 (see Table 1).

Test Statistics

Rule 1	N	281
	Kendall's W[a]	,159
	Chi-Square	89,298
	df	2
	Asymp. Sig.	,000
Rule 2	N	151
	Kendall's W[a]	,071
	Chi-Square	21,493
	df	2
	Asymp. Sig.	,000
Rule 3	N	29
	Kendall's W[a]	,207
	Chi-Square	12,000
	df	2
	Asymp. Sig.	,002
Rule 4	N	248
	Kendall's W[a]	,192
	Chi-Square	95,386
	df	2
	Asymp. Sig.	,000
Rule 5	N	71
	Kendall's W[a]	,476
	Chi-Square	67,633
	df	2
	Asymp. Sig.	,000

a. Kendall's Coefficient of Concordance

Table 1. Kendall's Coefficient of Concordance amongst the listeners in terms of rule

7. Conclusions

In conclusion, our results for production indicate that in words fitting into Monroy's Rules 1, 3 and 4 (2008-2009) this phonetician's claim of there being a syllabic consonant (Rule 1: almost 70% and 4: almost 60%) and a schwa (Rule 3: 100%) is followed to a great extent. In contrast, in Rule 2 the production results in a greater occurrence of syllabic consonants, as opposed to Monroy's claim for a schwa. As for the possible relationship between the rule and the listeners' level of agreement and disagreement, our results reveal that there is wider disagreement for Rules 1 and 4 (around 6 out of 10 cases), maybe due to their wider spectrum of sequences, which causes larger discrepancies. More striking are the findings of Rule 5 (a sequence in which /m/ precedes a potential schwa), which has only 15.5% agreement but would have been expected to behave like Rule 3. Kendall's Coefficient of Concordance reflects the fact that for the variable rule, the listeners are not independent in their assessment.

An increase in the amount of information for those data which have a low frequency in this study is also suggested because, obviously, the results would be more reliable. Statistical analyses should be enriched by means of other tests of inferential nature. The effect of the phonemic context surrounding the weak vowel (syllabic consonant or schwa) or the final consonant on the realization of schwa vs. syllabic consonant is also suggested to be examined on a perceptual basis.

REFERENCES

Abel, James W. (1962): "Syllabic [m, ŋ] ". *American Speech* 37/2: 106-113.

Cruttenden, Alan (2001): *Gimson's pronunciation of English*. London, UK: Arnold.

Alvarez, Juan Antonio (1980): "Syllabic and non-syllabic /l/ and /n/". *Atlantis* 2/1: 41-48.

Arboleda, Inmaculada de Jesús & Monroy, Rafael (2010): "Gender and accent in the perception of English syllabic consonants". R*evista de Lingüística y Lenguas Aplicadas* 5: 17-28. doi: http://dx.doi.org/10.4995/rlyla.2010.750

Arboleda, Inmaculada de Jesús (2012): "Do people selectively interpret their schwa vs. syllabic consonant perception on the basis of their background, experience and expectations?" *Revista de Lingüística y Lenguas Aplicadas 7*/1: 23-36. doi: http://dx.doi.org/10.4995/rlyla.2012.1122

Auran, Cyril & Bouzon, Caroline (2003) : "Phonotactique prédictive et alignement automatique: application au corpus MARSEC et perspectives". *In : Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA), 22*. 33-63.

Auran, Cyril, Bouzon, Caroline and Hirst, Daniel (2004): "The Aix-MARSEC project: an evolutive database of spoken English". *In: Proceedings of the Second International Conference on Speech Prosody*. 561-564.

BBC learning English website (2009): <www.bbc.co.uk/worldservice /learningenglish/index.shtml> [downloaded 02/01/2012]

Bhatia, Tej. K. (1973): "A computational investigation on the perception and acquisition of aspiration". *Proceedings of the 5th Conference on Computational Linguistics: Vol. 2.* Pisa, Italy, 33-35.

Becker, Thomas (1998): *Das Vokalsystem der Deutschen Standardsprache*. Frankfurt am Main, Germany: Peter Lang.

van Bergem, Dick (1995): "Perceptual and acoustic aspects of lexical vowel reduction, a sound change in progress". *Speech Communication* 16/4: 329-358.

Booij, Geert E (1983): "French C/0 alternations, extrasyllabicity and lexical phonology". *The Linguistic Review* 3: 102-134.

Cohen, Antonie (1957): *The phonemes of English*. The Hague, Netherlands: Mouton.

Cruttenden, Alan (2001): *Gimson's pronunciation of English*. London, UK: Arnold.

Cruttenden, Carney, Edward (1994): *Survey of English spelling*. London, UK: Routledge.

Deschamps, Alain, Fournier, Jean-Michel, Duchet, Jean-Louis & O'Neil, Michael (2004) : *English phonology and graphophonemics*. Paris, France: Ophrys.

Deterding, David (2006): “Deletion of final /t/ and /d/ in BBC English: Implications for teachers in Singapore”. *STETS Language & Communication Review* 5/1: 21-24.

García, Érica (2006): *The perception of lateral versus nasal syllabic resonants*. Haskins Laboratories, New York City. <http://www.haskins.yale.edu/SR/SR013/SR013_06.pdf> [downloaded 20/03/2012].

Giegerich, Heinz J. (1992): *English phonology.* Cambridge, UK: Cambridge University Press.

Gick, Bryan (2002): “An X-ray investigation of pharyngeal constriction in American English schwa”. *Phonetica* 59: 38–48.

Gimson, Alfred Charles (1970): *An introduction to the pronunciation of English*. London, UK: University of London Press.

Gussmann, Edmund (1991): “Schwa and syllabic sonorants in a non-linear phonology of English”. *Anglica Wratislaviensia*: 25-39.

Heselwood, Barry (2007): “Schwa and the phonotactics of RP English”. *Transactions of the Philological Society* 105/2: 148-187.

Jones, Daniel (1976) [1918]: *An outline of English phonetics*. Cambridge, UK: Cambridge University Press.

Jones, Daniel (1992) [1909]: *The pronunciation of English*. Cambridge, UK: Cambridge University Press.

Kager, René (1989): *A metrical theory of stress and destressing in English and Dutch*. Dordrecht, Netherlands: Foris publication.

Kenyon, John S. (1956): “Syllabic consonants in dictionaries”. *American Speech* 31/4: 243-251.

Lehiste, Ilse (1964): “Acoustical characteristics of selected English consonants”. *International Journal of American Linguistics* 30/3: 194-197.

MacCarthy, Peter (1965): *A practice book of English speech*. London, UK: Oxford University Press.

Mompeán, Pilar & Mompeán, José Antonio (2006): *Intrusive /r/ usage in RP: The case of BBC newsreaders*. Paper presented at 5th International Conference of the Spanish Cognitive Linguistics, Murcia, Spain.

Mompeán, Pilar & Mompeán, José Antonio (2007): “Phonetic factors in /r/-liaison usage: A first report”. In: Trouvain, Jürgen & Barry, William J. (eds.). *Proceedings of the 16th International Congress of Phonetic Sciences*. Universität des Saarlandes, Saarbrücken, Germany, 1397-1400.

Monroy, Rafael (1980): *La pronunciación del inglés RP para hablantes de español*. Madrid, Spain: Paraninfo.

Mora, Joan Carles (2003): “The formation of syllabic consonants and their distribution in Southern British English”. *Atlantis* 25/2: 97-112.

Mora, Joan Carles & Mott, Brian (2003): *Length as an acoustic cue in the perception of consonant syllabicity*. Paper presented at the 15th International Congress of Phonetic Sciences. Barcelona, Spain.

Monroy, Rafael (2008-2009): *Fonética inglesa* [Class Handout]. Murcia, Spain: UMU.

van Oostendorp, Marc (1995): *Vowel quantity and phonological projection* (Unpublished Doctoral Dissertation). Tilburg University, Netherlands.

van Oostendorp, Marc (1998): “Schwa in phonological theory”. *GLOT International* 3: 3-8.

van Oostendorp, Marc (2000): *Phonological projection: A theory of feature content and prosodic structure*. Berlin, Germany: Mouton de Gruyter.

Reese, Hayne W. & Lipsitt, Lewis P. (eds.) (1979): *Advances in child development and behaviour: Vol. 13.* New York, NY: Elsevier.

Roach, Peter (1986): *English Phonetics and Phonology: A Practical Course*. Cambridge, UK: Cambridge University Press.

Roach, Peter (2000): *English phonetics and phonology: A practical course*. Cambridge, UK: Cambridge University Press.

Roach, Peter (2002): *A little encyclopedia of phonetics.* Cambridge, UK: Cambridge University Press.

Schaeffer, Natalie & Eichorn, Nechama (2001): “The effects of differential vowel prolongations on perceptions of speech naturalness”. *Journal of Fluency Disorders*, 26 (4): 335-348.

Sourceforge (2013): *Audacity*. <http://audacity.sourceforge.net/> [down loaded 02/04/2012]

SPSS, Inc., an IBM company (1989, 2010): *IBM SPSS statistics version 19.0.0..* Chicago, IL: SPSS, Inc., an IBM company.

Trask, Robert Lawrence (1996): *Dictionary of phonetics and phonology*. London, UK: Routledge.

Wells, John Christopher (1965): “The phonological status of syllabic consonants in English RP”. *Phonetica* 13: 110-113.

Wells, John Christopher (1995): “New syllabic consonants in English”. In*:* Windsor Lewis, Jack (ed.). *Studies in general and English phonetics: Essays in honour of Professor J. D. O'Connor.* London, UK: Routledge, 401-412.

Wiese, Richard (2000): *The phonology of German*. Oxford, UK: Oxford University Press.

Windsor Lewis, Jack (1990): “Twentieth-century changes in British pronunciation”. *Windsor Lewis*’ web page*.* <http://www.yek.me.uk/ changestwe.-html> [downloaded 05/05/2012].

Acknowledgement

I am sincerely grateful to the listeners in the study: Sonya Ross, Stephanie Bremner and Charlotte Sarah Walker, for their kind collaboration. Very special words go to Prof. Monroy, for his valuable advice and Dr. Francisca J. Serrano, for her statistical support. Thanks as well to Prof. Upton, Prof. Wells and Dr. House. We are also indebted to a research grant from the *Fundación Séneca*, which has allowed us to delve into this topic.

APPENDICES

Appendix 1

A.

PRE-TASK QUESTIONNAIRE

1) What do you do in life?
2) What accent do you have?
3) Do you have any knowledge of English phonetics? If so, please tell me a bit about your experience in the area.
4) Do you have a good ear for music? And for sounds? Tell me a bit about it.

B.

TASK QUESTIONNAIRE
SPEAKER 15: FEMALE 15 (LESLEY CURWEN)

- LISTENER F1

1) Which one do you hear in the last syllable of each word, a) a syllabic consonant (S) or b) a schwa (W)?
troubles (Thai Airways' troubles are mounting.) a
internal (...and plagued by internal labour disputes.) a
fallen (...- it's fallen by thirty percent...,) b
global (...the global economy is slowing down.) a
question (The question is, how to...) b
barrels (...of 1.5 million barrels a day, if non-0PEC nations...) a
action (...to discuss possible action.) b
question (The question is –was it...) b
driven (...to insider trading, driven by false and misleading financial statements.) a
assemble (…difficult to assemble a criminal...) a

Appendix 2

VARIABLES (qualitative because the scale used was nominal)

- PREDICTOR, GROUPING OR INDEPENDENT VARIABLES

RULE (for the rest of rules see section 2.1.1)
EXTRA RULES (THEY ARE NOT MONROY'S. THEY INCLUDE MIXTURES OF SEVERAL OF HIS RULES AND EXCEPTIONAL CASES)
↓
RULE 5: words which belong to two rules at the same time
RULE 6: words which belong either to one rule or another one

- CRITERION OR DEPENDENT VARIABLES

PRODUCTION

Agreeement 3 listeners S (syllabic consonant)
Agreement 3 listeners W (schwa)

PERCEPTION
Disagreement
Agreement 3 listeners S and W

ANSW. (ANSWER) L.F1 (Listener -Female- 1):

No /l/
S (syllabic consonant)
TW (tiny schwa)
W (schwa)

ANSW. (ANSWER) L.F2 (Listener -Female- 2):

No /l/
S (syllabic consonant)
TW (tiny schwa)
W (schwa)

ANSW. (ANSWER) L.F3 (Listener –Female 3):

No /l/
S (syllabic consonant)
TW (tiny schwa)
W (schwa)

--GROUPED

RULE

Rule 6 (only visual inspection)

ANSW. (ANSWER) L.F1 (Listener -Female- 1):

S (syllabic consonant)
W (schwa)

ANSW. (ANSWER) L.F2 (Listener -Female- 2):

S (syllabic consonant)
W (schwa)

ANSW. (ANSWER) L.F3 (Listener -Female- 3):

S (syllabic consonant)
W (schwa)

CHAPTER 3. FRICATIVES REVISITED

Silvia C. Barreiro Bilbao
UNED- Madrid

1. Introduction

The process of language acquisition involves speech production as well as speech perception, with a complex relationship between these two (Llisterri, 1995; Newman, 2003). Leaving aside the theoretical discussion concerning the nature of the connection between production and perception, in this chapter we will focus on the perceptual aspect and, especially, on a concept that is crucial to the interaction of phonological processes and perception, the *cue.* According to Wright (2001), “the information provided by the acoustic cues contained in the signal may allow the listener to apprehend the existence of a phonological contrast”. In other words, as recently defined by McMurray & Jongman (2011), a *cue* is “a specific measurable property of the speech signal that can potentially be used to identify a useful characteristic like the phoneme category or the talker”. This theory-neutral definition allows us to avoid the current debate, tackled by speech perception theories, on how to solve the problem of the lack of invariant cues in the signal for categorical distinctions (McMurray & Jongman, 2011). This, therefore, will be our definition as well.

Descriptions of the acoustic characteristics of speech sounds started to emerge as soon as speech technology provided the relevant equipment, such as the sonograph. As early as 1947, Potter, Kopp and Green presented their classification of sounds, including fricatives, based on their analysis of spectrograms. Since then, there has been a huge amount of work done both on the articulatory and perceptual description of acoustic cues for fricatives in isolation as well as in connected speech. It is far beyond the scope of the present work to explain in detail all the previous studies on fricatives, in particular, on English fricatives. The main core of this chapter will be devoted to the presentation of a comprehensive peer-reviewed research on the acoustic characterization of English fricatives in terms of place of articulation and voicing, and, to a lesser extent, to manner of articulation –based on the most significant, and recent, publications in the field. As a secondary goal, we will point out those disputed areas in which more investigation is clearly necessary and which, therefore, are the subject matter of current research.

Let us start by saying that the acoustic properties that are reported to influence fricative perception can be categorised as spectral, temporal and amplitudinal (or power) properties (Silbert & de Jong, 2008), and as being of two kinds (Steriade, 1997). On the one hand, those internal cues that occur during the articulation of the fricative itself, which would include the intrinsic spectral properties of the *frication noise*, the *frication duration* and the *overall frication amplitude*. These cues do not depend entirely on the phonetic context. On the other hand, the external cues that occur in adjacent elements as a result of coarticulation, which would comprise the *relative amplitude* between the frication and the abutting vowels, the *relative duration* between the consonant and the adjacent vowel, and the *formant transitions*. These cues depend on the phonetic environment, and, therefore, are highly variable across contexts.

Broadly speaking, all these defining properties are either static or dynamic by nature. The former pertain to acoustic information that is measured at one location of the speech signal, whereas the latter concern changes in acoustic information during the fricative and/or adjacent elements (Jognman *et alii*., 2000).

Our analysis of the different acoustic cues associated with fricative perception in English will include their perceptual weighting in native fricative contrasts as well as their interactions with certain variables (i.e., prosodic context, vowel quality, speaker and/or gender). Information on fricative articulation will be provided merely for a better understanding of the acoustic side. All of these cues will be explained in depth in the following sections, discussing the fricative manner of articulation, followed by place and voicing.

2. Manner of articulation: Articulatory description and acoustic cues

English fricatives, as well as stops, are *obstruent* sounds, that is, during their production there is a closure, partial or total, of the vocal tract, stopping or interfering with the airflow coming from the lungs. This occlusion or constriction generates noise, either transient or turbulent, depending on the manner of articulation. In the case of fricatives, turbulence is created (Shadle, 1990) either:

> (1) In a constriction, when a rapid flow of air goes through a very narrow constriction in the supraglottal cavities, whose position depends on the particular fricative. Constriction is made by the approaching of two organs (teeth, tongue, lips and/or palate) close enough to produce friction without completely stopping the airflow.
>
> (2) At (or around) an obstacle, that is still and sharp (such as the upper teeth), which is hit by the high-velocity air jet created by a narrow constriction.

As sounds with a noise component, with the exception of /h/, most of the acoustic energy of obstruent sounds occurs at higher frequencies than those with harmonic structure (such as vowels, approximants or nasals). Therefore, listeners have to differentiate among these aperiodic sounds using various acoustic cues, including the duration of the noise component, the energy distribution along the frequency scale, or the presence or absence of silence (Gil, 2010). Fricatives, as opposed to other sounds such as stops, are defined as *continuous* sounds since they require a non-stop air flow. This articulatory feature results acoustically in the presence of noise for a relatively long period, the main acoustic cue to fricative perception of manner of articulation. Jongman (1985) suggested a minimum duration of about 20 ms was required in order to perceive fricative manner of articulation.

According to the Acoustic Theory of Speech Production (Fant, 1960), any speech sound that emerges at the end of the vocal tract is the result of the combination of a *source* and a *filter*. In the case of English fricatives (Hugues & Halle, 1956; Fant, 1960 & 1973; Fant & Mártony, 1960; Heinz, 1961; Heinz & Stevens, 1961; Stevens et *alii.*, 1968; Stevens, 1971; Scully & Allwood, 1985; Shadle 1985 & 1990; Badin, 1989; Scully *et alii.*, 1992), the source (i.e. the turbulence) is located in the vocal tract, in particular, at the place where the constriction occurs. Fricatives can have a secondary source if the vocal folds vibrate at the same time as the turbulence is being generated, as in the voiced fricatives.

The narrow constriction, which generates noise, limits the amount of airflow (or energy) coming from the back cavity to the front. Consequently, the final spectral shape (or overall contour) is mainly determined by the response of the cavity in front of the constriction. In fact, the front cavity is the most effective resonator, although the presence of an obstacle in this cavity seems to affect fricative resonant characteristics too. Both the constriction and the front cavity function as the filter.

In sum, the fricative spectrum is the result of the resonant responses of the cavities above the larynx (filter) that shape the sound spectrum generated by turbulence (source). In the case of voiced sounds, the aperiodic noise is modulated by the pattern of periodically-interrupted air-flow caused by the vibration of the vocal folds.

An exception to this is the case of the glottal fricative /h/. Firstly, the aperiodic source is not generated in the vocal tract but at the glottis where the vocal folds get close enough to produce turbulent noise without vibration. As for the filter, once again /h/ does not fit neatly into the pattern of the other fricatives. The vocal tract takes the shape of whatever vowel is to follow, which has caused some linguists, mainly phoneticians, to characterize /h/ as a voiceless vowel (Cruttenden, 2008; Raphael *et alii.*, 2011). In fact, its spectral pattern, with relatively low intensity, can be defined as a mirror of the formant structure of the following vowels (Lehiste, 1964; Barreiro, 1994). It is usually voiceless, although it can be voiced, especially if followed by a stressed syllable, when it occurs between voiced sounds: "*Aparece un componente de baja frecuencia (sobre 500 Hz), […], especialmente cuando la consonante se halla en posición intervocálica*" (Barreiro, 1994: 295). This voicing is similar to, but not exactly the same as the normal voicing of vowels, being "a weak, slightly fricative sound called breathy voice" (Roach, 2009: 52). Therefore, as phonetically the consonant /h/ can be considered a voiceless vowel with the quality of the voiced vowel that comes after, it has been of no interest to most research on the acoustic cues used for the perception of English fricatives, and thus references to this sound in this chapter will be scarce.

3. Place of articulation: Articulatory description and acoustic cues

English fricatives are usually divided into five classes according to their place of articulation: labiodental /f, v/, (inter-)dental /θ, ð/, alveolar /s, z/, palato-alveolar /ʃ, ʒ/, and glottal /h/. Based on the presence or absence of an obstacle during their articulation, English fricatives can also be grouped into another major division, that is, into sibilants /s, z, ʃ, ʒ/ and non-sibilants /f, v, θ, ð, h/. In the former, the turbulent air created by a narrow constriction hits the upper teeth and, as a result, creates a louder, higher-pitched noise than non-sibilants.

A large number of previous studies aimed at determining the acoustic cues (and their relative weighting) used by the listeners to identify 'place of articulation' in fricatives have been mainly devoted to the intrinsic properties of the frication itself mentioned above. In the late eighties, however, research studies extended the analysis. In an attempt to solve the issue of mapping from continuous acoustic cues to categories, statistical measurements (such as spectral moments and locus equations) were included in order to ascertain if the cues are homogeneously distributed throughout the fricatives, or if there are some regions that provide more information than others.

The main acoustic cues for the perception of place of articulation in fricatives are the spectral properties of the *frication noise*, the *frication amplitude*, and, to a lesser extent, *frication duration* and *formant transitions.* Other properties of the acoustic signal, with contradictory, inconsistent, or not yet fully understood results, will be analysed in the section called *other cues.*

The following sections will cover the review of previous research on this topic, showing firstly, that place (as well as voicing) distinctions in fricatives can be perceived by using multiple cues; and secondly, that there is significant *speaker* variation in the use of English fricative perceptual acoustic cues, as already observed by Bush in 1964. Furthermore, reports will show a clear interaction between place (as well as voicing) cues and other variables, such as *vowel quality, prosodic context* or even *gender*. All this manifest variability may account for the contradictory results shown by previous literature.

Finally, all these studies used laboratory speech in which fricatives appear in nonsense words either in CV, VC or VCV syllables, which were occasionally embedded in a carrier sentence.

3.1. Frication noise

The most enlightening spectral properties of the frication noise related to fricative place of articulation include the *overall spectrum contour* and the *location of spectral peaks*. *Spectral moments* will be described in *other cues*.

3.1.1. Overall spectrum contour

Differences in place of articulation have strong effects on fricative spectrum (Harris, 1956; Strevens, 1960; Delattre 1966; Jassem, 1968; Fant 1973; Behrens & Blumstein, 1988a; Shadle & Mair, 1996), largely because the noise source mechanism changes drastically with place (Silbert & de Jong, 2008). Labiodental and dental frication noise is produced by airflow going through a narrow constriction, whereas alveolar frication noise is also produced by the channel of the air stream into the lower teeth (Shadle, 1990 & 1991). Consequently, non-sibilants (labiodentals and dentals) show a low level of energy spread over a broad frequency band starting below 2 kHz (around 1500-1700 Hz, being a bit higher for dentals), while the sibilants (alveolars and palato-alveolars) exhibit their high-energy noise over a narrower band starting at higher frequencies. The narrow constriction for sibilants prevents the back cavity resonances from being salient, causing the energy below 2 kHz to be attenuated by the back cavity anti-resonances.

As explained above, the shape and size of the cavity in front of the constriction mostly determines the final spectral shape. Thus, those fricatives with a rather short front cavity –namely, labiodentals and dentals–

show a relatively flat spectrum with no clear spectral peaks in any frequency region, whereas those fricatives with a longer front cavity –that is, alveolars and palato-alveolars– exhibit distinct spectral peaks at frequency regions above 2 kHz and 4 kHz, respectively. Therefore, this spectral difference divides English fricatives into two general categories of place: anterior (non-sibilants) and posterior (sibilants) (Raphael *et alii*, 2011).

It is fair to say, however, that a few studies analysing the spectrum of non-sibilants have shown that it is not as flat as reported, but with a difference of about 20 dB across the frequency range band when voiceless, and a wider range when voiced (Shadle & Mair, 1996: 194). Furthermore, there can be slight differences between them in the spectral shape in the transition regions: "[f] tends to have a lumpier spectrum, with noticeably higher amplitude at low frequencies than [θ]" (Shadle & Mair, 1996: 195). It has even been suggested that the relationship between the spectral energy around 3 kHz and above 6 kHz may account for the /f-θ/ distinction: the spectrum of the labiodental is characterised by a marked concentration of energy at low frequencies (below 3 kHz) whereas that of the dental shows spectral energy both below 3 kHz and around 6 kHz (Feijóo & Fernandez, 2002). Whether the spectrum is flat or peaked, spectral energy below 6 kHz seems to contain the most relevant acoustic cues for the perceptual identification of fricative place of articulation (Feijoó & Fernández, 2002).

As far as the effects of context are concerned, it is clear that the *quality* of the adjacent *vowel* has some effect (great in the case of rounded and high vowels, Whalen, 1983) on the articulatory and spectral characteristics of fricatives (LaRiviere *et alii*., 1975; Shadle & al. 1996), especially for non-sibilants (Jesus & Shadle, 2002). Contradictory results have been reported regarding the influence of vowel quality on listeners' perception of voiceless sibilants. Some studies claim that the presence of round vowels biased listeners towards perception of the alveolar (Kunisaki & Fujisaki, 1977), but others state the opposite effect, that is, round vowels increase the perception towards the palato-alveolar (Nittrouer & Studdert-Kennedy, 1987). This is a clear example of the individual variation in the use of perceptual cues (Casserly, 2010). Furthermore, vowel context may aid identification of the *speaker* from the speech signal (McMurray & Jongman, 2008), and even *gender* although not always with reliable results (Bolt *et alii*, 1973).

3.1.2. Spectral peak locations

The location of the spectral peaks in fricatives has been widely analysed in previous research studies (K.S. Harris, 1954; Whalen, 1991; Pickett, 1980; Jongman, 1985; Behrens & Blumstein, 1988a; Jognman *et alii*, 2000; Kent *et alii,* 2002; Raphael *et alii*, 2011, among others).
English sibilants can be clearly distinguished by the spectral peaks, in particular, by the frequency and the amplitude of the lowest peak, i.e., the spectrum shows a peak around 4-5 kHz for English alveolars and around 2.5-3 kHz for English palato-alveolars (Abdelatty Ali *et alii*, 2001). A farther back production together with lip rounding and protruding seem to account for the lower frequencies associated with palato-alveolar sounds.
The spectra of non-sibilants contain distinctive spectral peaks, although they are very similar due to the dental constriction that characterises them. However, these peaks are not predominant compared to the well-defined peaks shown for sibilants. It is, therefore, hard to find reliable spectral differences between labiodentals and dentals (Tabain, 1998). Some studies, however, (Jassem 1965; Shadle *et alii,* 1996; Jongman 2000) have reported that the location of the most prominent spectral peak could be a useful cue that may help distinguish between them, located at around 7700 kHz for labiodentals, and around 7400 kHz for dentals. In sum, it can be claimed that spectral peak locations can differentiate all four fricative places of articulation in English, although this is not always statistically significant in the case of non-sibilants.
Regarding to the interaction with context variables, spectral peak location is noticeably influenced by the quality of the adjacent vowel and by the speaker who utters the consonantal sound. In other words, it is vowel-dependent (Carney & Moll, 1971; LaRiviere *et alii,* 1975; Soli, 1981; Yeni-Komshian & Soli, 1981; Whalen, 1983; Behrens & Blumstein, 1988a; Maniwa *et alii.*, 2008) as well as speaker-dependent (Hughes & Halle, 1956; Behrens & Blumstein, 1988a, McMurray & Jongman, 2008; Maniwa *et alii*, 2009;

McMurray & Jongman, 2011). Furthermore, it has been reported that the actual location of the spectral peaks differs across syllable position (Seitz el al., 1987), with higher values in initial position than in mid and final positions.

In general, *vowel context* mainly affects the frequencies of those spectral peaks in the fricative spectra that are associated with the F2 of the adjacent vowel, that is, it occurs between 1.5 and 2 kHz are around 100-300 Hz higher before front vowels than before back vowels (Mann & Repp, 1980; Soli, 1981; Yeni-Komshian & Soli, 1981; Shadle *et alii*, 1996, Seitz *et alii*., 1987).

There can be slight differences between labiodentals and dentals in the spectral shape, showing [f] a broad peak in the region 6-12 kHz in some [vowel] contexts (Shadle & Mair, 1996), however, Jongman (1989), and a more recent research study by Jongman *et alii,* (2000), showed that spectral peaks were significantly lower only for the alveolar fricatives. The physiological reason is evident; vowel influence is greater upon those articulatory regions that do not intervene in the formation of a fricative constriction. In those fricatives in which the tongue is not the main articulator, coarticulatory effects are observed in the whole tongue whereas in fricatives in which the tip of the tongue is used as the main articulator, the main effects are seen in the back of the tongue (Carney & Moll, 1971; Recasens, 1990). Furthermore, close vowels leave more clear articulatory effects than central or open vowels: “since the narrow constriction necessary for producing [i] and [u] is close to that needed for the fricatives, the two gestures can affect each other more easily than with the more open [o] and [a]” (Whalen, 1983: 91).

In relation to the *speaker*-dependent effects, results from the literature show that the fricative spectrum varies widely across speakers, as mentioned above, and clear intra-speaker differences have also been reported (Seitz *et alii*, 1987). Furthermore, a few studies have claimed that it is possible to identify speaker *gender* from the spectral analysis of fricatives, including /h/ (Schwartz, 1968; Ingerman, 1968; Repp & Mann, 1980; Seitz *et alii,* 1987; Jongman *et alii*, 2000; Silbert & de Jong, 2008). Gender differences in the location of the spectral cues are due not only to the differences between men and women regarding the shape and size of the vocal tract, “but also to a larger ratio of pharyngeal cavity to oral cavity for adult males as compared to women…" (Raphael *et alii*, 2011:110). In this regard, spectral peaks tend to occur in higher frequencies for sibilants when the speaker is a female than when the speaker is male (Jongman *et alii*, 2000). For instance, the spectral peak for /s/ occurs around 5,5 kHz for male speakers while for females it occurs around 6,7 kHz (Yeni-Komshian & Soli, 1981), consistent with a shorter resonator for the female speakers (Silbert & de Jong 2008). The lack of well-defined peaks in the spectrum of the non-sibilants also makes it impossible to perceive gender distinctions (Schwartz, 1968).

3.1.3. Other cues: Spectral moments

The great variability in frication noise makes it very difficult to provide conclusive results. Only a few cues can clearly be attributed to one feature (place, voicing or sibilance) over others, and none can be associated with a single feature (Jongman & McMurray, 2011). Therefore, since the late eighties, and based on statistical models of speech perception, different techniques have been used (and have proved useful) to investigate other relevant differences in fricative spectra in relation to place of articulation. A statistical method commonly used to analysis the spectral characteristics of frication noise is the *spectral moments analysis* in which the power spectrum is considered as a probability distribution in order to calculate different mathematical moments.

Foster *et alii,* (1988) pioneered its use to study spectral differences in voiceless obstruents with clear results for stops but less reliable for fricatives. Tomiak (1990) reported differences among fricatives, although they were not very consistent in the case of non-sibilants. A few years later, the study by Shadle & Mair (1996), and especially that of Jongman *et alii,* (2000) extended the use of the spectral moments analysis to a larger set of place of articulation distinctions in English fricatives. These works were followed by other revealing studies (Jesus & Shadle, 2002; Maniwa *et alii*, 2008; Silbert & de Jong, 2008; Fangfang *et alii*, 2009, to mention a few).

The statistical analysis of spectral moments, based on one or multiple regions of the speech signal, calculates a series of FFT (Fast Fourier

Transform) every certain amount of time (10 ms, for instance) from the onset of the sound (Jongman *et alii.*, 2000). For each FFT *mean, frequency* (first moment), *variance* (second moment), *skewness* (third moment) and *kurtosis* (fourth moment) are computed, taken as measures of the *central tendency* (the mean or average energy concentration), *variation* (or energy range), *symmetry* (or distribution asymmetry) and *peakedness of the spectrum*, respectively (Silbert & de Jong, 2008).

The first spectral moment in the case of sibilants is negatively correlated with the length of the front cavity and therefore roughly describes where the constriction is made relative to the length of the oral cavity (Fangfang *et alii,* 2009). Most studies, concentrating on the spectral mean, have been able to perceive distinctions between the different sibilants, showing that palato-alveolars have a lower mean than alveolars (Tomiak, 1990; Nittrouer, 1995). The non-sibilants' spectral mean falls in between (Jongman *et alii*, 2000).

The second spectral moment does not seem to be very useful for distinguishing places of articulation in fricatives, only for differentiating between a flat diffuse spectral shape, as in a labiodental, and a peaky, compact distribution as in alveolars (Fangfang *et alii.*, 2009). Therefore, variance is low for the sibilant fricatives and high for the non-sibilants. Tomiak (1990) had obtained similar results with standard deviation measurements.

The third spectral moment refers to the overall slant of the energy distribution within the spectrum. Positive skewness suggests a negative tilt with a concentration of energy in the lower frequencies [below the mean value]. Negative skewness is associated with a positive tilt and a predominance of energy in the higher frequencies [above the mean value] (Jongman *et alii*, 2000). This cue shows clear distinctions among the sibilants, the palato-alveolars, unlike the alveolars, having a positive value. In the case of non-sibilants, the effectiveness of skewness is relative, although it seems to increase substantially at the fricative-vowel transition, "reflecting the predominance of low-frequency over high-frequency energy as the vowel begins" (Jongman *et alii,* 2000: 1257). Tomiak (1990) had reported that the dentals displayed a greater skewness than labiodentals.

The fourth spectral moment is an indicator of the peakedness of the distribution. “Positive kurtosis suggests a clearly defined spectrum with well-resolved peaks, while negative kurtosis indicates a flat spectrum without clearly defined peaks” (Jongman *et alii,* 2000: 1253). Therefore, it may be useful for differentiating among fricatives with different tongue postures such as the sibilants, showing a large positive value for alveolars and small positive or negative values for palato-alveolars (Tomiak, 1990). On the other hand, it failed to distinguish significantly between dentals and labiodentals, although /θ/ tends to show a greater kurtosis than /f/ (Tomiak, 1990).

On the whole, the first four moments of the spectral energy distribution may help distinguish places of fricative articulation. Labiodental and dental English fricatives tend to have lower spectral means and kurtosis than alveolar fricatives, higher spectral means and kurtosis than palato-alveolars, higher spectral variance than sibilant fricatives, and skewness that is higher than alveolars, but lower than palato-alveolars (Jongman *et alii*, 2000; Silbert & de Jong, 2008). Nevertheless, across the moments, only skewness (or the third moment) seems to be significantly different for all four places, whereas the other moments are only significantly different for two or three place-groups (Forrest *et alii*., 1988; Jongman *et alii*., 2000). Both noise onset and fricative-vowel transition region seem to contain the most distinctive information (Jongman *et alii*, 2000).

As spectral moments have proved to be very useful cues for the identification of the place of articulation within sibilants, but not that reliable in the case of non-sibilants, a classification based on these measurements alone would probably have a high failure rate (Jesus & Shadle, 2002). Consequently, other techniques have been proposed. Since an in-depth discussion of other statistical techniques would exceed the scope of a chapter of this nature, we will briefly mention another technique (Jesus & Shadle, 2002) -based on Evers *et alii*, 1998-), in which the slopes of two *regression lines* are fitted to spectra in lower and higher frequency regions. One is fitted from 500 Hz to the mean frequency of the highest amplitude peak, and the other, from that mean frequency up to 20 kHz. Their results, although based on European Portuguese languages, were similar to those of Jongman *et alii,* (2000) and Maniwa *et alii,* (2008) which were based on English, showing consistent changes in overall spectral shape across place of articulation in fricatives.

It is worth mentioning that all previous studies claim that spectral moments are moderately affected by the speaker and context variables, such as position in the word, and even vowel quality, although very little information is provided about their effects on the values of each moment, as claimed by Jesus & Shadle (2002). *Gender* analysis gives us more details: females exhibit significantly higher values than males for spectral mean, variance and kurtosis, and lower skewness values than those of males (Nittrouer, 1995). Consequently, compared to males, the spectra of female speakers had clearer peaks and a concentration of energy towards higher frequencies (Jongman *et alii*, 2000). The interaction between place and gender is also manifest although with no conclusive results. The study by Silbert and de Jong (2008) reported that female speakers produced coronal [alveolar] fricatives with high first and fourth but low second and third spectral moments, and labiodental fricatives with low first, third, and fourth spectral moments, and high spectral standard deviation (p. 2775), although in the case of male speakers, results were sometimes unexpected in their coronal fricative productions: low first and fourth spectral moments and high second and third spectral moments.

As for the *position* of the fricative *in the word*, the study by Silbert & de Jong (2008) indicated that fricatives in onset position show higher spectral means, and also that place interacts with *prosodic* context: the difference in spectral kurtosis between non-sibilants (labiodentals) and sibilants (coronals) is larger in onset position than in coda position.

3.2. Frication amplitude

The most relevant amplitude cues of the frication related to fricative place of articulation comprise the *overall frication amplitude* and the *relative amplitude*. *Comparisons of acoustic power between frequency regions* will be described in *other cues*.

3.2.1. Overall frication amplitude

Most research studies that focused on *frication* (or *noise*) *amplitude* have investigated the overall amplitude of English fricatives (Strevens, 1960; Delattre 1966, Stevens, 1971; Han-Yong, 1979; Behrens and Blumstein 1988b), showing that it is possible to distinguish sibilance, that is, to make distinctions between sibilants and non-sibilants: the former have a substantially greater amplitude (around 10–15 dB) than the latter. This generally lower level of intensity in non-sibilants is due to the lack of a noticeable resonating cavity in front of the dental constriction.

In general, although the relative importance of this cue in the perception of fricative place of articulation is similar to that shown by the spectral differences, no further distinctions can be made within each group. Jongman *et alii,* (2000), in an attempt to find a more relevant parameter, analysed what is called *normalised amplitude* or "the difference between noise amplitude minus vowel amplitude measured in dB (p. 1259), which normalised intensity differences among speakers. Following previous procedures (Behrens & Blusmtein, 1988b), first the rms (root-mean-square) amplitude was measured for the entire noise portion of each fricative token, and then, the vowel amplitude ("rms amplitude […] averaged over three consecutive pitch periods at the point of maximum vowel amplitude", Jongman *et alii,* 2000: 1256) was subtracted from the noise amplitude. Their results showed that normalized amplitude could be distinguished within each group. Thus, in the group of sibilants, palato-alveolar /ʃ, ʒ/ had a greater frication amplitude than alveolar /s, z/, while for the nonsibilants, labiodental /f, v/ had a greater amplitude than dentals /θ, ð/. In all cases there were also clear differences between voiced and voiceless terms of normalised amplitude, as will be discussed later.

The same study reported the main effects of the *quality* of the adjacent *vowel* on the overall normalized frication amplitude, in particular, "only the amplitude difference for /i/ and /ɑ/ differed from that for the rest of the vowels" [/o,u, e, æ/] (p. 1259). Also, no main effect of *gender* has been reported (p. 1259).

3.2.2. Relative frication amplitude

In 1985 Stevens suggested the existence of a change in amplitude of the frication noise in relation to the vowel in a specific region, adding that this parameter could be a more relevant acoustic cue than the overall amplitude of the frication noise. *Relative amplitude* measure has also been called "*frequency specific relative amplitude* (FSRA) to distinguish it from the overall normalized amplitude" (Maniwa *et alii,* 2009: 3966).

The studies on the role of the relative amplitude, defined as the difference between fricative and vowel amplitude in the same frequency region (in the third formant frequency region) focused on its effect on perception. These studies (Harris, 1958; Pickett & Rueenstein 1960; Stevens, 1985; Hedrick & Younger, 2003) showed that manipulation of frication amplitude relative to vowel amplitude in that frequency region could affect the labelling of place of articulation for the sibilant fricative contrasts. It was stated, however, that its effect was not as essential as other cues (Mann & Repp, 1980; Stevens, 1985) and was even claimed (Behrens & Blumstein, 1988b) that, when the spectral properties of the fricative noise and formant transitions were compatible, the perceptual effects of the manipulation of the amplitude of the noise had little effect on the overall identification of sibilant place of articulation.

The introduction in the analysis of two specific frequency regions instead of one (i.e., the third formant frequency region F3 for sibilants and the F5 region for non-sibilants) brought promising results (Hedrick & Ohde, 1993; Hedrick, 1997; Hedrick & Carney, 1997). Jongman *et alii,* (2000) carried out a study to find out the magnitude of differences in relative amplitude as a function of place of articulation, showing that all four places of articulation were significantly different. Relative amplitude in the region corresponding to F3 of the following vowel decreased as place moved further back in the oral cavity except for the palato-alveolars. These sounds showed a major concentration of energy in the region that corresponded to the F3 of the subsequent vowel; therefore the relative amplitude was small. The great relative amplitude of the alveolars in the F3 region for alveolars supports the notion that these fricatives have their major energy in a frequency region well above F3. Furthermore, fricative amplitude in the F5 region was reported to be smaller for /θ, ð/ than /f, v/.

McMurray & Jongman (2011) used a modified relative amplitude measurement, what they called *narrow-band amplitudes* in the frication to avoid possible artificial distinctions between sibilants and non-sibilants. So far, the amplitude in the F3 region in the vocalic portion used for sibilants was almost certainly greater than in the F5 (used for non-sibilants), therefore, they measured both F3 and F5 amplitude for all fricatives and treated them as two separate cues. Results with sibilants reported that the amplitude at F3 had a strong relationship to place of articulation (/s/ vs. /ʃ/) while at F5 it was related to sibilance but not place within each class.

In relation to the effects of *speaker* and *context* variables, it can be claimed that, in general, all these measurements are moderately affected by the context (McMurray & Jongman, 2011). As an example, the study by Jongman *et alii,* (2000) reported a slight effect of *vowel*, in particular of /e/ and /u/. They also proved a certain amount of vowel by place interaction: "while vowel-intrinsic differences in relative amplitude were similar across most places of articulation, labiodentals deviated from this pattern, showing much lower values for the back vowels as compared to the front vowels." (p. 1260). The effect of *gender* on relative amplitude seem to be moderate, with smaller values for females than for males, and being more evident on dentals than the rest of fricatives, according to Jongman *et alii,* (2000).

3.2.3. Other cues: Comparisons of acoustic power between frequency regions

In mechanical models of human vocal tracts, an increase in the volume velocity of air that flows through the constriction was shown to increase the power more at higher frequencies than at lower frequencies (Shadle, 1990 & 1991). Comparisons between the acoustic power in higher frequencies and lower frequencies, or what it is known as *noise power* (power above 750 kHz) versus *voice* (or *glottal*) *power* (below 750 kHz), have revealed a significant difference between sibilants (alveolars) and non-sibilants (labiodentals): the former show a higher noise power than the latter (Silbert & de Jong, 2008; McMurray & Jongman, 2011). These results are consistent with Wilde's statement (1995) that the amplitudes of fricative noise in restricted frequency regions can distinguish sibilants from non-sibilants.

The analysis of the effects of context variables shows the interaction of place by *prosodic* context, with an overall higher power for onset coronals (alveolars) relative to coda coronals, while there is roughly the same power across prosodic contexts for labiodentals (Silbert & de Jong, 2008). The place by prosodic context interaction is also evident in the voice power analysis. In labiodentals it decreases slightly over the three measurements (to a greater extent in coda than in onset position) whereas in alveolars (coronals) low frequency power increases in onset position and decreases in coda (Silbert & de Jong 2008).

3.3. Frication duration

Several studies have analysed the role of frication duration on fricative place of articulation. There is no agreement on its relative importance, whether this temporal property is a primary robust cue (Barrs, 1966; Jongman, 1989; Hedrick & Ohde, 1993) or a secondary cue (Lariviere *et alii,* 1975; Jongman, 1985; Gordon, 1989; Whalen, 1991; Jongman *et alii,* 2000).
Noise duration seems to be a perceptual cue with less weighting for the identification of place of articulation than other cues mentioned above, although it does serve to distinguish sibilants from non-sibilants, with /s, ʃ/ being longer than /f, θ/ (Han-Yong, 1979; Álvarez-González, 1982; Baum & Blumstein, 1987; Behrens & Blumstein, 1988a; Pirello *et alii*, 1997). Thus, as the place of articulation moves posteriorly, frication duration increases (Jesus & Shadle, 2003). These distinctions, however, are not always significant (Silbert & de Jong, 2008) whether we measure the absolute frication duration (Behrens & Blumstein, 1988a) or the *normalised duration* (as a proportion of the syllable duration) so as to mitigate the effects of the variations in speaker rate across speakers (Jongman *et alii*, 2000; Silbert & de Jong 2008).
The frication duration required in order to correctly perceive a fricative depends on the particular consonant (Stevens *et alii*, 1992), ranging from approximately 30 ms for [s, z] to 50-60 ms for [f, v], “while [θ, ð] are identified with reasonable accuracy in only the full frication and syllable conditions” (Jongman, 1989).

In relation to the effect of context variables, fricative duration seems to vary, although not always significantly (Klatt, 1974) with the *prosodic* context, in particular, with the fricative *position within the word* (Jesus & Shadle, 2003), the shortest being in mid position, followed by the initial and final, respectively (Álvarez-González, 1982). Other studies have investigated the influence of *stress* on fricative durations with inconclusive results, as usually the fricative duration values overlapped significantly (Crystal and House, 1988b), although there is a tendency for fricatives to be longer (15% more) in accented syllables in comparison with unaccented ones (Klatt, 1974).

The influence of *vowel quality* on frication duration has also been studied in depth (Denes, 1955, Cole & Cooper, 1975; Baum & Blumstein, 1988; Jongman, 1989, Jongman *et alii*, 2000), and it was found that, although with minimal effects (Behrens and Blumstein 1987) frication duration can be affected by the duration of the neighbouring vowel (Denes, 1955, Cole & Cooper, 1975; Baum & Blumstein, 1987; Jongman, 1989; Jongman *et alii*, 2000) decreasing with decreasing vowel height. It has also been suggested that noise duration varies according to the vocalic context and the fricative place of articulation (Denes, 1955; Isenberg, 1980). The *speaker* and, most noticeably, *gender*, also have an effect on frication duration. In particular, female speakers produce fricatives with slightly shorter duration than those produced by male speakers (Jongman *et alii*, 2000, Silbert & de Jong, 2008).

3.4. Transitions cues

Transition cues in relation to fricative perception of place of articulation mainly refer to *F2 & F3 formant transitions*. *Locus equations* will be described in *other cues*.

3.4.1. Formant transitions

Second and third formant frequencies have long been known to cue place of articulation distinctions in a variety of consonants (e.g., Delattre, 1963; Lariviere *et alii*, 1975; Stevens & Blumstein, 1978; Whalen, 1981; Behrens & Blumstein, 1988a). Nevertheless, formant transitions, especially those of the second formant (T2), seem to exert a much poorer effect on fricative place of articulation than has been observed with stops.

Individual differences in the use of transitions for the identification of fricatives, in particular, in sibilants, can account for the apparent contradictory results reported in previous works (Casserly, 2010). It has even been suggested that formant transitions might be most useful for listeners of languages with spectrally similar fricatives (Wagner *et alii*, 2006).

On the one hand, there are studies that claim that formant transitions, although not always present (Shadle *et alii*, 1996), can help distinguish non-sibilants (Harris, 1954 & 1958, Heinz & Stevens, 1961; McCasland, 1978 & 1979; Carden *et alii*, 1979; Pickett, 1980; Nittrouer, 2002). In fact, they take on more weight when spectral cues are ambiguous (Whalen, 1981; Hedrick & Ohde, 1993; Hedrick & Younger, 2003). The differences identified consisted of dentals having higher formants near the transition regions than labiodentals. Furthermore, differences in formant transitions, especially T2, have been reported to correspond to place distinctions between sibilants, either voiced or voiceless (Soli, 1981).

More specifically, in English, F2 onset for a given vowel context seems to be progressively higher as the place of constriction moves back in the oral cavity, for instance, F2 onset frequency for /ʃ/ is around 100-300 kHz higher than for /s/ (Wilde, 1993; Mann & Repp, 1980; Whalen, 1981; Nittrouer, 1992, among others), and it is clearly higher than /θ/ and /f/, although this difference between sibilants alveolars and non-sibilants dentals is not always significant (Jongman *et alii*, 2000). In addition, “the range of F2 onset is progressively smaller as place of constriction moves further back” (Wilde, 1993). These results, consistent with Recasens’ findings (1985), show that those consonants with a greater degree of tongue-body raising (with usually a more posterior place of articulation) are more resistant to co-articulation.

On the other hand, formant transitions have traditionally been thought to be secondary or alternative cues for fricative place of articulation (Harris, 1958; Sharf & Hemeyer, 1972; Repp & Mann, 1980; Whalen, 1981 & 1991; Hedrick & Ohde, 1993; Kent *et alii,* 2002; Raphael *et alii,* 2011). In fact, there are doubts about their role for fricative place of articulation perception, especially by more recent studies (e.g., Jongman *et alii,* 2000; McMurray &

Jogman, 2011). They claim that F2 transition (T2) properties do not reliably distinguish all places of articulation in fricatives, although T2, and even T4 and T5, had moderate effects of place of articulation for both sibilants and non-sibilants (McMurray & Jongman, 2011). There are other cues, such as frication spectrum, or even frication duration, that override context-dependent formant transition cues (Hedrick & Ohde, 1993; Nittrouer, 2002).

There is a clear effect of the fricative *position in the word* on formant transitions: VC transitions seem to have more perceptual weighting than CV transitions (Sharf & Hemeyer, 1972), although CV transitions can also assist speakers to identify place of articulation in fricatives (Zeng & Turner, 1990).

Vowel quality has a mild effect on transitions, more pronounced in voiceless fricatives (Mann & Repp, 1980; Repp & Mann, 1980; Whalen, 1981). The study by Jongman *et alii,* (2000) showed that “F2 onset values are higher for front vowels compared to back vowels and that F2 onset values significantly increased as a function of increasing vowel height” (p. 1259). They also reported a significant place by vowel interaction for labiodentals and alveolars, while for the other fricatives differences in the F2 onset were restricted to /i, e/. It has also been noted that *gender* has a major effect on transitions (Repp & Mann, 1980). It seems that the onset of F2 is significantly higher for female speakers than for males (Jongman *et alii*, 2000).

3.4.2. Other cues: Locus equations

Locus equation analysis is a method that has been used by many researchers to find other relevant cues for the identification of fricative place of articulation. This analysis consists of “finding the slope and the intercept of a line between the same acoustic feature in a vowel and an adjoining consonant” (Jesus & Shadle, 2002). In particular, the analysis of locus equations is based on the second formant frequency (F2) at both vowel onset and vowel midpoint, “since it seems that the starting frequency of the F2 of a vowel preceded by an obstruent could provide information about the articulatory configuration used to generate the consonant” (Jongman *et alii,* 2000: 1254). In a way, locus equation may cue place of articulation indirectly by showing the degree of coarticulation between consonant and vowel (Fowler, 1994).

Sussman *et alii,* (1991) used locus equations as statistically powerful phonetic descriptors of place or articulation with stops. Wilde (1993) applied this method for voiceless fricatives with promising results, claiming that it could provide additional place information for these consonants too. However, more recent studies reported clearly contradictory results. Jongman & Sereno (1995) used locus equation to look for differences among non-sibilants, showing a greater spectral dispersion over the range of 0-11 kHz for labiodentals, although predictions of these fricatives based on this measure were considered poor. Jongman *et alii*, (2000) also found out that, after averaging across vowel context, the *y* intercepts were only distinct for labiodentals and palato-alveolars and that the slope value obtained for labiodentals was significantly different from the other three places of articulation. Therefore, they considered that locus equation was not always reliable for distinguishing all four places of fricative articulation, nor even for making clear distinctions between sibilants and non-sibilants. It can be claimed that locus equations do not accurately reflect consonantal place of articulation (Fowler, 1994), unless combined with other parameters (Jesus & Shadle, 2002).

Not much information has been provided in prior literature regarding the effect of speaker and other context variables. *Gender* has been observed to have a major effect on these measurements, showing that the locus equation (*y* intercept, in particular) is significantly higher for females than for males (Jongman *et alii,* 2000).

4. Voicing: Articulatory description and acoustic cues

It is well known that English fricatives can be also classified in terms of production according to the presence or absence of voicing: /f, θ, s, ʃ, h/ vs. /v, ð, z, ʒ/, but in some contexts, in particular, in initial and final position, the so-called voiced fricatives have very little voicing (or none). However, the voiced/voiceless pairs are also distinguished as *fortis* and *lenis,* referring to the force of articulation, that is, the degree of muscular effort and breath involved in their articulation. So /f, θ, s, ʃ/ tend to be pronounced with a relatively more muscular effort and breath force than their voiced counterparts. /h/ is usually included in the former.

Fewer studies have examined which acoustic cues seem to influence the perception of fricative voicing as shown in the following sections. The main acoustic cues for the perception of voicing in fricatives are the *presence of phonation*, the *relative durations of vowel and fricative segments*, and the *frication amplitude.* The spectral properties of *frication noise* and *formant transitions* will be analysed in the section called *other cues.*

4.1. Phonation

The presence of phonation (that is, glottal pulsing or glottal excitation) during frication that produces energy primarily at low frequencies (e.g. below 750 kHz) has traditionally been considered the main acoustic cue of voicing.

Many studies have analysed the important role played by the duration (and amplitude) of glottal vibration to make distinctions between voiced and voiceless fricatives (Hughes & Halle, 1956; Ingerman, 1960; Glicksman & Stevens, 1988; Stevens *et alii*, 1992; Pirello *et ali*, 1997; Maniwa *et alii,* 2008, to mention a few), showing that the glottal pulsing is extended over a longer period for the voiced than for the voiceless fricatives.

Nevertheless, voicing in fricatives is not only characterized by the presence or absence of the glottal excitation (Stevens *et alii*, 1992). In fact, although in those contexts where the fricative is not fully voiced, (namely, in initial and final position) the presence or absence of glottal excitation at the acoustic boundaries of the fricative noise can classify a majority of the fricative consonants in terms of voicing (Pirello *et alii,* 1997), this acoustic property tends to overlap with many other cues (Maniwa *et alii*, 2008; Raphael *et alii*, 2011), especially the temporal and amplitudinal ones as will be discussed later.

Devoicing of voiced fricatives has been studied in depth showing a great variability among *speakers* and *contexts*, although not always with consistent effects (Smith, 1997). In general, voiced fricatives are devoiced in

initial and final position, occurring more often in the later than in the former (Stevens *et alii*, 1992; Smith, 1997; Jesus & Shadle, 2002 & 2003).
Furthermore, the *place* of articulation has been found to have a significant effect on the devoicing of fricatives. Labiodental fricatives are reported to remain fully voiced more often (Stevens *et alii*, 1992) or to have a greater preponderance of voicing throughout than alveolar fricatives (Pirello *et alii*, 1997). It seems then that as the place of articulation moves further, back the amount of devoicing increases proportionally (Jesus and Shadle, 2003).

4.2. Frication duration

The most salient temporal properties related to voicing in fricatives involve the *overall frication duration* and the *relative duration. Vowel duration* will be described in *other cues*.

4.2.1. Overall frication duration

Frication duration influences voicing perception (Glicksman & Stevens, 1988, Giavazzi & Cho, 2008), with greater effects than place of articulation (Maniwa *et alii*, 2008). The difference in duration between voiceless (or unvoiced) and voiced fricatives has been studied in previous works. In general, voiceless fricatives at a given place of articulation are substantially longer than their voiced counterparts (Álvarez González, 1982; Docherty 1992; Stevens *et alii*, 1992; Pirello *et alii*, 1997; Smith, 1997; Jesus & Shadle, 2003) as "the maintenance of voicing during the articulation of a fricative constriction results in later onset and earlier offset" (Silbert & de Jong, 2008: 2770). Nevertheless, the study by Crystal & House (1988a) claimed no correlation between voicing and duration in fricatives. What is certain is that when the overall frication duration, (or the normalized duration) is shortened, there is a clear perceptual shift from voiceless to voiced (Flege & Hillenbrand, 1986; Jongman *et alii*, 2000). This effect is more pronounced for the non-sibilants than for the sibilants (Harris, 1958; Jongman *et alii*, 2000).
There is a patent interaction between frication duration and the *prosodic*

context, in particular, the position within the word. Frication duration seems to be an important cue when the fricative is in a partially or fully devoiced position, and is particularly significant for the voicing distinction in syllable-initial positions, with voiceless fricatives having longer noise durations than voiced fricatives (Álvarez González, 1982; Behrens & Blumstein, 1988a; Baum and Blumstein, 1987; Giovazzi & Cho, 2008). This observation of fricatives in isolated syllables also holds in connected speech (Crystal and House, 1988a). In addition, temporal (as well as spectral) characteristics of voiceless fricatives have been reported to be more dependent on *vowel context* than those of their voiced counterparts (Wilde, 1993 & 1995).

4.2.2. Relative duration

The fact that the mean frication duration values between voiced and voiceless fricatives frequently overlap (Baum & Blumstein, 1987; Crystal & House, 1988a & 1988b) has made researchers point out that frication duration is not the only acoustic cue when making voicing contrasts in fricatives. Another temporal measurement has been proposed, namely the *relative durations* of the vocalic and fricative segments (Denes, 1955; Harris, 1958; Cole & Cooper, 1975; Soli, 1982; Baum & Blumstein, 1987; Glicksman & Stevens, 1988; Jongman, 1989; Gordon, 1989; Raphael *et alii*, 2011; McMurray *et alii*, 2011), which seems to be a more salient cue for identifying the voicing class of a fricative than the frication duration itself (Hughes & Halle, 1956; Raphael *et alii*, 2011). Voicing perception seems to increase as fricative duration decreases in relation to the adjacent *vowel*, although the relative weighting of this cue depends on the *position* of the consonant (Giovazzi & Cho, 2008).

4.2.3. Other cues: Vowel duration

The relative weighting of the cue *vowel duration* is determined by the context. Vowel duration has a a strong effect on fricative voicing in syllable-final position (Denes, 1955; Álvarez González, 1980; Raphael, 1972, 1981; Maxwell & Weismer, 1982; Smith, 1997; Broersma, 2010): When the preceding vowel is relatively long the consonant is usually perceived as voiced whereas when the vowel is relatively short it may be perceived as voiceless. In contrast the effect produced is weak when the duration of the adjacent vowel is shortened in syllable-initial position (Cole & Cooper, 1975; Baum & Blumstein, 1987).

Nevertheless, other works claim that syllable affiliation (onset versus coda position) does not to correspond to reliable frication duration differences (Gordon, 1989; Pirello *et alii*, 1997; Silbert & de Jong, 2008).

4.3. Frication amplitude

The most pertinent amplitude properties of the frication relative to voicing in fricatives comprise the *overall frication amplitude* and the *relative amplitude*. *Comparisons of acoustic power between frequency regions* will be described in *other cues*.

4.3.1. Overall frication amplitude

Frication amplitude has also been considered as a potentially salient cue for voicing in fricatives. The maintenance of voicing during the articulation of a fricative constriction tends to lower overall amplitude of the frication noise (Silbert & de Jong, 2008). Voiced fricatives have significantly smaller overall frication amplitude than their voiceless counterparts (Ingerman, 1960). The study by Jongman *et alii* (2000) shows a *place* (more precisely, *sibilance*) by voicing interaction, with a greater difference between voiceless and voiced fricatives for the non-sibilants than for the sibilants. No *gender* effects were observed.

4.3.2. Relative amplitude

The relative intensity difference between the frication and the vowel has also been suggested as a useful cue when making voicing distinctions among fricatives. In general, consistent with the observations of overall frication amplitude, voiceless fricatives seem to have significantly greater amplitude relative to the vowel than their voiced counterparts (Ingerman, 1960). Moreover, a certain amount of significant interaction has also been reported between voicing and *place*: the difference in relative amplitude between the voiceless and voiced labiodental and palato-alveolars is significant, while it is not so for dentals and alveolars (Jongman *et alii*, 2000).

In relation to the effects of *speaker* and other *context* variables, it has been claimed that, in general, relative amplitude is moderately affected by all these variables (McMurray & Jongman, 2011). Furthermore, *gender* produces voicing distinctions in relative amplitude, with greater values for males than for females (Jongman *et alii*, 2000).

4.3.3. Other cues: Comparisons of acoustic power between frequency regions

Outstanding research carried out by Silbert & de Jong (2008) compared the acoustic power in higher frequencies (or 'noise power' above 750 kHz) and lower frequencies (or 'voice power' below 750 kHz taken at three intervals for each fricative). Their results showed voicing differences among English fricatives: phonologically voiceless fricatives have higher noise power than phonologically voiced fricatives, with a significant interaction with *gender*, and showing a greater difference between voiceless and voiced fricatives in the low-frequency power measurements of the male speaker values than those of the female ones.

Moreover, it seems that the *position in the word* has a significant effect on the voicing power of fricatives. The difference between voiced and voiceless segments within each prosodic context is larger in onset than in coda position. The voicing power by *prosodic* context interaction is evident with a higher degree of voicing near the vowel in all syllable positions. In particular, while coda fricatives exhibit considerably more low-frequency power in the first third than at either of the other two thirds (that is, near the adjacent vowel), onset fricatives showed a more or less constant low-frequency power with a slightly increasing level commensurate with their proximity to the vowel (Silbert & de Jong, 2008).

Other amplitude-relative measures that have lately been associated to voicing in fricatives include harmonics-to-noise ratio (Maniwa *et alii*, 2008), dynamic amplitude (Jesus & Shadle, 2002) and low-frequency energy (McMurray & Jongman, 2011).

4.4. Other cues

In this section we will include other cues that have less potential perceptual weighting for fricative voicing, specifically, *frication noise* and *formant transitions*.

4.4.1. Frication noise

The analysis of *spectral moments* in relation to fricative voicing has shown a different distribution of energy across fricative spectra between voiceless and voiced fricatives. Differences have been reported to be significant, although with a rather small effect (Jongman *et alii*, 2000), showing that voiceless fricatives are characterized by higher values for spectral mean, skewness and kurtosis than voiced fricatives, and, in addition, they have significantly less variance than voiceless ones (Jongman *et alii*, 2000; Jesus & Shadle, 2002; Silbert & de Jong, 2008; Maniwa *et alii*, 2009). Consequently, compared to voiced fricatives, the spectra of voiceless fricatives have a concentration of energy towards slightly lower frequencies and slightly better defined peaks at a significantly higher frequency (Jongman *et alii*, 2000).

The study by Silbert and de Jong (2008) shows significant interactions between voicing and *place* and voicing and *prosodic* context. Particularly, voiced labiodental fricatives show higher spectral variance (standard deviations) than their voiceless counterparts, an effect not found in alveolars (coronals). Also, spectral mean in voiceless onset fricatives have higher spectral means, with a larger differences in variance (standard deviations) between these two places of articulation in onset position than in coda.

The technique of *spectral slopes of two regression lines* (Jesus & Shadle, 2002) reveals that the average regression always overlaps between voiced and voiceless fricatives, thus supporting the idea of a very mild effect of voicing on frication noise.

4.4.2. Formant transitions

The first formant frequency of a vowel, in particular, the *extent of F1 transition*, has been claimed as another potential cue for voicing in fricatives (McMurray & Jongman, 2011). It seems that there are more wide transitions of the first formant adjacent to voiced fricatives than for the voiceless cognates, at least in intervocalic position (Stevens *et alii*, 1992).

There are no conclusive results on the effect of fricative voicing on other transitions. F2 onset exhibits a very mild interaction between *place* and voicing, with a significantly higher onset for voiced palato-alveolars than for the voiceless counterparts (Jongman *et alii*, 2000) and no voicing difference among the other places of articulation. The analysis of F2 and F3 onset frequencies between voiceless and voiced fricatives has shown that the former are more dependent on *vowel* context than the latter (Wilde, 1993).

5. Conclusions

The aim of this chapter was to offer a critical survey of previous studies in the characterization of English fricatives. In particular, we have concentrated only on those speech patterns that could be potential acoustic cues for distinctive features of English fricatives, analyzing in depth their role (or perceptual weighting) in fricative perception of manner, place and voicing.

Spectral, temporal and amplitudinal properties have been analysed by researchers in an attempt to find a consistent mapping between these acoustic properties and phonetic features, such as place of articulation, manner or voicing. This information is of great importance as it can help us, firstly, to understand and explain the classification of English fricatives according to these phonetic features, despite variation in speaker or phonetic context, and secondly, to interpret a particular acoustic property with its articulatory correlate(s), if any.

Summing up, in the first place, it is clear that place (as well as voicing) distinctions in fricatives can be realised by multiple cues. Up to 24 cues have been proposed for distinguishing place of articulation, voicing and sibilance (McMurray & Jongman, 2011). Some cues are clearly more salient than others, although it is extremely difficult to attribute a few cues to one phonetic feature, and almost none of the cues can be associated with a single feature, as McMurray & Jongman (2011) have recently pointed out. Therefore, new measures have been included in the analysis, some of them still showing contradictory or inconclusive results. These areas clearly need a great deal more study, offering plenty of scope for future research. It is hoped, therefore that future researchers will be stimulated to carry out more comprehensive work, enabling us to gain a deeper understanding of the issues under discussion in this chapter.

And secondly, and even more revealing, is that there is a significant speaker variation in the use of English fricative perceptual acoustic cues. Moreover, significant effects of gender and prosodic context variables (including vowel quality, position within word and stress) have been observed in almost every cue, although still more work needs to be done so as to provide more conclusive results.

REFERENCES

Abdelatty Ali, A. M., Van der Spiegel, J. & Mueller, P. (2001): "Acoustic-phonetic features for the automatic classification of fricatives". *Journal of the Acoustical Society of America*, 109/5, 2217-2235.

Álvarez-González, José Antonio (1980): "Voicing in English consonants". *Atlantis* 1/2: 44-52.

Álvarez-González, José Antonio (1982): "Duration of English consonants". *Anglo-American Studies,* II/1: 49-62.

Badin, Pierre (1989): "Acoustics of voiceless fricatives: production theory and data". *Speech Transmission Laboratory Quarterly Progress and Status Report* 3: 33-55.

Barreiro, Silvia Carmen (1994). *Análisis acústico descriptivo de fricativas español-inglés: estudio comparativo*. Unpublished PhD thesis. León: University of León.

Barrs, James T. (1966): "Identification of voiceless fricatives versus modification of their duration". *Linguistics* 21: 5-23.

Baum, Stephen, R. & Blumstein, Sheila E. (1987): "Preliminary observations on the use of duration as a cue to syllable initial fricative consonant voicing in English". *Journal of the Acoustical Society of America* 82: 1073-1077.

Behrens, Susan & Blumstein, Sheila E. (1988a): "Acoustic characteristics of English voiceless fricatives: a descriptive analysis". *Journal of Phonetics* 16: 295-298.

Behrens, Susan & Blumstein, Sheila E. (1988b): "On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants". *Journal of the Acoustical Society of America* 84: 861-867.

Bolt, Richard H. Cooper, Franklin S., David, Edward E., Denes, Peter B., Pickett, James M. & Stevens, Kenneth N. (1973): "Speaker identification by speech spectrograms: some further observations". *Journal of the Acoustical Society of America* 54/2: 531-534.

Broersma, Miriam (2010): "Perception of final fricative voicing: Native and non-native listener's use of vowel duration". *Journal of the Acoustical Society of America* 127/3: 1636-1644.

Bush, Clara N. (1964): *Phonetics Variations and Distinctive Features.* The Hague: Mouton.

Carden, Guy A. L., Levitt, Andrea & Jusczyk, Peter (1979): "Manner judgments affect the locus of the labial/dental boundary for stops and fricatives". *Journal of the Acoustical Society of America* 65/S1: S79.

Carney, Patrick .J. & Moll, Kenneth L. (1971): "A cinefluorographic investigation of fricative consonant-vowel coarticulation". *Phonetica* 23: 193-202.

Casserly, Elizabeth (2010): “Individual differences in use of English fricative perceptual cues”. *Journal of the Acoustical Society of America* 127/3: 1957-1957.

Cole, Roland A. & Cooper, William E. (1975): “Perception of voicing in English affricates and fricatives”. *Journal of the Acoustical Society of America* 58/6): 1280-1287.

Cruttenden, Alan (2008): *Gimson’s Pronunciation of English* (7th ed.). London: Hodder Arnold Publication.

Crystal, Thomas H. & House, Arthur S. (1988a): “Segmental durations in connected-speech signals: Current results”. *Journal of the Acoustical Society of America* 83: 1553–1573.

Crystal, Thomas H. & House, Arthur S. (1988b): “Segmental durations in connected-speech signals: Syllabic stress”. *Journal of the Acoustical Society of America,* 84: 1574–1585.

Delattre, Pierre C., Berman, A.M. & Cooper, Franklin S. (1963): “Formant transitions and loci as acoustic correlates of place of articulation in American fricatives”. *Studia Linguistica* 16/1/2: 104-122.

Delattre, Pierre C. (1966): *Studies in French and Comparative Phonetics: Selected Papers in French and English.* The Hague: Mouton.

Denes, Peter B. (1955): “Effect of duration on the perception of voicing”. *Journal of the Acoustical Society of America* 27/4: 761-764.

Docherty, Gerard J. (1992): *The Timing of Voicing in British English Obstruents*. Berlin: Foris Publications.

Evers, Vincent, Reetz, Henning & Lahiri, Aditi (1998): “Crosslinguistic acoustic categorization of sibilants independent of phonological status”. *Journal of Phonetics* 26: 345–370.

Fangfang, Li, Edwards, Jan & Beckman, Mary E. (2009): “Contrast and convert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese”. *Journal of Phonetics* 37/1: 111-124.

Fant, Gunnar (1960): *Acoustic Theory of Speech Production*. La Haya: Mouton.

Fant, Gunnar (1973): *Speech Sounds and Features*. Cambridge, Massachusetts: The MIT Press.

Fant, Gunnar & Mártony, J. (1960). Pole-Zero matching techniques. *Speech Transmission Laboratory Quarterly Progress and Status Report* 1/1: 14-16.

Feijóo, Sergio & Fernández, Santiago (2002): “Location of spectral cues for the perceptual identification of fricatives”. <http://www.idsia.ch/~santiago /papers/forum2002b.pdf> [downloaded 20/06/2012].

Flege, James E. & Hillenbrand, James (1986): “Differential use of temporal cues to the /s/ -/z/ contrast by native and non-native speakers of English”. *Journal of the Acoustical Society of America* 79/2: 508-517.

Forrest, Karen, Weismer, Gary, Milenkovic, Paul & Dougall, Roland. N. (1988): “Statistical analysis of word-initial voiceless obstruents: Preliminary data”. *Journal of the Acoustical Society of America,* 84: 115-123.

Fowler, Carol. A (1994): “Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation”. *Perception and Psychophysics* 55: 597–611.

Giavazzi, Maria & Cho, Hyesun (2008): “On the perception of voicing in strident fricatives”. <http://web.mit.edu/magia/www/docs/ ChoGiavazzi _2008.pdf.> [downloaded 04/07/2012].

Gil Fernández, Juana (2010): *Los Sonidos del Lenguaje.* Madrid: Sintesis.

Glicksman, Laura B. & Stevens, Kenneth. N. (1988): “Some parameters affecting the perception of voicing for fricatives”. *Journal of the Acoustical Society of America* 84/S1: 58-59(A).

Gordon, Peter C. (1989): “Context effects in recognizing syllable-final /z/ and /s/ in different phrasal positions”. *Journal of the Acoustical Society of America* 86: 1698-1707.

Han-Yong, You (1979): *An Acoustic and Perceptual Study of English Fricatives.* MSc Thesis. Edmonton, Canada: University of Edmonton.

Harris, Katherine. S. (1954): “Cues for the identification of the fricative of American English”. *Journal of the Acoustical Society of America* 26: 952.

Harris, Katherine. S. (1956): “Some acoustic cues for the fricative consonants”. *Journal of the Acoustical Society of America* 28: 160-161.

Harris, Katherine S. (1958): “Cues for the discrimination of American English fricatives in spoken syllables”. *Language and Speech* 1: 1–7.

Hedrick, Mark S. (1997): “Effect of acoustic cues on labeling fricatives and affricates”. *Journal of Speech, Language and Hearing Research* 40: 925–938.

Hedrick, Mark S. & Ohde, Ralph. N. (1993): “Effect of relative amplitude of frication on perception of place of articulation”. *Journal of the Acoustical Society of America* 94: 2005-26.

Hedrick, Mark S. & Carney, Arlene E. (1997): “Effect of relative amplitude and formant transitions on perception of place of articulation by adult listeners with cochlear implants”. *Journal of Speech, Language and Hearing Research* 40: 1445–1457.

Hedrick, M. S. & Younger, Mary S. (2003): “Labeling of /s/ and /ʃ/ by listeners with normal and impaired hearing, revisited”. *Journal of Speech, Language and Hearing Research* 46: 636-648.

Heinz, John M. (1961): “Analysis of fricative consonants”. *MIT Research Lab of Electronics Quarterly Progress Report* 60: 181-184.

Heinz, John M. & Stevens, Kenneth N. (1961):“On the properties of voiceless fricative consonants”. *Journal of the Acoustical Society of America* 33: 589-596.

Hughes, George W., & Halle, Morris (1956): “Spectral properties of fricative consonants”. *Journal of the Acoustical Society of America* 28: 303–310.

Ingerman, Frances (1960): “Formants as voicing cue in the perception of /s/”. *Journal of the Acoustical Society of America* 66: 1700.

Ingerman, Frances (1968): “Identification of the speaker’s sex from voiceless fricatives”. *Journal of the Acoustical Society of America* 44: 1142-1144.

Isenberg, David (1980): "A reaction-time study of the production of [s] and [ʃ]". *Journal of the Acoustical Society of America* 67/S1: S38.

Jassem, Wiktor (1965): "The formants of fricative consonants". In*:* Fry, Dennis B. (ed.), *Language and Speech*. Teddinton, Middlesex: Robert Draper Ltd., 1-16.

Jassem, Wiktor (1968): "Acoustic description of voiceless fricatives in terms of spectral parameters". *Speech Analysis and Synthesis* 1: 189-206.

Jesus, Luis M. T. & Shadle, Christine H. (2002): "A parametric study of the spectral characteristics of European Portuguese Fricatives". *Journal of Phonetics* 30: 437–464.

Jesus, Luis M. T., & Shadle, Christine H. (2003): "Temporal and devoicing analysis of European Portuguese fricatives". In: *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, Spain, 779–782.

Jongman, Allard (1985): "Duration of frication noise as a perceptual cue". *Journal of the Acoustical Society of America* 77(S1): S26(A).

Jongman, Allard (1989): "Duration of frication noise required for the identification of English fricatives". *Journal of the Acoustical Society of America* 85/4: 1718-1725

Jongman, Allard & Sereno, Joan A. (1995): "Acoustic properties of non-sibilant fricatives". In: *Proceedings of 13th International Congress of Phonetic Sciences* 4. Stockholm, 432-435.

Jongman, Allard, Sereno, Joan A., Wayland, Ratree, & Wong, Serena (2000): "Acoustic characteristics of English fricatives" *Journal of the Acoustical Society of America* 118: 1252–1263.

Kent, Raymond D., Read, Charles & Kent, Ray (2002): *The Acoustics Analysis of Speech* (2nd ed.). San Diego, CA: Singular.

Klatt, Dennis H. (1974): "The duration of [s] in English words". *Journal of Speech, Language and Hearing Research* 17: 51–63.

Kunisaki, Osamu & Fujisaki, Hyroya (1977): "On the influence of context upon perception of voiceless fricative consonants". *Annual Bulletin of the RILP* 11: 85–91.

LaRiviere, Conrad, Winitz, Harris & Herriman, Eve (1975): "The distribution of perceptual cues in English prevocalic fricatives". *Journal of Speech Hearing* 18/4: 613-622.

Lehiste, Isle (1964). *Acoustical Characteristics of Selected English Consonants.* Indiana University, Bloomington, Indiana & The Hague.

Llisterri, Joaquim (1995): "Relationships between speech production and speech perception in a second language". In: *Proceedings of 13th International Congress of Phonetic Sciences* 4. Stockholm, 92-100.

McCasland, George P. (1978): "Formant transition place cues of intervocalic fricatives". *Journal of the Acoustical Society of America* 63(S1): S21(A).

McCasland, George P. (1979). "Noise intensity and spectrum cues of spoken fricatives". *Journal of the Acoustical Society of America* 65(S1): S78(A).

Maniwa, Kazumi, Jongman, Allard & Wade, Travis (2009): "Acoustic characteristics of clearly spoken English fricatives". *Journal of the Acoustical Society of America* 123: 1114–1125.

Maniwa, Kazumi, Jongman, Allard & Wade, Travis (2009): "Acoustic characteristics of clearly spoken English fricatives". *Journal of the Acoustical Society of America* 125/6): 3962-73.

Mann, Virginia A & Repp, Bruno H. (1980): "Influence of vocalic context on perception of the [ʃ]-[s] distinction: III Acoustic observations". *Haskins Laboratories Status Report on Speech Research* 62: 177-85.

Maxwell Edith M. & Weismer Gary (1982): "The contribution of phonological, acoustic, and perceptual techniques to the characterization of a misarticulating child's voice contrast for stops". *Applied Psycholinguistics* 3: 29–43.

McCasland, George P. (1978): "Formant transition place cues of intervocalic fricatives". *Journal of the Acoustical Society of America* 63(S1): S21(A).

McCasland, George P. (1979). "Noise intensity and spectrum cues of spoken fricatives". *Journal of the Acoustical Society of America* 65(S1): S78(A).

McMurray, Bob & Jongman, Allard (2008): "From production to acoustics to perception: The case of fricatives". *Journal of the Acoustical Society of America* 124/4: 2437-2437.

McMurray, Bob & Jongman, Allard (2011): "What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations". *Psychological Review* 118/2: 219-246.

Newman, Rochelle S. (2003): "Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report". *Journal of the Acoustical Society of America* 113/5: 2850-2860.

Nittrouer, Susan & Studdert-Kennedy, Michael (1987): "The role of coarticulatory effects in the perception of fricatives by children and adults". *Journal of Speech, Language and Hearing Research* 30: 319-329.

Nittrouer, Susan (1992): "Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries". *Journal of Phonetics* 20: 351–382.

Nittrouer, Susan (1995): "Children learn separate aspects of speech production at different rates: Evidence from spectral moments". *Journal of the Acoustical Society of America* 97: 520-530.

Nittrouer, Susan (2002): "Learning to perceive speech: How fricative perception changes, and how it stays the same". *Journal of the Acoustical Society of America* 112/2: 711-19.

Pickett, James M. (1980): "Acoustic cues in speech perception". *The Sounds of Speech Communication*. Baltimore: University Park Press.

Pickett, James M. & Rueenstein Herbert (1960): "Perception of consonant voicing in noise". *Language and Speech* 3: 155-163.

Pirello, Karen, Blumstein, Sheila E., & Kurowski, Kathleen (1997): “The characteristics of voicing in syllable-initial fricatives in American English”. *Journal of the Acoustical Society of America* 101: 3754–3765.
Potter, Ralph K., Kopp, George A. & Green, Harriet C. (1947): *Visible Speech*. New York: Nostrand Company, Inc.
Raphael, Lawrence J. (1972): “Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of Word-Final Consonants in American English”. *Journal of the Acoustical Society of America* 51: 1296-1303.
Raphael, Lawrence J. (1981): “Durations and Contexts as Cues to Word-Final Cognate Opposition in English.” *Phonetica*: 38: 126-147.
Raphael, Lawrence J., Borden, Gloria J. & Harris, Katherine S. (2011): *Speech Science Primer: Physiology, Acoustics, and Perception of Speech* (6th ed.). Baltimore: Lippincott Williams & Wilkins.
Recasens, Daniel (1985): “Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences”. *Language and Speech* 28: 97–114.
Recasens, Daniel (1990): “An Electropalatographic and Acoustic Study of Consonantto-Vowel Coarticulation”. *Haskins Laboratories Status Report on Speech Research*, *SR*-101/102, 130-148.
Repp, Bruno H. & Mann, Virginia A (1980): “Influence of vocalic context on perception of the [ʃ]- [s] distinction: II spectral factors”. *Haskins Laboratories Status Report on Speech Research* 61: 65-84.
Roach, Peter (2009): *English Phonetics and Phonology* (4th ed.). Cambridge: C.U.P.
Schwartz, Martin F. (1968): “Identification of speaker sex from isolated, voiceless fricatives”. *Journal of the Acoustical Society of America* 43: 1178-1179.
Scully, Celia & Allwood, Emily (1985): “Production and perception of an articulatory continuum for fricatives of English”. *Speech Communication* 4: 237-245.
Scully, Celia, Castelli, Eric, Brearley, Eric & Shirt, Marion (1992): “Analysis and simulation of a speaker’s aerodynamic and acoustic patterns for fricatives”. *Journal of Phonetics* 20: 39-51.
Seitz, Philip F. D., Bladon, Richard A. W. & Watson, Ian M. C. (1987): “Across-speaker and within-speaker variability of British English sibilant spectral characteristics”. *Journal of the Acoustical Society of America* 82(S1): S37.
Shadle, Cristine H. (1985): “The acoustics of fricative consonants”. Massachussetts Institute of Technology, *Research Laboratory of Electronics Technical Report*, *506.*
Shadle, Cristine H. (1990): “Articulatory-acoustic relationships in fricative consonants”. In: Hardcastle, William J. & Marchal, Alain. *Speech production and speech perception*. Dordrecht: Kluwer Academic, 187–209.

Shadle, Cristine H. (1991): “Source parameters for the fricative consonants /s, ʃ, ç, x/”. *Journal of the Acoustical Society of America* 89/4: 3SP10.
Shadle, Cristine. H. & Mair, Sheila J. (1996): “Quantifying spectral characteristics of fricatives”. In: *Proceedings of the International Conference on Spoken Language Processing*, Philadelphia, U.S.A., 1517–1520.
Shadle, Cristine H., Mair, Sheila J. & Carter, John N. (1996): “Acoustic characteristics of the front fricatives [f, v, θ, ð]”. In: *Proceedings of the 1st ESCA Tutorial and Research Workshop (ETRW) on Speech Production ModelingF4th Speech Production Seminar*, Autrans, France, 193–196.
Sharf, Donald J. & Hemeyer, Thomas (1972): “Identification of place of consonant articulation from vowel formant transitions”. *Journal of the Acoustical Society of America* 51: 652-658.
Silbert, Noah & de Jong, Kenneth (2008): “Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production”. *Journal of the Acoustical Society of America* 123/5: 2769-2779.
Smith, Caroline L (1997): “The devoicing of /z/ in American English: Effects of local and prosodic context”. *Journal of Phonetics* 25/4: 471–500.
Soli, Sigfrid D (1981): “Second formant in fricatives: Acoustic consequences of fricative-vowel coarticulation”. *Journal of the Acoustical Society of America* 70: 976-984.
Soli, Sigfrid D (1982): “Structure and duration of vowels together specify fricative voicing”. *Journal of the Acoustical Society of America* 72: 366–378.
Steriade, Donca (1997): *Phonetics in phonology: The case of laryngeal neutralization* Ms., UCLA.
Stevens, Kenneth N. (1971): “Airflow and Turbulence Noise for Fricative and Stops Consonants: Static Considerations”. *Journal of the Acoustical Society of America* 50: 1180-1192.
Stevens, Kenneth N. (1985): “Evidence for the role of acoustic boundaries in the perception of speech sounds”. In: Fromkin V. A. (ed.), *Phonetic Linguistics*. New York: Academic, 243-255.
Stevens, Kenneth N. & Blumstein, Sheila E. (1978): “Invariant cues for place of articulation in stop consonants”. *Journal of the Acoustical Society of America* 64/5: 1358-1368.
Stevens, Kenneth N., Blumstein, Sheila E., Glicksman, Laura, Burton, Martha & Kurowski, Kathleen (1992): “Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters”. *Journal of the Acoustical Society of America* 91: 2979-3000.
Stevens, Kenneth N., Halle, Morris, Henke, William L., Klatt, Dennis H., Opperheim, Alan V., Ishida, Hideaki, Perkell, Joseph S. (1968): “Acoustic correlates of place of articulation for stops and fricative consonants”. *MIT Research Lab of Electronics Quarterly Progress Report* 89: 199-205.
Strevens, Peter (1960): “Spectra of fricative noise in human speech”. *Language and Speech* 3: 32-49.

Sussman, Harvey M., McCaffrey, Helen A. & Matthews, Sandra A. (1991): "An investigation of locus equations as a source of relational invariance for stop place categorization". *Journal of the Acoustical Society of America* 90: 1309–1325.

Tabain, Marija (1998): "Nonsibilant fricatives in English: Spectral information above 10 kHz". *Phonetica* 55: 107–130

Tomiak, Gail R. (1990): *An acoustic and perceptual analysis of the spectral moments invariant with voiceless fricative obstruents*, PhD dissertation, SUNY Buffalo.

Wagner, Anita, Ernestus, Mirjam. & Cutler, Anne (2006): "Formant transitions in fricative identification: the role of native fricative inventory". *Journal of the Acoustical Society of America* 120/4: 2267-2277.

Whalen, Douglas H. (1981): "Effects of vocalic transitions and vowel quality on the English [s]-[ʃ] boundary". *Journal of the Acoustical Society of America* 69: 275-282.

Whalen, Douglas H. (1983): "Vowel information in postvocalic fricative noises". *Language and Speech* 26/1: 91-100.

Whalen, Douglas H. (1991): "Perception of the English [s]-[ʃ] distinction relies on fricative noises and transitions, not on brief spectral slices". *Journal of the Acoustical Society of America* 90/4: 1776-1785.

Wilde, Lorin F. (1993): "Inferring articulatory movements from acoustic properties at fricative vowel boundaries". *Journal of the Acoustical Society of America* 94: 1881.

Wilde, Lorin F. (1995). *Analysis and synthesis of fricative consonants*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, U.S.A.

Wright, Richard (2001): "Perceptual cues in contrast maintenance". In: Hume Elizabeth & Johnson Keith (eds.). *The role of speech perception phenomena in phonology*. San Diego: Academic Press, 251-277.

Yeni-Komshian, Grace H. & Soli, Sigfrid D (1981): "Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation". *Journal of the Acoustical Society of America* 70: 966-975.

Zeng, FanGang. & Turner, Cristopher W. (1990): "Recognition of voiceless fricatives by normal and hearing-impaired subjects". *Journal of Speech, Language and Hearing Research* 33: 440–44.

CHAPTER 4. ENGLISH PLOSIVES: BEYOND THE [±VOICE] DISTINCTION

Joaquín Romero & **María Riera**
Universitat Rovira i Virgili

1. Introduction

1.1 Classification of voicing

Traditional phonemic approaches to sound classification rely on binary oppositions as a way to minimally distinguish between contrastive units. Phonemes and phonological features are identified and described precisely on the basis of their capability to enter into such semantically meaningful binary distinctions. Phonological features are therefore discrete units that are either present [+] or absent [–] in a particular sound, as in the case of voicing: sounds are *voiced* [+voice] or *voiceless* [–voice], but the theory does not contemplate the possibility of degree of voicing, that is, a sound cannot be more or less voiced than another. While there are clear advantages to such a phonemic view of sound structure, e.g., the possibility to characterize a wide variety of sounds in many languages using the same basic parameters, a binary opposition often falls short of its goal when applied to sounds that are phonetically complex. In fact, in many cases the binary approach does not just fail to describe phonetic detail appropriately, it also misrepresents reality by excessively simplifying the true nature of sound contrasts.

Most phonological features make more or less direct reference to speech articulation. In the case of voicing, a [±voice] specification implies the presence vs. absence of laryngeal activity in the form of modal vocal fold vibration. This distinction is probably found universally, though the extent to which it is used varies from language to language. A [+voice] specification is characteristically found in vowels and sonorant consonants such as approximants and nasals, whereas obstruent consonants—stops and fricatives—are often primarily [–voice]. However, [+voice] obstruents are common as well and, in fact, many languages distinguish, at least partially, between voiced and voiceless plosives, while the presence of voiced fricatives appears to be more restricted. According to the World Atlas of Linguistic Structures Online (http://wals.info), of 567 languages considered, only 158 contrast voicing in both plosives and fricatives and 182 have no voicing contrasts in plosives and fricatives at all, meaning that the voiced counterparts are typically absent.

While these figures are interesting and indeed show the wide occurrence of the voicing contrast, it is the case that it is often hard to tell exactly what counts as voiced and voiceless. For example, cross-linguistic observation shows that a voiced plosive in one language can actually be perceived as voiceless by speakers of a different language. Also, voicing often varies contextually within the same language, as shown by the common occurrence of processes such as final devoicing –voiced obstruents become devoiced in word final position– and a variety of voicing assimilations, by which the voicing of an obstruent will depend on the [±voice] specification of adjacent sounds. Finally, voicing can also vary within the same language depending on a variety of non-segmental issues, such as speaking rate, prosodic environment, emphasis, etc.

Given these considerations, it is important, when attempting to produce a detailed description of voicing in a language, to not be restricted to the limitations imposed by the binary phonemic opposition between [±voice]. This is indeed crucial in the case of languages like English. As will be shown in the remainder of this chapter, English does indeed show much of the variability pointed out above with respect to the realization of voicing. It is for this reason that a different, non-binary approach is proposed which can better describe the true nature of voicing in English without resorting to a degree of phonetic complexity that would make it inaccessible to most university students of the language.

The chapter is organized as follows. The remainder of section 1 provides an introduction to the inadequacies of a traditional phonemic account when it comes to describing the plosives of English. Section 2 introduces the notion of VOT and how it applies to the consonants under study. In section 3 other related processes affecting English plosives are presented and described. Section 4 offers some pedagogical and methodological appreciations in relation to the advantages of incorporating these phonetic aspects of English plosives in the teaching of the sound system of the language at university level. Finally, Section 5 wraps up the paper with some general conclusions.

1.2 Voicing in English plosives

In a standard binary phonemic classification of English plosives /b, d, g/ are classified as voiced while /p, t, k/ are identified as their voiceless counterparts. As regards place of articulation, /p, b/ are described as (bi)labial, /t, d/ as alveolar, and /k, g/ as velar. Without any further detail, it would also be assumed that these plosive sounds appear without variation in all phonotactic contexts: word initially, as in *bee*, *tea* and *go*, word internally as in *obey*, *adore* and *locker*, and word finally as in *up*, *lot* and *rag*. In addition, prosodic aspects such as the presence vs. absence of stress are not reflected in this simple classification, and therefore the same phonetic symbols would be used to represent, for example, /t/ in *atom* vs. *atomic*, or /k/ in *soccer* vs. *occur*. While this kind of very basic description might be appropriate as a first approximation into the nature of English plosives, it soon becomes apparent that the /t/ of *atomic* sounds conspicuously different from the /t/ of *atom*, just as the /k/ of *occur* is unequivocally distinct from the /k/ of *soccer*.

Thus, a description of English plosives that fails to recognize these obvious differences clearly seems insufficient for anyone who is interested in exploring the true nature of these sounds in all their complexity. A clear understanding of the rules underlying the variation of English plosives is also a necessary step towards achieving a proficient command of the spoken language for those who are learning English as a second or foreign language. More importantly even, it is essential that those who are being trained as future teachers of English do not simply rely on their phonetic perception and/or production abilities when it comes to teaching the sound system of the language. A sound knowledge of the nature of English plosives will undoubtedly facilitate their own command of the spoken language while, at the same time, making the teaching task much more effective.

1.3 Phonemic vs. phonetic descriptions

As we mentioned above, the binary distinction which is implicit in the use of the simple voiced/voiceless contrast is typical of traditional phonemic accounts. These accounts are more concerned with determining which phonological features can be identified to play a role in minimally distinguishing pairs of words in any particular language, than in describing the full range of possible sounds that occur in that language, whether they are phonologically contrastive or not. A more comprehensive description which considers the sounds of a language in all their fine detail necessarily has to move beyond the realm of binary phonemic classification and move into the phonetic component.

When it comes to dealing with English plosives, authors who are not satisfied with the basic binary phonemic classification often expand on their description by introducing a range of distributional allophonic variants which are conditioned by phonotactic and/or prosodic factors. These typically make reference to non-contrastive parameters such as aspiration, types of release, weakening and deletion, glottalization, etc. Of all these, aspiration is perhaps dealt with most commonly, since it is such a salient and easily identifiable feature of English plosives from an auditory point of view. The discussion of the presence vs. absence of aspiration, however, is often restricted to the set of voiceless plosives, as illustrated for the velar plosive /k/ by pairs such as *occur* vs. *soccer*, where the first one would be aspirated and the second one would be unaspirated; this distinction hinges on whether the stop appears in a stressed vs. an unstressed syllable. This, however, misses an important generalization, which is that aspiration is not an isolated phenomenon but rather one of the outcomes of a more far-reaching principle

of articulatory organization, in particular, the temporal coordination between onset of laryngeal vibration and supralaryngeal consonantal release.
Thus, as we have seen, the inclusion of the allophonic or phonetic level adds much valuable depth to the description of the plosives of English. Given the complexities of these sounds, a finer description is not only desirable but essential if we are to offer a faithful picture of the sound system of the language.

2. VOT

2.1 Voice Onset Time as spatiotemporal parameter

In their seminal 1964 paper on initial stop voicing, Leigh Lisker and Arthur Abramson introduced the term *Voice Onset Time* (commonly abbreviated to VOT) to refer to a parameter that measured the temporal relationship between the beginning of vocal fold vibration for the vowel and the release of the oral constriction for the consonant in CV sequences. The study found that different languages showed different patterns of coordination for the different plosives in their systems and that these patterns strongly correlated with well-known phonemic/allophonic categories traditionally acknowledged for these consonants. They investigated languages with two sets of plosives (Spanish, Cantonese, English), three sets of plosives (Korean, Thai, Armenian) and four sets of plosives (Hindi, Marathi). The study concluded that a single dimension, VOT, could in fact account for the differences in most of the categories in the languages under study, independently of whether they were considered phonemic or allophonic. In addition, VOT could easily be used to describe the phonetic details of most of the consonants and to show how even languages with the same number of plosive categories differed in their phonetic realizations.
Figure 1 shows a schematic representation of how VOT works. The different panels provide an approximation to the temporal coordination between onset of laryngeal voicing and oral release for the following four commonly found stop categories: voiceless unaspirated, voiceless aspirated, voiced unaspir-

ated and voiced aspirated. The categories are illustrated for the labial set in each case but the general principle applies equally to other possible places of articulation. The top left panel shows the simplest case, where the onset of voicing for the vowel and the oral release occur more or less simultaneously, resulting in the category identified as voiceless unaspirated and represented phonetically as [p], [t] and [k]. The top right panel shows a configuration where the oral release precedes the onset of voicing for the vowel. In this situation, the time between these two events is usually referred to as *voice lag*, implying precisely that the onset of voicing lags behind the release of the oral constriction. It is assumed that air from the supraglottal cavity begins to flow through the larynx immediately following the oral release and before the onset of vocal fold vibration, which results in a period of audible voiceless noise that is usually defined as aspiration.

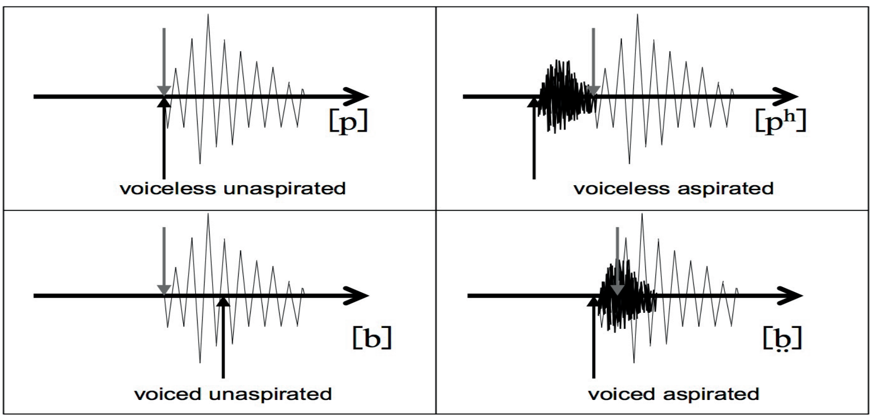


Figure 1. Exemplification of stop consonant category distribution in CV sequences based on *Voice Onset Time*. The horizontal arrow indicates time. The downward arrow identifies the onset of laryngeal vibration for the vowel while the upward arrow indicates the point of oral release. The darker areas represent the presence of audible laryngeal noise.

Hence the identification of this category as voiceless aspirated, which is customarily represented in transcription with as superscript ^{h}, as in [p^h], [t^h] and [k^h]. The two bottom panels show categories that are usually identified as *voiced*. On the left hand side the figure shows a situation where the onset of voicing for the vowel precedes the release of the oral constriction, which results in a short period of vocal fold vibration while the supraglottal constriction is still firmly sealed. This situation is known as *voice lead* and is represented by the standard voiced unaspirated symbols [b], [d] and [g]. The fourth configuration, bottom right, also shows a certain degree of voice lag and audible noise, as in the voiceless unaspirated consonants above. However, the audible noise occurs to a large extent concomitantly to the vibration of the vocal folds for the vowel. This is an uncommon situation which is usually perceived as containing varying degrees of breathy or murmured voice, depending on the specific language. According to Lisker & Abramson (1964) this breathy or murmured voice is achieved by the fact that a small opening is maintained between the arytenoid cartilages which allows air to flow through while simultaneous regular modal voice is produced at the vocal folds. These consonants are known as voiced aspirated and are represented in standard IPA symbols as [b̤], [d̤] and [g̈].

The four configurations illustrated in Figure 1 correspond to the most common plosive consonant phonemic categories that are found in languages of the world. However, not all languages with the same number of categories in their inventory necessarily employ the same two VOT configurations when it comes to their phonetic implementation. This issue becomes particularly revealing when we attempt to describe the plosives of English in comparison to those of languages like Spanish or French. As was mentioned in the Introduction, the standard phonemic description of English plosives opposes voiceless (/p/, /t/, /k/) to voiced (/b/, /d/, /g/) consonants. Following this classification, one might be tempted to conclude that the VOT configurations that correspond to these categories are the voiceless unaspirated [p], [t] and [k] and the voiced unaspirated [b], [d] and [g], respectively. This would in fact be identical to the situation in Spanish or French. However, even the least sophisticated listener would clearly perceive that the initial consonants in English *pan*, *tan* and *can* do not sound like those of Spanish *pan*, *tan* and *can*. Someone with a slightly more trained ear would equally recognize that the initial plosives in English *ban, dad and gas* also do not usually sound like the initial plosives in Spanish *van*, *dad* and *gas*. In the following section we analyze in detail how VOT is useful to establish the true phonetic nature of English plosives and how these differ from the plosives of Spanish.

2.2 English aspiration and devoicing

The following words are examples of English and Spanish voiceless plosives in word initial position[1]:

	English	Spanish
Labial	*pore* /pɔr/	*pon* /pon/
Alveolar	*tore* /tɔr/	*ton* /ton/
Velar	*core* /kɔr/	*con* /kon/

Here the standard phonemic transcription (voiceless vs. voiced) is used to illustrate that these words are all minimally distinguished by the initial consonants. As stated above, however, we do not need a particularly phonetically sophisticated listener to perceive a clear difference between the pronunciation of the initial stops in English *pore*, *tore* and *core* and those in Spanish *pon*, *ton* and *con*. Consideration of the four VOT configurations shown in Figure 1 and described above clearly indicate that the plosives in English *pore*, *tore* and *core* are best characterized as *voiceless aspirated*. Indeed, their production involves a relatively long period of voice lag, that is, the oral release takes place well before the onset of voicing for the vowel; in addition, there is clear audible noise between the two events, which is what we usually describe as *aspiration*. The Spanish consonants, on the other hand, do not show much or any voice lag at all and certainly do not carry any concomitant aspiration; for these reasons these would indeed be best classified as *voiceless unaspirated*. Thus, a phonetic representation ought to reflect these differences:

	English	Spanish
Labial	*pore* [p^hɔr]	*pon* [pon]
Alveolar	*tore* [t^hɔr]	*ton* [ton]
Velar	*core* [k^hɔr]	*con* [kon]

Consider now the following set of examples of English plosives:

Labial	*spore* /spɔr/

[1] Throughout this chapter examples and pronunciations are given in General American English. Accordingly, transcriptions are provided following the system used in Kenyon & Knott (1953).

Alveolar	*store* /stɔr/
Velar	*score* /skɔr/

The similarity between these words and the previous set of English words is clear, the only difference being the presence of an /s/ in front of the plosives. However, careful pronunciation of these sequences reveals a fundamental difference: the plosives in *spore*, *store* and *score* do not show signs of voice lag and/or aspiration and therefore cannot be classified as *voiceless aspirated*. Instead, these plosives would be better characterized as *voiceless unaspirated*. The reasons why /p/, /t/ and /k/ are realized as unaspirated plosives after /s/ are complex and have to do with the presence of strong turbulent air and resulting noise associated with the fricative /s/. A more detailed explanation of this phenomenon would be beyond the scope of this chapter, but readers are referred to Catford (1977) for a thorough account of the interaction between friction and aspiration in /sC/ sequences in English. We can conclude then that a phonetic representation of these sequences would be as follows:

Labial	*spore* [spɔr] not *[spʰɔr]
Alveolar	*store* [stɔr] not *[stʰɔr]
Velar	*score* [skɔr] not *[skʰɔr]

It is interesting to note, incidentally, that words with /s/ plus voiced plosives are not attested in English. Thus, there are no such words as **sbore*, **sdore* or **sgore* in the language. We will argue below that these are not only just accidental gaps in the English lexicon, but they are in fact impossible words and that this phonotactic restriction is narrowly related to the nature of VOT.
From the examples we have seen so far one might be inclined to conclude that English /p, t, k/ are realized as voiceless aspirated plosives in word initial position except when they are preceded by /s/, in which case the realization is as voiceless unaspirated. The reality, however, is more complex than that, as illustrated by the following sets of words:

Labial	*appear* /ə'pir/ → [ə'pʰir]

Alveolar	*attack* /əˈtæk/ → [əˈtʰæk]
Velar	*accord* /əˈkɔrd/ → [əˈkʰɔrd]

These examples show an important aspect of the phonetic implementation of plosives in English: its intricate relationship with stress. The plosives in these words are clearly realized as voiceless aspirated even though they do not appear in word initial position. It is the fact that they appear in the onset of a stressed syllable that is the crucial aspect in their realization as aspirated stops. Notice the exact correspondence between the words above and the phonologically identical sequences below:

Labial	*appear* /əˈpir/ → [əˈpʰir]	vs.	*a pier* [ə ˈpʰir]
Alveolar	*attack* /əˈtæk/ → [əˈtʰæk]	vs.	*a tack* [ə ˈtʰæk]
Velar	*accord* /əˈkɔrd/ → [əˈkʰɔrd]	vs.	*a chord* [ə ˈkʰɔrd]

The phonetic description of English phonemically voiceless plosives is completed with the pattern illustrated by the following examples:

Labial	*supper* /ˈsʌpɚ/ → [ˈsʌpɚ]
Alveolar	*lettuce* /ˈlɛtəs/ → [ˈlɛtəs]
Velar	*locket* /ˈlɑkɪt/ → [ˈlɑkɪt]

In these words the plosive consonants appear in the onset of unstressed syllables and are, therefore, realized as voiceless unaspirated, which points at the tight relationship between stress and the presence of aspiration in English voiceless plosives. This constitutes a key issue in the phonology of English which, unfortunately, is often overlooked in general manuals of English phonetics. While the relationship between stress and vowel quality is generally given ample attention, the consequences of the presence vs. absence of stress for the phonetic realization of consonants is seldom dealt with in depth and often fully ignored. This circumstance is very unfortunate, as a thorough understanding of the intimate connection between stress and the phonetic nature of both vowels and consonants in English is crucial in the development of a proficient command of both the production and the perception of the language by foreign learners.

We now turn our attention to the so-called voiced plosives /b, d, g/. Let us begin by comparing similar words in English and Spanish, just as we did for the voiceless ones:

	English	Spanish
Labial	*boar* /bɔr/	*vas* /bas/
Alveolar	*door* /dɔr/	*das* /das/
Velar	*gore* /gɔr/	*gas* /gas/

Again, just as we observed for the voiceless plosives, at first sight the two sets of consonants are equivalent in the two languages. Unlike the situation with voiceless /p, t, k/, however, the differences between the two languages may not be as obvious, since the presence vs. absence of aspiration is perceptually very salient. Closer observation reveals consistent divergences, though, which are more easily understood with reference to VOT. Referring back to the patterns in Figure 1, a comparison between the /b/ in Spanish *vas* vs. English *bore* shows that in the Spanish consonant there is a significant amount of voice lead, i.e., a short but easily noticeable period of vocal fold vibration prior to the release of the oral constriction. In English, on the other hand, there is a significantly shorter voice lead or none at all. The logical conclusion, then, is that we should classify Spanish /b, d, g/ as *voiced unaspirated*, whereas the corresponding English consonants are better identified as *voiceless unaspirated*. Accordingly, an accurate phonetic transcription of the items above should be as follows:

	English	Spanish
Labial	*boar* [pɔr]	*vas* [bas]
Alveolar	*door* [tɔr]	*das* [das]
Velar	*gore* [kɔr]	*gas* [gas]

The appropriateness of this transcription and its effectiveness in reflecting the phonetic reality of English phonemically voiced plosives is further supported when we compare these observations with the situation shown above in the /sC/ sequences illustrated by the words *spore*, *store* and *score*.

It was shown there that in these contexts /p, t, k/ are realized as voiceless unaspirated plosives. According to this, then, there should be no difference between /p, t, k/ in *spore*, *store*, *score* and /b, d, g/ in *bore*, *door*, *gore*, respectively. This is indeed the case, as illustrated by the homophonous sequences in the following examples:

Labial	*this spore* ['ðɪs 'spɔr]	vs.	*this boar* ['ðɪs 'pɔr]
Alveolar	*this store* ['ðɪs 'stɔr]	vs.	*this door* ['ðɪs 'tɔr]
Velar	*this scale* ['ðɪs 'skel]	vs.	*this gale* ['ðɪs 'kel]

Further perceptual confirmation of the equivalence between the two sets of consonants can be obtained with a quick and simple experiment consisting in instrumentally removing the /s/ portion of the acoustic signal corresponding to a word like *store* and asking subjects to identify the remaining portion of the signal: invariably they will identify it as *door*. Incidentally, this equivalence is at the root of the lack of */sb/, */sd/ and */sg/ sequences in English that we mentioned above, since, given the situation just presented, there would be no possible differentiation between */sp/ vs. /sb/, */st/ vs. /sd/ and */sk/ vs. /sg/. The fact that pairs of words like *discussed* vs. *disgust* are completely homophonous for many speakers of English perfectly illustrates this situation.
Having said this, however, and in order to avoid overt confusion when transcribing /b, d, g/ phonetically, it is customary to represent these consonants as *devoiced* voiced stops rather than as voiceless unaspirated stops, as illustrated below:

Labial	*spore* [spɔr]	vs.	*boar* [b̥ɔr]
Alveolar	*store* [stɔr]	vs.	*door* [d̥ɔr]
Velar	*score* [skel]	vs.	*gore* [g̊ɔr]

Although this may sound like an unnecessarily convoluted way of approaching the phonetic nature of /b, d, g/ in English, on further reflection it makes very good sense, as it again relates the realization of English plosives to stress. Compare the examples we have just seen with the following:

Labial *abound* [ə'b̥aʊnd] vs. *habit* ['hæbɪt]
Alveolar *adore* [ə'd̥ɔr] vs. *edit* ['ɛdɪt]
Velar *again* [ə'g̊ɛn] vs. *soggy* ['sɑgi]

Now we see how, just as the presence vs. absence of stress determined whether /p, t, k/ surfaced as aspirated vs. unaspirated plosives, respectively, equally stress is at the root of the alternation between devoiced or fully voiced /b, d, g/. When these appear in the onset of a stressed syllable, regardless of whether the syllable is word initial or not, they are devoiced.

Table 1. Summary of allophonic variation of English plosives as a function of word position and stress

		voiceless			Voiced			
		initial	not initial	/sC/		initial	not initial	/sC/
lab	Stressed Aspirated	*pore* [pʰɔr]	*appear* [ə'pʰir]	*spore* [spɔr]	stressed devoiced	*boar* [b̥ɔr]	*abound* [əb̥aʊnd]	*
	Unstressed Unaspirated	*parole* [pə'rol]	*supper* ['sʌpɚ]	*Aspen* ['æspən]	unstressed fully voiced	*below* [bɪ'lo]	*habit* ['hæbɪt]	*
alv	Stressed Aspirated	*tore* [tʰɔr]	*attack* [ə'tʰæk]	*store* [stɔr]	stressed devoiced	*door* [d̥ɔr]	*adore* [əd̥ɔr]	*
	Unstressed Unaspirated	*terrain* [tə'ren]	*lettuce* ['lɛtəs]	*Easter* ['istɚ]	unstressed fully voiced	*decide* [dɪ'saɪd]	*edit* ['ɛdɪt]	*
vel	Stressed Aspirated	*core* [kʰɔr]	*accord* [ə'kʰɔrd]	*score* [skɔr]	stressed devoiced	*gore* [g̊ɔr]	*again* [ə'g̊ɛn]	*
	Unstressed Unaspirated	*convey* [kən'veɪ]	*locket* ['lɑkɪt]	*asking* ['æskɪŋ]	unstressed fully voiced	*galore* [gə'lɔr]	*soggy* ['sɑgi]	*

Conversely, if the syllable is not stressed, /b, d, g/ are realized as fully voiced, that is, the VOT category is identified as voiced unaspirated.
This can be taken as further evidence of the crucial role that stress plays in the phonology of English. Just as the presence of stress in a syllable makes

its vowel stronger and more salient in all aspects (louder, longer, higher in pitch and with a well-defined spectral pattern) as opposed to vowels in unstressed syllables, which are characteristically weak, short and centralized, plosives in stressed syllables are equally strong, i.e., aspirated if they are phonemically voiceless and devoiced if they are phonemically voiced. This contrasts with the realization of plosives in unstressed syllables, where /p, t, k/ are unaspirated (and, as we will see below, can at least in some instances undergo further processes of weakening) and /b, d, g/ are fully voiced. Table 1 summarizes the phonetic realization of English plosives according to the parameters discussed here, i.e., stress and word position.

Before moving on to less general phenomena affecting a more restricted subset of plosives, it is worth taking a look at the situation in yet one more context that is not so directly affected by stress: absolute final position. Consider the examples below:

	Voiceless	Voiced
Labial	*sap* /sæp/	*sab* /sæb/
Alveolar	*sat* /sæt/	*sad* /sæd/
Velar	*sack* /sæk/	*sag* /sæg/

The VOT configurations shown in Figure 1 clearly cannot be applied to these cases straightforwardly, since the plosives here are timed not with the onset of the vowel but rather with its offset. This fact precisely, the timing of the plosive closure with the end of the preceding vowel, determines in large part the phonetic realization of these consonants in absolute final position. Since there is no following vowel, there is likely to be no audible release (or no release at all), which is why plosives in these consonants are often described as *unreleased* and transcribed as shown in the following set of words:

	Voiceless	Voiced
Labial	*sap* /sæp/ → [sæp̚]	*sab* /sæb/ → [sæb̚]
Alveolar	*sat* /sæt/ → [sæt̚]	*sad* /sæd/ → [sæd̚]
Velar	*sack* /sæk/ → [sæk̚]	*sag* /sæg/ → [sæg̚]

Thus, even though unreleased stops in absolute final position do not strictly follow the basic patterns of VOT distribution, we can see that the same underlying principle is at play here, i.e., the relative timing and coordination of glottal and supraglottal events.
This section has shown how the complexity of plosive consonant phonetic realization in English can be described in a rather simple and elegant manner by using a single phonetic parameter, Voice Onset Time. In addition to that, we have shown how this spatiotemporal dimension provides equally accurate explanations as to how similar consonants are distributed allophonically in other languages. This cross-linguistic generalization is of particular interest, as we will see below, when speakers of languages with a different distribution of VOT categories attempt to acquire a proficient command of English plosives.

3. Other processes affecting plosives

The special status of alveolar (or more generally *coronal*) consonants in many languages of the world has long been recognized (Paradis & Prunet, 1991; Hall, 1997) both phonologically and morphologically. For example, in English they play a crucial role in nominal as well as verbal inflexion: alveolar fricatives /s/-/z/ are responsible for plural, third person singular and genitive inflexions; alveolar plosives /t/-/d/ indicate past tense formation in weak verbs. From a phonological point of view English /t/ (and to a lesser extent /d/) is special in the sense that it behaves in ways that are exclusive to it and different from the other plosives, as will be shown below.
In the previous section we showed how the phonotactic variability of English plosives is intricately related to stress. The connection between these two aspects is particularly relevant when it affects the alveolar subset of plosives, i.e., /t/ and /d/. In addition to all the allophonic variations described above, these consonants, especially /t/, undergo further processes of articulatory weakening and reduction which result in some of the most perceptually salient features of (at least some) varieties of English. We will explore now three of those processes: flapping, glottalization and deletion.

3.1 Flapping

Flapping is a process by which an alveolar plosive, /t/ or /d/, becomes an alveolar flap, represented by the standard IPA symbol /ɾ/. Different theories have been proposed to explain how exactly the change from plosive to flap occurs, but it seems reasonable to believe that it is an instance of articulatory weakening or reduction, whereby the tight, long occlusion associated with the plosive is replaced by a looser, shorter contact of the tongue tip against the alveolar ridge. Even though the exact nature of the resulting flap can vary as a factor of surrounding context, style, dialect, etc., the most common result in standard General American English is a sound that resembles the intervocalic single *r* of Spanish, as in *cara* /'kaɾa/.
Understanding flapping as a weakening process does not only seem appropriate from an articulatory point of view, it also makes sense if we consider the contexts where it occurs in English. Below are the general conditions that result in flapped /t/ and /d/:

(1) /t/ or /d/ must appear in the onset of an unstressed syllable[2]
(2) they must be preceded by a vowel or /r/
(3) they must be followed by a vowel

Condition (1) illustrates best the weakening nature of the flapping process: it can only occur in unstressed syllables. Compare the following pairs of words:

atomic [ə'tʰɑmɪk]	vs.	*atom* ['æɾəm]
metallic [mə'tʰælɪk]	vs.	*metal* ['mɛɾəl]
medallion [mə'd̥æljən]	vs.	*medal* ['mɛɾəl]
modality [mo'd̥æləɾi]	vs.	*modal* ['moɾəl]

In the words on the left hand side /t/ and /d/ appear in the onset of stressed syllables and are thus realized as aspirated and devoiced, respectively,

[2] This condition may not apply when /t/-/d/ occur in word-final position followed by a stressed word, especially in phrasal verbs such *get in* or *shut up*, where /t/ is also commonly flapped.

according to the distributions that we saw in Section 2 above. The related words on the right hand side, on the other hand, show /t/ and /d/ in the onset of unstressed syllables, which is why they are flapped and turned, in all instances, into [ɾ]. In all of these words, conditions (2) and (3) are also fulfilled, since /t/ and /d/ are always preceded and followed by vowels.

A couple of interesting observations can be made regarding the words above. First, notice that in the word *modality* /d/ is not flapped because, as we said, it appears in the onset of a stressed syllable. The /t/ in the final syllable of the same word, however, is flapped by virtue of appearing in an unstressed syllable preceded and followed by vowels. Thus, flapping can occur regardless of whether the preceding syllable is stressed or not. In fact, more than one instance of flapping can occur within the same word, as in the words *editor* or *competitor*, which, in colloquial speech, would be pronounced [ˈɛɾɪɾɚ] and [kəmˈpʰɛɾɪɾɚ], respectively. Second, because both /t/ and /d/ undergo flapping in the contexts mentioned earlier, cases of neutralization arise, as in the words *metal* vs. *medal* above. For a majority of speakers of General American English, these two words are completely homophonous. For others, a difference is maintained which is not based on the plosives themselves but on the length of the preceding vowel which, as expected, is greater in the case of the original voiced consonant (*medal*) than in the original voiceless consonant (*metal*) (Yavaș, 2006). Pairs of this type are not unusual in the language, as for example *writer* vs. *rider*, *petal* vs. *pedal*, *liter* vs. *leader*, *latter* vs. *ladder*, etc.

In all the examples we have seen so far flapped /t/ or /d/ occur after vowels. The only other possible consonant that satisfies the flapping conditions is /r/, as illustrated by the following examples:

articulate [ɑrˈtʰɪkjələt]	vs.	*article* [ˈɑrɾɪkl]
particular [pɚˈtʰɪkjəlɚ]	vs.	*particle* [ˈpɑrɾɪkl]
ordain [ɔrˈd̥en]	vs.	*ordinal* [ˈɔrɾɪnəl]
gardenia [gɑrˈd̥iniə]	vs.	*garden* [ˈg̊ɑrɾən]

Now consider the following examples, where flapping does not apply:

alternative [ɔlˈtʰɝnəɾɪv]	vs.	*alternate* [ˈɔltɚnət]
instantiate [ɪnˈstænʃiˌet]	vs.	*instance* [ˈɪnstəns]

Andean [æn'd̥iən]	vs.	*Andes* ['ændiz]
vindictive [vɪn'd̥ɪktɪv]	vs.	*vindicate* ['vɪndɪˌket]
Britannia [brɪ'tʰænjə]	vs.	*Britain* ['b̥rɪtn]
sentential [sɛn'tʰɛnʃl]	vs.	*sentence* ['sɛntns]

In the first four cases flapping does not occur because, even though /t/ and /d/ appear in the onset of unstressed syllables, the preceding sounds (/l/, /s/, /n/) all involve movement of the tongue tip to create a complete constriction (in the case of /l/and /n/) or a narrow constriction (in the case of /s/), which prevents the formation of the characteristic flapping movement of [ɾ]. In the last two examples, it is the fact that the following sounds also involve a complete occlusion by the tongue tip (syllabic /n/ in this case) that prevents flapping from taking place. In this respect, an interesting asymmetry occurs with /l/. As we have just seen, /l/ blocks flapping when it precedes unstressed /t/ or /d/. However, in syllable final position /l/ (syllabic /l/ in this case), unlike syllable final /n/, may not bock flapping, as in *subtle* ['sʌɾl] or *paddle* ['pʰæɾl]; this would equally apply, in a relaxed pronunciation, to the words *metal* and *medal* that we saw earlier[3].

Finally, it is worth noticing that flapping can also occur in very common unstressed monosyllabic words, most conspicuously in the weak particle *to* or the auxiliary verbs *do* and *did*, when they occur in a phrasal context that fulfills the flapping requirements, as in the following examples:

go to work	['go ɾə 'wɝk]
far to see	['fɑr ɾə 'si]
how do you say?	['haʊ ɾə jə 'se]
where did you stay?	['wɛr ɾɪd jə 'ste]

Even though flapping, especially in the extensive sense presented here, is most commonly identified with North American English, it is by no means exclusive of these varieties of the language. Australian and New Zealand English flap consistently in pretty much the same contexts as North

---

[3] In those varieties of the language in which flapping is not as common or widespread, as in standard British English, /t/, and to a lesser extent /d/, in these words would be pronounced with a lateral release, which, in the case of /t/, can be accompanied by varying degrees of affrication.

American English does, except, of course, after /r/, these being non-rhotic dialects. But flaps are also not uncommon in the speech of many British speakers, especially of the younger generations, for which a flapped /t/ in words like *butter* represents a compromise between the formal pronunciation with a complete plosive and the markedly colloquial pronunciation with a glottal stop, which we will briefly review in the following section.

3.2 Glottalization

Similarly to flapping, glottalization is a phenomenon that is closely related to the absence of stress and, as such, can also be considered as an instance of articulatory weakening or reduction. Unlike in flapping, though, how a glottal stop is a weaker version of a coronal plosive is not necessarily as obvious. A detailed description of how the relationship between glottal adduction and oral constriction combine in the process of glottalization and how this can indeed be seen as an example of articulatory weakening would require more phonetic detail than can be provided here. However, in so far as we can see that glottalization involves the reduction or loss of the oral component of a plosive, we can feel comfortable classifying this process as an instance of weakening. Also, unlike in the case of flapping, glottalization affects only /t/, not /d/, which follows naturally from the fact that glottal closure and modal voicing are in principle incompatible. Thus, glottalization in its full realization involves the substitution of a closure or narrowing of the glottis for the alveolar plosive /t/ (Docherty & Foulkes, 1999).
Consider again some of the examples that we saw in Section 3.1 as cases where flapping was blocked by the presence of a following syllabic /n/:

Britannia [brɪˈtʰænjə]	vs.	*Britain* [ˈb̥rɪtn]
sentential [sɛnˈtʰɛnʃl]	vs.	*sentence* [ˈsɛntns]

In the words in the right hand side column /t/ cannot become a flap because it is followed by syllabic /n/ ([n̩]). This is exactly the context for glottalization in standard General American English: /t/ appears in the onset of an unstressed syllable followed by syllabic /n/. The exact phonetic outcome of glottalization can actually vary depending on a number of factors such as speech rate, style, geographic and social variety, etc., as illustrated below:

Britain ['b̥rɪtn̩] → ['b̥rɪtˀn̩] / ['b̥rɪʔn̩]
sentence ['sɛntn̩s] → ['sɛntˀn̩s] / ['sɛnʔn̩s]

These examples show two common realizations of glottalized /t/ in American English, the first one being a case of glottalized /t/, in which the release of the plosive is accompanied by a narrowing or closure of the vocal folds but still with a clearly noticeable alveolar contact, whereas the second one shows a more extreme case of glottalization in which the oral component has disappeared and it has been replaced completely by a glottal stop. Both realizations are common in everyday colloquial American English, even though the latter is often associated with careless speech and, as such, often frowned upon by prescriptive grammars.
Another relevant aspect of glottalization is that, unlike flapping, it can apply independently of the nature of the preceding segment, as illustrated by the following examples, in which glottalized /t/ occurs after a variety of obstruent and sonorant consonants:

carton ['kʰɑrtn̩] → ['kʰɑrtˀn̩] / ['kʰɑrʔn̩]
Bolton ['b̥oltn̩] → ['b̥oltˀn̩] / ['b̥olʔn̩]
fountain ['faʊntn̩] → ['faʊntˀn̩] / ['faʊnʔn̩]

Finally, even though the focus of this chapter is primarily with American English, it is worth noticing that glottalization of /t/ is a much more widespread phenomenon in other varieties of English, particularly among speakers of southeastern British English and especially London English, for whom pronunciations like *butter* ['b̥ʌʔə] or *forty* ['fɔːʔi] are not uncommon. Moreover, the enormous influence of this variety of English in the media is causing a generalization of this phenomenon among younger speakers in many other parts of the country.

3.3 Deletion

The last type of weakening of alveolar plosives that we will be reviewing is deletion of /t/. In fact, this represents the ultimate stage in the process of weakening or reduction associated with lack of stress, since, as its name indicates, it involves a complete disappearance or deletion of /t/. The context for /t/ deletion is again the onset of an unstressed syllable where the consonant is preceded by /n/ and followed by a vowel. Below are some examples that are extremely common in colloquial American English:

twenty ['tʰwɛnti] → ['tʰwɛni]
center ['sɛntɚ] → ['sɛnɚ]
internet ['ɪntɚˌnɛt] → ['ɪnɚˌnɛt]
Toronto [tə'rɑnto] → [tə'rɑno]
sentimental [ˌsɛntɪ'mɛntəl] → [ˌsɛnɪ'mɛnəl]

It is important to emphasize that these pronunciations do not constitute overly colloquial forms in the everyday speech of most speakers of North American English. At the same time, however, it is true that, unlike the allophones of the plosives that we saw in Section 2, deletion is an optional change that speakers can control at will, perhaps to a higher degree than flapping or glottalization. Thus, whereas the pronunciation of the word *party* as ['pʰɑrti] instead of the more common ['pʰɑrri] would sound clearly emphatic and contrived, the pronunciation of *winter* as ['wɪntɚ] instead of ['wɪnɚ] has no such connotations.

4. Pedagogical considerations

The account of English plosives that has been presented in this chapter is clearly much more extensive than those provided by most common introductory manuals of English phonetics and phonology. In addition, it requires familiarity with some aspects of speech production that, though basic, may appear unnecessarily complex to instructors of English phonetics and phonology courses at university level. Also, in order to take full advantage of the details of this account, students need to become familiar with narrow transcription, which certainly adds an extra layer of complexity to the learning process. We believe, however, that the extension and the complexity represent a small price to pay in exchange for a much more realistic view of English plosives than the oversimplified voiced/voiceless opposition that most manuals provide. In our experience, none of these apparent drawbacks constitute a problem at all for university level students.

One of the major frustrations that instructors of English phonetics and phonology experience when teaching this discipline to learners of the language is the students' inability to bridge the gap between the theoretical phonological description of the sound system and their own capacity to pronounce the language according to this description. Thus, even though students often have no major difficulty learning the mechanics of English transcription, they often fail to apply those principles to their speech. This, in our opinion, constitutes a significant flaw in the teaching system, since the purpose of a course in English phonetics and phonology should not be just theoretical, it should also be useful to improve the students' command of the spoken language.

The introduction of narrow transcription in the English phonetics and phonology class along the lines of what has been presented in this chapter is a tool that can help alleviate the shortcomings we just mentioned. Too often we leave it up to our students' aptitude or *good ears* to be able to infer the type of systematic allophonic variation that occurs in the plosives of English and to subsequently apply it to their own oral productions of the language. Too often that is a recipe for failure. An explicit review of the specific allophones and their distribution as well as intensive practice in narrow transcription facilitates this process of awareness and internalization tremendously. A student's ability to distinguish between aspirated, unaspirated, devoiced and unreleased stops in their productions of English is a major factor in their intelligibility, probably comparable to being able to differentiate between the various vowel categories of the language.

A collateral advantage of dealing with the allophonic variation of English plosives in detail is the reinforcement of the crucial importance of stress in the phonology of the language. Most manuals of English phonetics and phonology do indeed emphasize stress as a key aspect and building block in the understanding of the sound system of the language. However, as we mentioned earlier, most of these manuals spend a considerable amount of time reviewing the effects of the presence vs. absence of stress in vowels and in the overall stress patterns of strong and weak forms, complex word formations, etc., but few devote much attention to the correlates of stress in the consonantal system, especially the plosives. By emphasizing the segmental implications of stress in the plosives, we are able to provide an even more far reaching and comprehensive view of the relevance of stress in the phonology of the language, while at the same time improving our students' chances at a proficient command of spoken English.

Finally, an account of English plosives that also takes into consideration allophonic variations that may be more dialect specific or paradigmatically restricted provides a more realistic description of the language, in line with the oral input that our students are constantly exposed to in their interactions with English, be it via movies, music videos, documentaries, TV series, video games, etc. The fact that this huge amount of input is so readily available nowadays is a tremendous asset that instructors of English phonetics and phonology need to take advantage of and exploit in their attempt to improve their students' use and understanding of the spoken language. Any minimally perceptually adept student watching an American movie or TV show necessarily becomes aware of the overwhelming presence of flapping, glottalization and deletion in the speech of native speakers. Being exposed to this type of phenomena in the English phonetics and phonology class and becoming familiar with the distribution and nature of these sounds maximizes the chances that students will not only be more easily understood by native speakers, they will also improve their listening skills significantely. For that reason, it is important to provide as accurate a description of the true nature of plosives in English, even down to the detail of geographic and social variation.

5. Conclusions

In this chapter we have advocated for a description of English plosives that takes into consideration phonetic as well as phonological factors. Thus, we have shown how certain simple phonetic parameters, specifically voice onset time (VOT), can be used to provide a more faithful description of the true nature of English plosives than is presented in most current introductory manuals of English phonetics and phonology aimed at university students.

We have shown how characteristic properties of these sounds such as aspiration and devoicing can be easily explained with reference to this simple spatiotemporal parameter. A comparison with plosives in other languages further reveals the usefulness of the VOT parameter to illustrate the variability found in English. In addition, we have emphasized the relationship between the variability in the phonetic realization of English plosives and stress, a factor that is known to play a crucial role in many aspects of the language's sound system.

This intimate relationship between segmental and suprasegmental aspects of the language has been further reinforced by describing other characteristics of English plosives that are seldom dealt with in much detail, such as flapping, glottalization and deletion. As we mentioned above, it is believed that an overview of these phenomena helps to give students a more realistic vision of English plosives than the oversimplified voiced/voiceless opposition. The obvious drawback of this approach is the inevitable introduction of a certain degree of phonetic detail and complexity. In our experience, however, this does not represent a problem at all for university level students who are capable of bridging the gap between the theoretical description of the language and the phonetic detail and thus benefit from a more thorough description. The pedagogical benefits of introducing a certain degree of fine phonetic detail in the description of English plosives clearly outnumber the difficulties in terms of maximizing the development of a more accurate command of English pronunciation.

Acknowledgments

Research Group in Experimental Phonetics (Universitat Rovira i Virgili, Tarragona, Spain)

Research Groups 2005-SGR00864 and 2009-SGR003 (Generalitat de Catalunya, Spain: Institut d'Estudis Catalans and Universitat Autònoma de Barcelona)

Projects HUM2005-02746 and FFI2010-19206 (Ministerio de Educación y Ciencia, Spain: Universitat Autònoma de Barcelona)

REFERENCES

Catford, John C. (1977): *Fundamental problems in phonetics*. Bloomington, IN: Indiana University Press.

Docherty Gerard J. & Foulkes, Paul (1999): "Sociophonetic variation in 'glottals' in Newcastle English". In: John J. Ohala, Yoko Hasegawa, Manhari Ohala, Daniel Granville & Ashley C. Bailey (eds.). *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS*). Berkeley, CA: University of California, Berkeley, 1037-1040.

Hall, T. Alan (1997): *The phonology of coronals*. Philadelphia, PA: John Benjamins.

Kenyon, John S. & Knott, Thomas A. (1953): *A pronouncing dictionary of American English*. Springfield, MA: Merriam Webster.

Lisker, Leigh & Abramson, Arthur S. (1964): "A cross-language study of voicing in initial stops: Acoustical measurements". *Word* 20: 384–422.

Paradis, Carol & Prunet, Jean-François. (eds.) (1991): *The special status of coronals: Internal and external evidence. (Phonetics and Phonology 2)*. San Diego, CA: Academic Press. Pp. xvii + 231.

Yavaș, Mehmet. (2006): *Applied English phonology*. Oxford: Blackwell.

PART II
Suprasegmental Aspects of English

CHAPTER 5. LEXICAL STRESS

Jean-Louis Duchet

University of Poitiers

1. Introduction

The **nature** of stress is a phonetic problem which is still unresolved. The association of stress with the intensity (or amplitude) of the sound wave of the spoken chain was questioned by Fry who showed in 1955 that duration and intensity are both cues for judgments of stress but that duration "is a more effective cue than intensity". His *Experiments* in 1958 went further and showed the role of pitch (frequency): there is a "tendency for a higher syllable to be heard as stressed in preference to a lower one." The debate is still going on[1].

The **placement** of stress is a phonological issue which concerns linguists and language teachers and learners, because it is specific to each language and plays a role in analysing, identifying and learning language units.

Stress is a prosodic feature applying to the word as opposed to *accent* which applies to an utterance or part of an utterance. A word is recognized not only through its consonants and vowels, but also through its stress pattern which plays a crucial role in its acoustic image and therefore in its identification in the spoken chain. This is all the more obvious as English, like Russian and Catalan, alters the vowel quality of unstressed syllables: "weak" vowels are centralized. Although semantically related the noun *majesty* ['mædʒəsti] and the adjective *majestic* [mə'dʒestɪk] have different stress patterns which in turn alter the quality of the the first two vowels. This prosodic feature provides a contrast between the stressed syllable or syllables and the

---

1 See Ferragne (2011) Corrélat acoustique des degrés de l'accent de mot, *12e Colloque d'avril sur l'anglais oral,* Université de Paris-Nord.

unstressed syllables. The notion of syllable will therefore be relevant for stress placement. The question of which syllable is going to be stressed is determined by the various ways in which the stress helps identify the form of the word by emphasizing the word stem, or prefixes and suffixes, or marking the beginning or end of the word.

Languages which make use of stress placement as a feature of the lexical form of the word may have straightforward rules such as “always stress the first syllable”, or it may be the last one or the penultimate, or “always stress the suffix”.

English also has such rules but they are many, and each one of them applies to limited sets of words. It is the aim of this chapter to formulate and account for, the rules which determine stress placement in English.

Before doing so, we introduce the topic with a short overview of stress placement as dealt with by phoneticians in recent history.

2. A fairly recent realization

The British phonetic tradition was long reluctant to accept that stress placement was rule-governed. Daniel Jones's sceptic attitude is expressed in his *Outline of English Phonetics*: “The rules regarding the position of the stress in English words of more than one syllable are very complicated and most of those which can be formulated at all are subject to numerous exceptions” (§579, p. 111). But he does provide the rules already popularized by textbooks. It is easy to realize how his formulation of Rule I placing stress on the second syllable of prefixed two-syllable words (*arrive, collapse, defence, precise*) without consideration of the part of speech concerned inevitably leads him to a long list of exceptions (*access, congress, expert, etc.*), thereby confirming the impression of “complexity”. It is only in a footnote (p. 112) that Jones concedes that “many of these words have corresponding verbs which are stressed on the last syllable according to the rule.”

Jones's pronouncement was echoed by many other British phoneticians. Jack Windsor Lewis writes in his *Guide to English Pronunciation,* published in 1969 for Scandinavian students, that “the fixed place for the accentual syllables in English words is not predictable any more than it is in Scandinavian languages. Each word has to be learnt individually” (p. 54).

As late as 1971, J. D. O’Connor devotes only two pages to stress placement in his *Better English Pronunciation* and little more is needed since “there is no simple way of knowing which syllable or syllables in an English word must be stressed, but every time you learn another word you must be sure to learn how it is stressed” (p. 115).

In America, Prator and Robinett (1972) insist that “stress is the key to the pronunciation of an English word” and therefore conclude that its location “should always be learned with the word” since “there are no infallible rules for determining which syllable of a word should be stressed” (p. 19). Roger Kingdon started breaking new ground with his *Groundwork of English Stress* (1958), comforting a tradition of teachers of English abroad[2] who had always been eager to provide their students with stress-placement rules: “It is also possible to draw up rules showing how suffixes influence stress; in most cases these rules are subject to remarkably few exceptions” (p. 13). Kingdon provides the student with lists of lexical units distributed among Romanic-type compounds (including derived words with suffixes and opaque endings), Greek-type compounds (combining forms), and English-type compounds. His book also includes a study of the stress differences between British and American English (p. 195-221).

Lionel Guierre followed in Kingdon's footsteps. He based his linguistic claims on a computer treatment of Jones's 12th edition of the *English Pronouncing Dictionary* and his *Drills in English Stress-Patterns*, intended as a practical textbook, was at the same time an influential description providing complete inventories of stress-patterns to be found in the English lexicon. His *Essai* in 1979 systematized his findings on stress patterns and complemented them with phonographemic rules to predict the vowel quality of stressed and unstressed syllables.

2 There is evidence of stress placement rules being taught in schools, relying on the orthoepic principles made explicit by John Walker in his *Critical Pronouncing Dictionary*, published in 1791 and regularly republished for more than a century. See Beljame, Alexandre (1891) *Cours pratique de prononciation anglaise*, Paris: Hachette, p. 181-194.

In the early 80s, books devoted to stress-patterns reinforced the realization that stress-placement was predictable, especially those by I. Poldauf and E. Fudge, both of whom quote Guierre as a source of inspiration. Poldauf points out that the apparent complexity of stress placement in English is motivated by the fact that its “lexicon is hybrid in that it is built from disparate word structures” (p. 7). The notion that the English lexicon is made up of several layers of vocabulary each of which has its own sound and stress rules is a common feature of Guierre's and Poldauf's approaches.

In the late 60s, Chomsky & Halle's *Sound Pattern of English* and their followers[3] devoted many efforts to achieve a satisfactory formulation of stress-placement rules in English.

But because of the vowel weakening process which affects unstressed vowels mentioned above, it seems preferable to avoid formulating stress-placement rules on the basis of the phonetic quality of syllables as Chomsky and Halle have done. We wish to do the reverse and base the prediction of vowel quality on the basis of stress and base stress placements on rules combining morphology and phonology. Such an approach has been attempted by Kreidler (1989) in a broad generative framework (p. 197-218) and by Teschner & Whitley (2004) in their “stress-based approach” to English pronunciation.

The diachronic study of stress placement, as we shall see, shows that classes of lexical units have been prone to analogy: a particular pattern has often been assigned to a class of words analysed as morphologically similar, the morphological criteria gaining priority over phonetic properties.

3. Basic principles of stress placement

Each lexical unit pronounced in isolation takes a full stress (which is associated with a tone, since there is no utterance without intonation) and only one. Monosyllabic words are treated as being stressed on their only syllable which shares the properties of stressed syllables in longer words:

---

3 See John Robert Ross (1972), A Reanalysis of English Word Stress, in Michael Brame ed., *Contributions to Generative Phonology,* Austin: University of Texas Press, p. 229-323.

stand ['stænd]	*withstand* [wɪð'stænd],
sign ['saɪn]	*signing* ['saɪnɪŋ],
assign [ə'saɪn]	*assignment* [ə'saɪnmənt]

may also carry secondary stress. Secondary stress occurs in words with more than three syllables whose main stress is placed after the second syllable:

understand [ˌʌndə'stænd]	*understandable* [ˌʌndə'stændəbḷ]
reassignment [ˌriːə'saɪnmənt]	

A basic constraint on stress placement in English is that a word cannot begin with more than one unstressed syllable. We shall discuss secondary stress placement further on.

The few examples above provide cases of derivation in which stress placement remains unchanged.

Both *-ment* and *-able* are suffixes referred to by Fudge (1984) and Wells (1990) as **stress-neutral**.

This is the case of all inflections: the verb morphemes *-ed* and *-ing*, the noun morpheme for the plural -*(e)s* or the genitive *'s,* although they may add one syllable to the stem, do not affect stress placement:

enjoy [ɪn'dʒɔɪ]	*enjoyed* [ɪn'dʒɔɪd]
enjoying [ɪn'dʒɔɪɪŋ]	*enjoyment* [ɪn'dʒɔɪmənt]
animate ['ænɪmeɪt]	*animated* ['ænɪmeɪtɪd] *animating* ['ænɪmeɪtɪŋ].

But the noun connected with the latter verb is *animation*, with the primary stress on *-mat-* and secondary stress on *an-:* [ˌænɪ'meɪʃᵊn], because *-ion* imposes stress placement on the previous syllable.

4. Stress-neutral and stress-imposing endings and suffixes

Stress-imposing endings and suffixes are mostly unstressed but they determine stress placement on one particular syllable of what Fudge (1984) calls the stressable portion of the lexical unit. Thus the verb *animate* has an -

ate ending which places stress two syllables before the ending: *'animate.* The *-ion* ending of *animation* in turn imposes stress on the syllable preceding the ending*: ani'mation* and secondary stress on the first syllable. Here is a summary of stress-neutral and stress-imposing endings.

Word category	Stress-neutral endings	Stress-imposing endings
Nouns	inflectional *-(e)s: 'churches* *-ness: dis'interestedness* *-ment: establishment*	*-ity:* *-iV*, *-uV* in the ending: *-ion, -ia, -uous, etc.: i'nertia, phy'sician, indi'vidual*
Verbs	inflectional *-(e)d, -ing: 'dedicated* *-ize : 'nationalize*	*-ify: syl'labify, so'lidify* *-ish: es'tablish, dis'tinguish* *-ate, -ent: de'liberate, 'implement*
Adjectives	inflectional *-(e)d, -er, -est* *-y, -ly : 'tasty, 'weekly* *-able : in'terpretable* *-ish: 'yellowish*	*-ic, -ical* *-ible/-igible: com'patible, 'eligible* *-ose: : 'bellicose, ve'siculose* *-al, -ar, -an, -ant, -ous*
Adverbs	*-ly: 'scrupulously, indi'vidually*	

Stress-imposing endings with -i ou -u followed by a vowel, the noun ending *-ity*, the verb ending *-ify*, and the adjective ending *-ic/-ical* place stress **one syllable before** the ending.

Endings	Examples
-ia, -ion, -ial, -ious, -ual	*inertia, attention, official, curious, residual*
-ity	*an'xiety, ˌauthen'ticity, deˌsira'bility*
-ic/-ical	*ro'mantic, ˌscien'tific, his'torical, ˌalpha'betical*
-ify	*'justify, di'versify*

If a word has no other syllable than -*ity* or -*ion/-ia/-ie* or -*able*, the first syllable is stressed:

city ['sɪti], *lion* ['laɪən], *trial* ['traɪəl], *diet* ['daɪət], *able* ['eɪbl].

The noun suffix *-ity* is spelt *-ety* if there is no consonant in the syllable before the ending: *so'ciety, an'xiety, pro'priety,* in which case the vowel spelt <i> is stressed and diphthongized.

The verb ending *-ish* is to be found in some forty verbs which have two or three syllables: *perish, nourish, flourish, demolish, abolish, extinguish*. Only one verb has four syllables and seems to be an exception to the rule: *impoverish* [ɪm'pɒvərɪʃ]. It is probably more appropriate to consider this verb and other verbs like *demolish* and *extinguish* as prefixed verbs which regularly bear stress on the first syllable of the stem.

Some examples are in order to illustrate the stress-patterns occurring in grammatical inflections and lexical derivations and the way in which the pattern of the deriving word is, or is not, affected by stress-neutral and stress-imposing endings and suffixes.

4.1 Stress-neutral derivation:

Interest ['ɪntrɪst]	*disinterested* [dɪs'ɪntrɪstɪd]
	disinterestedness [dɪs'ɪntrɪstɪdnəs]
critic ['krɪtɪk]	*criticize* ['krɪtɪsaɪz]
	criticism ['krɪtɪsɪzm]
implement ['ɪmplɪment]	*implementing* ['ɪmplɪmentɪŋ]
	implemented ['ɪmplɪmentɪd]

4. 2 Stress-imposing derivations:

tonic ['tɒnɪk]	*tonicity* [tə'nɪsəti]
magic ['mædʒɪk]	*magician* [mə'dʒɪʃᵊn]
technical ['teknɪkᵊl]	*technicality* [ˌteknɪ'kæləti]
specific [spə'sɪfɪk]	*specificity* [ˌspesɪ'fɪsəti]
specification [ˌspesəfɪ'keɪʃn]	

4.3 Stress-imposing and stress-neutral derivations combined:

ingenuous [ɪn'dʒenjuəs]	*ingenuousness* [ɪn'dʒenjuəsnəs]
	ingenuity [ˌɪndʒə'nju:əti]

commerce ['kɒmɜːs] — *commercial* [kə'mɜːʃᵊl]

commercialize [kə'mɜːʃᵊlaɪz]

If a word has more than one stress-imposing ending, the last one prevails:

digest [daɪ'dʒest] — *digestible* [daɪ'dʒestəbl]

digestibility [daɪˌdʒestə'bɪləti]

associate [ə'səʊʃieɪt] — *association* [əˌsəʊʃi'eɪʃᵊn]

The verb endings *-ate, -ent* place stress **two syllables before** the ending: *hesitate, devastate, accelerate, contemplate, accommodate, inaugurate, document, complement, etc.*

Adjectives and nouns in *-ate* follow the same pattern, even though the vowel of the ending may be weakened to a central vowel:

adjectives: '*delicate, 'desperate, 'accurate, 'intricate, de'generate, de'liberate, le'gitimate;*

nouns: *'magistrate, 'caliphate, 'consulate, cer'tificate, pa'latinate, per'manganate.*

Verbs ending in *-ute* are being regularized along the same pattern: *execute*, *prosecute*, *persecute*; *institute*, *constitute*, *substitute*, *prostitute*, *destitute; 'contribute* now wins the majority of votes, *'distribute* is gaining ground and *'attribute* is attested but still considered as non-RP: it remains the last verb in *-ute* to be stressed on the stem. This small class of verbs provides a showcase of stress placement change in progress, the conflict of rules being resolved in favour of a stress-imposing ending against the prefix-stem hierarchy becoming less and less perceptible.

Verbs and nouns in *-iate* and *-uate* have to comply with two rules, since they combine two stress-imposing endings: i+V places stress on the syllable before the ending and *-ate* imposes stress two syllables before itself. The rule conspiracy is made possible by the syllabic pronunciation of <i> (a diaeresis process):

associate [ə 'səʊ ʃi eɪt], *negotiate* [nɪ 'gəʊ ʃi eɪt], *perpetuate* [pə 'pet ju eɪt].

The interplay of stress placement rules has an interesting rhythmic consequence on the form of verbs. The verb ending *-ify* imposing stress one syllable before itself has two syllables whereas the ending *-ate* imposing stress two syllables before itself has only one. The stress pattern of both derivations is therefore the same:

ratify *legislate*

['rætɪfaɪ] ['ledʒɪsleɪt], with the same stress-pattern which may be symbolized thus:

/100/ /100/

The verb *legalize*, which has a stress-neutral suffix added to the adjective base *legal* has the same pattern but produced by totally different rules:

legalize

['liːgᵊlaɪz]

/100/.

The adjective ending *-ose* also places stress two syllables before the ending: *bellicose* ['belɪkəʊs], *vesiculose* [və'sɪkjʊləʊs].

But most adjectives in *-ose* have only two syllables. In such an occurrence, it is the ending *-ose* itself which bears stress: *jocose, fibrose, morose, verbose* [vɜː'bəʊs].

The pattern we have applied to words of at least three syllables is therefore not always applicable because there are shorter words.

5. Stress placement and the number of syllables

If a verb ending in *-ate, -ute* or *-ent* has only two syllables, stress is placed on the the ending:

cre'ate, dic'tate, nar'rate, di'lute, fre'quent, pre'sent, la'ment, as opposed to *'celebrate, 'implement, etc.*

If a noun ending in *-ate* or *-ent* has only two syllables, stress is placed on the the stem:

'senate, 'curate, 'climate, 'legate, 'magnate, 'talent, 'ferment.

The number of syllables may also interfere in the placement of stress with verbs derived by a stress-neutral suffix. This is the case of verbs in *-ize/-ise*.

Verbs in *-ize* of at least **four** syllables are always derived from adjectives (or sometimes nouns) of at least three syllables. This pattern is very productive and the derivatives produced are very frequent. It is a **stress-neutral** derivation, as shown by the following table:

/100/ → /1000/		/010/ → /0100/	
General	**Gen**eralize	com**merc**ial	com**merc**ialize
decimal	**Dec**imalize	ma**ter**ial	ma**ter**ialize
Popular	**Pop**ularize	fa**mil**iar	fa**mil**iarize
/100(0)/ → /1000/		/0100/ → /01000/	
Category	**Cat**egorize	pe**des**trian	pe**des**trianize

Similarly, on the basis of the adjective *le'gitimate*, regularly stressed two syllables before the ending *-ate*, a verb in -*ize* has been formed; *le'gitimize,* stressed on the same syllable.

But verbs in *-ize* of **three** syllables all follow one and the same stress-pattern: /100/.

Many of them follow a stress-neutral derivation pattern /10/ → /100/: *civil → civilize, central → centralize, tyranny → tyrannize, theory → theorize, emphasis → emphasize, fertile → fertilize.*

The same pattern also applies when the derivation is opaque or doubtful as in the case of *organize, mechanize, ostracize, pulverize.* Only two words of **two** syllables and stressed on the **second** syllable have derived a verb in -ize: *ca'nal, im'mune.* The statistical[4] weakness of this class seems to have prompted an analogical alignment of the two verbs on the predominant pattern /100/ hence: *'canalize, 'immunize*.

The conclusion we are led to is that when a transparent and productive derivation is obvious, stress-neutral derivation is the rule. But when there is no such transparency and productivity in the would-be derivation, the tendency is to impose a pattern which applies regularly without any consideration of the base it is applied to.

Complementary illustration of this is provided by the very rare examples of verbs in which the -*ate* ending has remained stress-neutral, because of the transparent derivation it reflects: 'oxygenate is not stressed on the penultimate syllable like all other verbs in *-ate* (with their opaque derivation),

4 This can be provided as a further example of the frequency effect in the emergence of linguistic structure. See Bybee & Hopper (2001).

but on the same syllable as 'oxygen. This transparent, stress-neutral derivation is the same as that in 'oxygenize, which also exists and is probably the cause of the analogy exerted on the stress-pattern of 'oxygenate.

6. Stress placement and the number of pre-final consonants

English adjectives may be derived with the typical English stress-neutral suffix *-y* (*rainy, salty, tasty, itchy, homey*) or *-ly* (*friendly, scholarly, daily*). But English also has many adjectives with endings of French and Latin origin. We have already dealt with the endings *-ic/-ical*, which determine stress placement on the syllable before the ending as in *nostalgic* [nɒs'tældʒɪk], *scientific* [ˌsaɪən'tɪfɪk]; *-ese* which is a stressed ending as in *Vietnamese* [viˌetnə'mi:z]; and *-ose* which places stress two syllables before the ending or on itself in adjectives of two syllables.
The other endings are *-al, -ar, -an, -ous, -ant, -ent.*
Stress is placed either on the syllable before the ending (the penultimate syllable) or two syllables before the ending (the antepenultimate syllable), depending on the consonant environment before the ending:

Example	*universal*	*particular*
	<--CCal> (2 consonants)	<--Car> (1 consonant)
Transcription	[ˌju:nɪ'vɜ:sl]	[pə'tɪkjʊlə]
Stress-pattern	/2010/	/0100/

Further examples illustrate this principle:

au'tumnal	*mo'lecular*
pa'rental	*'popular*
conti'nental	*me'dicinal*
senti'mental	*o'riginal*

tre'mendous	*'fabulous*
mo'mentous	*ri'diculous*
re'luctant	*'competent*
in'dignant	*'tolerant*
U'gandan	*cosmo'politan*

Many adjectives in *-ant* correspond to verbs in *-ate*, which have the same stress-pattern, have only one consonant before the ending, and significantly display a spelling of the ending with the letter <a>:

/100/	/100/
Tolerate	*Tolerant*
Hesitate	*Hesitant*
Celebrate	*Celebrant*
Expectorate	*Expectorant*
Protuberate	*Protuberant*
Exuberate	*Exuberante*

Significantly, adjectives in *-ant* with two consonants before the ending and conforming to the rule like *re'dundant, re'pugnant, re'luctant* are not derived from current verbs.

It should be noted that in words like *consequence, consequent,* and *subsequent,* the letter <u> is not a pre-final vowel but is part of consonantal digraph <qu> pronounced [kw]. Therefore the ending which is here stress-imposing is the *-ent* ending which places stress two syllables before the ending: *'subsequent* /100/.

7. Adjectives ending in -ive

Adjectives in *-ive* follow the same patterns depending on the presence or absence of a consonant cluster before the ending:

/100/: *sensitive, putative, negative, fricative, genitive, tentative.*

/010/: *defective, addictive, vindictive,*

But most adjectives in *-ive* are connected to verbs in *-ate*, the ending of which is stress-imposing. And the *-ive* ending tends to be treated as stress-neutral in this environment:

/100/ → /1000/		/0100/ → /01000/	
cumulate	*cumulative*	*communicate*	*communicative*
speculate	*speculative*	*assimilate*	*assimilative*
imitate	*imitative*	*approximate*	*approximative*

There are a few exceptional words borrowed at an early date and which are not only adjectives but also nouns:

de'monstrative, in'dicative, pre'dicative[5]*, al'ternative, cor'relative,* the stress-pattern of which is different from the verbs they seem to be derived from: '*demonstrate, 'indicate, 'alternate, 'correlate*.

Finally there are adjectives in *-ative* which are derived from verbs which do not end in *-ate,* like '*talkative* and *de'clarative*. The stress-pattern which applies indicates that the *-ative* derivation is here again treated as stress-neutral: the adjective *declarative* has stress on the same syllable as the verb *declare* and does not behave as if there existed a verb *declarate. This shows that the derivational motivation plays a role in stress placement processes.

8. Adjectives ending in -ory

The same is true of derivations of learned adjectives in *-ory* such as *congratulatory*. Many of them, derived from a verb in *-ate,* treat the ending as a stress-neutral ending:

/100/ → /10000/		/0100/ → /010000/	
***ded**icate*	***Ded**icatory*	*Con**gra**tulate*	*con**gra**tulatory*

5 Also regularized as 'predicative in American English.

***os**cillate*	***Os**cillatory*	*Re**verb**erate*	*re**verb**eratory*
***comp**ensate*	***Comp**ensatory*	*An**tic**ipate*	*an**tic**ipatory*

The same derivational principal treats the ending as stress-neutral when it is applied to verbs which do not end in *-ate*:

/01/ → /01000/			
Exclaim	*exclamatory*	*Explore*	*exploratory*
Confirm	*confirmatory*	*explain*	*explanatory*
oblige	*obligatory*	*Respire*	*respiratory*[6]

The *-ate* ending is still there and it adds to the number of syllables, but it is neutralized in terms of stress placement.

Nouns and adjectives devoid of any transparent derivational pattern also have stress placed two syllables before the *-ory* ending or the *-ary* ending, if only one consonant precedes the ending:

'predatory, 'purgatory, de'rogatory,

'ordinary, 'commissary, 'emissary, 'military, 'solitary, etc.

And if there is a consonant cluster, stress is placed on the syllable before the ending, which is the regular stress-imposing pattern of adjectives we have mentioned above:

re'fectory, di'rectory, con'sistory, ˌcontra'dictory, ˌsatis'factory, pe'remptory.

Here again the same ending is used as a stress-neutral suffix in productive and transparent derivations whereas it follows a regular stress-imposing pattern in other cases.

9. Stressed endings

Stressed endings are typical of nouns, mostly borrowed from foreign languages at various stages in the history of the language:

6 American English has adopted another pattern: *'respiratory*, as if it was derived from a verb form *respirate.

Ending	Examples
VCCe	ga'zette, ga'zelle, gi'raffe, fi'nesse, cas'sette, silhou'ette
-oo, -oon, -ee, -een, -eer	bam'boo, car'toon, de'gree, can'teen, engineer
-ade	bri'gade, ˌlemo'nade, ˌharlequi'nade, ˌprome'nade
-V(s)que: -esque, -ique -aque	ˌpictu'resque, an'tique, tech'nique, ob'lique, o'paque
-ese	ˌJapa'nese, ˌPortu'guese, ˌSia'mese
-osis, -itis	hyp'nosis, ˌdiag'nosis, prog'nosis, bron'chitis, ˌmenin'gitis

The ending is stressed. Therefore if the word has three or more syllables, secondary stress is required on the first or second syllable.

There are exceptions but they are not very numerous and concern rather frequent words, which provides an indication of their nativization into the English sound system:

'omelette, 'vaudeville, 'coffee, 'toffee, com'mittee, 'igloo, 'cuckoo.

'decade, 'marmalade.

10. The role of prefixes in stress placement

The role of prefixes is far more limited than that of suffixes and endings. But in prefixed words which have no ending, prefixes do play a role.

There are two different sorts of prefixes, which play a different lexicological role and have a different impact on stress placement. **Separable** prefixes of verbs or adjectives have a discernible semantic content and although the main stress will bear on the stem the prefix also tends to be stressed, i.e. to receive secondary stress: in ˌde'populate, the antonym of 'populate, secondary stress is placed on the first syllable, thus placing stress on two adjoining syllables which is otherwise avoided by the stress alternation rule. But in the verb de'nominate, the prefix has no semantic content of its own and is therefore **inseparable**. As a consequence it does not receive secondary stress unless it has to, because of stress alternation, as in ˌdevo'lution.

Inseparable prefixes are generally unstressed but their presence determines stress-placement in words of two and three syllables.

Separable prefix	Inseparable prefix
ˌde'regulate	de'liberate
ˌde'vitalize	de'teriorate
ˌdis'relish	dis'cover
ˌun'settle	in'spire
ˌre'route	re'quire

It should be noted that de'preciate, which could be interpreted as the antonym of ap'preciate, is not directly derived from an identifiable stem: the derivation was borrowed from French and Latin: *de-* and *ap-* are not productive prefixes in this case.

(i) Verbs of two syllables with an inseparable prefix are stressed on the stem:

be'come, for'get, for'bid, ac'cept, com'pose, de'ceive, de'sign, of'fend, pre'fer, re'sent.

Inseparable prefixes may be of Latin or Germanic origin. A Germanic prefix may be affixed to a Romance stem as in *besiege*. But the placement of stress with all inseparable prefixes follows the same rules.

(ii) Nouns of two syllables with an inseparable prefix are stressed on the prefix:

'refuge, 'concert, 'colleague, 'commerce, 'district, 'insect, 'income, 'province, 'preface, 'precinct, 'premises, 'suburb, 'outskirt, 'outcome.

(iii) Words of two syllables with an inseparable prefix which are both verbs and nouns are caught in a conflict of rules. They have been for generations a favourite class of words with textbooks of English: the stress alternation between the two is usually presented as the rule. This is true for a good number of examples but it is not the most frequent pattern:

'import, *n.* - im'port, *v.* ; 'export, *n.* - ex'port, *v.* ; 'insult, *n.* - in'sult, *v.* ; 'suspect, *n.* - sus'pect, *v.*

'transport, *n.* - trans'port, *v.* ; 'refuse, *n.* - re'fuse, *v.* ; 'protest, *n.* - pro'test, *v.* ; 'progress, *n.* - pro'gress, *v.*

'permit, *n.* - per'mit, *v.* ; 'rebel, *n.* - re'bel, *v.* ; 'convict, *n.* - con'vict, *v.* ; 'object, *n.* - ob'ject, *v.*

The most frequent pattern in the lexicon is that of verbs which are also nouns and stressed on the stems on the same pattern as verbs: /01/

account, amend, award, command, concern, consent, control, decline, default, defeat, design, decease, dissent, distrust, disgust, embrace, exhaust, recruit, repair, reply, respect, result, regard, regret, remark, reserve, resort, report, return, retreat, revolt, rebuke, supply, surprise, mistake, etc.

Not only are they the most numerous in the lexicon, but they are also the most frequently occurring in texts and discourse.

There is a limited number of verb-noun pairs which follow the opposite pattern like combat, *n.* and *v.*: /10/

combat, comfort, comment, exile, incense, invoice, purchase, offer, promise, profit.

One should add to this list noun-verb pairs such as *pardon, traffic, distance, rescue, profit,* which, although historically prefixed, are probably no longer analyzed as such and are therefore re-analyzed as nouns, which can undergo conversion and be used as verbs.

Alternating pairs of verbs and nouns are also produced with separable prefixes:

re'fill, *v*. - 'refill, *n.*; re'wind, *v*. - 'rewind, *n.*; re'play, *v*. - 'replay, *n.*; re'write, *v*. - 'rewrite, *n.*

This seems to indicate that this pattern is still productive.

There is a tendency, detected by Trevian (2003), to create new alternating pairs in American English. This is also the case in British English: the noun *defect* which was regularly transcribed like the verb as [dɪ'fekt] until recently is now predominantly ['di:fekt], a pronunciation still stigmatized as non-RP by Wells (1990) and which has become the recommended pronunciation since the second edition of Wells's *LPD* in 2000 and the 16th edition of Roach's *EPD* in 2003. The verb *purport* offers another example: until the 15th edition of the EPD, it used to be stressed on the prefix, like the noun. It is now stressed /01/ thus creating a new alternating pair. Conversely, an alternating pair of British English is disappearing: rather than dis'pute, *v.* - 'dispute, *n.,* the prevailing tendency is now to pronounce both stressed on the stem:

dis'pute /01/ like *disgust* or *dissent*. Another British English alternating pair has disappeared: con'tact, *v.* - 'contact, *n.,* are now stressed on the prefix like the noun: 'contact, *n. and v.* This change was first documented by Gimson in the 14th edition of *EPD.* The fact that the noun is far more frequent may provide further evidence here of the role of frequency in the emergence of regularity.

In American English, *combat* and *research* have become alternating pairs. But the British English pair pros'pect, *v.* - 'prospect, *n.* has one and the same pattern in American English for the verb and the noun, both stressed like *account* and *concern.* It seems therefore difficult to conclude that American English is more particularly prone to alternating pairs.

The situation seems to be rather unstable and shows a degree of variation which may last for a long time.

Derivational processes play an important role in stress assignment. Verbs like *retail*, *abstract* and *process* which had long existed in the language had a regular /01/ pattern. The corresponding nouns had the expected /10/ pattern. But, due to the semantic distance which polysemy and semantic change have introduced between the verbs and the nouns, new verbs have been created by conversion from the noun to refer to a verb notion closer to the meaning of the noun: 'abstract, *v.* means “to write an abstract” as opposed to ab'stract, *v.* “remove”; 'process, *v.* means “submit to a process”, as opposed to pro'cess, *v.* “walk in procession”. This shows again that the semantic transparency of a morphological derivation is likely to preserve the motivated relationship between both words through conversion which amounts to a stress-neutral derivation.

(iv) Adjectives of two syllables with an inseparable prefix are mostly stressed on the stem: /01/

abrupt, absurd, adept, content, correct, complete, compact, direct, distinct, extinct, exact, exempt, extant, intact, intense, remote, concise, precise, immune, obscure, reverse, perverse, diverse, replete, obscene, profound, succinct, sublime.

Among the exceptions, some have a tendency to be regularized in American English, as shown by Trevian (2003, p. 164): abstract, *adj.*, compound, complex, occult, converse now have the stress-pattern /01/ like *abrupt* and *compact*. Other apparent exceptions are adjectives which are also nouns and share the stress-pattern of their noun counterparts: /10/ *instant, constant, present, subject, expert, concrete, common.*

The only remaining exceptions are: */10/ distant, absent, oblong, prolix.*

It should be noted that, despite appearances to the contrary, the adjectives *dismal, decent* and *recent* are not prefixed but accidentally have a first syllable which looks like a prefix. They are therefore stressed on the first syllable of their stem: /10/.

3. Adverbs and prepositions of two syllables with an inseparable prefix are stressed on the stem like verbs and adjectives: /01/ *before, besides, below, ago, again, etc.*

4. Verbs of three syllables with an inseparable prefix are stressed on the first syllable of the stem: /010/

   de'velop, de'liver, de'posit, con'sider, con'tinue, de'termine, e'xamine, pre'destine, en'title, en'deavour,

5. Verbs of three syllables with two inseparable prefixes are stressed on the stem and have a secondary stress on the first syllable: /201/

   ˌcoin'cide, ˌcorres'pond, ˌdisal'low, ˌrepre'sent, ˌrecom'mend, etc.

There are only two exceptions: 'improvise, 'reconcile.

11. Secondary stress

The placement of secondary stress occurs in words of three or more syllables. There are constraints on the stress pattern of the words which prevent sequences of unstressed syllables at the beginning of a word and tend to produce rhythmic alternation between strong and weak syllables.

The question whether secondary stress should appear is relevant only when primary stress is placed neither on the first nor on the second syllable.

It is mostly the case of words in which a stress-imposing ending has attracted stress close to the end of the word. English has a lexical stock in which words of Germanic origin mostly stress the word on the first or second syllable and words whose origin goes back to Latin or a Romance language

which have stress on the last, penultimate or antepenultimate syllable.
Secondary stress therefore counterbalances the impact of their original stress in words derived from Romance languages.
Secondary stress is shown by a mark placed below the line at the beginning of the syllable.
For lack of a main stress on the first or second syllable, a secondary stress will therefore be placed on one of the two.
A sequence of two unstressed syllables at the beginning of a word is prohibited by the rhythmic constraints already mentioned.
And in case there are more than one syllable stressed in a word, a principle of stress alternation should be respected.
Therefore, if the main stress is placed on the third syllable, a secondary stress must be placed on the first syllable: *satisfaction* [ˌsætɪs'fækʃn].
But if the main stress is placed on the fourth syllable, either the first or the second may be stressed. The choice between the two is based on a derivational principle: the main stress of the deriving word becomes the secondary stress of the derived word. Here is an illustration of this process:
characterize ['kærəktəraɪz] *characterization* [ˌkærəktəraɪ'zeɪʃn]
discriminate [dɪs'krɪmɪneɪt] *discrimination* [dɪsˌkrɪmɪ'neɪʃn]
The application of this principle to other examples requires special attention.
(i) If we consider again the noun *majesty* and its derived adjective *majestic*, no room is left for the secondary stress since there is only one syllable before the main stress in *majestic*. Stress alternation requires an unstressed vowel in the first syllable. The main stress of *majesty* cannot be echoed in a secondary stress of the adjective derived from the noun.
(ii) In the case of a noun like *ˌdecla'ration* the main stress on the third syllable of the derived noun leaves no other possibility than placing the stress on the first syllable to comply with stress alternation. The main syllable bearing the stress of the verb *de'clare* is reduced to the status of an unstressed syllable.
(iii) In the case of a noun like *habilitation*, derived from the verb *habilitate*, the stress shifts to the fourth syllable and the choice being open for the placement of secondary stress on either the first or second syllable, the primary stress of *ha'bilitate* becomes the secondary stress of *haˌbili'tation*.
(iv) In the case of a noun like *acceptation*, the main stress also shifts to *'-*

ation. But there are only two syllables before the main stress of the derived noun, which prevents the application of the derivational principle because the stress alternation principle has priority over the derivational principle. Secondary stress is placed on the first syllable *ac-*, despite the fact that it is unstressed in the deriving verb.

(v) A verb like *pre'sent* serves as the base of a prefixed verb: *represent*, with the main stress on the same syllable and secondary stress on the first syllable which happens to be the prefix. If a further derivation is formed to generate the noun *representation*, the alternation principle makes it impossible to place the secondary stress on *-sent-* which bears stress in the deriving verb. Therefore the only avenue is to place the secondary stress on the first syllable which already bears secondary stress in the deriving verb:

present	represent	representation
[prɪ'zent]	[ˌreprɪ'zent]	[ˌreprɪzen'teɪʃn]

We therefore reach the conclusion that in dealing with the placement of secondary stress, rhythmic constraints have priority over the derivational principle which is applicable only within the bounds of the flexibility allowed by the stress alternation principle. When considering suffixation, on the contrary, we have noticed that derivational pressure often imposes an analogical pattern which turns stress-imposing suffixes or endings into stress-neutral ones or vice versa as we shall see later.

12. Tertiary stress

An intermediate level of stress is often felt necessary to distinguish between secondary stress and unstressed syllables. Tertiary stress was used consistently by John Wells in the first edition of *LPD* (1990).

Wells defines it as a "potential rhythmic beat either after the primary stress or between the secondary and the primary." (p. 683).

A word like *examination* has a typical stress-pattern with pre-final primary stress and secondary stress on the second syllable : /02010/. The word *cross-examination* places secondary stress on *cross-*, without shifting it from its previous position. Hence the pattern /203010/: ˌcross-e˳xami'nation.

In compounds, tertiary stress is also typical of the relative destressing of the second element of compounds:

/130/: cliffhanger, woodpecker, beefburger, bluestocking, box-office, book-binding;

/1030/: busybody, choirmaster, copywriter, cotton-picking, crease-resistant;

/10030/: cabinet-maker.

In words with more than one derivation, tertiary stress is a marker of the hierarchy of the various stages of derivation: *habilitate* has /1/ on *-bil* two syllables before the stress-imposing ending *-ate; habilitation* has /2/ on *-bil* after the primary stress has be moved to the penultimate syllable due to the effect of the stress-imposing ending *-ion; rehabilitation* has /3/ on *-bil* due to the placement of secondary stress on the prefix. Similarly: *confidentiality, comprehensibility, constitutionality* have /2030100/; *circumnavigation, contraindication, autocorrelation* have /203010/.

Tertiary stress also appears regularly in **variant** forms of words where it echoes the primary stress of the deriving word but goes counter to stress alternation, as in *containerization* /230010/, *cohabitation* /23010/, *congratul-ation* /23010/, *cooperation* /23010/.

Finally an ending like *-ism* with its two syllables tends to mark a stress hierarchy between them. To do this tertiary stress is used: *Anglicanism* /10030/, *capitalism* /10030/, *communism* /1030/, etc.

John Wells now marks tertiary stress with the same symbol as secondary stress, because "the separate symbol for tertiary stress was not well received" (Foreword to the second edition). This editorial argument does not alter the linguistic reality which tertiary stress marking conveyed in a rather convincing way.

13. Stress in compounds

As we have seen in the paragraph above about tertiary stress, compounds reduce the stress level of their second element. Lexicalized compounds, which correspond to a general agreement of the speakers to associate two lexical elements into one lexical unit (see Huart 2010, p. 183), place their primary stress on the stressed syllable of their first component: 'eyelid, 'shortbread, 'textbook, 'grammar school, 'grocery store, 'drinking water[7].

Some N1+N2 combinations do not correspond to a lexicalisation of a new unit made up of two elements, but N1 rather expresses a certain quality or property of N2 as would an adjective and therefore each element retains its lexical stress: this is the case of nouns indicating the substance the head noun is made of: 'brick 'wall, 'paper 'plate, 'glass 'table, 'silk 'shirt; of nouns indicating location in time and space: 'April 'showers, 'weekend 'traffic, 'London 'buses, 'village 'green, 'middle 'class; and of nouns indicating a relationship within a social group: 'club 'secretary, 'team 'captain, 'family 'joke, etc. (Huart 2010, p. 184-187).

The noun *yard* enters into a variety of combinations which illustrate the various stages mentioned here: *backyard* with equal stress on both elements corresponds to a location in space, opposed to *front yard*. *Churchyard* is a compound forming a lexical unit which has long been recorded as such in dictionaries: hence the placement of stress on the first element: 'churchyard. The fact that it is a compound is marked by the strong vowel retained by the second element, a feature which might be treated as a minimal level of stress: ['tʃɜːtʃjɑːd]. The noun *vineyard* as it is spelt has the appearance of a compound (*vine* + *yard*), but it is no longer one. Its pronunciation shows that it is no longer analyzed as such, as evidenced by the short vowel under stress in the first element and the weak vowel of the second element: ['vɪnjəd]. It has become an opaque unit of the lexical stock of English. The noun *orchard*, which was etymologically a compound (*wort* + *yard*) is no longer perceived or even written as one: the first element has altered its pronunciation, the second element has a weak vowel and the contact between the final [t] of the first element has been palatalized by the initial [j] of the second to produce the affricate now spelt <ch> thus hiding the spelling of the original elements: ['ɔːtʃəd]. This shows that the process of lexicalisation can reach the stage of complete integration. But the stable nature of a large number of compounds is due to the fact that compounding

7 It is obvious that spelling is not a reliable indication of the status of compounds.

is a productive process which associates two units into a new lexical unit without removing the traces of its compositional formation and meaning. New compounds not formally lexicalized are regularly observed: 'mouse pad, com'puter table, 'blog addict (Huart 2010, p. 185). Tertiary stress is in order on the first syllable of the second elements *table* and *addict*.
Compound adjectives have equal stress on both elements: 'red-'handed, 'long-'legged, 'hot-'blooded. They remain closely connected to the syntactic structure they are derived from, a noun phrase the head noun of which is the second element.

14. Stress in learned compounds

Learned compounds (Kingdon's "Greek-type compounds") associate combining forms which are not morphologically autonomous but behave lexically as meaningful units. A word like *autonomy* is made up of *auto-* and *-nomy*. The latter does behave like a stress-imposing ending and it is treated as such by Wells under *-nomy* (Wells 2000, p. 517). But it is more than that lexically and it belongs to a different system which has the characteristics of compounds rather than those of derived forms.
Stress is placed on the antepenultimate syllable (two syllables before the end): au'tonomy.
If the second element has only one syllable the resulting compound has three syllables with a /100/ stress-pattern. But if the second element adds one more syllable the primary stress shift to the right:
photograph ['fəʊtəgrɑːf] /100/ → photography [fə'tɒgrəfi]. /0100/
Like a standard compound *photograph* retains a strong vowel in its second element [grɑːf]. But in *photography* stress alternation makes it weak after stress has been shifted to *-tog-*.
Further derivation may shift primary stress to the right as in ˌphoto'graphic.
Many examples of the same pattern are to be observed:
/100/: autocrat, pentagon, thermostat, synonym, dinosaur;
/0100/: monopoly, analogy, monogamy, synonymy, periphery, bureaucracy (on the same pattern as *autocracy* and *aristocracy*);

/20100/: methodology, physiology, bibliography.

In compounds like *hydrocarbon*, the second element is an autonomous lexical unit. And therefore it receives primary stress, leaving only secondary stress on the combining form which precedes it: /2010/. This is also the case of ˌtele'photograph, ˌbio'chemistry, ˌphoto'synthesis, due to the autonomous existence of *photograph, chemistry, synthesis*.

It is true, however, that some derivations do not follow the same pattern. When the second element is -gog-, -gor-, -mon-, -leps-, stress is by default placed on the first element (i. e. the first syllable) following the more general rule of compounds:

'demagogue, 'demagogy, 'pedagogy; 'category, 'allegory; 'ceremony, 'testimony, 'acrimony, 'parsimony; 'epilepsy, 'catalepsy; 'melancholy; 'chiromancy.

Other such learned terms follow the stress-pattern of Latin through which they have been borrowed from the Greek, placing stress on the syllable before a pre-final consonant cluster:

-andr- : poly'androus	-derm- : epi'dermal
-hedr- : ˌdodeca'hedron, ca'thedral	-morph- : ˌanthropo'morphism

15. Stress shift

Stress in context may be shifted to a syllable which does not bear lexical stress. The reason for this is that stress alternation tends to avoid stress clash.

Stress-shift is often applied in the case of compound adjectives for example: a compound adjective like *dark-blue* with equal stress on both elements takes the primary stress on the first element and reduces the stress level of the second element to avoid a stress clash between its second element and the stressed noun which follows: a 'dark-ˌblue 'shirt (Deschamps *et alii,*. 2004, p. 83).

The requirement for stress shift is that a stressable syllable should be available to receive the shifted stress. This is the case with numerals (*thirteen* to *nineteen*) and adjectives in -*ese*, for example:

Fif'teen → 'fifteen 'years a'go; Chi'nese → 'Chinese 'restaurant;

Compounds which enter further compounding also mark the hierarchy of elements with a stress hierarchy:

ˌSouth 'London → 'South ˌLondon 'accent (Deschamps & al. 2004, p. 82).

The tendency to stress shift can reach even further. Wells 2000, p. 742 writes that “native speakers of English usually switch round the stress levels of the first element” to reach a better balance in speech rhythm. Although there is no direct stress clash between stresses in ˌfunda'mental 'frequency, a better balance is found if stress levels are shifted in the adjective : 'fundaˌmental 'frequency. The result is a two-level stress alternation: a regular stress alternation between stressed and unstressed syllables is accompanied by an alternation of primary stress and secondary stress between stressed syllables: /1020100/.

The same type of stress level reduction occurs in sequences of adjectives. Huart 2010, p. 142 provides several examples of this, with a clear alternation in monosyllabic words and a no less clear alternation in longer words even though it may not be so easy to perceive:

the 'big ˌbad 'wolf, a 'beautiful ˌold 'painting.

16. Stress variation in the past and nowadays

Stress variation although not so widespread as segmental variation is easily observed. It often reflects a situation in which conflicting rules have no clear priority over each other. We have already mentioned the variability and changes of stress placement in words of two syllables.

Other examples show the nature of rules involved in stress placement. An important change in stress placement took place during the late nineteenth century: verbs in -ate of at least three syllables have all adopted the same stress-pattern: /-100/. In the eighteenth and early nineteenth century there prevailed a dual set of rules similar to that still applying to adjectives in -al/-ar: pa'rental vs 'popular (see above).

J. Walker 1791 provides evidence of the older stress-patterns of verbs in -ate:

con'fiscate, con'template, il'lustrate as opposed to XX[th] century English 'confiscate, 'contemplate, 'illustrate.

This placement of the main stress was determined by the consonant cluster and not by an alleged stress-repellent prefix: 'connotate was stressed like 'tolerate in the XVIII[th] century.

It seems obvious that the whole class of verbs in *-ate* have followed the analogy of the statistically dominant pattern of a verb like *tolerate*. This type of analogy which we have already come across with when dealing with verbs in *-ize* of three syllables (see above) is a powerful promoter of language change.

Similarly ongoing changes which are being recorded nowadays show that the stress-neutral ending -ly, associated with the *-y* ending of adjectives in -*ary* is creating a new stress-imposing ending:

'necessarily and 'primarily /1000/ have been the first adverbs to develop a new stress-pattern of their own[8] in the middle of the XX[th] century, /0100/, which, by analogy, is now expanding to all adverbs in *-arily*.

These two examples among many others show that the observation of past changes and current developments in the patterns of stress placement help interpret the synchronic network of rules by which it is governed, which includes morphological derivation, semantic motivation and demotivation as well as phonetic and rhythmic constraints.

---

8 J. Castanier, *Vers un nouvel équilibre rythmique en anglais contemporain: l'évolution du placement accentuel dans les mots terminés par de longues séquences post-toniques,* Master's degree thesis, Poitiers, 2010, p. 74-75.

BRIEF GUIDE TO STRESS PLACEMENT

1°) The first question is where to place the main stress.

a) Does the word have stress-imposing endings ? If so it corresponds to one of the following cases:

- *pol'lution, ter'rific, a'bility, a'bolish, 'terrify :*

  The endings *-ion, -ic(al), -ible, -ity, -ish (verbs), -ify/-efy* require stress on the syllable before them.

- *'concentrate / cre'ate:*

  Words in -ate are stressed two syllables before the ending. But disyllabic verbs like create, are stressed on the ending. Disyllabic nouns in -ate are stressed on the stem: 'senate.

- *ˌuni'v**e**r**s**al, par'**tic**ular :*

  Suffixed adjectives of Latin or French origin in -al, -ar, -an, -ous, -ent/-ant are stressed on the pre-final syllable if it ends in a consonant cluster but on the antepenultimate if the pre-final ends with one consonant only.

Inflections (tense, plural morphemes), however, are stress-neutral :

- dis'criminate, dis'criminating, dis'criminated.

  Some endings are stressed: *-ee, -eer, -oo, -ese, -V(C)que, -ade, -aire* : tu'tee.

b) If there is no stress-imposing ending

Is there a prefix ? If so,

(i) a noun with a pseudo-prefix (inseparable) prefix is stressed on the pseudo-prefix: ***'pro****vince,* ***'ref****uge,* ***'circ****umstances*

(ii) a verb with a pseudo-prefix is stressed on the first syllable of the stem: *re**'sign**, dis**'cov**er*

(iii) so are pseudo-prefixed adjectives: *com'**plete**, e'x**act**, dis'**tinct**, pre'**cise**, cor'**rect,*** although there are a few exceptions: ***com****mon,* ***com****plex.*

(iv) Words of two syllables which are both nouns and verbs are caught in a conflict between rules. Most of them follow the pattern of verbs: con'cern (*my con'cern, it con'cerns me*); some alternate *('record n.,* re'cord *v.;* 'insult *n.,* in'sult *v.;* 'increase *n.,* in'crease *v.*); very few follow the noun pattern: *'**com**ment, '**o**ffer, '**prom**ise, '**pro**cess*.

c) If there is no prefix, stress the leftmost syllable: 'manifest, 'labyrinth, 'miserable, *etc.*

d) Stress in compounds

Usually the first element is stressed but there is no vowel reduction of the second element: *'blackboard, 'textbook, 'matchbox, 'sheep dog, 'flower-bed.*

A distinction must be made between compounds and adjective+verb sequences:

'dark '**room** vs '**dark**room; 'French '**teach**er vs '**French** teacher; 'brass '**door**knob vs '**brass** plate

Names of streets and roads are special: Mornington '**Cres**cent, Oxford '**Cir**cus, but '**Ox**ford Street.

e) In compounds with combining forms: stress is on the penultimate

words of three syllables: 'telegraph, 'thermostat; words of four syllables: te'legraphy, au'tonomy; of five.

2°) Secondary stress

The basic principle is alternation between strong and weak syllable: a stressed syllable is followed by a weak one. Another principle is that no word of English begins with two unstressed syllables. Therefore:

(i) when the main stress is placed on the third syllable, secondary stress is required on the first: *ˌindi'gestion.*

(ii) when the main stress bears on the fourth syllable, secondary stress is required either on the first or on the second syllable, on the basis of a derivational analogy (such a long word is almost necessarily derived rather than primary): the placement of stress in the *deriving* word determines secondary stress placement in the *derived* word:

dis'criminate > disˌcrimi'nation; 'characterize > ˌcharacteri'zation.

Other cases: ˌrepresen'tation, ˌrehaˌbili'tation.

REFERENCES

Bybee, Joan & & Paul, Hopper (2001): *Frequency and the Emergence of Linguistic Structure*. Amsterdam/ Philadelphia: John Benjamins.

Chomsky, Noam & Morris, Halle (1968): *The Sound Pattern of English.* New York: Harper & Row.

Deschamps, Alain & Jean-Louis Duchet, Jean-Michel Fournier, Michael O'Neil (2004): *English Phonology and Graphophonemics,* Paris: Ophrys.

Fry, D. B (1955): "Duration and Intensity as Physical Correlates of Linguistic Stress". *Journal of the Acoustic Society of America,* vol. 27, Issue 4, pp. 765-768.

Fry, D. B (1958): "Experiments in the Perception of Stress". *Language and Speech* April/June 1958 vol. 1 no. 2 126-152.

Fudge, Erik (1984): *English Word-Stress.* London: G. Allen & Unwin.

Guierre, Lionel (1970): *Drills in English Stress-Patterns.* London: Longman.

Guierre, Lionel (1979): *Essai sur l'accentuation en anglais contemporain.* Université de Paris 7.

Huart, Ruth (2010): *Nouvelle grammaire de l'anglais oral*. Paris: Ophrys.

Jones, Daniel (1914, 1922): *Outline of English Phonetics.* Cambridge: Heffer & Sons.

Kingdon, Roger (1958): *The Groundwork of English Stress.* London: Longman.

Kreidler, Charles (1989): *The Pronunciation of English. A Course Book in Phonology,* Oxford: Blackwell.

Lewis, Jack Windsor (1969): *A Guide to English Pronunciation for Users of English as a Foreign Language.* Oslo, Bergen, Tromsø: Universitet-sforlaget.

O'Connor, J. D (1967, 1971): *Better English Pronunciation.* Cambridge: University Press,

Poldauf, Ivan (1984): *English Word-Stress.* Oxford: Pergamon Press.

Prator, Clifford H and Betty W. Robinett (1972): *Manual of American English Pronunciation.* New York: Holt, Rinehart & Winston.

Ross, John Robert (1972): "A Reanalysis of English Word Stress". In*:* Michael Brame ed., *Contributions to Generative Phonology.* Austin: University of Texas Press, p. 229-323.

Teschner, Richard V & M. Stanley Whitley (2004): *Pronouncing English, A Stress-Based Approach with CD-ROM.* Washington, D.C.: Georgetown University Press.

Trevian, Ives (2003): *Morphoaccentologie et processus d'affixation de l'anglais*. Bern: P. Lang.

Walker, John (1791): *A Critical Pronouncing Dictionary and Expositor of the English Language.* London. A critical edition by Nicolas Trapateau, Université de Poitiers, forthcoming.

Wells, John C. (1990, 2000, 2008): *Longman Pronunciation Dictionary* Harlow: Longman-Pearson Education Ltd.

CHAPTER 6. STRESS SHIFT IN ENGLISH: THE CASE OF TEEN NUMBERS.[1]

José Antonio Mompeán
Murcia University

1. Introduction

The study of the phonetic and phonological features of stress has a long history in the specialized literature. In this respect, researchers have long distinguished between lexical stress and sentence stress. Lexical stress, also referred to as word stress, is the emphasis placed on syllables within phonological words –or a syllable's potential to receive prominence. Sentence stress refers to a certain natural stress pattern or 'rhythmic' characteristic of a given language –it is then often referred to as rhythmic stress. Sentence stress may also involve the placing of emphasis on particular syllables because of their importance in phonological phrases or intonation units (prosodic stress).
The use of lexical stress and sentence stress is language-specific, with noticeable differences across languages. In English, for example, lexical stress is a contrastive property of words (e.g. ˈ*increase* n. vs. (to) *inˈcrease* v.). Furthermore, the placement of lexical stress in words is variable although researchers have identified so-called stress 'rules'. Authors typically ground stress rules in phonology or morphology principles. Thus, Chomsky and Halle (1968), for example, outline phonological stress rules that rely mainly on syllable weight and the notion of lax vs. tense vowels. In contrast, authors like Fudge (1984) or Kingdom (1958) describe stress rules with respect to

[1] This paper has benefited from the invaluable feedback and assistance of Margaret Bermes (University of Murcia), Antonio Maurandi (SAI, University of Murcia), the editors of the volume (Inmaculada Arboleda, Rafael Monroy) as well as two anonymous reviewers.

affixation. As for sentence stress placement, certain tendencies can also be observed. For example, rhythmic stresses tend to fall on lexical (or content) words, while grammatical function words are usually rhythmically unstressed. Similarly, the most important prosodic (or 'nuclear') stress is typically found on the potentially stressed syllable of the last content word in an intonation unit. However, this tendency can shift for various pragmatic functions like, for example, focus or contrast depending on the portion of an utterance that is more important for the speaker or that the speaker wants the listener to concentrate on (see e.g. Wells, 2006 for an account).

One area of research that deserves attention is the relationship between lexical stress and rhythmic/prosodic stress. These often coincide, but not always. Traditionally, lexical stress has been considered as the input to sentence stress, whose rules may determine the redistribution or 'shift' of lexical stress within certain words and its suppression in others. The output of sentence stress becomes then an input to pitch-accentuation (i.e. intonation) rules, which yield the final configuration of prominence within and between intonation units. A case in point of the relationship between lexical and sentence stress is the phenomenon often referred to as 'stress shift', or the tendency in some languages to avoid two primary stresses on adjacent syllables. This is achieved by moving the primary stress from its canonical position to a previous syllable carrying secondary stress.[2] A simple example is *afternoon* /ˌɑːftəˈnuːn/, with canonical primary stress on the third syllable and secondary stress on the first syllable, when it is part of the compound *afternoon tea*.[3] This compound is usually pronounced [ˈɑːftə(ˌ)nuːn ˈtiː] –with the primary stresses two syllables apart from each other– rather than [ˌɑːftəˈnuːn ˈtiː] –with the primary lexical stresses of the two words falling on adjacent syllables– (Ashby & Maidment, 2005: 161). In stress shift, the primary and secondary stresses of the word undergoing stress shift are

[2] Apart from the extensive literature on English stress shift, several studies have identified equivalent phenomena in languages such as German (e.g. Bohn *et al.* 2011), or Dutch (e.g. Gussenhoven, 1983).

[3] Most traditional approaches to English stress assume that each syllable in polysyllabic words is ascribed one of three degrees of stress: *primary*, *secondary* or *unstressed*. Each such word has one primary stress, possibly one syllable having secondary stress, and the remainder unstressed. This is the approach taken in this chapter.

reversed, with the syllable that had the canonical primary stress pronounced then with secondary stress or losing stress altogether. Research on stress shift in English has been carried out at both theoretical and empirical levels. In phonological theory the phenomenon of stress shift, often referred to as the 'Rhythmic Rule', was first discussed within classical Generative Phonology (e.g. Chomsky & Halle, 1968) and it has also been the focus of more recent treatments in Optimality Theory (e.g. Cho, 2002). Stress shift has also been dealt with within Metrical Phonology (e.g. Giegerich, 1985; Gussenhoven, 1991; Hayes, 1984; Nespor, 1990). This body of work suggests that the motivation for stress shift is to be found in rhythm. Stress shift would be an attempt to avoid a so-called rhythmically disharmonic 'stress clash' (e.g. Liberman & Prince, 1977) and/or to achieve eurhythmy, a more even distribution of primarily stressed syllables in prosodic units (e.g. Hayes, 1984; Prince, 1983; Selkirk, 1984). Empirical research has focused, amongst other issues, on the acoustic correlates of stress shift and its perception. As for the acoustic correlates, syllable duration seems to be the main indicator for the perception and production of stress shifts (Bohn *et alii,* 2011; Grabe & Warren, 1995; Grela & Gandour, 1999), although other cues are also used (see e.g. Shattuck-Hufnagel, 1991). Moreover, irrespective of the acoustic correlates of stress shift, all researchers agree that native speakers –and, in some cases, foreign language learners– can perceive stress shift very accurately (e.g. García-Lecumberri, 2006; Grabe & Warren, 1995).

2. Stress and stress shift in teen numbers in English

One point of agreement between the different theoretical accounts and empirical studies carried out so far is that stress shift in English is not a categorical phenomenon but a variable one. The empirical studies available reveal the existence of variation in normal populations (e.g. Quené & Port, 2002) and even in special populations like aphasics (e.g. Grela & Gandour, 1999). Moreover, studies have identified some of the potential variables that condition variation in the use of stress shift. These include speaking rate and number of intermediate unstressed syllables between stresses (e.g. Quené

& Port, 2002, 2003), word frequency (Hammond, 1999) or stylistic variables such as degree of rhetoricity (e.g. Gussenhoven, 1983). One problem with the research carried out so far is that, although variation is generally acknowledged, little is known about its actual occurrence or frequency. This is mainly due to the fact studies typically involve subjects reading sentences containing experimental target words. However, comprehensive production data for stress shift are scarce (Grabe & Warren, 1995). Speakers seem to vary in the extent to which they use stress shift but, as Vogel *et alii,* (1995) put it, "…the small sizes of the data sets examined may cause more general patterns to be obscured by the variability…." (p. 113).

Given the scarcity of extensive production data on stress shift, the aim of the present study is to shed light on the variation in the use of stress shift in English. In this respect, one paradigmatic case of stress shift involves the compound teen numbers 13 to 19, both cardinal (e.g. ˌ*thir*ˈ*teen*) and ordinal (e.g. ˌ*thir*ˈ*teenth*) in attributive position. Lexical stress in teen numbers requires the primary stress to be on the last syllable whereas a secondary stress is placed earlier on the first syllable. However, as Celce-Murcia *et alii,* (1996: 138) point out, two different stress patterns are possible with teen numbers. On the one hand, the pattern with stress on –teen is more common in phrase-final or utterance-final position or when speakers are trying to make a deliberate distinction between the ten and teen digits (e.g. ˈ*thirty* vs. ˌ*thir*ˈ*teen*). On the other hand, the pattern with primary stress on the first syllable is more common when counting (e.g. ˈ*thir*ˌ*teen,* ˈ*four*ˌ*teen,* ˈ*fif*ˌ*teen*…) and before nouns in attributive position (e.g. ˈ*four*ˌ*teen* ˈ*women,* ˈ*thir*ˌ*teenth* ˈ*century*). This latter case, i.e. teen numbers in attributive position in noun phrases, is a potential context for stress shift. Although not mentioned by Celce-Murcia *et alii*, anecdotal observation of stress shift seems to suggest that stress on the first syllable in teen numbers is also common in compound words such as ˈ*eigh*ˌ*teen*-ˈ*wheeler*, ˈ*nine*ˌ*teenth* ˈ*hole*. (Jones *et alii,* 2011; Wells, 2008). However, compounds seem to be relatively infrequent except for the names of years (e.g. 1972, i.e. *nineteen seventy-two*), decades (e.g. 1980's, i.e. *the nineteen eighties*) and centuries (e.g. 18^{th} c., i.e. *eighteenth century*) as well as compound adjectives and nouns referring to age or age groups respectively (e.g. *a seventeen-year-old* *boy*, *fifteen-year-olds*, etc.).

Given the variation in the use of stress shift in English in general and in noun phrases and compounds including teen numbers as the focus in the current study, different questions can be asked regarding such variation. This study addresses one of those questions, namely how variable stress shift is in teen numbers in attributive position in noun phrases or as the first constituent of a compound word in a corpus of spoken English. Based on previous studies, it is hypothesized that stress shift will be very frequent in those contexts.

3. Method

3.1 Data

The study focused on the non-rhotic accent of English known as Received Pronunciation (henceforth RP). This accent was chosen as it is relatively easy to find considerable recorded speech that exemplifies it, and because it was considered important to rule out speakers' accent as a source of variation in the results.

The data were obtained from a corpus of newscasts from the BBC World Service website, available in the *Words in the News* link. Versions of this corpus have been previously used in the study of variation in phenomena such as syllabic consonants (Arboleda-Guirao & Monroy-Casas, 2010), phonological free variation (e.g. Mompeán, 2010) or /r/-liaison (e.g. Mompeán & Gómez, 2011). The study comprised the years 1999-2009. The corpus amounts to a total of 1503 newscasts, 1263 eventually included in the analysis and 240 ruled out because the speaker was not a newsreader, the speech was not RP, the text was technically unavailable, or the speaker's identity was unknown. Observing these criteria, 344 newsreaders were considered and 86 were left out due to the absence in the latter's speech of phonological/phonetic features typical of RP. The relevant texts analyzed amount to around 250,000 words and 32 hours of audio material. Moreover, out of the 258 speakers considered, potential cases of stress shift in teen numbers obtained as described in the Procedure section below were found in the speech of 142 speakers (104 males and 38 females). The number of potential cases of stress shift varies between one and 18 for each specific speaker, although most speakers have very few cases, with 73.1% of speakers having only between one and two potential cases in their newscasts, and 22.8% of speakers having between three and six cases.

3.2 Procedure

Once the relevant newscasts had been identified, the procedure involved two stages: a) identification of potential cases of stress shift involving teen numbers in compounds and in attributive position; and b) auditory and acoustic analysis of the potential cases.

For the identification of potential cases, automatic searches were made on a database created with the written version of the newscasts. The searches identified the ordinal numbers 13 to 19 as well as their ordinal counterparts ending in the suffix –th and other lexical items containing the combining form –teen such as *umpteen*. The searches also targeted figures in Arabic or Roman numbers as well as other abbreviations (e.g. cm., km., hrs., etc.) and symbols likely to attract teen numbers such as percent signs (%) or currency signs ($, €, £, etc.) to make sure no teen numbers were missed. Potential cases of stress shift excluded the use of teen numbers when a shift in stress could be due to pragmatic/discourse factors such as contrastive focus, i.e. the coding of information that is contrary to the presuppositions of the interlocutor, and requires the overt availability of a pair of alternatives (Riester & Baumann, 2011), as in *Did you say 'fifteen or 'sixteen?* (Wells, 2006: 133). With these considerations in mind, the identification process yielded 343 potential cases of stress shift.

After the identification of potential contexts, the cases were analyzed auditorily by two independent judges for the occurrence or non-occurrence of stress shift. Most decisions about the occurrence of stress shift were quickly made and coincidences between judges were above 95%. When in disagreement, the opinion of a third judge was sought after and a spectrographic inspection of the audio material was made with SFS and WASP (version 1.54), i.e. the *Speech Filing System* and related *Windows*

Tool for Speech Analysis, two free speech research software developed at University College London (Huckvale, 2013).

3.3 Results and Discussion

The analysis of the corpus provides empirical evidence to suggest an informed answer to the research question of the present study, namely how variable stress shift is in expressions involving teen numbers in a corpus of spoken RP speech. The analysis shows that out of the 343 potential cases of stress shift identified, 329 actually involved stress shift (95.9%) whereas stress shift did not apply in the remaining 14 potential cases (4.1%). This shows that stress shift is the rule, rather than the exception, in potential cases involving teen numbers as the first constituent of a compound or as the modifier of a head noun in a noun phrase. Apart from these general figures, it is also interesting to see the results by different types of items (e.g. compounds vs. noun phrases) and even by different types of lexical sets (e.g. decades' names) as well as by different types of noun phrases as defined semantically by the type of head noun.

As far as compounds are concerned, the corpus contains 247 compounds in which teen numbers were potential cases for stress shift. The results reveal that stress shift applied in 244 out of the potential cases (i.e. 98.8%). In this respect, eight compounds are names of centuries (e.g. *eighteenth century*), 67 compounds are names of decades (e.g. 1980s, 1990s, etc.), and 157 compounds are names of specific years such as 1880, 1992, etc. All these potential cases exhibit stress shift. This suggests that stress shift is categorical or nearly categorical in potential cases of stress shift when the compound designates a year, decade or century. The analysis of the corpus yielded 13 potential cases of teen numbers referring to age in compound adjectives (e.g. *sixteen-year-old boy*) or nouns (e.g. *eighteen-year-olds*), with ten cases exhibiting stress shift. The corpus included two miscellaneous compounds placed before noun phrases (*eighteen-month prison sentence*, *eight-hour working days*), both of which with stress shift.

Regarding noun phrases with teen numbers in attributive position (e.g.

fifteen days), the data reveal that stress shift is again the rule rather than the exception. Stress shift applied in 85 of the 96 potential cases (88.5%). The data also reveal a general tendency, namely the common use of head nouns relating to units of measurement (weight, distance, time, etc.), particularly in the case of units of time. These include the head noun *years* (29 cases) as well as other 'time' head nouns such as *hours, days, months*, and *minutes* (15 cases), with stress shift applying in all potential cases except three. In addition, head nouns often designate fractions of 100 such as *per cent* (14 cases) or multiples of ten such as *hundred*, *thousand, million* and *billion* (31 potential cases), with stress shift applying in all potential cases except six. The corpus also includes a small set of seven miscellaneous noun phrases with head nouns such as *people*, *brands*, etc., all of which exhibit stress shift.

One question that may be asked is whether there is any difference between compounds and noun phrases in English regarding the use of stress shift. In this respect, it could be argued that usage-based factors may determine the variability in the application of stress shift. These factors relate to the type and token frequency of use of expressions, types of units, etc., leading to the lexicalization of stress shift in certain compounds, types of expressions, etc. (see e.g. Langacker, 1999). It could be argued then that stress shift is more frequent in compounds as its application can have become lexicalized, whereas stress shift is more variable in noun phrases.

The data obtained seem to suggest that stress shift is more common in compounds as opposed to noun phrases, i.e. 97.6% (243) vs. 88.5% (85 cases). Lexicalization may well play a role in the frequency of use of stress shift, particularly in the case of compounds. In fact, a Chi-Square (Pearson's test of independence) analysis reveals that the difference is statistically significant (χ^2 (1)=16, p=0.0000689, ϕ = 0.23). However, it should be taken into account that there are considerable differences in the number of potential cases that each of the speakers has in the corpus, ranging from one single case to 18 cases. Thus, the effect of the number of potential cases per speaker has to be taken into account as it can affect the results.

In order to rule out the 'speaker' effect and provide a more informed answer to the question of whether there is an association between the grammatical status of the target items (compounds/noun phrases) and the presence or absence in them of stress shift (yes/no), a different approach was taken. This approach implies making random observations of single potential cases by individual subjects and calculating, using a Chi-Square test, whether there is an association between grammatical status and presence/absence of stress shift for each subject. It should be noted that the results obtained could certainly depend on the specific choice made in the observation of each subject. Thus, to rule out this effect of the random choice, a large number of observations of random subsets of speakers and subsequent chi-square calculations of them (N = 1,000) were made. The statistical software R 3.0.1 was used for this purpose. Next, the distribution of the p-values of the 1,000 chi-square tests was analyzed. If the confidence interval (CI) for the mean of the p-value distributions happened to be above 0.05, we could then conclude that there is no association between the two variables (grammatical status and stress shift), ruling out the subject effect. In this respect, figure 1 shows the distribution of the p-values for 1000 random subsets of speakers.

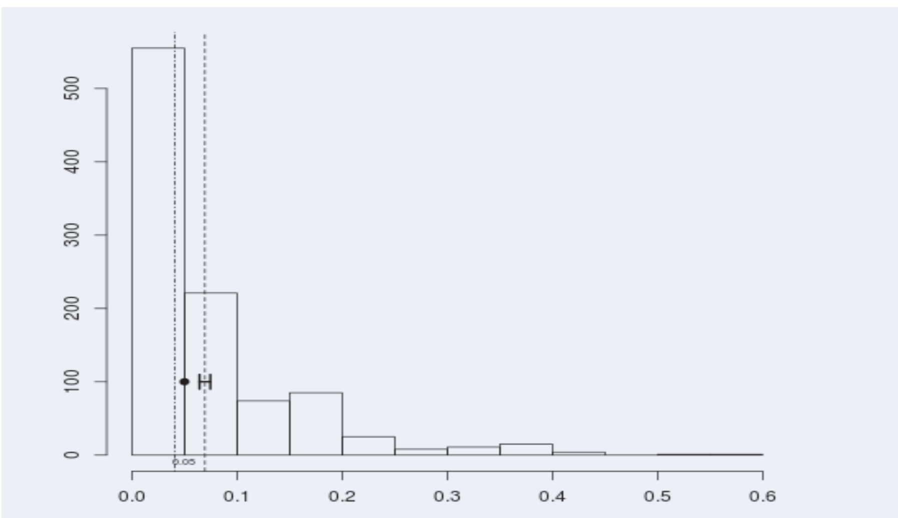


Figure 1. Histogram of the distribution of the p-values for 1000 chi-square tests with mean and median and a 95% confidence interval for the mean: (0.0643, 0.0746).

The figure shows that mean value of the p-value distribution (0.069) and the median (0.041). More interestingly, as the lower end of the 95% CI (0.064, 0.075) is higher than 0.05, we cannot confirm an association between the two variables (i.e. grammatical status and stress shift). Thus, the high rate of stress shift in both compounds and noun phrases may be most likely due to rhythmic reasons rather than the lexicalization or not of stress shift in the two subgroups. Alternatively, some sort of lexicalization might also have been present in some of the noun phrases analysed. For example, noun phrases including teen numbers are often part of larger expressions involving a teen number plus 'a(n) day/hour/month/year.." (e.g. *fifteen minutes a day*), 'ago' (e.g. *eighteen months ago*), 'later' (e.g. *eighteen years later*), etc. These phrases are relatively frequent in language and they could have a rhythmic pattern associated with them involving stress shift, which would contribute to the statistically non-significant differences between compounds and noun phrases in this study. In fact, there are nine examples of such phrases in the corpus, all of them exhibiting stress shift. Clearly, further research should look into this matter as well as the possible influence of other usage-based factors in the application of stress shift in compounds as opposed to that of noun phrases.

4. Conclusion

The relationship between lexical stress and sentence stress has received relatively little attention in empirical studies. In this respect, one interesting phenomenon is that of stress shift, or the tendency to avoid stress on adjacent syllables in prosodic units. The phenomenon has so far been the focus of some discussion in the phonological theoretical literature and in some empirical phonetic studies. All accounts of stress shift in English agree that stress shift is a variable phenomenon, and some variables that may condition such variability have been identified. However, little is known about how variable stress shift is in English. One main reason for this is the limited set of data analysed in empirical studies. In an attempt to answer this question the current study has looked at stress shift in a specific type of lexical item recurrently mentioned as exhibiting stress shift, namely teen numbers in attributive position or as the first constituent of a compound.

The study was carried out using a corpus of spoken RP English gathered from newscasts available on the BBC World Service website and comprising the years 1999 to 2009. The procedure involved the identification of potential cases of stress shift involving teen numbers in attributive position or compounds. The procedure also involved the auditory –also acoustic if necessary– analysis of the potential cases for the occurrence of stress shift. The analysis of the results reveals that stress shift is the rule rather than the exception in compounds and noun phrases involving teen numbers in the corpus studied. The analysis of the data also suggests that there are no differences in the occurrence of stress shift associated with the variable grammatical status (compounds vs. noun phrases). Thus, not only was stress shift overwhelmingly used in compounds, where its lexicalization could be expected, but also in attributive noun phrases. This suggests that stress shift is a rhythmic process that takes place irrespective of the grammatical status of the expression it can occur in.

Despite the results obtained, the current study has several limitations that should be acknowledged. In this respect, one important limitation is that the corpus was not extensive enough to study some word combinations exemplified by very few items; as a result, the conclusions drawn from the evidence obtained may be considered to be inconclusive. This suggests that the corpus should be enlarged to obtain more evidence. This is possible since the BBC World Service website continues to update its Words in the News section, so the amount of RP speech of the same characteristics continues to grow. Alternatively, a set of data elicited from informants in experimental settings could be used. This kind of data may allow better control of the number of potential cases of, for example, some of the lexical sets under investigation. Moreover, corpus-based and experimental approaches to stress shift research may complement each other and provide a better understanding of its variability.

Another limitation of the study is that it only looked at one accentual variety. This is a positive aspect in that the variable ‘accent’ was controlled for. However, the study cannot offer potential differences between different accentual varieties or even cross-linguistic differences between languages that exhibit stress shift. In this respect, future studies could be carried out to compare stress shift across different accents. A further limitation of the study is that it did not look at potential diachronic changes or sociolinguistic variables such as gender (e.g. male vs. female), age (younger vs. older speakers), register/degree of formality (e.g. colloquial vs. formal style), etc. Since the analysis of stress shift in the study is based on formal, careful, declarative style, where speakers tend to be speech-conscious, the study cannot offer a picture of stress shift in informal, colloquial, spontaneous speech. Thus, further research with non-scripted speech is needed. Similarly, the corpus does not allow the study of stress shift in data before 1999. However, other larger corpora could be used to study the use of stress shift from a diachronic perspective.

All these limitations suggest directions for future research that addresses these or other questions. For example, research could look into potential stress shift in teen numbers in expressions other than noun phrases or compounds. In the corpus analyzed, for example, the days of months occasionally show stress shift, although the data available are not extensive. As is well-known, month dates can be expressed in two different ways, i.e. mentioning the name of the day of the month, preceded by ‘the’ and followed by ‘of + the name of the month’ (e.g. the 13^{th} of November) or mentioning the name of the month and the ordinal number of the day preceded by ‘the’ (e.g. March the 17^{th}). Research could also look at compounds in which the teen number is the last constituent (e.g. the year 2015, i.e. *twenty fifteen*), hardly present in the corpus studied. Similarly, further research could look at related cases of potential stress shift. One such case is that of hyphenated numbers 21 to 99, with canonical primary stress on the second constituent (e.g. ˌ*thirty-*ˈ*three*). However, as Celce-Murcia *et alii* (1996:139) claim, hyphenated numbers in attributive position also undergo stress shift (e.g. ˈ*thirty-*ˌ*seven* ˈ*dollars*). Finally, future studies should also look at potential cases of stress shift in lexical items with suffixes other than –teen. These include adjectives with suffixes ending in -ic (e.g. ˌ*photo*ˈ*graphic* but ˈ*photo*(ˌ)*graphic* ˈ*record*), -ese (e.g. ˌ*Chin*ˈ*ese* but ˈ*Chin*(ˌ)*ese* ˈ*restaurant*), etc.

Despite the need for further research, the results obtained can have several practical applications in the teaching of English suprasegmental pronunciation as well as in lexicographic work for general or pronunciation dictionaries. The latter sometimes include not only alternative variants for lexical items, but also statistics on the different variables (e.g. Wells, 2008). The results can also be used to update introductory books describing English phonology and its variation or comparing it with the phonology of other languages. Finally, we would like to conclude this paper by saying that despite the limitations of the current study and the possible directions for future research, we believe our results show that stress shift in teen numbers can be considered to be a nearly categorical phenomenon in English.

REFERENCES

Arboleda-Guirao, Inmaculada & Monroy-Casas, Rafael (2010): "Gender and accent in the perception of English syllabic consonants vs. schwa". *Revista de Lingüística y Lenguas Aplicadas* 5: 17-28.

Ashby, Michael & Maidment, John (2005): *Introducing phonetic science*. Cambridge: Cambridge University Press.

Bohn, Karen, Wiese, Richard & Domahs, Ulrike (2011): "The status of the Rhythm Rule within and across boundaries in German". In: Lee, Wai-Sum & Zee, Eric (eds.). *Proceedings of the 17th International Congress of Phonetic Sciences*. City University of Hong Kong Press: Hong Kong, China, 332-335.

Celce-Murcia, Marianne, Brinton, Donna & Goodwin, Janet M (1996): *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. New York: Cambridge University Press.

Cho, Hyea-Sung (2002): "Stress clash revisited". *The Linguistic Association of Korea Journal* 10/1: 101-115.

Chomsky, Noam & Halle, Morris (1968): *The sound pattern of English*. New York: Harper & Row.

Fudge, Erik C (1984): *English word-Stress*. London & Boston: Allen & Unwin.

García-Lecumberri, María Luisa (2006): "Perception of double stress by Spanish learners of English". In: Baptista Barbara O. & Whatins Michael A. (eds.). *English with a Latin beat: Studies in Portuguese/Spanish-English interphonology*. Amsterdam: John Benjamins, 185–198.

Giegerich, Heinz J. (1985): *Metrical Phonology and phonological structure*. Cambridge: Cambridge University Press.

Grabe, Esther & Warren, Paul (1995): "Do speakers do it or do listeners hear it?" In: Connell, Bruce & Arvaniti, Amalia (eds.). *Phonology and phonetic evidence: Papers in Laboratory Phonology*, Vol. 4. New York: Cambridge University Press, 95-110.

Grela, Bernard & Gandour Jack (1999): "Case study stress shift in aphasia: A multiple case study". *Aphasiology* 13/2: 151-166.

Gussenhoven, Carlos (1983): "Stress shift in Dutch as a rhetorical device". *Linguistics* 21: 603-619.

Gussenhoven, Carlos (1991): "The English rhythm rule as an accent deletion rule". *Phonology* 8: 1-35.

Hammond, Michael (1999): "Lexical frequency and rhythm". In: Darnell, Michael, Moravcsik, Edith, Newmeyer, Frederick, Noonan, Michael & Wheatley, Kathleen (eds.). *Functionalism and formalism in linguistics*, vol. 1. General Papers. Amsterdam/Philadelphia: John Benjamins, 329-358.

Hayes, Bruce (1984): “The phonology of rhythm in English”. *Linguistic Inquiry* 15: 33-74.
Hoey, Michael (2005): *Lexical priming: A new theory of words and language*. London: Routledge.
Huckvale, Mark (2013): *SFS/WASP Version 1.54* <http://www.phon.ucl.ac.uk /resource/sfs> [downloaded 01/09/2013]. DO Consulted 1 September 2013
Jones, Daniel, Roach, Peter, Setter, Jane, & Esling, John (2011): *Cambridge English Pronouncing Dictionary*. [18th ed.]. Cambridge: Cambridge University Press.
Kingdom, Roger (1958): *The groundwork of English stress*. London: Longman, Green and Co.
Langacker, Ronald W (1999): *Grammar and conceptualization*. Berlin & New York: Mouton de Gruyter.
Liberman, Mark & Prince, Alan S. (1977): “On stress and linguistic rhythm”. *Linguistic Inquiry* 8: 249-336.
Mompeán, Jose A. & Gómez, Alberto (2011):“Hiatus-resolution strategies in non-rhotic English: The case of /r/-liaison”. In: Lee, Wai-Sum & Zee, Eric (eds.). *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, China, 1414-1417.
Mompeán, Jose A. (2010): “A corpus-based study of phonological free variation in English”. In: Henderson, Alice (ed.). *English pronunciation: Issues and practices (EPIP): Proceedings of the First International Conference*. June 3-5 2009, Université de Savoie, Chambéry, France. 13-36.
Nespor, Marina (1990): “On the separation of prosodic and rhythmic phonology”. In: Inkelas, Sharon & Zec, Draga (eds.). *The phonology-syntax connection*. Chicago: Chicago University Press, 217-242.
Prince, Alan S. (1983): “Relating to the grid”. *Linguistic Inquiry* 14: 19-100.
Quené, Hugo & Robert, Port (2002): “Stress shift in rhythmical speech”. *Acoustical Society of America Journal* 111/5: 2477.
Quené, Hugo & Robert, Port (2003): “Produced speech rhythm depends on predictability of stress patterns”. In: Solé, María J. & Recansens, Daniel (ed.). *Proceedings of the 15th International Congress of Phonetic Sciences.* Adelaide, South Australia: Causal Productions, pp. 2445-2448.
Riester, Arndt & Baumann, Stefan (2011): “Information structure annotation and secondary accents”. In: Dipper, Stefanie & Zinsmeister, Heike (eds.). *Beyond semantics: Corpus-based investigations of pragmatic and discourse phenomena*. Volume 3 of BLA (Bochumer Linguistische Arbeiten). Göttingen: Ruhr-Universität Bochum, Sprachwissenschaftliches Institut, 111-127.
Selkirk, Elisabeth O. (1984): *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.
Shattuck-Hufnagel, Stefanie (1991): “Acoustic correlates of stress shift”. *Proceedings of the XIIth International Congress of Phonetic Sciences*,

Aix-en-Provence: Université de Provence, Service des Publications. Vol. 4, 266-269.

Vogel, Irene, Bunnell, H. Timothy & Hoskins, Steven (1995): “The phonology and phonetics of the Rhythm Rule”. In: Connell, Bruce & Arvaniti, Amalia (eds.). *Phonology and phonetic evidence: Papers in laboratory phonology*, Vol. 4. New York: Cambridge University Press, 111-127.

Wells, John C. (2008): *Longman Pronunciation Dictionary* [3rd edition]. Harlow, UK: Longman.

Wells, John C. (2006): *English intonation: An introduction*. Cambridge: Cambridge University Press.

CHAPTER 7. YOU GOT THE BEAT: RHYTHM AND TIMING[4]

Robert Fuchs

University of Münster

1. Introduction

It is a truism that one learns most about one's own country when living abroad. Habits and values that seemed self-evident up to the point of being hardly noticeable at all, suddenly come to the fore through the confrontation with another culture. In the same manner, how a certain language works may become clearer through the comparison with another language or another dialect of the same language. Pitch (or intonation), for example, is a feature that in English is not used on single phonemes, but on larger units made up of several syllables. It is thus called a **supra-segmental** feature of English, i.e. applying above (Latin *supra*) the segment.[5] What pitch movement is used on an English utterance influences its pragmatic meaning. In Southern British English, a falling pitch movement is used at the end of a sentence to mark it as a declarative sentence (e.g. "In the Middle Ages people believed that the sun revolves around the earth."), but rising pitch indicates a question (e.g. "People believed that the sun revolves around the earth?").

---

4. The author would like to thank Ulrike Gut and the anonymous reviewer for their comments on an earlier version of this paper.

5. How phonemes combine into syllables and even larger units is often shown in tree diagrams that display smaller units (e.g. phonemes = segments) below larger units (e.g. syllables). These "trees" have their roots at the top and the branches extend downwards. Because of this convention it is customary to refer to features that apply to units larger than phonemes/segments as applying above the segment or being supra-segmental.

However, rising intonation is not always and exclusively used to mark questions. In California, for example, rising intonation in declarative sentences has become common among young women in the last decades. Initially derided as “valley girl speak” because it originated in the San Fernando Valley, it can now be found also among older and male speakers and in other regions. Speakers unfamiliar with this use of rising intonation may find it initially puzzling. They will come to realise that they only ever used rising intonation to encode declarative utterances, a fact that until that point might have seemed so self-evident that they were not consciously aware of it.
Apart from intonation, another supra-segmental feature is **rhythm**. In its earliest definition, (speech) rhythm was considered to distinguish two groups of languages, those that are **syllable-timed** (such as Spanish) and those that are **stress-timed** (such as English). Spanish rhythm was described as that of a machine gun, with impulses of equal strength (Fig 1 upper part). English rhythm, by contrast, was compared to that of Morse code (Fig. 1 lower part), a succession of weak and strong impulses.

Syllable-timing
(“Machine gun rhythm”)

Stress-timing
(“Morse code rhythm”)

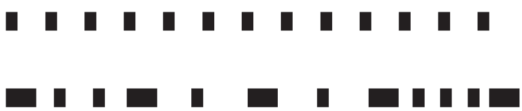


Figure 1. Schematic representation of idealised syllable-timing (above) with equally spaced impulses of the same strength (“syllables”), and stress-timing (below), with impulses of lesser strength (“syllables”) that are not equally spaced, and impulses of greater strength (stressed syllables/”feet”) that are equally spaced.

Although the terms rhythm and syllable- and stress-timing have been used for probably more than 100 years, they continue to be hotly debated concepts. Part of the reason for this is that just like generalisations about differences between cultures and nations, the distinction between syllable- and stress-timed languages is attractive for its simplicity but upon closer inspection glosses over many details. I will thus try to show in this chapter how British English rhythm can be characterised by comparison with other languages and other dialects of English, but will also focus on the details that such sweeping generalisations conceal.

2. Feet and syllables

First explanations of rhythm were quite simple. In syllable-timed languages, all **syllables** were thought to be of equal duration or **isochronic** (*iso* = Ancient Greek for equal, *chronos* =Ancient Greek for time). In stress-timed languages, on the other hand, all **feet** were said to be isochronic.[6] This came to be known as the isochrony hypothesis. The concept of feet and their rhythmic significance is familiar to many non-linguists from poetry, where trochees (=trochaic feet) for example consist of a stressed (also called strong) syllable followed by an unstressed (also called weak) syllable. Generally speaking, a foot consists of a stressed syllable followed by an arbitrary number of unstressed syllables. Most feet are made up of a stressed and one or two unstressed syllables, although there are also feet with more than two unstressed syllables or no unstressed syllable at all.

The crucial point in this definition is that if feet consist of a variable number of syllables, feet and syllables cannot be isochronic at the same time. If feet are isochronic, syllables in feet consisting of one or two syllables have to be compressed, and in longer feet they have to be stretched. Imagine an utterance consisting of three feet, each 300 ms long (somewhat less than one third of a second). If in the first foot there is one syllable, two in the second, and three in the third, then the syllable in the first has to be 300 ms long, the two syllables in the second foot are on average 150 ms long and the remaining three syllables in the third foot are 100 ms long, on average (see Tab. 1).

6. A third group, mora-timed languages (such as Japanese) was later added, where the mora, a sub-syllabic unit, is isochronic. Mora-timing has received much less attention than syllable- and stress-timing and will not be discussed any further here.

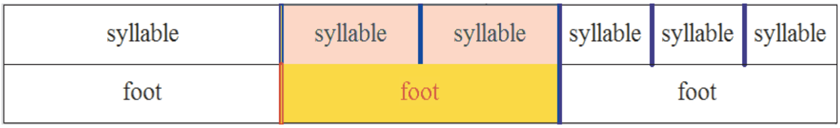


Figure 2. Schematic representation of stress-timing. All feet are of equal duration and the length of the syllables depends on how many other syllables are contained in the same foot.

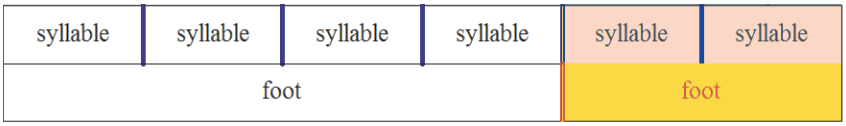


Figure 3. Schematic representation of syllable-timing. All syllables are of equal duration (isochronic) and the length of each foot is a function of the number of syllables it contains.

Likewise, if syllables are isochronic then the feet have to be of variable duration since all feet do not contain the same number of syllables. A foot with four syllables will be twice as long as one with only two syllables (see Tab. 2).

3. Measuring rhythm

3.1 Pairwise Variability

The idea of rhythm classes came up in the first half of the 20th century, when acoustic equipment was not sophisticated enough to verify the isochrony of feet in stress-timed and of syllables in syllable-timed languages. Technological advances in the earlier second half of the 20th century provided a boost to acoustic research, and the isochrony hypothesis turned out to be inaccurate. Neither feet nor syllables were found to be of equal length in the languages examined.

Instead, later research showed that if syllables were not isochronic in so-called syllable-timed languages, at least durational differences between successive syllables were smaller in syllable-timed than in stress-timed languages. The basic idea of this new approach is quite simple. If we want to determine whether the differences in duration between syllables matter for a listener, we only need to focus on those differences that are easily perceived. We can thus restrict ourselves to comparing syllables that directly follow each other. Thus, in an utterance, the first and second syllable will be compared, then the second and third, followed by the third and fourth and so on.

How does this work in practice? Let us take a look at a part of a sentence read by a male student from the South-East of England (taken from the Dynamic Variability in Speech Database - DyViS, for details see Nolan *et alii*. 2006 or the project website at http://www.ling.cam.ac.uk/dyvis/). Step by step, we will develop a suitable measure of how much the syllables in an utterance differ in duration from each other. Readers who are not used to working with numbers may find the prospect a bit daunting, but the important point is not to focus on the numbers, which are given here for illustration only, but to grasp the principle.

In our sentence fragment ("Police announced last night", see Fig. 2), the first syllable, [pə], is 97 ms long and the second, [li], is 154 ms long, which makes the difference between the two 97-145= - 48ms. The difference between the second and the third ([sə], 180 ms) syllable is 145-180= - 35ms. These values are negative because in both cases the second syllable was longer than the first and we subtracted the duration of the second from that of the first. Since we are not interested in the direction of the difference but just in its size, we take the absolute of the difference and disregard its sign. The differences between the remaining syllables are calculated similarly (see the third row of Table 1 for details).

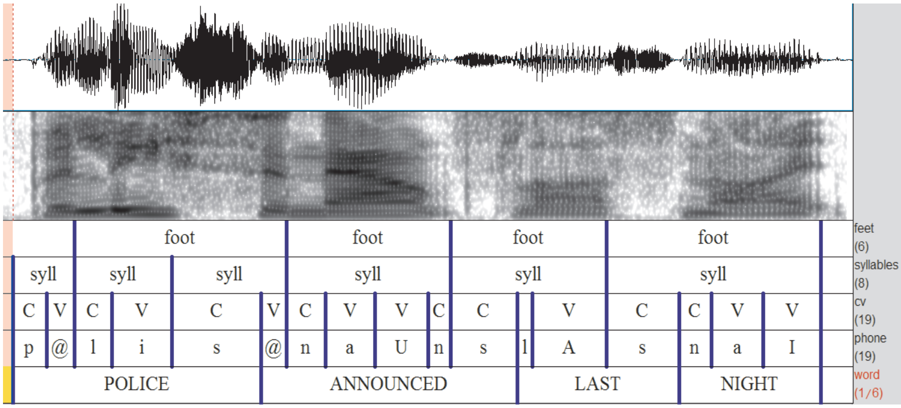


Figure 4. A recording and transcription of the phrase "Police announced last night" read by a speaker of Southern British English. The top part shows the waveform of the utterance, below that the spectrogram is shown, and several levels of transcription follow (feet, syllables, consonantal and vocalic intervals, phonemes, words). Note that the final /t/ of 'announced' and 'last' are elided, which is common in connected speech.

Table 1. Durations of the syllables in the phrase from Fig.2, their absolute and normalised pairwise differences, and the average of those.

<table>
<tr><td>Syllables</td><td>pə</td><td>li</td><td>sə</td><td>naʊn</td><td>slɑ</td><td>snaɪ</td><td></td></tr>
<tr><td>Syllable durations (ms)</td><td>97</td><td>154</td><td>180</td><td>259</td><td>245</td><td>338</td><td></td></tr>
<tr><td>Difference (absolute)</td><td>|97-154| = 57</td><td>|154-180| = 26</td><td>|180-259| = 79</td><td>|259-245| = 15</td><td>|245-338| = 93</td><td>Average:</td><td>54</td></tr>
<tr><td>Difference (normalised)</td><td>|97-154| /(97+154) *200 = 45</td><td>|154-180|/ (154+180) *200 = 16</td><td>|180-259|/ (180+259) *200 = 36</td><td>|259-245|/ (259+245) *200 = 6</td><td>|245-338|/ (245+338) *200 = 32</td><td>Average:</td><td>27</td></tr>
</table>

Now that we know how much the individual syllables differ from each other in duration, the next step consists in determining the average difference between successive syllables. That is, we take all the differences computed in the above manner, and determine the average. Since there are six syllables in the fragment, there are only five comparisons to be made: (57+26+79+15+93)/5=54. Thus, the average difference between successive syllables in this utterance is 54ms.

This fragment was spoken with average speed. Now imagine the speaker had uttered the same phrase much faster, in half the time. We will assume that every syllable is now half as long as before, but the relative differences between them remain the same. The rhythm of this utterance has not changed through the acceleration, so the index computed before should remain the same if it is useful for measuring rhythm. In fact, it is halved to 32ms, because as the durations of all syllables were halved, so were the differences between them. For example, before the acceleration the difference between the first and second syllable was 97-145= - 48ms. After doubling the speed, the difference is only 48.5-72.5 = - 24. This is contrary to the assumption that rhythm remains the same irrespective of how fast a speaker talks.

How can this problem be solved? Instead of just computing the absolute differences between syllables, it may be better to normalise this difference by the added durations of both syllables. In this way, relative differences are used as the basis of our syllabic rhythm index. For the first two syllables, this is (97-145)/(97+145) = -0.20. This is a very small value, and for reasons of practicality it is customary to multiply it by 200, which makes it 40 (taking the absolute as above). For the second and third syllables, this makes (145-180)/(145+180)*200=-22. After doing this for all pairs of consecutive syllables (see row four of Table 1 for details) we can again compute the average difference between consecutive syllables: (45+16+36+6+32)/5=27. Of course, a serious investigation of rhythm in British English could not make do with just one fragment of an utterance. Longer stretches of speech from a number of speakers would have to be considered. In fact, the variability found in this short utterance is quite small and not representative of British English rhythm in general. A more common value would be between 45 and 55. It is normal that values of individual utterances deviate from this average and it will be discussed below what consequences this has for referring to British English rhythm as a whole.

To summarise the approach, we have computed an index of durational variability between successive syllables. In addition, this index was normalised for speech rate. The index is known as the **normalised Pairwise**

Variability Index for Syllables (nPVI-S). Slightly different versions and different names are used by scholars, but the idea of a pairwise comparison of durations with subsequent normalisation is common to most of them (for an overview see Fuchs 2013).

Next, consider the feet found in this utterance. The same method will now be used to compute how much they differ in duration from each other in this utterance. To determine what the feet are, we need to find out which syllables are stressed. We assume that a syllable containing a full vowel is stressed. The first and third vowels are not full, so that the first foot starts with the second syllable. The first syllable remains unassigned to any foot as it is not headed by any stressed syllable. The four feet in this utterance are then [li.sə], [naʊn], [slɑs] and [naɪ]. The first two are 334 and 260 ms long, respectively, and thus (334-260)/(334+260)*200=25. Doing that for all pairs of consecutive feet, we get the average difference between them (25+6+32)/3=21. Note that the average difference between the feet (21) is smaller than the average difference between the syllables (27) in this utterance. This indeed lends support to the idea that (Southern British) English is a stress-timed language or variety of a language, as feet tend to differ less in duration from each other than syllables.

Such a conclusion of course needs to be based on a sample of a number of different sentences and speakers, and research in this vein has found similar tendencies. The example utterance considered in this chapter is not typical of Southern British English in the specific values computed here, but it is typical in the tendencies shown. Average values using the same approach would be around 45 for differences between feet and around 60 for differences between syllables. British English speech tends to show higher variability in the durations of syllables than feet. This is not the case in syllable-timed languages. Here, either both syllables and feet are equally variable in duration, as has been shown for Castellano/Castilian Spanish, or variability in foot duration is higher than in syllable duration, as has been shown for Mexican Spanish (Nolan and Asu 2009).

Whereas early concepts of speech rhythm posited distinct categories (syllable-timed languages vs. stress-timed languages), the more recent quantitative approach implies a gradient understanding of speech rhythm.

Now, languages (or varieties of a language) can be compared. For example, Spanish can be said to be more syllable-timed than British English, and British English in turn is more stress-timed than Spanish. On such a scalar understanding of rhythm, stress-timing and syllable-timing are poles of a continuum. These poles may even be understood as ideals that no language ever reaches, but languages can be located on this scale with respect to the poles and to other languages.

3. 2 Methodological problems

The preceding section demonstrated that comparing variability in the durations of feet and syllables can account for widespread intuitions about which languages are syllable-timed and which are stress-timed. However, we have not defined how exactly we determine where a syllable or foot ends and where the next begins in an utterance. There is indeed widespread agreement in linguistics that feet and syllables are useful supra-segmental categories. But when faced with a recording of a sentence, actually cutting up this recording into syllables (a process known as **syllabification**) or feet will produce different results depending on who you ask.

To many readers such considerations may seem either unsettling or unnecessary, as the syllable is a concept known to most adolescents at least on lower secondary level and the idea of the foot is often introduced to upper secondary school students or earlier when discussing poetry. Nevertheless, any empirical science is fraught with such problems, including more established fields such as physics, and they can even occur in everyday situations. For illustration, consider a train operating company and an airline, both of which claim that 95% of their passengers arrive on time at their destination. The train operating company may define “on time” as not more than five minutes behind schedule. Some passengers may find a delay of five minutes too much, but even those who would happily accept it will find that a train running 20 minutes behind schedule cannot by any definition be considered “on time”. Those same passengers will in all likelihood happily accept the same delay if they took a plane, perhaps because going through security takes so much time that 20 minutes more or less don't matter, or perhaps because they are so used to airplanes being delayed that 20 minutes seem like a good deal in comparison. That is, their definition of what delay is still acceptable and what can be considered to be on time depends on the mode of transportation, even with such a common sense notion.

Definitions of syllable boundaries, too, may differ between languages, and even for one and the same language different but equally sensible definitions may exist side by side –just as train operating company A may consider delays of up to three and company B of up to five minutes to be within the margin of tolerance. In the following, some methods of syllabification in English will be considered as well as how they may influence the definition of rhythm given above. All have their justification, just as different definitions of what an acceptable delay is may exist side by side.

Syllables consist of a nucleus (usually a vowel) and optional onset and coda (made up of consonants). Syllabification starts with vowels, which function as syllable nuclei or the center of a syllable. A syllable nucleus can be a monophthong or a diphthong. Nasal consonants can also sometimes function as syllable nuclei but this is less frequent and will be disregarded here. Once the syllable nuclei in an utterance have been identified, any consonants in the utterance have to be attached to one of the nuclei. One common method of doing this rests on the assumption that any consonant sequence occurring at the beginning of an English word is a possible onset. Other sequences are not allowed as onsets (see Fallows 1981: 52-63, and Duanmu 2009 for a discussion of syllabification methods).

In the above example, [pə.li.sə.naʊn.slɑ.snaɪ] (dots indicate syllable boundaries), [s] is attached to the third syllable, [sl] to the fifth and [sn] to the last syllable because words with these onsets occur in English (for example 'sun', 'slate' and 'snail', respectively). The fourth and fifth syllables could have also been arranged as [naʊns.lɑ] with the [s] belonging to the coda of the preceding syllable, since both [l] and [sl] are possible onsets in English, if it were not for another crucial point in this syllabification method: The aim is to make the onsets as long as possible, and [sl] is a longer onset than just [l]. The method is thus known as the **Maximum Onset Principle (MOP)**.

There is, however, another reason why the [s] might be attached to the preceding syllable, which is that a word boundary intervenes between [s] and [l]. Arguments in favour of maintaining word boundaries during syllabification are that higher level (word) boundaries should be maintained when assigning lower level (syllable) boundaries, and that some consonants have different allophones depending on pre- or postvocalic position (such as clear and dark /l/). The major argument in favour of syllabification across word boundaries is that linking is a feature of British English and that linking triggers resyllabification, such that [naʊn.slɑ] is to be preferred (see Davis 2000: 27-60, and references cited in Gick 2003: 222).

Both choices are justifiable, but they influence the measurement of rhythm. When every word boundary is automatically also a syllable boundary, syllables tend to differ more from each other in duration, and hence the variability index becomes higher. If word boundaries are maintained, in the example clause [pə.lis.ə.naʊn] the third syllable now consists of [ə] only and the preceding [s] becomes the coda of the preceding syllable. The third syllable is now much shorter than the second and fourth syllable. To a lesser degree, maintaining word boundaries during syllabification probably also increases variability in feet durations.

A compromise between both options is to consider single intervocalic consonants as belonging to both syllables, which is called **ambisyllabicity** (*ambi* = Latin for both). For the purposes of measuring syllable durations, the [s] could be split with the first half belonging to the preceding and the second to the following syllable. When more than one consonant occurs between two nuclei, half of them should belong to the preceding and the other half to the following syllable if there is an even number of consonants. For uneven numbers the middle one is to be split, all this under the condition that the resulting onset is a possible onset (Pulgram 1970).

Proponents of ambisyllabicity argue that it is propped up by psycholinguistic studies where native informants, when asked to spell out the syllables of a sentence, sometimes indicate that a consonant should belong to both adjacent syllables or else there is frequent disagreement between informants as to which syllable a consonant should belong to. A counter-argument is that splitting of phonemes for syllabification introduces a boundary on a higher level (syllable) where there was none on the lower level (phoneme). In terms of the consequences for the measurement of rhythm, ambisyllabicity will even out differences in duration between adjacent syllables and thus lead to a reduction of values for the syllabic index.

That such different methods of syllabification exist may sound utterly defeating for the whole enterprise of basing rhythm measurements on the durations of syllables and feet; especially since there is not much more agreement when it comes to defining feet. On the other hand we should not throw out the baby with the bathwater. Going back to the train operator metaphor, no matter whether a train is defined as "on time" if it arrives no later than three, five or ten minutes, in most comparisons of two operators if A has more trains arriving below three minutes behind schedule than B, it probably also has fewer trains arriving more than ten minutes late than B. Likewise, it is quite possible that different syllabification methods might produce similar results, as long as what is of interest is whether syllables or feet show more durational variability and, if they do, by what factor, roughly, feet are more (ir)regular in duration than syllables. One condition, though, has to be maintained at all times, which is that absolute values produced by the rhythm measures can only be compared when similar definitions for syllables and feet were used. This is important when comparing results from different studies.

3.3 From syllabic to vocalic variability

As the previous section has shown, problems occur when defining syllables even within a single language. Maintaining consistent definitions across languages is even more difficult since concepts such as the MOP are designed to differ between languages. This is because a possible onset in one language is not necessarily allowed in another language. A way out of this dilemma is to abandon the syllable as the basic unit of duration measurements for rhythm studies.

Instead, vowels, or more specifically, vocalic intervals and their durations can be used. A vocalic interval is a stretch of speech consisting of vowels only, which means that it can consist of one, two or three consecutive vowels in English. Immediately adjacent to each vocalic interval there is

either a consonantal interval or a pause.
For a number of reasons, rhythm measures based on vocalic intervals have been used more often and for longer than those based on syllables. The first reason, as already indicated, is that vocalic intervals can be identified with less recourse to the phonology of languages and one can instead rely more often on phonetic/acoustic landmarks since the articulation of vowels is quite different from that of most consonants. This makes comparisons between languages more reliable. Second, although the initial definitions of syllable- and stress-timing did not refer to vowels or vocalic intervals, vowel reduction has relatively early been identified as a frequent correlate of stress-timing. That is, syllable-timed languages tend to not show vowel reduction (Spanish for example does not), but stress-timed languages do have vowel reduction (English and German, for example, do). In the latter group, the vowels in unstressed syllables are often reduced in both quantity (shorter) and quality (centralised vowels, close to schwa, e.g. the first vowel in 'announced' /ənaʊnst/).
Variability in durations of consecutive vocalic intervals can be measured in the same way as between syllables and feet. The **Pairwise Variability Index for Vocalic Intervals** is also normalised, and is abbreviated nPVI-V (Low 2000, Grabe and Low 2002). In difference to the syllabic and foot indices, vocalic intervals do not follow each other immediately, as a matter of fact, but are separated by intervening consonantal intervals.
In the example clause [pəlisənaʊnslɑsnaɪ] , the first vocalic interval consists of a schwa only and is 44 ms long, and the next is 95 ms long. The normalised difference between the two is (44-95)/(44+95)*200=-73, but the sign will again be ignored in all cases. The average durational variability of vocalic intervals in this clause is then, together with the other syllables, (73+81+121+31+38)/5=69, which is a fairly high value for British English. Average values for Southern British English are around 57.
The nPVI-V is the most widely used rhythm measure and has often substantiated intuitions about which languages or varieties of a language are syllable- or stress-timed.[7] Spanish, for example, shows lower vocalic

7. Other rhythm measures are also used, for example %V, which is the percentage of all vocalic parts in the whole utterance. %V is higher in syllable-timed than in stress-timed languages. Other measures rely on the durations of consonantal intervals, but are much less reliable than

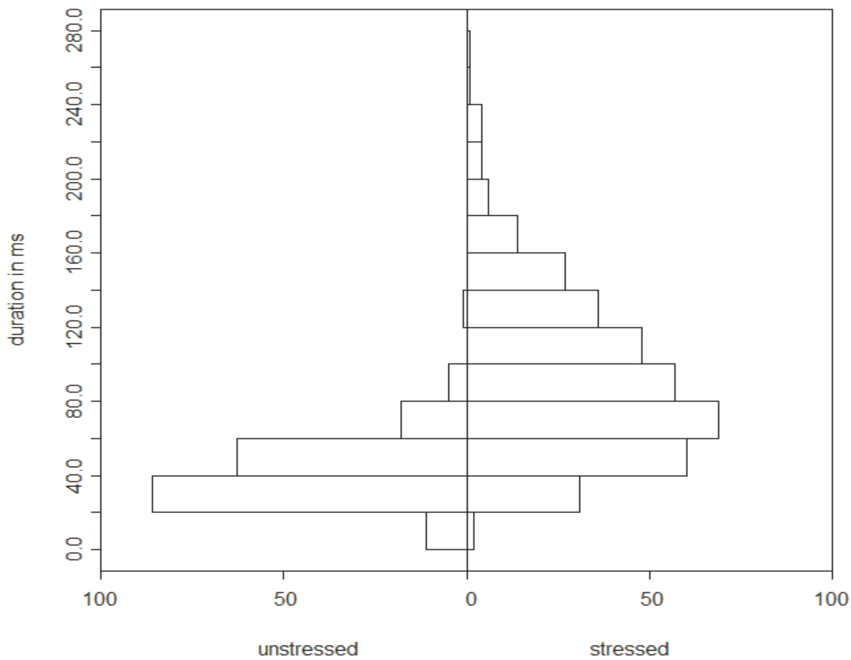


Figure 5. Histogram of the durations of all stressed and unstressed vowels in a passage read by a speaker of Southern British English. The length of each horizontal bar, measured from the center, indicates the number of vowels whose duration falls in each of the 20 ms ranges shown on the vertical axis. Values for unstressed vowels are shown in the left panel, values for stressed vowels in the right panel.

variability than British English, varieties on the Celtic fringes of the British Isles (Irish, Cornwall, and Shetland English) show lower variability than Southern British English, and nativised varieties such as Nigerian, Singapore, and Indian English also show lower values than British English.

One of the main reasons why there is so much variability in vocalic interval durations (as indicated by high nPVI-V values) in Southern British English is **vowel reduction** in unstressed syllables. For example, the first vowel in the example phrase is less than half as long (44 ms) as the second vowel (95 ms), which occurs in a stressed syllable. This is because the first is an unstressed vowel, reduced to schwa, and the second is a full vowel. Just because this is the case in the example phrase it does not mean it applies to

---

consonantal measures. See Wiget *et alii* (2010) for a comparison of consonantal and vocalic measures and their ability to discriminate stress-timed from syllable-timed languages.

Southern British English in general, so let us take a look at a longer stretch of speech.

In a recording of a text of around 400 words read by a male speaker of Southern British English (taken from the DyViS database), all stressed and unstressed vowels were marked. Fig. 3 shows a histogram of the durations of all vowels, with unstressed vowels on the left and stressed vowels on the right. The length of each bar represents how many vowels fall in a certain range. For example, the second bar from the bottom in the left panel represents all vowels between 20 and 40 ms (as shown on the vertical axis). This bar is the longest in the left panel, it represents more than 80 vowel tokens. In the right panel, the longest bar is the one representing the 60 to 80 ms range. More than 70 stressed vowels fall into this range. Also note that there are some very long stressed vowels more than 160 ms long, but no unstressed vowels in this range. It is clear from this chart that stressed vowels in this recording are, on average, much longer than unstressed vowels (and this difference is highly significant), a tendency well documented for speakers of British English in general.

Unstressed (short) and stressed (long) vowels alternate in many utterances, thus giving rise to large differences in duration and hence high nPVI-V values. This alternation is not perfect however, sometimes two or more unstressed vowels (separated by consonants) occur in succession, and sometimes two or more stressed vowels occur in a row, just like the fourth, fifth and sixth vowels in the example phrase above. Typically, function words such as articles and prepositions contain reduced vowels, as well as syllables of lexical words that do not bear lexical stress (the first vowel in 'police' is an example we already came across).

Vowel reduction, however, is not the only phonological feature of British English that gives rise to its stress-timed rhythm. Vowel lengthening or shortening effects also arise in other contexts, and produce more variation in duration that increases nPVI-V values and also variation in syllabic duration. Other factors increasing variation in vocalic interval duration are that vowel duration acts as a cue for phonological voicing in British English and that vowel lengthening also occurs in syllables that are the focus of a sentence. These two phenomena will be discussed in turn.

Vowel duration acts **as a cue for phonological voicing** in English, which means that vowels preceding a voiced plosive or fricative are longer than those preceding a voiceless plosive or fricative (Klatt 1976: 1214). This may sound surprising, since the terms voiced and voiceless seem to imply that the absence or presence of voicing is what distinguishes voiced and voiceless plosives/fricatives from each other. The confusion arises because the terms voicing/voiced/voiceless are sometimes used to refer to a phonological category (the contrast between /p/ and /b/ for example –note the slashes which indicate phonemes) and sometimes to refer to the phonetic realisation of this category. Crucially, a bilabial plosive may be perceived as /b/ when vocal fold vibration (i.e. periodic opening and closing of the soft tissue at the entrance to your windpipe) is involved in its realisation, but other so-called cues can also trigger this perception even when vocal fold vibration is absent. In other words, an English bilabial plosive can be perceived as /b/ even when not vocal fold vibration (voicing) is involved.

The difference between the phonological category [+/- voice] and its phonetic realisation appears counter-intuitive at first glance, but such phenomena may also occur with other senses. The perception of heat, for example, is without doubt partially caused by high air temperature. When the latter increases, we feel it has become hotter. However, we will also feel it is hotter when air humidity is increased and the perceived temperature decreases when a breeze sets in. In other words, perceived temperature has multiple cues, among them of course actual physical temperature, but also air humidity and air speed.

Just as perceived temperature, phonological voicing also has multiple cues. If and when glottal fold vibration sets in is an important cue, but equally important is the duration of a preceding vowel. In many languages that have a voicing contrast, longer vowels occur before voiced plosives and fricatives and shorter vowels before voiceless plosives and fricatives. Take for example the words 'reaches' and 'reads', which occur in a set of sentences that speakers participating in the DyViS project had to read. The two words are similar from the start of the word until the plosive (the first part of the affricate in 'reaches' is phonetically similar to the plosive [t]), and thus are ideal for comparison. Two speakers were chosen at random, and the length of [i] was measured across four repetitions of each sentence.

Table 2 shows the length of each vowel token for the two speakers with a following voiced or voiceless plosive and their averages. Average vowel duration before a voiced plosive is 133 ms for speaker 1 and 110 ms for speaker 2, whereas before a voiceless plosive vowel duration is only 60 ms for speaker 1 and 26 ms for speaker 2. It is clear that both speakers produce much longer vowels before the voiced than before the voiceless plosives. What is more, it has been shown that British English speakers do not only produce longer vowels before voiced plosives and fricatives, but that this

Table 2. Vowel length (in ms) before voiced and voiceless plosives in similar contexts.

	Speaker 1		Speaker 2	
	/rid/	/riʧ/	/rid/	/riʧ/
	140	66	142	27
	129	52	110	20
	124	56	84	24
	140	66	105	32
Average	133	60	110	26

also acts as a cue for the perception of (phonological) voicing of the following plosive or fricative. This has been tested by playing recordings of minimal pairs (such as 'dog' and 'dock') to native speakers where the vowels have been artificially lengthened or shortened. When asked which word they heard, participants tended to opt for the one with a following voiceless plosive if the vowel is shorter and for the following voiced plosive when the vowel is longer.

For British English speech rhythm (and also other varieties of English) this means that there is, in addition to vowel reduction, another source of variation in the duration of vowels and consequently also syllables (since a major part of most syllables is a vowel). That vowels are longer before voiced plosives and fricatives seems to be a cross-linguistic tendency. Interestingly, in some syllable-timed languages this contrast is smaller (e.g. French; Chen 1970), or very small and rarely used as a cue (e.g. Spanish; Morrison 2002), or there are no syllable-final plosives that could give rise to such a phenomenon (e.g. Chinese; Wang and Wu 2001).

A third source of variation in vocalic interval duration in British English is lengthening of syllables that are in focus. The focus is the element of a sentence that answers a (hypothetical) question. For example, the DyViS database contains variations of sentences such as 'We don't want you doing that'. One version answers the question 'Who doesn't want you doing that?' – 'Wé don't want you doing that' ('we' is in focus), another answers 'Who don't we want doing that'? –'We don't want yóu doing that' ('you' is in focus) and a third is a reply to 'What don't we want you doing?'– 'We don't want you doing thát' ('that' is in focus).

Table 3. Duration of the word 'we' (underlined) in ms in the sentences '**Wé** don't want you doing that' (answering 'Who doesn't want you doing that?') and 'We don't want you doing **thát**' (answering 'Want don't we want you doing?'), spoken by two speakers of Southern British English.

	Speaker 1		Speaker 2	
	wé don't … that	we don't … **that**	**wé** don't … that	we don't … **thát**
	194	39	209	40
	226	34	196	36
	239	22	178	42
	190	30	200	42
Average	212	31	196	40

Table 5 shows the duration of 'we' in the first version answering the 'who doesn't want' question and in the third version answering the 'what' question. In the condition where 'we' is in focus (left columns for each speaker) it is much longer than in the condition where 'that' is in focus (right columns). This shows what a strong influence sentence focus can have on the duration of syllables, and consequently, vocalic intervals in British English. Together with vowel reduction and vowel lengthening before voiced plosives and fricatives this factor is responsible for much variation in the durations of vocalic intervals and syllables.

3.4 Variation between speakers and sentences

Having laid out factors responsible for the stress-timed rhythm of British English, it is time to critically examine the evidence. We have seen that vocalic interval duration (as measured by nPVI-V) is, on average, much more variable in British English than in syllable-timed languages such as Spanish. But when calculating nPVI-S and nPVI-V for the example phrase 'Police announced last night', it was pointed out that syllable duration variability is lower in this phrase than in Southern British English in general. On the other hand, the variability of vocalic interval duration was found to be somewhat higher than the average for Southern British English. If there are such differences between individual sentences, does it make sense to speak of the rhythm of Southern British English? And what about differences between individual speakers? Do all speakers of Southern British English use a similar rhythm or are there also differences with regard to this dimension?

It is possible to construct English sentences that contain almost only full vowels. If such sentences are read by native speakers, and when their vocalic intervals are measured and compared to calculate nPVI-V, one arrives at very low values. Such sentences could be called syllable-timed sentences. An example is “These three large bears swam too soon” (Gibbon and Gut 2005). Particularly stress-timed sentences that lead to very high nPVI-V values can also be constructed, such as “The production increased by three fifths in the last quarter of two thousand and seven” (Arvaniti 2009).

Also for syllable-timed languages such as Spanish, similar sentences with (for the respective languages) unusually high or low nPVI-V values can be constructed and this has been taken by some to suggest that it makes little sense to make general statements like 'Language x is stress-timed' if indeed there is so much variation within each language.

Two arguments preclude such a conclusion. First, such sentences, notwithstanding their occasional occurrence even in natural speech, are unusual and not representative of each language as a whole. To draw a comparison, let us consider two groups of students, one of which receives, on average, very good grades, and another, whose average grades are worse. It proves nothing to point out that in the group with good grades there are some students who received abysmal results, and that in the bad group there are some with excellent grades. Whether two groups differ significantly from each other is a matter of statistical significance and not of what the maximal and minimal values are in each group. In other words, we need to look at randomly chosen sentences if we want to draw conclusions on the rhythm of a language. In comparisons of a number of languages, it was shown that rhythm measures such as nPVI-V can reliably distinguish between stress-timed and syllable-timed languages such as British English and Spanish, respectively (Wiget *et alii*, 2010).

The second argument is that the phonology of each language only allows so much manipulation in terms of rhythm. For example, English stress-timed sentences show higher nPVI-V values than average English sentences. “Stress-timed” sentences can also be constructed for Spanish, and these in turn show higher nPVI-V values than average Spanish sentences. But the Spanish “stress-timed” sentences still lead to much lower nPVI-V values (based on lower variability in vocalic interval durations) than English stress-timed sentences. In other words, the phonology of each language allows manipulations of speech rhythm, but only to a certain degree. It still makes sense to describe British English as more stress-timed than Spanish because the average British English sentence has a more stress-timed rhythm than the average Spanish sentence.

The same is true of variation between speakers. When rhythm measures such as nPVI-V are applied to the speech of a number of speakers of British English (or other languages), a certain degree of variation between speakers becomes apparent. Some speakers produce speech with more variability in duration, and some with less. But it is still possible to identify a range where most British English speakers can be situated. This, in other words, is the average speech rhythm of British English. Variation between individual speakers is normal, and has long been recognised for many features of English and other languages.

Speakers often consciously choose certain pronunciations or certain words because of their identity value. T-glottaling, the realisation of intervocalic /t/ as a glottal stop in words such as 'potter', has long been regarded as a working class feature in England, but has now become fashionable with young speakers from many social levels. Not every speaker uses it and those who do usually not all the time. Young Britons may use t-glottaling when talking to friends of a similar age, but will use it less often in formal situations, such as a job interview.

Likewise, variation in speech rhythm is to be expected in any language, including English, and may be motivated by a speaker's identity-construction or by factors depending on the communicative situation. It has for example been suggested that British English news readers tend towards a very stress-timed pronunciation, but that the speech of caretakers directed towards children tends towards more syllable-timing. To date, such variation has not been investigated in any detail but it is possible that speech rhythm is a feature just like t-glottaling that speakers adapt with regard to the communicative situation.

4. Summary

This chapter started by introducing the distinction between syllable-timed languages, which were thought to have isochronic syllables, and stress-timed languages, which were said to have isochronic feet. Although these claims did not hold up to empirical verification, they turned out to be true as a tendency. In stress-timed languages such as (Southern British) English, feet tend to differ less from each other in duration than syllables. In contrast, in syllable-timed languages such as Spanish, syllables differ less from each other in duration than feet.

The comparison of durations of syllables and feet was formalised as the average speech rate normalised difference between syllables/feet. This index is called the Normalised Pairwise Variability Index. However, defining the beginning and end of syllables and feet is not a trivial task, and rules of syllabification differ between languages.

A more reliable way of measuring rhythm is based on the durations of vocalic intervals, stretches of vowels in an utterance that are not interrupted by any consonants but enclosed by consonants (or pauses) on each side. The normalised Pairwise Variability Index for Vocalic intervals (nPVI-V) is the most widely used measure of speech rhythm. Stress-timed languages such as English show higher values because there is more variation in the duration of vocalic intervals in these languages than in syllable-timed languages.

Comparisons of the speech rhythm of varieties of English have also been conducted. Southern British English is more stress-timed than some other British dialects of English and many nativised varieties of English that have established themselves in former British colonies. Singapore English, Nigerian English and Indian English are some examples.

Three factors have been identified that lead to strong alternations in vocalic interval durations in Southern British English. Vowels in unstressed syllables are reduced, which means they are shorter than vowels in stressed syllables. Vowels before voiced plosives or fricatives are longer than before voiceless ones. Finally, syllables, including their vowels, that bear sentence focus tend to be longer than syllables that do not bear focus.

Characterising Southern British English as more stress-timed than other varieties of English or than other languages does not mean that its rhythm is uniform. There is considerable variation between individual sentences and speakers. Variation occurs with many features in languages, so it is not surprising that it should also occur with regard to speech rhythm. The crucial point in such comparisons is that Southern British English is, on average, more stress-timed than other languages and other varieties of English.

REFERENCES

Arvaniti, Amalia (2009): “Rhythm, timing and the timing of rhythm”. *Phonetica* 66:46-63

Chen, Matthew (1970): “Vowel length variation as a function of voicing of the consonant environment”. *Phonetica* 22:129-59.

Davis, Matthew Harold (2000). *Lexical Segmentation in Spoken Word Recognition*. Unpublished PhD dissertation, University of London. Available online at http://citeseerx.ist.psu.edu/viewdoc/ download?doi =10.1.1.6.4449&rep=rep1&type=pdf.

Duanmu, San (2009): *Syllable Structure*. Oxford: Oxford University Press.

Fallows, Deborah (1981): “Experimental evidence for English syllabification and syllable structure”. *Journal of Linguistics* 17:309-317.

Fuchs, Robert (2013). *Speech Rhythm in Educated Indian English and British English.* Unpublished PhD dissertation, Westfälische Wilhelms-Universität Münster.

Gibbon, Daffydd & Ulrike Gut (2005): “Measuring speech rhythm in varieties of Englishw. *Proceedings of Eurospeech* 2001, Aalborg, 91-94.

Gick, Brian (2003): “Articulatory correlates of ambisyllabicity in English glides liquids”. In: J. Local, R. Ogden & R. Temple (eds.), *Phonetic Interpretation - Papers in Laboratory Phonology VI*, Cambridge: Cambridge University Press, 222-236.

Grabe, Esther & Ee Ling Low (2002): “Durational variability in speech and the Rhythm Class Hypothesis”. In: C. Gussenhoven & N. Warner (eds.), Laboratory Phonology 7, Berlin/New York: Mouton de Gruyter, 515-546.

Klatt, Dennis H. (1976): “Linguistic uses of segmental duration in English: Acoustic and perceptual evidence”. *Journal of the Acoustical Society of America* 59(5): 1208-1221.

Low, Ee Ling (2000): *Prosodic Prominence in Singapore English*. Unpublished PhD dissertation, University of Cambridge.

Morrison, Geoffrey Stewart (2002): “Spanish listeners' use of vowel spectral properties as cues to post-vocalic consonant voicing in English”. CD-ROM paper collection of the *First Pan-American/Iberian Meeting on Acoustics*, 2-6 December 2002, Cancún, Mexico.

Nolan, Francis, Kirsty McDougall, Gea de Jong & Toby Hudson (2006): “A Forensic Phonetic Study of 'Dynamic' Sources of Variability in Speech: The DyViS Project”. In: P. Warren & C.I. Watson (eds.), *Proceedings of the 11th Australasian International Conference on Speech Science and Technology*, 6-8 December 2006, Auckland: Australasian Speech Science and Technology Association, 13-18.

Nolan, Francis and Eva Liina Asu (2009): “The pairwise variability index and coexisting rhythms in language”. *Phonetica* 66: 64-77.

Pulgram, Ernst (1970): *Syllable, Word, Nexus, Cursus*. The Hague: Mouton.

Wang, Hsu Samuel and Jui-Chun Wu (2001): “The effect of vowel duration on the perception of postvocalic voiced/voiceless consonants”. *Concentric: Studies in English Literature and Linguistics*, 27(2): 35-52.

Wiget, Lukas, Laurence White, Barbara Schuppler, Izabelle Grenon, Olesya Rauch and Sven L. Mattys (2010): “How stable are acoustic metrics of contrastive speech rhythm?” *Journal of the Acoustical Society of America* 127(3): 1559-69.

Olesya Rauch and Sven L. Mattys (2010): ”How stable are acoustic metrics of contrastive speech rhythm?” *Journal of the Acoustical Society of America,* 127(3):1559-69.

CHAPTER 8. STUDIES ON THE INTONATION OF ENGLISH: A CRITICAL REVIEW

Luisa Granato
University of La Plata (Argentina)

1. Introduction

Various positions have been adopted by linguists as regards the role of sound in language, depending upon the linguistic theories to which they ascribe, but beyond the fundamental differences existing between the numerous approaches described in the literature, certain basic notions and assumptions are shared. In this respect, Halliday and Greaves, for example, point to the basic concept on which most analysts agree, namely that the sound of a language is a resource to construct meanings, so that sounds are regarded as integrating the complexity of language: phonology, they hold, "is a stratum or level in the overall architecture of language" (2008: 10-11).
The interest of phoneticians was centered for a long time on the study of the segmental units of the sound systems of languages, the result being the design of detailed descriptions of the phonemic and allophonic features of these units and of their phonotactic possibilities and phonetic variations in connected speech. Later, and to some extent following theoretical interests, and partly motivated by the need to teach pronunciation in the English as a Foreign Language classroom, intonational phonology became the object of study of academics that produced descriptions and models related to the melodic contours used by the speakers, and the meanings expressed with them. However, the status of this level as well as its independence from other sub-systems of the language is still a controversial matter. The purpose of this paper is to present a brief account of the development of the

studies of the prosodic characteristics of English in Europe, and particularly in England but, since a comprehensive review would require the length of a whole book, we will try to assess the relevance of only a small, but representative selection of some influential works and investigations produced in the twentieth, and beginning of the twenty-first centuries. Studies in which there are close connections between the attitudes of the speakers and the grammatical organization of utterances will be first considered and then, the focus will be on the intonation of discourse (Brazil 1997), the notions developed in this perspective, the differences with other viewpoints and the evaluation of some criticisms made to the approach. Our own conclusions and viewpoints about the strengths and weaknesses of this model will also be presented, with the aim of envisaging possible future improvements and fostering research on the independent contribution of intonation patterns to the meaning of language in use.

2. Fundamental concepts of intonation

2.1 Prosodic features

As non-tonal languages, English as well as Spanish are analyzed at a post-lexical level, the focus being mostly on the study of pitch, considered basic in the definitions of *stress*, *tone*, *accent*, and *pitch level*, notions that are common to most studies, although they may receive different names and treatment.

The perception of *stress* is related to the features of fundamental frequency and duration, and it is associated with the *foot*, defined as a sequence of two syllables, one weak (unstressed) and the other strong (stressed) (Gussen-hoven 2004). Feet with a strong-weak structure are called *trochees* and those with a weak-strong structure are called *iambs*, and each foot contains only one stressed syllable in whose realization vowel quality, duration and quantity play an important role. Although one could possibly differentiate between countless degrees of stress, so that the indication that a syllable is stressed may mean any of a number of values, most analysts of English recognize two that correspond to weak and strong syllables.

Tone is also realized by pitch: "Languages use pitch variations contrastively for the expression of discoursal meaning and for marking phrases" (Gussenhoven 2004: 22). This author states that an intonation contour is doubly structural: on the one hand, it exhibits a morphological structure that makes the contour meaningful, and on the other hand, a phonological structure that gives its tones. Tones that occur on accented syllables or near them are considered *pitch accents*; or they occur at the beginning or at the end of intonational phrases, in which case they are *boundary tones* (Pierrehumbert 1980 ; Pierrehumbert and Beckman 1988, in Gussenhoven op. cit: 22).

Pitch hight is in some models analyzed together with pitch direction, so that distinctions are made, for example, between high, mid and low falls, as in O'Connor and Arnold (1968) and Wells (2006). In other models, the contribution of meaning of different pitch heights is analysed separately, as in Brazil's study of discourse intonation (1997).

3. The study of sound from diverse disciplines

Halliday and Greaves (2008) present an adequate summary of the fields from which speech sounds can be thought of: in *physical terms*, in *biological terms*, in *engineering terms*, as *expression of meaning* and as *construction of meaning*. When we look at sound in physical terms, we place ourselves in the *acoustic* field since sounds are physical phenomena that can be studied form the point of view of their acoustic characteristics. In this discipline, *sound waves* can be analyzed in instrumental ways which makes it possible to obtain information about *amplitude* –the changes in air pressure– and *frequency* -the rate at which the air moves. This indicates the *fundamental frequency*, which is not formed by harmonic frequencies: it triggers them. The glottal tone is made up of the fundamental frequency and the harmonics. Finally, physicists describe the *phase* as the mapping of waves into one another that help determine the wave form. The parameters of *duration* and *timbre* are also considered in this view, and they refer to the time during which a sound wave lasts and to the quality of the sound wave respectively.

Biologists, on the other hand, centre their attention on the *source* and the *target* as the places of production and reception of sounds, that is to say, on the neurophysiologic activities of the brain processes involved in speech. Here, the brain, the so-called organs of speech and the ear play a crucial role at both the initiation and the reception stages closely connected to the physical properties of sounds in quite a complex way.

Engineers deal with the *recording*, *transmission* and *reproduction* of sounds by means of speech synthesis and speech recognition, and as far as speech science is concerned, they have contributed greatly to the understanding of these processes.

For a large number of specialists like philosophers, social scientists, dialectologists, psychologists, and clinicians, among others, sound is an instrument that conveys meaning and is used with diverse aims, but linguists, particularly those specialized in phonetics, think of sound as the "object under attention", not as an instrument (Halliday and Greaves op. cit: 10). However, for these authors, grammarians also deal with the phonological characteristics for the sake of grammatical descriptions because sound is the medium of expression in spoken language and it can also signal grammatical meanings.

Analysts of the prosodic characteristics of a language have been interested in describing the roles that intonation plays within a language and the attitudes and meanings it can convey. In all cases, they have expressed a concern with the perception of intonation contours by the human ear and with the acoustic characteristics of the sound waves produced. Although it may seem the best approximation to the topic given that pitch is a perceptual process (Gussenhoven 2004), the most convenient way is to obtain auditory representations of intonation making use of adequate computer programs -*pitch trackers*- that can show visual representations of the acoustic reality. Vocal fold vibration, measured in fundamental frequency values, is usually expressed in hertz (Hz), but it is important to remember that some changes in F0 may not be perceived by the human ear. It is also worth noting that the systematic use of instrumental measurements gained ground after the introduction of computer software that made it possible to carry out acoustic research, not only in specialized laboratories but also through the use of personal computers.

4. A shift in the study of suprasegmental features

Two names deserve special mention in a review of this kind: Nikolai Trubetzkoy and Roman Jakobson, both interested in the sound structure of language. Their insights into the notion of phonology are regarded as highly valuable and influential. Anderson (1985) presents an excellent summary of the work and contributions of these two linguists, as well as of other members of the same group from which I will try to extract the main concepts. The author holds that up to the beginning of their work in the field at the start of the nineteenth century, studies on the phonetic aspect of languages had centered their attention on details related to the articulatory and, to a lesser extent, the acoustic characteristics of sounds, rather than on providing a linguistic description of sound systems. A group of linguists soon joined Jakobson and Trubetzkoy in advocating for a new direction in linguistic research, and their first public presentations gained a number of followers who believed in the need for a fresh look into the study of all aspects of language which would throw light on the basic goals of phonology. Among the main contributions by this group was, on the one hand, the adoption of a phonological view, and on the other hand, closely related to this one, the consideration of a system of distinctive features. The cooperation of these academics found an institutional space when the Linguistic Circle of Prague was founded in 1926, and whose main activity was “the development of a structural perspective on language, and particularly on phonology” (Anderson op. cit: 88).

These studies, holds Anderson, produce a true revolutionary change in this area of research through the identification of phonological systems, the study of those features that differentiate pairs of elements or phonemes and ignore non significant phonetic variations, and the initiation of acoustic analyses which is considered a great advance as compared to the traditional articulatory descriptions. The spread and acceptance of the Prague Circle newly developed ideas on language, and phonology in particular, exerted an increasing influence in the direction of the development of linguistic studies and these ideas made analysts realize that phonological studies should be understood as “the science which deals with the functional role of sounds within a linguistic system.” (Anderson op.cit: 91)

Phonemic differentiation is, in this view, attained through a process of decomposing sound material into smaller units on the basis of functional contrasts (Trubetzkoy in Anderson op. cit), conception that is based on certain social norms or conventions existing within a system of communication. The phoneme will be later defined by Twadell as a “mental or psychological reality” (Twadell (1939) in Anderson 1985: 94) who drew upon Trubetzkoy’s earlier work. Another crucial contribution was the identification of a system of structured oppositions, that is, one in which each component element relates specifically to the other elements which it opposes. This system of structures is regarded as being functional to the whole and encompassing all the elements in opposition with each other. The School of Prague went beyond the description of the elements empirically observed in a language, because their purpose was to describe all the contrasts found in natural languages; their main goal was to produce a phonological explanation and to establish general laws that were followed by empirical observations.

To differentiate phonemes diverse dimensions are taken into account. Some pairs are distinguished by a combination not observed in the distinction of any other pair of phonemes, like /d/ and /m/ in English while others are recurrent, as is the case with voicing in English /p/ and /b/, which also appears in the other two plosive pairs in the language, /t/ and /d/ and /k/ and /g/. Trubetzkoy puts the emphasis on the status of the phoneme but he himself, together with other members of the Prague Circle, were the first to look at the prosodic features of language, and to claim that the characteristics of tone affecting the whole syllable should not be considered as properties of the segment or phoneme, but should be regarded as realized on a part of the syllable, namely the nucleus. This process led to the separation of syllabic and non syllabic forms. He engages in a detailed analysis of the behaviour of long and short vowels and diphthongs and describes the sub-units called “moras”, an example of which is the long vowels formed by two moras. This decomposition of sounds is made evident in that long syllables can bear either falling or rising accent, while short syllables can only be accented or unaccented. A difference is also noticed in the phonological interpretation of sounds with diverse quantity values, qualities that were extended to the capacity of some consonants to interrupt the preceding vowel articulation, and the treatment of these consonants as belonging or not to the nucleus of the syllable.

When it comes to the analysis of the pitch of a segment, Trubetzkoy attributes to it a relative value, and here again his view is phonological rather than phonetic, for he looks at the contrasts found in a system ignoring phonetic variations, which means that there is not a single realization of a *high tone* for it might occur at different points within the pitch range of the voice. He then distinguishes *pitch register* (relative pitch) and *tone movement* contrasts (falling and rising), and holds that this only applies to long syllables in which the first mora is higher than the second and recognizes three tone registers: a *normal* register, and two more values, one above and the other below it. In his view, this applies to all languages and variants of these only represent phonetic differences. The work of Hjelmslev, especially carried out between the years 1935 and 1970, Anderson points out, attributes importance to the segments that form a sentence and which together constitute a text. He holds that some characteristics are realized over the length of these units, and refers to *intonations* realized over whole utterances, for example, intonation patterns which characterize questions and *accents* as stress, pitch accent over a syllable, among others; in this way, phonological properties are associated with units at different levels of a hierarchical organization.

The invaluable contribution made by the above mentioned scholars can be regarded as the starting point towards the maturity of phonological studies in several European languages.

5. Studies of intonation in the United Kingdom

For reasons of space, we have made a selection of only some influential models that were established in the last decades with the hope that this selection will show the way in which this area of analysis has developed in the United Kingdom.

As a member of the staff at the University of London School of Oriental and African Studies (SOAS) and the Department of Phonetics, John R. Firth developed his career between the third and fifth decades of the nineteenth century. He produced novel views on theoretical and practical notions of semantics and phonology adopting a polysystemic approach to language. He considered that meaning was essential in linguistic analysis, for it could be associated with the functions of utterances in context (Anderson 1995). Regarding his *theory of prosodic analysis*, Anderson argues that it plays a central role in his analytical views and he observes:

> It represents the first substantial challenge within twentieth-century linguistics to the notion that division of the utterance into phonetic segments provides the essential basis for further analysis, and that the analysis can proceed exclusively as a matter of assigning particular properties of the phonetic material to particular segments (Anderson 1985: 185)

Firth also marked a clear difference between *phonological structure* formed by the syntagmatic relation between the different parts of a sentence like syllables, words, intonational phrases, and *phonological system*, repres-ented by the paradigmatic choices available at any given point, which allows for a distinction between *syllabic structure, prosody*, occurring over different segments of the structure and the *phonematic units* or sounds organized in systems (Anderson, op. cit.: 186-7) This perspective, based on phonetic material, incorporated grammatical analysis into the picture so that the two areas became interdependent. Firth's approach exerted a strong influence on his followers' prosodic analysis, whichever their lines of research.

5.1 The School of London

From the 60's to the present time, several descriptions of the prosodic characteristics of the English language originated in different academic circles in Europe and North America, and among these, the School of London, at the University College of London University, has always played a central role in the development of the field. Heads of the Phonetic Department like Daniel Jones and Alfred Gimson, among many others, ensured the high standard of the research done from diverse perspectives of analysis as well as of the teaching activities carried out in it.

The Department of Phonetics was established in 1912 with Daniel Jones as its head. He started the study of the intonation of English but did not arrive at devising a system (Gussenhoven 2004). While in 1926, Armstrong and Ward introduced innovations and claimed that theirs was a scientific analysis of intonation not based on anyone else's model but on their own observations, Palmer (1992) contributed to the area by establishing the *unit of intonation*, within which he distinguished a *head,* a *nucleus* and a *tail*. Kingdon, (1958a) adds Gussenhoven, further elaborated on Palmer's findings and subdivided the *head* into *pre-head* and *body*, associated intonation with grammar and referred to the attitudinal meanings of the different melodic patterns. He was the first to advocate a link between stress and intonation and the analysis of the functions of intonation in single contexts thus going beyond the association of intonation with grammatical form. The findings of these studies were published in books addressed to the areas of EFL and ESL, widely used all over the world, and especially designed to describe the characteristics of phonological patterns. They also contributed to the improvement of the foreign speaker oral production. Moreover these materials became obligatory reading in Teacher Education courses in England as well as abroad.

Crystal ([1969] 1976) too refers to the work of other academics who contributed to the maturity of the field, like Schubiger (1958), Lee (1960), Pike (1945), Abercrombie (1956), Gimson (1956), among others, and he adopts a critical stance and notices in all these works the absence of a well-defined theory of intonation, the lack of depth in the analysis of the features considered in the different descriptions and the lack of systematicity in the description of the semantics of intonation.

One of the best known publications among followers of the English tradition was O'Connor, J. D and G. Arnold's *Intonation of Colloquial English*, first published in 1961 and followed by several impressions and new editions. This book was the first choice in a large number of teacher education courses for many years. It aimed, on the one hand, to help students acquire a metaphonological knowledge of the formal properties of the intonation patterns used in English, and to show how these patterns affected or contributed to the meanings expressed by the lexico-grammatical expressions of the language, and on the other hand, it offered foreign students of English and future non-native teachers material to improve their own use of suprasegmentals through the inclusion of a session of Intonation Drills (1968: 72; 79). The authors present a detailed account of the phonetic shape of the different intonation patterns or *tunes* observed in native speakers' speech, and then analyze the meanings provided by them in association with utterances of different sentence types: *statement, question, command* and *interjection.* Ten *tone units* are distinguished in their study, each of them grouping different variations or *tunes*, resulting from the variations in pitch movement and the occurrence of stressed and unstressed syllables that may, or may not precede the *nuclear tone* in the *Head* and the *Pre-Head,* or after it, in the *Tail*. The stressed syllables could be *High, Mid* or *Low* in pitch, and might or might not be preceded or followed by unstressed syllables. Each of these variants is said to project different *attitudes* on the part of the speaker when used on the grammatical categories of statements, special questions, general questions, commands or interjections. As a way of exemplification I will reproduce the description of the possible phonetic shapes and meanings expressed by *Tone Group 3*.

The attitude shown by this Tone Group is described in the different sentences as follows:

> In *statements*, conveying personal concern or involvement, lively, interested, more airy and lighter in mood than when said with Tone Groups 1 and 2; (with Low Head) querulous or disgruntled protest.
>
> In *special questions*, lively, interested; (with Low Head) somewhat

unpleasantly surprised.

In *general questions*, mildly surprised acceptance of the listener's premises , willing to discuss though sometimes impatient that such discussion should be necessary.

In *commands*, warm, often with a note of critical surprise.

In *interjections*, more emotional but less portentous and less weighty than when said with Tone Group 2; (with Low Head) affronted surprise" (O'Connor and Arnold 1968: 109).

This *tone group* can be produced with 3 different *Tunes* having three different obligatory and optional components as is shown below:

Tune 1, High Fall (+Tail);
Tune II, Low Pre-Head + High Fall (+Tail) and
Tune III, (Low Pre-Head +) Low Head + High Fall (+Tail)

The book also offers brief, made up segments of interactive speech for exemplification and practice, which appear slightly contextualized as can be seen in the following fragments that only offer the immediately preceding co-text without making reference to other contextual features like the identity of the participants, the common knowledge between them, the time and place of utterance, the speech event within which they were produced:

Tone Group III. Tune II. Low pre-head + high fall (+tail). In Statments

Verbal context	*Drill*
|| He| promised to a↘pologise ||	|| That's the |least he can ↘do||
	(O'Connor and Arnold 1968: 116)

Tone Group VII. Tune I. (Low pre-head+) Stepping head + low rise (+ tail). In Special Questions.

Verbal context	*Drill*
|| I |don't think I'll ↘go ||	|||↗Why not? ||
	(O'Connor and Arnold 1968: 197)

It is clear that the utterances preceding the fragments for practice are far from offering a clear picture of the contextual characteristics in which the second utterance occurs. However, in their teaching activities at the University College Department of Phonetics addressed to foreign students, the authors used to read out plays and analyze the possible intonation contours for each utterance, this time incorporating features of the context which included more details than those given in the fragments for practice quoted above and offered practical activities which allowed for a much better understanding of the circumstances of the interactive events than was the case with the drills.

Although the effect attributed to the intonational patterns fitted closely the local meaning of the units in the interactions examined, generalizations were difficult to make, since often the same phonological shape might contribute to the expression of opposing attitudes according to the grammatical organization of the utterance which they accompanied.

This brief outline shows the complexity of the description and is also a clue to the long list of speakers' attitudes that are included. These add up to over a hundred if we consider the ten tone groups and their variants together.

Another outstanding and contemporary researcher, John Wells, publishes his book *English Intonation. An Introduction* ([2006] 2007) and points out that that he wrote it form a descriptive-linguistic and teaching perspective and that it is addressed to native and non-native speakers and to university students. He describes three intonation systems: *tonality, tonicity* or *nucleus placement* and *tone* (terminology already used by Halliday in his 1967 and 1970a publications) terms which correspond respectively to the chunking of speech into what he calls *intonation phrases* and within which intonation patterns are produced, to the accentuation of those words the speaker wishes to highlight for the sake of the listener's understanding, and to the tone or pitch variation chosen.

Wells attributes three basic functions to intonation: the *attitudinal function*, realized through *tone*, the *grammatical function,* since through intonation we can demarcate units by means of *tonality* and the *focusing function*, because a choice at the level of tonicity will allow speakers to show what is new and what is known to the interactant(s). To these, he adds the *discursive* or *cohesive function* conveyed by suprasegmentals that can help signal the way in which adjacent clauses go together, the *psychological function* because prosodic choices can facilitate the chunking of speech into units which are easy to construct and decode, and the *indexical function* since intonation characteristics can mark personal and social identities. Not only does Wells recognize three more functions to intonation as compared to O'Connor and Arnold's description, but he also characterizes the grammatical function in a different way, that is, not making a reference to the syntactic organization of utterance and its association with certain phonological contours at this stage.

Two basic tones are acknowledged in Well's description and they are the *falling tone*, tending "to indicate that the information conveyed is, or could be, complete" and which can be realized by a *high fall,* a *low fall* and a *rise-fall*, and the *non-falling tone,* used to show that "there is something more to come", that can show the contours of a *high rise,* a *low rise,* a *mid level,* and a *fall rise* (Wells 2006: 10 to 15). The author holds that there is no permanent coincidence between certain tone choices and sentence types in natural speech, but he claims that each type can be uttered using a natural or unmarked tone which he refers to as the *default tone: "*a fall for statements, exclamations, wh- questions and commands, and a rise for yes-no questions" (Wells op. cit: 15). He also differentiates between *independent and dependent* tones, according to whether they appear on a single intonation phrase utterance or on two intonation phrase utterances, in which case the main part is pronounced with a fall and the subordinate part with a non-fall. In this model, tones are also said to convey attitudinal meanings as well as meanings related to the contents of the unit. However, when dealing with the types of falls and rises, the author organizes his presentation under the headings *statements*, *questions* and *other sentence types,* and he closes the topic with a table that lists the general meanings of tones also in relation to the different grammatical structures with which they can be used. (See Wells op. cit: 91) The tones (fall, fall-rise, rise, and rise-fall) include simple and complex heads and pre-heads, and distinctions are made according to the pre-nuclear segments formed by stressed and unstressed preceding syllables; if there is a tail, the contour chosen spreads out over it and in all cases, the meanings conveyed by these variants are discussed. We notice here an improvement in relation to O'Connor and Arnold's description of tone, namely the overt differentiation between the intonation of utterances in isolation and that of utterances in discourse.

When dealing with *tonicity,* Wells refers to the accentuation of a word by means of making one of its syllables prominent and he presents quite a complete description of the topic offering explanations about different types of focus and looking at the classes of words and parts of an utterance which may become nuclear, as he holds that it is content words –as opposed to function words– that are generally accented. However, the allocation of nucleus placement is associated to what is new information to the interlocutor of an utterance, which represents a discoursal view of the phenomenon, already found in Halliday in 1970a.

In what concerns the segmentation of speech into intonations units, the author observes that this is usually done according to the grammatical structure, so that there is a coincidence between grammatical and phonological units. Utterances show the most diverse structuring, and although intonation phrases are said to contain a chunk of information, their length depends on the number of words and syllables the speaker wishes to accent, and it is often the case that these units do not include any kind of information at all. The style of speech may also have an incidence on the segmentation of discourse, an aspect which escapes the lexico-grammatical characteristics of the language and is related to the area of genre studies.

The book offers ample exemplification and drills, clearly addressed to students and is accompanied by a CD-ROM with recordings of the examples and drills in the text.

In short, we observe that this is a complete and detailed description of different intonation patterns used in English for the expression of the speaker's attitudes, and even though the author makes it clear that grammatical structure does not condition intonation choices, the presentation of most topics in relation to the syntactic organization of utterances and word classes does not seem helpful in considering these as independent levels of analysis. In spite of the fact that this work represents a step ahead in some ways such as making use of a different taxonomy to classify tones with functional labels like the *definitive fall* and the *implicational fall-rise*, and extending the discussion of tonicity as compared with previous work in the Department of Phonetics, similarities outnumber differences and from the point of view of the approach, there are only a few substantive changes.

5.2 The work of Peter Roach

Outside the London School, Peter Roach from the University of Leeds publishes *English Phonetics and Phonology. A practical course* ([1983] 1984), a book also meant to be used by native and non-native speakers who study English at university level, or are training to teach English. Roach is interested in a linguistically significant account of intonation and holds that what should be taken into consideration are only those features that can be perceived by the human ear, excluding characteristics that are only detectable in laboratory analysis, and that a linguistic description that establishes contrasts between these features with others should be designed.

He focuses on the form and the function of intonation and analyses the intonation choices within *tone units* –equivalent to phonological units or intonation phrases in other models– whose structure is described in the same way as in O'Connor and Arnold and Wells. They have an obligatory element, the *nucleus,* and the optional elements *head, pre-head*, preceding the nucleus and *tail* following the nucleus. The possible forms that the patterns can exhibit are described in great detail in his work, in which Roach makes a distinction among the *fall*, the *rise,* the *fall-rise*, the *rise-fall* and the *level* tones and describes the impact of *pitch height* at which these tones can be produced. The melodic patterns resulting from all the possible variations are said to fulfill four main functions: the *attitudinal function,* for they make it possible to express attitudes and emotions; the *accentual function,* through which one can exploit the effect of making syllables or words more prominent than others, the *grammatical function* given that intonation provides elements for the listener to recognize the grammatical organization of utterances, and the *discourse function* likely to help differentiate the *new* from the *given* in natural speech*.* The author warns about the danger that

the analyst may attribute hir/her own labels to intonation meanings, and claims the convenience of working with groups whose members would be asked to provide adjectives or adverbs to refer to the attitudes they detect in the recordings they are asked to listen to. Variations in *loudness, speed,* and *voice quality* may also project different attitudes. Pitch variations resulting in *high, mid* and *low key* will also have an incidence on the effect of an utterance, together with other semiotic resources available in interaction such as *gesture, vocal effects* (like laughs and sobs) and *body movements*.
The accentual function of intonation is discussed in relation to the placement of tonic stress in the unit. The explanation of the dislocation of accent to an item other than the last lexical item, responds to discourse meanings in this perspective, and is not limited to questions of contrast. As regards the grammatical function, the unit is said to generally coincide with the grammatical structure, although the author observes the possibility for speakers to make boundaries at other places with different aims.
Roach refers to the tendency to assign greater importance to the study of intonation in relation to discourse, the main areas being the use of intonation features to (a) help listeners focus on the most relevant parts of a message, through the use of subordination, the choice of tonic syllables and the choice of tone, which can mark shared or not-shared knowledge between the participants and (b) to regulate conversational behaviour in relation to the use of intonation to indicate, for example, end of a turn or a desire to continue holding the floor. The introduction of discourse and interactional meanings is a crucial innovation in the study of the functions of the suprasegmental characteristics.

6. The Autosegmental-metrical theory

More recently, Ladd (1996) opens the introduction of his book, *Intonational Phonology,* with a reflection about the present state of phonological theory, stating that: “Research on intonation has long been characterized by a number of unresolved basic issues and fundamental differences in approach” (Ladd op. cit: 1), with which he acknowledges the existence of diverse works in the area, but at the same time the need to construct a standard theory of intonational structure which, in his opinion, had not been developed so far. He aims at demonstrating that intonation, and particularly pitch, has a phonological organization –excluding tone languages in which pitch is part of the phonemic system– and notices that difficulties arise given that pitch is a relative concept largely dependent upon situations and speakers who do not always exploit pitch consciously. The author claims that the fact that language serves the purpose of expressing functions is universal, and considering that pitch can fulfil both linguistic and paralinguistic functions, he holds that intonation is regarded as linguistic, and adopts a phonological view in his study. He starts form an auto-segmental and metrical perspective, which, in his view, offers an innovative approach to the study of suprasegmentals, through the recognition of two types of syllables (stressed and unstressed) and a system of tunes formed by sequences of high and low tones (Chun 2002).

Likewise, Gussenhoven and Jacobs (2011) claim that a linear model of analysis in which segments have a list of features is inadequate because it does not allow for the study of characteristics which affect more than one segment and they advocate the representation of tonal features in separate strings form other features, that is to say, that the cuts for different features need not be made at the same places. This allows the analyst to apply an autosegmental approach and, for example, describe a word as being formed by six segments and said in only one tone, the result here being an analysis in two separate *tiers*.

> Tones are independent morphemes, whose meanings are somewhat vague, having to do with whether you are giving the listener some information or are asking for some, but no attempt to define them will be made here (Gussenhoven and Jacobs 2011: 155)

As can be observed, their description is not centred on the function that tones might fulfil within the phonological units. The focus of their interest is on the phonetic form of tones and the possibilities of occurrence in adjacent syllables, a non-linear model of analysis fully developed by Goldsmith (1976), which represented a breakthrough in the conception of phonology (Gussenhoven and Jacobs, op. cit.: 161)

Despite the valuable work of well known analysts, Ladd considers that a comprehensive account "available to the outsider" has not yet been produced, so he aims at devising a theoretical and empirical basis of the autosegmental-metrical approach to intonational phonology that might be used by specialists and non-specialists in the field that would constitute a universal framework.

Following Ladd, *Intonation* has three defining characteristics:

> Intonation, as I will use the term, refers to the use of *suprasegmental* phonetic features to convey postlexical or *sentence level* pragmatic meaning in a *linguistic structured* way. (Ladd 1996: 6).

By suprasegmentals, the author refers to the fundamental frequency (F0), intensity and duration as it has been traditionally described, and he acknowledges the pertinence of these three features to three different levels of analysis, namely *physical, psychophysical* and *phonetic*. The suprasegmental characteristics are analysed at a post-lexical level; they are linguistically structured and, as such, described in terms of categories and relations, that is, absolute and relative values leaving out the analysis of features like *tempo* and *loudness*.

Tunes are described both formally and functionally, so in his characterization of the two main intonation patterns, he attributes to the *falling tone* the value of a "straightforward answer", and the *rising tone*, that of "uncertainty and some other questioning modality", since it projects doubt and invites feedback (Ladd op. cit.: 9). He does not, however, attribute particular intonation tunes to particular syntactic organizations for he acknowledges, for example, that the two meanings can be found equally in statements and in questions,a position shared with other analysts like O'Connor and Arnold (1978) who at the beginning of the chapter on *Intonation and Meaning* state:

> It has often been pointed out, and rightly, that no tone group is used exclusively with this or that sentence type- question, statement and the like- and also that no sentence type always requires the use of one and only one tone group... Broadly speaking, any sentence type can be linked with any tone group, and in this chapter we shall consider the effect of our ten tone groups in association with each of the four main sentence types... O'Connor and Arnold (1978: 32)

In Ladd, two basic variants –*weak* and *strong*– are described as far as relative prominence is concerned, weak-strong being the normal stress pattern, the third aspect being the intentional segmentation of speech into intonational units, recognised by most researchers in the field.

Ladd points out that in the sixties and the seventies instrumental phonetics received special attention from areas like acoustic phonetics, speech perception and experimental psychology offering cues to intonational patterns; most of these studies pursued the aim of identifying the acoustic cues with syntactic and pragmatic meanings, with the emotional states of the interactants and with stress at the level of words and sentences. The result of these investigations showed only general findings, but lacked convincing answers to questions about important topics in the field. Parallel to this scientific analysis was the practice of impressionistic descriptions to which instrumental phoneticians attributed less rigour. In fact, the impressionistic approach was criticised for only attending to the most noticeable changes of the fundamental frequency, for contrasting invented taxonomies with empirical findings and for basing its results on auditory impressions without experimental evidence (Liberman and Michaels 1962; Ohala 1975 and Maeda 1976, in Ladd 1996:13). Ladd shows his disagreement with this criticism and states that the problem resides in the difference between the theoretical assumptions in the two views.

The author also mentions the work done at the Institute of Perceptual Research (henceforth IPO) in Denmark, where a general theory of intonation of Dutch was designed by t'Hart, Collier and Cohen in 1990, then extended to English among other languages, whose task was to combine "an abstract phonological level of description with a detailed account of the phonetic realization of the phonological elements" (Ladd op. cit: 14). From this perspective, the intonational contours are seen as sequences of pitch movements which connect line segments of which listeners interpret only some of the movements as relevant, and they are those which correspond to the discrete events in the pitch contours. Two distinctions in pitch variation are made through the technique of "analysis-through-resynthesis" (de Pijpern 1983, 5, in Ladd op. Cit) with the aim of working on the perception of sound as different to the original sound: relatively high and relatively low levels, which involve "the reconstruction of an intonation type with different values of its physical properties such as amplitude and fundamental frequency". (Halliday and Greaves 2008: 11).

In Ladd's description of pitch movement, differences are made between *prominence lending* and *non-prominence lending*, and these are distinguished according to whether the movement occurs on lexically stressed syllables, *pitch accents*, or at the end of prosodic constituents, *boundary tones*. The movement that contributes to show distinctions between questions and statements can be of two types: *rising* and *falling,* and it may extend over more than one unstressed syllable, contrary to most analysts' expectations that these distinctions should occur on dominant words or on prominent syllables only.

The description is done in phonological terms that take into consideration only the variations in pitch which help distinguish the above mentioned functions in the language. However, the IPO analysts offered very detailed phonetic characterizations of the acoustic variations and their occurrence on stressed and unstressed syllables as well and demonstrated which values, despite the phonetic differences they might exhibit, were perceptually perceived as the same. This replicates the approach to the analysis of segmental elements –the phonemes– and the possible allophones through which they are realized.

Another important notion developed in this model, and taken up later by most phoneticians is that of *declination* which refers to the general tendency in utterances to decrease the fundamental frequency towards the end.

As opposed to instrumental studies, the IPO adopts an approach in which the elements have two important properties, the first being their *phonetic definition* that relates to the fact that the elements that form the system are perceived, identified and defined according to the phonetic differences existing between them. Researchers have deliberately centred their attention on the melodic aspect only, without looking at the functional values added by the intonation patterns. The second property attributed to the approach is its *linearity* since a contour is considered to be formed by sequentially organized discrete intonational elements. Halliday and Greaves (2008: 11-12) point to the interest of the IPO in comparing the qualities of a constructed sound equivalent to a natural one, with the qualities of the same sound at the perception end. This type of analysis resulted in an endless list of phonetic contours. Six intonational patterns are described for English which differ phonetically and perceptually, and whose characterization is based on a phonetic analysis of the acoustic signals.

Gussenhoven (2004) follows Ladd's idea that intonation is structural as lexical tone and morphological paradigms are, and that intonation has a *morphological* structure that gives meaning to the contour through morphemes, and a *phonological* structure which gives its tones (Gussenhoven op. cit.: 22). Intonational contours express informational meanings in discourse, as well as the speaker's emotions and attitudes.

7. A Systemic Functional Approach

From a Systemic Functional Linguistics perspective (henceforth SFL), Halliday and Greaves (2008) expand on Halliday's previous ideas on the role of intonation (1967 and 1970a) whose phonological categories showed grammatical contrasts in a linguistic analysis of attitudinal meanings, in a study of the emotional aspect of talk. They work on the assumption that language is a single system that can be analysed in different ways, as was stated above.

According to the authors, language is a single system that can be conceived in four ways:

> ... meanings the language system can create, lexicogrammatical patterns which give form to those meanings, phonological patterns which enable the communication of those forms and phonetic substance through which the phonological patterns are uttered and perceived (Halliday and Greaves, 2008: 4).

and their aim is to analyse the phonological system within the complex of these four levels. In this line of thought, linguists' interest resides in language itself and not in the possibility of using it for other aims, and language itself is regarded as a set of strata or levels which form the architecture of language, responding to three interests such as sound, wording and meanings attended by phonetics, lexico-grammar, and semantics respectively. The theory points to an interesting distinction between the ways sounds are regarded by grammarians, lexicographers and semanticists, as opposed to phoneticians: the former look at sounds as instruments, while the latter take them as their object of study.

In this linguistic approach, a relationship is established between intonation and grammar which is not fixed when attempting to establish phonological characteristics, and grammar and phonetics should go together.

In their analysis of the different strata recognised in a language, Halliday and Greaves (op. cit.) hold that the speaker has to choose from a set of options in every stratum, which encapsulates the meaning potential of a language. (For more details, see Halliday and Greaves op. cit., Chapter One).

After a description of the representation of sound in physical terms, the book presents the prosodic patterns involved in the construction of the *foot* and the *tone unit* and matters of rhythm and stress are discussed in detail. Finally, the authors hold that the meanings intonation may project are *interpersonal, textual* and *logical.* (For types of meanings, see Halliday, M. A. K. and C. Matthiessen, 2004. *An Introduction to Functional Grammar.)*

7.1 Tonality and Tonicity

In the model, the systems of *tonality* and *tonicity* organise the flow of discourse.

A *unit of information* belongs to the system of *tonality* and relates to the lexicogrammatical level of language; it is described as a fragment of discourse organised in ‘new’ and ‘given’ material, typically, though not always realised by a tone unit, and whose distribution within the clause is discussed.

The *tone unit* is, on the other hand, a phonological element, the highest in the rank scale consisting of *tone unit or tone group, foot, syllable* and *phoneme.*(op. cit.: 13) and which forms the system of *tonicity* , due to its relation to the location of the tone. A spoken melodic contour with an obligatory element on which pitch variation occurs –the *tonic*– whose contour depends on the choice of tone, and an optional one –the *pre-tonic*–

characterise this unit.

A distinction between the *information unit* and the *tone unit* is made clear in this work by means of (a) the differentiation of their functions: the first contributes to the construction of meaning and the second, facilitates the organization of speech sound and (b) their pertinence to different strata, their boundaries marked according to different parameters. Some units can bear only one tonic element –simple tone units– or two tonic elements – compound tone units.

A combination of *tonicity* and *tonality* is an important resource in the construction of textual meaning.

7.1.1 Tone

Five different tones can be produced in a tone unit (Adapted from op. cit.: 44):

Labels	*Phonetic realization*
Tone 1	fall
Tone 2	rise
Tone 3	level rising
Tone 4	fall rising
Tone 5	rise falling

These, plus the two composite tones, 6 and 7, realised as 'fall + level rising' and 'rise-falling + level rising', represent seven choices altogether which are called *primary tones* to which delicate analysis can add other distinctions – the *secondary tones*– in which pitch height is combined with pitch direction; this allows for a distinction, for example, among *high, mid* and *low fall*, each variant expressing a different meaning.

Tones can realise *logical meanings* manifested in the choice of different tones that can indicate coordination, subordination and related meanings, and *interpersonal meanings*, that establish the relationships between speakers and listeners. The discourse unit *move* presents a system of two features: *commodity* (information and goods and services) and *orientation*. (*giving* and *demanding*), features expressed by a network of elements widely used in everyday communication such as *statements, questions, commands,* and *offers.* The three first are typically realised by declaratives, interrogatives, and imperatives; the last one being less tied to a particular form and all of them, however, being subject to lexicogramatical realisations which may not respond to the most frequently used forms.

The four basic functions mentioned above are attributed a neutral choice of tone: polar interrogatives are pronounced with tone 2; declaratives, non-polar interrogatives and imperatives, with tone 1; use of other tones can be observed, the result being a change of attitude on the part of the speaker towards the interlocutor and the content of the utterance. This is called the system of *key,* different form *tone,* associated with the grammatical categories of declarative, interrogative and imperative mood and not with the semantic categories of statement, question and command. It is also possible to change the meaning of an utterance by changing the neutral tone, exemplified by the authors with the following examples:

I like it, with tone 1 (fall); the meaning is neutral: an answer to a question.
I like it, with tone 2 (rise); the meaning is combative 'why do you assume otherwise?'
I do like it, with tone 1 (fall); the meaning is demanding 'I need to know'.
Do you like it, with tone 2 (rise); the meaning is a neutral question.
(op. cit.: 52)

Summing up, Halliday and Greaves conclude by saying that *tone* is a system of the phonological stratum that occurs within the tone unit and is made up of different choices: speakers first choose between simple and compound tones, secondly, among five simple tones and two compound tones which add up to seven possibilities realised at the lexico-grammatical level. These intonational systems that belong to the phonological level interact with other grammatical systems, both at the logical and the interpersonal levels.

As was said before, tone choice and mood often combine to realise speech functions, and it is through the marked choices of tone that the semantic meaning of utterances may change. These possibilities of adding tone choices to the meanings expressed by mood, enlarge the potential of expressing the interpersonal function of language. Similar analysis is carried out in relation to modality, system which apart from the lexico-grammatical choices can use tone for the expression of obligation and probability.

This condensed summary of Halliday and Greaves's model of intonation suffices to see that prosodic features are associated with the grammatical characteristic of utterances, the attitudes and emotions of the participants as well as other models reviewed in this article. The approach is, however, quite different since as all the other strata of language studied in SFL, intonation is also related to the metafunctions of language described in this theoretical perspective in which interfaces between intonation and lexico-grammar, intonation and semantics and intonation and contents of an utterance are part of the model. Tonicity and tonality, described as separate types of features are related to one another and their respective functions are clearly identified. This is a very sound and consistent model of analysis, a perfect complement to the grammatical descriptions of the functional view of language presented in Halliday and Matthiessen (2004).

The book offers sensitization exercises to help students gain awareness of the characteristics of the melodies used in English and the meanings they express, the rhythm, the emotional response each sample triggers and the particular melodies heard in each of the recorded examples.

8. The intonation of discourse

The innovative studies of the intonation of discourse realized by David Brazil in the 80's and 90's reflect a view that diametrically opposes that held by previous authors as regards the meanings attributed to the melodic patterns used in English, and represent a solid theoretical advance. In this respect, Chun asserts:

> Although much work has been done in the domains of both intonation and discourse, combining the study of intonation with the study of discourse has, until recently, been relatively rare ... (2002: 32).

Brazil (1997, 1992, 1983, and 1981, and Brazil in Coulthard, Brazil and

Johns 1980) does his research work as a member of a large project on the study of discourse at the English Language Research Department of the University of Birmingham. He offers a description of intonation as an independent meaning system and gives a different interpretation to the phonological characteristics already described in previous studies. From a discourse-pragmatic perspective, and also influenced by the SFL principles, Brazil starts from the analysis of the meanings expressed by interactants in a speech event and relates it to the interactional assessment of intonation, but showing no reference to the participants' attitudes or associations with the grammatical characteristics of the utterances. Couper-Kuhlen and Selting (1996b in Chun 2002) point to the need for an interactional analysis of intonation and prosody in general that would be based on a consideration of verbal forms and context, which also indicates a focus on pragmatic, rather than on semantic meaning that emphasizes the contextualizing function of prosodic components (Chun op. cit.). While not denying the contribution of intonation to the realization of local meanings, Brazil does not refer to them in his work, for he aims at devising a linguistic system by focusing on those invariable meanings that speech can carry through the chunking of discourse into *Tone Units* and the choices at the levels of *Prominence, Key, Tone* and *Termination*.

The tone unit is described as the locus where all the intonation choices take place; *minimal* units contain only one prominent syllable, and *extended* units have usually two or eventually more prominences. The *tonic segment* is that which starts at the first prominent syllable and finishes at the tonic syllable while the unstressed syllables occurring before and after this segment are called *proclitic* and *enclitic segments* respectively.

The speaker can make a decision according to his desire to present a concept as a selection between two options: as new to the recipient, or as a given element in the flow of discourse, effect attained through the system of *prominence* which offers a choice between making a word prominent through the accentuation of a stressed syllable in it, or making it non-prominent. Prominence occurs on the tonic syllable or word of a unit, and on accented words before the tonic, and it is the first non-tonic prominence in a unit the one that realizes the system of *key* which marks the pitch level of the whole tone unit. Three communicative values can be expressed through the choice of key related to the establishment of different kinds of relationships with the preceding unit: a *contrastive* effect attained when the key is comparatively higher than the previous one; an *additive* meaning expressed with a mid key, and a relationship of *equivalence* shown with the use of a low key. The level of the last prominence in the unit, which coincides with the realization of tone, creates an expectation of concordance with the following unit and this is called the system of *termination*. Through it, the speaker sets up expectations as regards the key in the following unit with an intention to constraining the next contribution to the discourse: if the termination is high, it expresses the expectation of a high key in the following unit, if the termination is mid, it expresses the expectation of a mid key and if it is low, it projects no expectations as to what is to be done in the coming unit. This shows that the values of *key* and *termination* are not absolute, but relative to the values of previous or following pitch levels. These two systems of key and termination contribute to the cohesion and the coherence of a text.

Table 1. Tones and their phonological realization

Basic tones		*Variants*	
Label	*Phonetic shape*	*Label*	*Phonetic shape*
Proclaming tone	Fall	Proclaming+	Rise-fall
Referring tone	Fall-rise	Referring+	Rise
Level tone	Sustained		

The system of *tone* affects the meaning of a unit in different ways; the labels and the phonological form of the tones described are shown in table 1.

Brazil holds that speakers have the possibility of producing *interactive* and and *non-interactive* speech and that these two forms project different positions of the discourse producer: in interactive situations, the speaker always addresses his discourse to an interlocutor, either present or in mind, and he chooses between *proclaiming* and *referring* tones and their variants. In non-interactive speech, that is to say when there is no consideration of an interlocutor, the choices available are *proclaiming* and *level* tones.

A. *Meanings conveyed by tone choice in non-interactive speech*

(a) Mark a point of possible completion

(b) Signal a point of non possible completion.

Possible completion is shown through a proclaiming tone and non-possible completion through a level tone.

B. Meanings conveyed by tone choice in interactive speech

1. Related to giving information

(a) Present information as not shared with the interlocutor

(b) Present information as shared with the interlocutor.

The first meaning is realised by means of a proclaiming tone and the second by means of a referring tone.

2. Related to eliciting information

(a) Elicit information that the speaker does not know.

(b) Elicit information that the speaker wants to confirm.

The first meaning is given by a proclaiming tone and the second by a referring tone.

3. Related to the social distance between participants

(a) Create a feeling of 'separateness'

(b) Create a feeling of 'togetherness'

Here (a) is realized with a proclaiming tone and (b) with a referring tone.

4. Related to linguistic dominance

(a) Project a linguistically dominant speaker

(b) Project a linguistically non-dominant speaker.

The meaning in (a) is attained by the use of an r+ tone and that of (b) by an r tone.

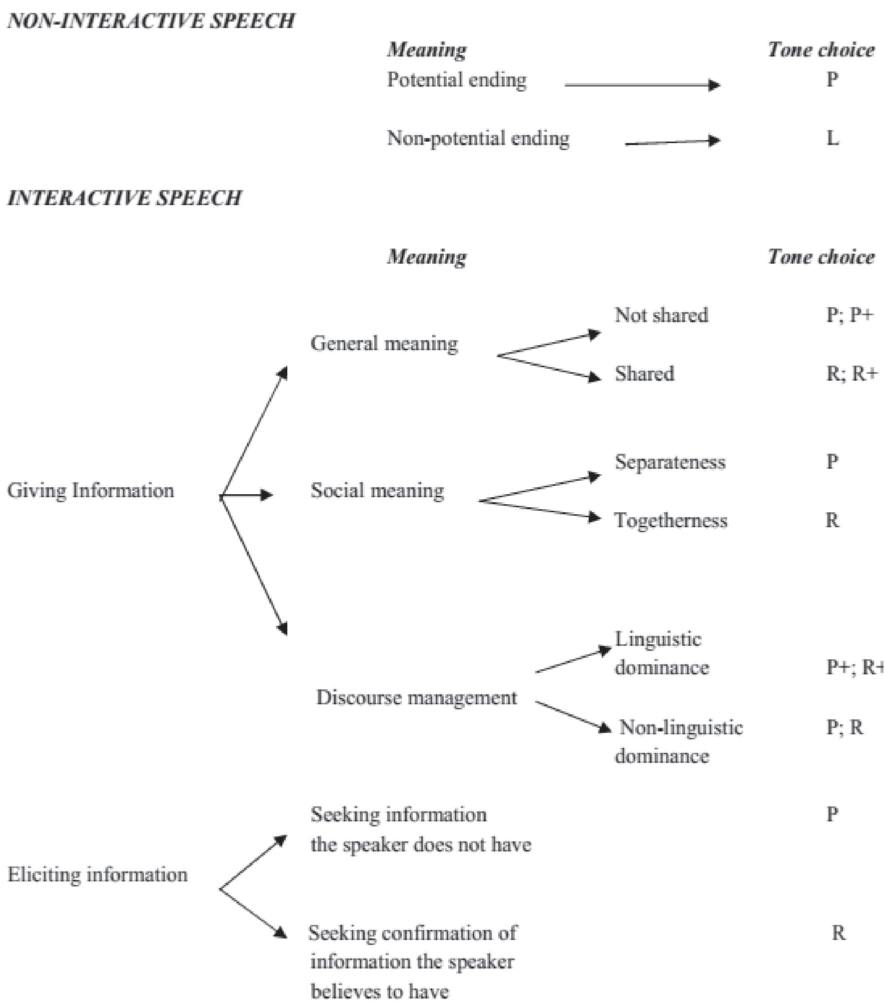
Although this is a very brief reference to the most salient meanings expressed by tone in Brazil's system, I hope it will suffice to show the systematic choices available to a speaker when he intends to produce the meanings referred to. (For a complete explanation of the model see Brazil 1997).

We would like to point out that when the '+ tones' are used, the meaning expressed by the basic tones is kept, so that if, for example, a speaker uses an r+ tone, the effect of linguistic dominance will be added to that of conveying information which is shared, expressed by means of r tones.

Brazil describes the possibility of 'exploiting' the system for the sake of

producing a different impact on the interlocutor, as could be the case of presenting information which is shared, as not-shared, and vice versa.
Non-interactive speech is associated with an orientation that Brazil calls *oblique,* and is used when the speaker orientates to the language used or to the text, as opposed to *direct orientation* in which case the discourse producer addresses his speech to an interlocutor.

Table 2. The communicative values of tone choice.



Examples of oblique orientation are some reading activities and the production of ritualistic language as in prayers, that do not require interpretation for the sake of a listener and consequently, as Brazil holds, the speaker is heard to repeat what the text 'says', while in direct orientation, the speaker says what the text 'means'. The proclaiming tone has different meanings depending upon whether it is used in oblique or direct orientation or to give or elicit information as stated above. The meanings described in this model present paradigmatic options that consequently form linguistic systems. For the sake of clarity, we show tone values in a diagrammatic way. (See Table 2)

While not denying the effects that intonation may have on the manifestation of emotions and attitudes, these are not dealt with in the model due to the impossibility of offering a systematic account of them. They are considered to be the expression of local meanings which can be accompanied by melodic contours that may also be used to express other different attitudes and emotions. Phoneticians who have tried to account for intonation in this way, like O'Connor and Arnold (1968), Wells (2006) and Beaken (2012) among others, have claimed it is impossible to produce a finite list of terms to describe attitudes. No external parameters can help devise a system of attitudes with a fixed number of elements each of which could be defined negatively vis-a-vis the others, a condition that a system must satisfy. As we said before, the models of intonation in which choices are associated with attitudes offer accurate descriptions of the meanings of each occurrence of a melodic pattern but they cannot form a closed system of oppositions: the interpretation of the meaning expressed will depend entirely on the context of the utterance.

Some dissenting voices have been raised against aspects of Brazil's theory while others subscribe to its principles. There is general agreement among various analysts in admitting that intonation patterns cannot be associated with specific attitudes or grammatical patterns. On the other hand, Halliday (1967) states that intonation can construct meanings when there are no other linguistic resources to produce them, as is the case of an utterance without subject-verb inversion likely to be categorized as a declarative, but which a rising intonation can turn into an interrogative. Mike Beaken from

Scheffield Hallam University in England, is one of the authors that has made critical comments on Brazil's work, in an article published on the web (2011). He also demonstrates the effect of different intonation contours on sentence structures through the following examples:

a. Mary´s reading in the garden. (falling intonation)
b. 1. Is Mary reading in the garden? (S-V inversion)
b.2. Mary´s reading in the garden. (rising intonation)

Example 'a' without S-V inversion and a falling intonation is a statement, 'b.1' with S-V inversion plus rising intonation is a question and 'b.2' in declarative mood with rising intonation is a question.

In the treatment of sentences considered in isolation, the grammatical function of intonation in b.2 turns a statement into a question, as a consequence of the change of tone, but this is not always the case in the production of utterances in natural discourse where falling and rising tones may express other meanings as well.

Bolinger (1982) and Cruttenden (1986) postulate the need for a more abstract kind of meaning than the ones obtained through the analysis of attitudes (Beaken 2012) and we believe this is in fact attained in the Discourse Intonation Model developed by Brazil.

Beaken also holds that Brazil's account of tones is not complete; he finds that the use of the modified basic tones –*falls and rises*– indicate the participant´s point of view on the topic discussed, claiming that the third meaning incorporated in relation to these two tones is that of showing the degree of engagement both with the interlocutor/s and the information presented in a tone unit. In my opinion, this distinction is made by Brazil through the possibilities of using a referring tone to indicate togetherness in terms of sharing a point of view –"talking on behalf of us"– and a proclaiming tone to the effect of creating separateness in the same sense –"talking on behalf of I".

In 2009, Beaken produced a model of the intonation of English that contained four nuclear tones, described on the basis of a combination of features (pitch level, pitch movement and accent or prominence): *Low-rise,* *Low-fall, High-fall* and *Fall-rise*, which proved itself to be sufficient in the teaching of reading of some types of genres (Beaken op. cit.). However, he did not offer a complete account of English intonation applicable to a higher number of registers.

Ladd (op. cit) also increases the number of tones to eight, adding to the four mentioned above, the *Level tone*, the *Rise-fall,* the *High-rise* and the *Rise-fall-rise* which Beaken adopts and continues to work on in an attempt to describe the discourse value of these eight tones together with tones produced on the pre nuclear stressed syllables. This shows a difference in the notion of tone as compared to that of Brazil's and of many other analysts.

Another problem that Beaken finds in Brazil's model is the way in which tone values are analyzed since he considers that the model lacks a detailed identification of the meaning of the elements that realize complex nuclear tones which he regards as formed by more than one pitch movement –as would be the case of the Falling-rising tone; in his opinion the analysis should be carried out separately. Each of these movements is said to fulfill different discourse functions, the second movement of the complex tone adds something to what was expressed by the first movement, indicating a change of stance with respect to the listener. In this view, if a Rise is followed by a Fall, the meaning is different from that of a Fall plus another Fall, because the second movement in either case would mark a difference in viewpoints between the interactants. In the rise-fall-rise produced over the tonic syllable, and over the syllables following it, the initial rise increases the engagement with the interlocutor, the fall, shows a difference in the participant´s views on the topic, and the final Rise indicates shared familiarity with the topic between the interactants. In Brazil, all those meanings seem to conflate in the choice of one of the available options: a rise, for example, will carry the meaning of shared information, and at the same time, togetherness with his interlocutor.

Brazil is credited with the idea that the onset in a tone unit is always uttered with a level tone (Beaken op. cit.), but this is not stated in his writings and cannot be inferred from any aspect of his theory. As said before, he clearly describes the three values of key as a resource to establish different kinds of relationships with the preceding unit in an interactive orientation. The above observation seems to be the result of a misinterpretation of the characteristics attributed to the onset syllables which, in the description, are not considered to be the syllables on which the most noticeable pitch movement is realized. The alternatives that Brazil does not consider in his model are those which, in his view, are not communicatively significant. As stated above, by definition a tone unit is said to contain only one tone realization which may or may not be preceded by an onset syllable, prominent but not tonic. Beaken considers that the meaningful distinction between different types of heads or onset patterns described by O´Connor and Arnold: Stepping, Falling, Sliding, Rising, Climbing and Low Head are missing in Brazil's work, but Brazil holds that, apart from the fact that their meanings cannot be systematized, they are not communicatively relevant, that is to say, they do not add anything or alter the basic meanings contributed by the Key, Tone and Termination choices. Besides, this onset syllable is, by definition, prominent but not tonic, so that it cannot be said to be always level; it is neither level nor anything else. Brazil refers to the use of level tones in his description of oblique orientation which is manifested through the use of level and falling or p tones, and because this non interactive speech is not addressed to an interlocutor, choices of pitch level are not analyzed.

Beaken, on the other hand, interprets that the falling head emphasizes the meaning of a Falling nuclear tone in relation to the signaling of diversity of opinion and he finds it is convenient to anticipate this meaning in the Head rather than postpone it to the end of the sentences. He presents this as a question of *timing*, and offers the following example:

a. || \ ↗Fortunately | we´d brought some ↘money ||

b. || We´d brought some ↘money | \↗fortunately ||

In (a) the change of viewpoint is here announced by the adverb ´fortunately´, said on a falling-rising tone that indicates contradiction with what was expected, then followed by an explanation of what was done to avoid the problem.

In (b) the order of the constituents is reversed and Beaken notices a difference in the effect produced which is, in this case, that of contributing to the listener´s understanding of the previous statement, all this treated again as a question of timing.
The interpretation of the real meaning of ´fortunately´ will, in our opinion, depend upon the context in which it occurs since it might not indicate contradiction to what was expected. Besides, it is not clear when a nuclear tone is to be considered as a fall- rise or as two nuclear tones, a Fall followed by a Rise, as in c).

c) Money fortunately (with a Falling rising tone on ´money' or a Falling tone on ´money´ and Rising tone on ´fortunately´)

These occurrences would carry differentiated meanings: the Fall- rise on ´fortunately´ is said to carry over tones of coincidence between participants. However, in the example given, it is said to carry the meaning of ´contrary to expectations´. The falling tone also indicates diversity, which would mean that different phonological forms stand for the same meaning which is a notion not accepted in Brazil´s systematic description of the meaning of intonation in which, regardless of the position in the utterance, an intonational contour always adds the same communicative value.
In example d):

d) || You could have waited ||

´could´ is, according to Beaken, pronounced with a Falling tone thus indicating lack of coincidence. In another example, typical of news reading, Beaken interprets that there is the need to mark as new to the listener or to the discourse more than one piece of information.

Example e):

e) ||The Meteoro logical Office has today issued se vere weather warnings ||

The only difference between d) and e) is the point about which there is lack of coincidence (´what was expected´ and ´the listener´s impatience´).
We do not disregard the possibilities of the right interpretation of the local meaning of the expressions of the example but at the same time, and as was said before, giving the same interpretation to different phonetic forms would not fit into a system. We wonder what the meaning attributed to ´fortunately´ would be if it was uttered in the middle of the tone unit as could be the case in the following made up example:

||We´d ↘ brought, ↘↗fortunately || a lot of ↘money with us ||

We find no resources in Beaken’s account to provide an answer to this question. The same difficulties arise when one tries to tell a tonic syllable from a prominent or accented one. This model seems to relate to the way in which O´Connor and Arnold treated the types of meanings that tones are assigned, since they describe local rather than general signification applicable to all occurrences of the same tone. For example, in different parts of the article the high rise is said to indicate contradiction, reassurance, throwing an accusation back to the accuser, expressing more feeling, engagement and attitude.
Beaken quotes Cruttenden:

> Any one tune can occur in a wide variety of contexts and hence result in a wide variety of conveyed attitudes. The basic meaning associated with a particular tune must therefore be of a more abstract kind than any of these individual conveyed attitudes (Cruttenden 1984:69, in Beaken 2012: 4)

We notice that Beaken himself evaluates Cruttenden’s conception as intellectually satisfactory but at the same time, finds it `of little use´ to teachers who deal with the meaning of intonation.
Another drawback of Brazil´s model, according to this author, is the limitation of common ground to the knowledge shared by speaker and interlocutor since he believes the expression of viewpoints should also be included in the picture. In my view, Brazil´s idea is not restricted to knowledge alone but he goes beyond it in his explanation of the social meaning of the basic tones when he claims that p tones are used to mean ´I am talking on behalf of ´I´,

and r tones to mean ´I am talking of behalf of ´us´´ (Brazil 1995). This can clearly fill up the gap that Beaken marks as missing in the model. In short, what in Brazil is explained through the analysis of separate intonation features (pitch level and tone direction) is in Beaken looked at as a combination of features. For example, a high fall would, in Beaken`s terms, be the result of the addition of two falls, the first signaling new to the hearer, and the second a divergence of viewpoint. In Brazil, the choice of a fall or Proclaiming tone is interpreted as signaling information which is not shared by the participants, whether it is high, mid or low, while the high pitch would be interpreted as the expression of a contrast, where contrast is attributed a very broad meaning of 'opposite to what is expected', which I think includes the idea of divergent views.

There seems to be general agreement on the interpretation of the meanings of the rise-fall between Brazil´s model and more traditional approaches: in all cases the tone is associated with the emotional or social aspect of the relationship between participants developed in detail in Brazil's (1985) work on the degrees of engagement. Brazil´s description, in my opinion, covers a wider field when he sustains that general and social meanings are present in every utterance whatever the choice of tone, and that engagement is a question of degrees that are signaled by intonation choices in combination with the style of speech being produced. When talking about the high-rise, (pp.10) Beaken (ibid., 11) holds:

> You cannot ask a question without engaging directly with your listener and like the rise-fall, the high-rise suggests, as Bolinger puts it, ´a keen interest in the issue or in other person´s opinion (Bolinger (1982:522).

It is evident that asking a question shows interest in the other´s knowledge but one might equally argue that producing a statement shows interest in satisfying the other´s desire to know. I believe it is a question of who says or asks what to whom, and in what circumstances; in other words, of the context of the interaction.

We find several strengths and some weaknesses in the perspective of analysis of the intonation of discourse. We can first observe that it seems to agree with the whole approach to the study of meaning in discourse when most of the other models reviewed often take into consideration utterances in isolation. Discourse Intonation presents an independent system without being subject to any constraints coming from the other strata of language in which the only relations established concern the functions of utterances, and where grammatical features do not condition choices at the level of intonation. Contrary to traditional approaches which centre on the individual speaker, this view looks at the behaviour and action of the two or more participants in an interactive event, in agreement with the widely accepted notion of dialogism (Linell 1998). This is in agreement with most discourse analysts who advocate the view of interaction as the result of a co-construction carried out by the participants, in which the minimal unit of interaction is the *dyad* (Arundale 1999) and not the individual. And whatever is said and interpreted in interactive discourse, is said and interpreted in consideration of all the other participants to a conversation.

In relation to the meanings attributed to intonation choices, the systems do not seem to show contradictory conclusions but rather, different ways of analyzing the same phonetic reality. What may constitute a problem is the reduced sets of meanings considered in Brazil's systematic model which makes those who would like to teach intonation using this system feel at a loss because they do not find all the desired answers in it and, as a consequence of this, one frequently observes the incorrect application of the model in an effort to account for any type of utterance produced in interactive speech, ignoring its limitations. It should be remembered that the systematic meanings considered in the model, only make it possible:

- ✓ to signal distinctions between new and given in relation to concepts in an utterance by means of a choice ***prominence*** options.
- ✓ to mark the contents of informative utterances as shared or not shared with the interlocutor/s in relation to general meanings by means of choices of ***tone***.
- ✓ to project separateness or togetherness from the social point of view by means of choices of ***tone***.
- ✓ to act as a dominant or a non-dominant speaker by means of choices of ***tone***.
- ✓ to mark elicitation utterances as meaning 'I do not know the answer'

or 'I think I know the answer, please confirm it' by means of choices of ***tone***.

- ✓ to orient towards the text or towards an interlocutor by means of choices of ***tone***.
- ✓ To establish relations with the preceding and following units in a discourse by means of choices of ***key*** and ***termination***.

To go beyond these distinctions is not yet possible and will never be unless the systematic description presented by Brazil is enlarged. It is my belief that if this system is used, one should look carefully at what can be accounted for following the principles and concepts developed in the theory. If one desires to find frequent relationships between prosodic melodies and grammatical organization of utterances or attitudinal meanings that can undoubtedly be granted by intonational resources, it is necessary to move away from a systematic description and look at the local meanings intonation helps to construct and that have been exhaustively described in other approaches.

Finally, the analysis at the level of the tone unit is relevant in a micro analysis of an interaction, though not always useful when one intends to look at the value added by intonation to a whole contribution which may be segmented into three or four tone units, but which expresses a single functional value. It is evident that not all the units are equally important in signaling this value, so in my opinion, it is necessary to make a selection of the fragment that is crucial in communicating what the speaker desires and concentrate on the intonation contour of this unit. Brazil does not consider this possibility, but the application of this procedure in our analysis of interactions has proved to be adequate and useful.

9. Conclusions

Beyond discrepancies in some details and in the terminology used, there seems to be general agreement between different approaches to the study of intonation on the chunking of speech in phonological units within which the intonation patterns are found, on the interpretation of accented or prominent syllables which are usually given the same function of highlighting a syllable or word from the rest. Most accounts regard intonation as multifunctional, but each of them emphasizes a different aspect. The school of London has produced models with a marked focus on attitudinal meanings, although recent descriptions have made substantial improvements making a different characterization of the grammatical function, the discursive or cohesive function, the psychological and the indexical functions (Wells 2006). In an SFL approach, Halliday and Greaves (2008) associate intonation with the lexico-grammatical level and hold that these two levels should not be separated in an analysis of language; for them, intonation contributes to the realization of the three functions of language, namely logical, interpersonal and textual. Brazil, in his pioneering work on the intonation of discourse, looks at the meanings intonation can help construct with no dependence on other levels of analysis, but with a close connection with the context in which the talk takes place.

All of the approaches outlined in this article provide insights into the effects of the suprasegmentals on the spoken mode of language, but at a time when discourse and pragmatics have reached maturity as separate and interrelated fields of research, we believe the analysis of intonation from a discourse pragmatic perspective deserves special attention. The model of discourse intonation has been applied to the study of the intonation of Spanish (Granato 2005) and has given satisfactory results.

More investigation could be done in the field with the aim of complementing the theory. It would be interesting, for example, to focus on the incidence of the meanings expressed by pitch level and tone, since observation of spoken discourse seems to make it evident that, in some cases, pitch level overrides the meaning added by the choice of tone. Efforts could also be made in the direction of finding more systematic meanings related either to general or interpersonal meanings. Addressing these, as well as other aspects of the theory not fully analyzed yet, might be useful for the development of new explanations and systematic descriptions of the meanings that prosodic characteristics can contribute to spoken interaction.

REFERENCES

Anderson, Stephen R (1985): *Phonology in the Twentieth Century. Theories of Rules and Theories of Representations.* Chicago and London: The University of Chicago Press.

Arundale, Robert (1999): "An alternative model and ideology of communication for an alternative to politeness theory". *Pragmatics.* Vol 9, Nº 1, 119-153.

Beaken, Mike. (2011): *Intonation: knowledge, viewpoint and engagement.* In: Blog: http://www.academia.edu/499946/Intonation_knowledge_viewpoint-_and_engagement. [Downloaded 15-11-2012]

Brazil, David.(1997): *The Communicative Value of Intonation in English*. Cambridge: Cambridge University Press.

Brazil, David (1992): "Listening to people reading". Coulthard, M. (ed.) *Advances in Spoken Discourse Analysis.* London: Routledge.

Brazil, David (1983): "Intonation and discourse: some principles and procedures". *Text* 3 (1) 39-70

Brazil, David. (1981) "Listening to people reading" In: Coulthard, Malcolm and Martin Montgomery (eds.) *Studies in Discourse Analysis.* London: Routledge. 209-241.

Brazil, David, Malcolm Couthard and Catherine Johns. (1980): *Discourse Intonation and Language Teaching.* London: Longman.

Chun, Dorothy M. (2002): *Discourse Intonation in L2. From Theory and Research to Practice.* Amsterdam/Philadelphia: John Benjamins.

Crystal, David ([1969] (1976): *Prosodic Systems and Intonation in English.* Cambridge: Cambridge University Press.

Gimson, Alfred C. ([1984] 1986): *An Introduction to the Pronunciation of English.* (Third edition). London and New York: Arnold.

Granato, Luisa (2005): "Aportes de la entonación al significado del discurso". *RASAL Revista de la Sociedad Argentina de Lingüística.* Nº 1. 85-109.

Gussenhoven, Carlos (2004): *The Phonology of Tone and Intonation.* Cambridge UK: Cambridge University Press.

Gussenhoven, Carlos and Haike Jacobs (2011): *Understanding Phonology:* London: Hodder Education

Halliday, Michael A. K (1967): *Intonation and Grammar in British English.* The Hage: Mouton (Janua Linguarum Series Practica 48)

Halliday, Michael A. K (1970a): *A Course of Spoken English: Intonation.* London: Oxford University Press.

Halliday, Michael A. K and William S. Greaves (2008): *Intonation in the Grammar of English.* London Oakville: Equinox.

Halliday, Michael. A. K and Christian M. I. M. Matthiessen (2004): *An Introduction to Functional Grammar.* (3rd. Edition) London: Arnold.

Kingdon, Roger. (1958): *The Groundwork of English Intonation*. London: Longman.

Ladd, D. Robert ([1996] 1998): *Intonational Phonology*. Cambridge USA:

Cambridge University Press.
Linell, Per (1998): *Approaching Dialogue. Talk, Interaction and Contexts in Dialogical Perspectives.* Amsterdam/Philadelphia: John Benjamins.
O'Connor, J. D and G. Arnold (1973): *Intonation of Colloquial English*. (2nd edition). London: Longman.
Palmer, Frank (1970): *Prosodic Analysis.* London: Oxford University Press.
Roach, Peter. ([1983]1984): *English Phonetics and Phonology.* Cambridge: Cambridge University Press.
Wells, John (2006): *English Intonation.* Cambridge UK: Cambridge University Press.

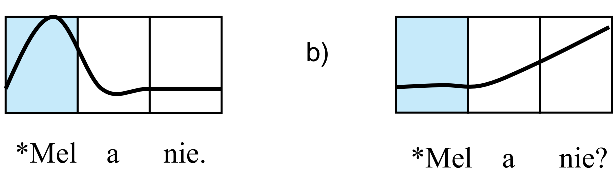
CHAPTER 9. PHONOLOGICAL MODELS OF INTONATIONAL DESCRIPTION OF ENGLISH

Eva Estebas Vilaplana
UNED- Madrid

1. Introduction

This paper presents a brief introduction to English intonation and, in particular, to the main tenets of two different traditions of intonational analysis, namely, the *British School* and the *American School*, including the latest proposals of the *Autosegmental-Metrical* framework and the *ToBI* system. Intonation deals with the pitch changes associated to utterances. Thus, any utterance can be produced with different pitch trajectories depending on the meaning the speaker wants to convey. The two graphs presented in (1), for example, show two different pitch movements associated to the word **Melanie.* Each box stands for a syllable and the shaded boxes represent stressed syllables that are accompanied with a relevant pitch movement. In the spelling stressed syllables are marked with an asterisk. Whereas the pitch trace in (1a) shows a falling contour, typical of a declarative intonation, the pitch trajectory in (1b) exhibits a rising movement, more common in questions.

(1) a.

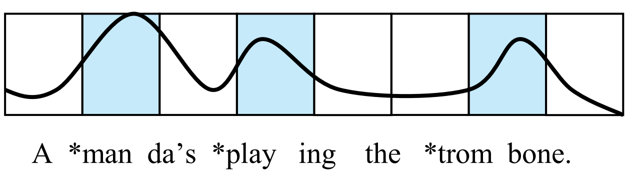


From a *phonetic* point of view, the articulatory parameter responsible for intonation changes is the rate of vocal fold vibration. Differences in the vibration of the vocal folds are perceived as differences in pitch. The relationship between vocal fold vibration and pitch perception is the following: the quicker the vibration of the vocal folds, the higher the pitch. Similarly, the slower the vibration of the vocal folds, the lower the pitch.

From a *phonological* perspective, the study of intonation is aimed at defining the inventory of tonal categories that are linguistically relevant in a given language. In other words, the phonology of intonation investigates which pitch movements trigger differences in meaning at the sentence level and therefore they should be interpreted as contrastive phonological entities or *tones.* For example, the two different pitch configurations of the word *Melanie* illustrated before should be categorized as two distinctive tones or phonological units since these pitch differences are responsible for changes in the meaning (declarative vs. interrogative) of the same sequence of sounds (*Melanie*). Therefore, we refer to a *falling* tone which contrasts with a *rising* tone.

Even though the intonation pattern of an utterance expands throughout the whole segmental string, the pitch movements that are linguistically relevant (*i.e.* the tones) are mainly associated to the stressed syllables. For example, in the following utterance (*A*manda's *playing the *trombone*) we can see that each stressed syllable, marked with an asterisk, becomes accented since it shows a relevant high pitch trajectory in the contour. As we will discuss in more detail in section 2, the right boundary of the utterance also shows pitch movements that are phonologically relevant. Thus, for example, in (2) the falling trajectory observed at the end of the utterance, after the last accented syllable, accounts for a declarative intonation[8].

(2)



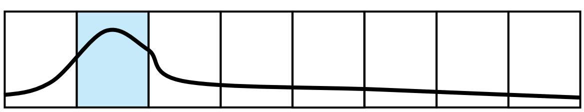
As we will examine in section 3, different theoretical frameworks, namely, the British School and the American School of intonational analysis, have different proposals to categorize the linguistically relevant entities (inventory

8 This graphic representation, based on Estebas-Vilaplana (2009), does not exhibit the interruptions of the pitch trace produced by voiceless sounds.

of tones) of English intonation. Before we actually describe the main tenets of such models, we will first review some of the main concepts used in the study of intonation, which tend to be common to most frameworks and will help us compare their descriptive proposals. In particular, two issues will be covered: 1) the differences between *stress*, *accent* and *focus*, and 2) the three parameters used to describe intonation (i.e. *tonality*, *tonicity* and *tone*).

2. Stress, accent and focus

The distinction between *stress* and *accent* was first proposed by Bolinger (1958) and has been used in most descriptions of English intonation ever since. Stress has to do with the degree of prominence with which a syllable is produced. A stressed syllable is more prominent than an unstressed syllable. This means that stressed syllables are produced with more energy and muscular effort than unstressed syllables. Hence, stressed syllables are usually longer and are perceived as being louder than unstressed syllables. Stressed syllables may also be produced with a relevant pitch movement. In this case, the stressed syllable becomes *accented*. Thus, whereas a *stressed syllable* is a syllable with rhythmic prominence, an *accented syllable* is a syllable which, apart from rhythmic prominence, also has pitch prominence. This means that all accented syllables have to be stressed but not all stressed syllables are accented. The following example shows the same sentence as in (2) but with a different accentuation pattern. The only syllable that is accented is the first stressed syllable (**man*). The other two stresses (**play* and **trom*) are not accented since no relevant pitch movement is associated to them.



A *man da's *play ing the *trom bone.

The decision of which syllables are accented and which syllables are not depends on the meaning the speaker wants to convey and on which part of the utterance (s)he wants to highlight or *focus*. Focalization has to do with the association of intonational prominence to those parts of the utterance that are more informative (*i.e*. carry the new information). Sentences can be of two kinds depending on their information or focal structure: 1) *broad focus* sentences (if the whole sentence is new information) or 2) *narrow focus* sentences (if part of the sentence is old information and part of it is new). If the whole sentence is new information (broad focus) the last lexical word will always get an accent, as in example (2), which can be the answer to a question such as *What happens?* The stresses preceding the last accent tend to be accented too. When the utterance is produced with narrow focus, the element that carries the new information must be accented, whereas the items that contain the old information are not, as in example (3), which can be the reply to a question such as *Who's playing the trombone?*

2.1 Parameters to describe intonation

The phonological description or modeling of the pitch contours is not an easy task to do given the continuous nature of intonation. Contrary to phonemes, which can be easily identified as discrete, single units, intonation is difficult to decompose into phonological units since it is not straightforward to decide which parts of the pitch contour are linguistically relevant and which parts are not. There have been various theories and schools which have proposed different conventions to analyze intonation (see section 3). Despite important differences in their proposals, they all agree that intonation has to be described according to three parameters: 1) the number of intonation units into which a speech chunk is divided (*tonality*), 2) the distribution of accents (*tonicity*), and 3) the kinds of accents (*tone* inventory). The terms *tonality, tonicity* and *tone* were first proposed by Halliday (1967) and have been used in most studies within the British School of intonational analysis. Even though the followers of the American School have not adopted those exact terms, their description of intonation is also based on the same parameters, namely, the division of the text in intonational phrases, the location of the accents within each phrase and the types of accents.

2.1.1. Tonality

Tonality has to do with the division of a chunk of speech into different *intonation units* or *phrases* (*i.e*. complete intonation patterns). An intonation unit is delimited by some kind of intonational boundary. Whereas sometimes a major break (pause) is produced between two intonation units, other times no actual break is observed. In such cases, speakers still perceive the presence of an intonation boundary which is signaled by means of a relevant pitch movement at the end of the phrase. Furthermore, other cues, such as syllable duration, can also help to identify the end of an intonation phrase. The duration of the last accented syllable and the post-accented syllables (if any) tends to be longer at the end of an intonation phrase (see Cruttenden 1986). The decision to divide a given speech chunk into a number of intonation units is speaker-dependent and it may vary according to the meaning the speaker wants to convey. Thus, tonality is the choice the speaker has of the placement of intonation boundaries in an oral production to create intonation units. For example, the following speech chunk can be produced with a different number of intonation phrases, as indicated with the double vertical lines.

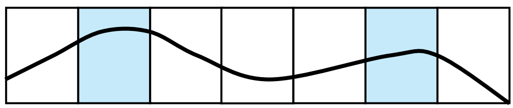
(4) ‖ I didn't know that Peter left the country. ‖
‖ I didn't know ‖ that Peter left the country. ‖
‖ I didn't know that Peter ‖ left the country. ‖
‖ I didn't know ‖ that Peter ‖ left the country. ‖

The concept of *intonation phrase* (term used in Pierrehumbert 1980) has received different names in the literature, namely, *tone unit* (Crystal 1969), *tone group* (Halliday 1967), *macrosegment* (Hockett 1958), or *word group* (O'Connor and Arnold 1973), among others.

2.1.2. Tonicity

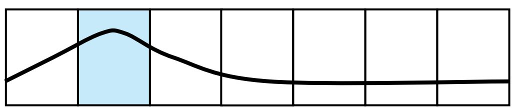
Once the speaker has decided the number of intonation phrases in which a given speech chunk is divided, the next step is to decide which syllables will become accented (*i.e.* will have pitch prominence) for each intonation phrase. Thus, tonicity deals with the distribution of accents within an intonation unit. Only stressed syllables can get an accent. However, as reported in section 2.1, not all stressed syllables must be accented and the choice of which syllables get a pitch prominence depends on the speaker's communicative intentions. The example in (5) contains a sentence with two stresses (*My *friend is a mu*sician*) which is produced with two different accentuation patterns: with two accents (5a) and with one accent (5b). The distribution of accents within an intonation phrase is a parameter which is closely linked to focalization. Whereas (5a) is an example of a broad focus sentence, (5b) shows a narrow focus structure.

(5) a.



My *friend is a mu *si cian.

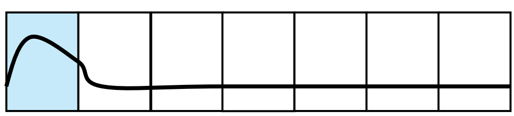
b.



My *FRIEND is a mu *si cian.

Normally, function words are not stressed and therefore they are not accented. Sometimes, however, function words may become informative (that is, they are focalized or highlighted) and hence they are subsequently stressed and accented. This is illustrated in (6) below where the possessive adjective receives an accent in a context in which *my* contrasts with *your* and therefore is focalized.

(6)



*MY *friend is a mu *si cian.
(not YOUR friend)

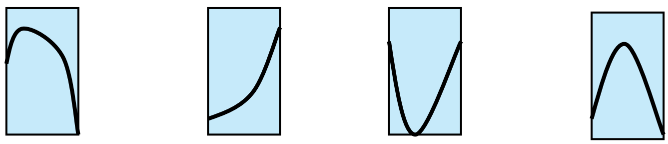
Once the speaker knows which syllables are accented for each intonation phrase, then (s)he must decide the kind of accent (s)he wants to use in order to convey a given meaning. The system of accentual choices of a given language is known as the *tone inventory* and it includes all the *tones* (linguistically relevant intonation entities) used in that particular language. In example (7), the word *no* is produced with different pitch trajectories (fall, rise, fall-rise and rise-fall) which can be categorized as different tones since each pitch movement is responsible for changes in the meaning of the utterance.

(7)

	*No	*No	*No	*No
Tones:	*Fall*	*Rise*	*Fall-rise*	*Rise-fall*
Meanings:	polite statement	question	statement showing reservations	angry statement

There have been different theoretical proposals to model intonation, that is, to find out and categorize the tones of a language. In the preceding example, for instance, we have described a falling pitch movement as a *fall* tonal category. This convention is basically used by the British tradition of intonational analysis which defines tones according to their pitch trajectories. As we will see in more detail in section 3, there are other frameworks, especially those derived from the American School of intonational modeling, that describe intonation not according to its trajectory (falling or rising) but according to pitch targets (high or low). Thus, for example, a falling contour will be categorized by means of a sequence of two tones: H (high) and L (low). The expected pitch movement between H and L is obviously a falling pitch trajectory. Thus, in a very simplified way, the categorization of the previous pitch contours according to a model based on tonal targets would be as follows.

(8)



	*No	*No	*No	*No
Tones:	H L	L H	H L H	L H L

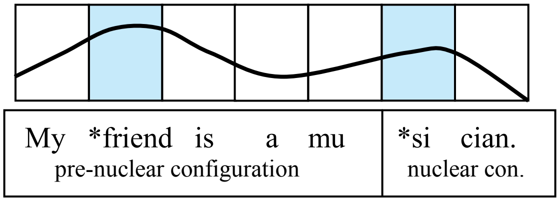
3. Intonational modeling

There are two main traditions for the modeling of English intonation: 1) The British School and 2) the American School. These two traditions are the main exponents of two different ways of intonational description, namely, the analysis of the pitch contours by means of *configurations* (as in the British tradition) or as a series of *level* tones (as in the American frameworks). The terms *configurations* vs. *levels* were first proposed by Bolinger (1951). In the next sections, we will briefly discuss the most important theoretical issues of the two proposals. Section 3.1 will be devoted to the works of the British School and section 3.2 will summarize the main contributions of the American tradition divided in two other sections, namely, 1) the early works and 2) the latest proposals, including the Autosegmental-Metrical framework and the ToBI system.

3.1. The British School

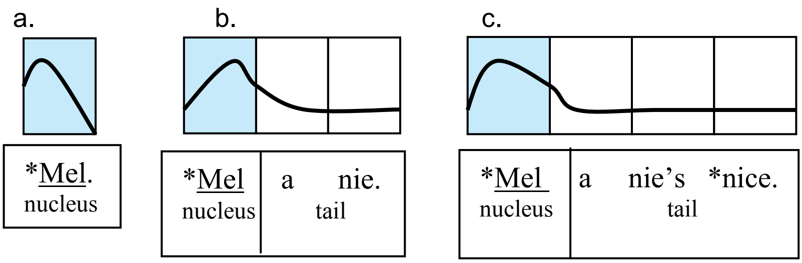
The intonational model proposed by the British School (see the works of Palmer 1922, Kingdon 1958, Crystal 1969, O'Connor and Arnold 1973, Gimson 1980, Brown *et al.* 1980, Brazil 1985, Couper-Kuhlen 1986, Cruttenden 1986, Tench 1996, Schubiger 1958, or Wells 2006) is also known as the *configurational* tradition since the intonation phrases are analyzed according to two configurations, the *nuclear* configuration and the *pre-nuclear* configuration (first proposed by Palmer 1922). Whereas the nuclear configuration includes the last accented syllable and all the subsequent unaccented syllables (if any), the pre-nuclear configuration includes all the syllables preceding the last accent. Example (9) reproduces the utterance in (5a) with an illustration of the nuclear and pre-nuclear configurations.

(9)



In order to model intonation, the British School proposes to divide each intonation phrase into the following parts: *pre-head*, *head*, *nucleus* and *tail*. The nucleus is the only obligatory element of an intonation phrase and it consists of the last (or only) accent. Thus, the nucleus is located on the last accented syllable of an intonation unit. This means that there might be other stressed or unstressed syllables after the nuclear one but there will never be other accented syllables. The tail includes the unaccented syllable(s) (if any) after the nucleus. The presence of a tail is optional. The following pitch contours illustrate: 1) an utterance containing only a nuclear syllable (**Mel*), as in (10a); 2) an utterance made up of a nuclear syllable (**Mel-*) and a tail with only unstressed syllables (*-anie*), as in (10b); and 3) an utterance containing a nuclear syllable (**Mel-*) and a tail which has both stressed and unstressed syllables (*-anie's *nice*), as in (10c). In all the cases the nucleus falls on the same syllable (*Mel-*) which is the only one that shows a pitch prominence and thus it is accented. The nuclear syllable is underlined. The nucleus and the tail make up the nuclear configuration.

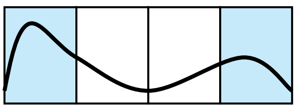
(10)



The head is an optional component of the intonation phrase and it consists of all the syllables from the first accent up to the syllable before the nucleus. The pre-head is also optional. It includes the unaccented syllables before the first accent. It is possible to have a pre-head without a head. The head and the pre-head make up the pre-nuclear configuration.

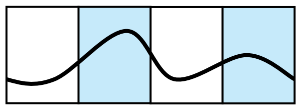
The following pitch contours illustrate: 1) a nucleus (**nice*) preceded by a head (**Melanie´s*), as in (11a); 2) a nucleus (**nice*) preceded by a pre-head (*My*) and a head (**sister´s*), as in (11b); and 3) a nucleus (**nice*) only preceded by a pre-head (*It's*), as in (11c).

(11) a.



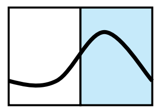
*Mel a nie's head	*nice. nucleus

b.



My p-h	*sis ter's head	*nice. nucleus

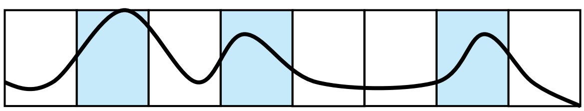
c.



It's p-h	*nice. nucleus

It is possible to have heads with more than one accent, as in (12), which exhibits a head with two accents (one on **sure* and the other one on **call*).

(12)



I'm p-h	*sure she'll *call you to head	*mor n	row. tail

Thus, the description of intonation within the parameters of the British School involves the division of each intonation phrase into four components, as schematized in (13). The elements in brackets are optional.

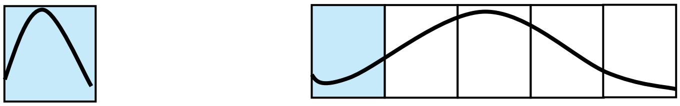
(13)

(Pre-nuclear configuration)	Nuclear configuration
(pre-head) (head)	**nucleus (tail)**

According to this framework, the tones, that is, the linguistically relevant entities that account for a pitch contour are associated to two elements in the segmental string: 1) the nuclear syllable and 2) the accented syllables in the head (if there is a head). The types of tones differ depending on their position: nuclear position or pre-nuclear (head) position.

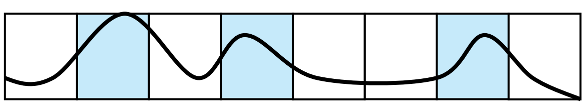
Nuclear tones are associated to the nuclear syllable. According to O'Connor and Arnold's (1973) proposal, the inventory of nuclear tones includes five simple tones (with one pitch trajectory) and two complex tones (with two pitch trajectories). The simple tones indicate the origin of the pitch (high, low or mid) and the direction of the pitch (fall, rise or level). The simple tones are: *high-fall, low-fall, high-rise, low-rise* and *mid-level.* The complex tones show the combination of two pitch trajectories (*fall-rise* and *rise-fall*). A high-fall (marked with the diacritic ↘ before the nuclear syllable) involves a high pitch which falls to low and then it remains low and level to the end of the intonation pattern. A low-fall (↘) is produced with a mid pitch falling to low and staying low and level to the end of the contour. A high-rise (↗) shows a mid pitch rising to a high pitch to the end of the intonation pattern. A low-rise (↗) involves a low pitch rising to a mid pitch to end of the contour. A mid-level (>) shows a mid pitch sustained to the end of the intonation pattern. A fall-rise (↘↗) includes a high pitch falling to low and rising again to the end of the intonation pattern. A rise-fall (↗↘) starts with a low pitch which rises to high (or mid), then falls to low and finally levels out to the end of the intonation pattern.

The realization of each tone may vary depending on whether there is a tail after the nuclear syllable or not. If there is no tail, the pitch movement will take place within the limits of the accented syllable. If there is a tail, the pitch movement will start on the nuclear syllable and expand over the syllables in the tail. (14) includes an example of the realization of a fall-rise nuclear tone with tail (14b) and without tail (14a). The small circle located on the first syllable of the word *singing* indicates that this syllable is stressed but not accented. The types of tones found in *pre-nuclear* position differ from those in nuclear position in that they do not account for the pitch trajectory at the end of the intonation phrase but they just describe the pitch movement(s) of the accented syllables before the nucleus, that is, in the head position.

(14) a. 

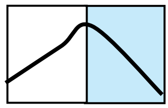
↗↘ No ↗↘ Mel a nie's ˚sing ing.

According to the British School of intonational analysis, there are two types of heads: 1) *simple heads*, which contain one accented syllable, followed by other unstressed syllables (if any), and 2) *complex heads,* which contain more than one accented syllable (along with other unstressed syllables). In simple heads, the pre-nuclear tones are: *high* (ˈ), *low* (ˌ), *rising* (↗), or *falling* (↘). A high tone indicates that the accented syllable of the head is produced with a high pitch. A low tone indicates that the accented syllable of the head is produced with a low or mid pitch (not high). A rising tone starts with a low pitch on the accented syllable followed by a peak on the next syllable. Finally, a falling tone involves a high pitch on the accented syllable followed by a fall in the following syllable. In complex heads, the accented syllable always has the same type of accent. In other words, we cannot have a head, for example, with a low accent and a high accent. Thus, if the first accent of a complex head is high, the subsequent accents will also be high. Complex heads are classified as: 1) *downstepping heads* with a high tone in each accented syllable, 2) *sliding heads* with a falling tone in each accented syllable, and 3) *climbing heads* with a rising tone in each accented syllable. Example (15) illustrates a complex downstepping head with two high accents followed by a high-fall nuclear accent.

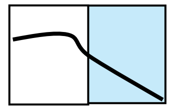
(15) 

A ˈman da's ˈplay ing the ↘trom bone.

Finally, pre-heads tend to be low but sometimes they can be high. In that case, they are marked with a ¯ symbol. Low pre-heads are not marked. Figure (16) includes two productions of the utterance *It's *nice* pronounced with a low pre-head (followed by a high-fall nuclear tone), as in (16a), and a high pre-head (followed by a low-fall nuclear tone), as in (16b).

(16) a. 

It's ↘nice.



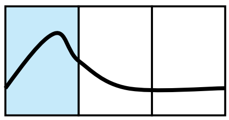
¯It's ↘nice.

3.2 The American School

The British and the American Schools of intonational analysis differ in two important characteristics: 1) the way of describing tones and 2) the interpretation of an intonation phrase. As we have stated, the British tradition describes tones by means of their pitch trajectories. Thus, the British School tonal inventory includes categories such as *high-fall*, *low-rise* or *fall-rise*. The American School, on the other hand, analyzes intonation by means of a series of tone levels (or tonal targets), which in the early works are identified by means of numbers (/1/- low, /2/-mid, /3/-high, and /4/-extra-high) and in the latest theories (such as the Autosegmental-Metrical approach) are specified as H (high) or L (low). Thus, according to these proposals, a falling pitch movement would be categorized as /3 1/ or H L since the pitch trajectory from /3/ to /1/ or from high (H) to low (L) involves a fall. Hence, /3 1/ or HL would be other ways of accounting for a high-fall.

The second difference between the two traditions is the way intonation phrases are interpreted. Whereas the British tradition decomposes an intonation phrase into a nuclear and a pre-nuclear configuration, the American tradition does not differentiate between nuclear and pre-nuclear accents. According to the American School, each stressed syllable that becomes accented is associated to a pitch number or to H and L tones, no matter whether the accented syllable is the last one of the contour or not. Thus, contrary to the British proposal, the American School has no specific tonal entities (such as nuclear tones) associated to the last accent of a pitch contour. However, in order to account for the final movement of an intonation phrase, the American School introduces the concept of *boundary tones*, that is, tones that occur at the final edge of an intonation phrase. These tones are not associated to stressed syllables but to the limits of intonation units. Thus, the falling contour of the example in (17) would be modeled as a /3/ or an H tone on the stressed syllable **Mel* and a /1/ or low boundary tone at the end of the phrase. For the specific diacritics used to signal a boundary tone in the two proposals see sections 3.2.1 and 3.2.2.

(17)



*Mel a nie
3 1
H L

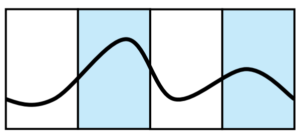
The American tradition of intonational analysis embraces a rather large variety of theoretical frameworks with different proposals to model intonation. All these models have two common features: 1) the analysis of the pitch contours by means of level tones and 2) the presence of boundary tones. In the following sections, we will have an overall view of the chronological development of such models.

3.2.1. The early works

The main goal of the early works of the American School of intonational analysis was to describe the intonation of American English as a series of distinctive units. Bloomfield (1933) was the first one to propose the analysis of the pitch contours as a sequence of *secondary phonemes* (also called *tonemes)*, identified through a contrastive analysis of sentences which were segmentally equal but differed in their pitch patterns. Bloomfield's proposal was followed by several authors, such as Pike (1945), Wells (1945), Trager and Smith (1951) or Liberman (1975), as well as Leben (1976) and Goldsmith (1976) within the Autosegmental Phonology framework.
Within the early works of the American School, pitch contours are described as a series of pitch level phonemes. In Trager and Smith (1951), for example, the speaker's pitch range is divided into four relative phonemic

pitch levels, /1, 2, 3, 4/, where /1/ means low, /2/ mid, /3/ high and /4/ extra-high. Pitch phonemes are associated to different sites of the segmental string, such as, the beginning of an utterance or the stressed syllables. Each pitch phoneme can have four allophonic realizations, from high to low, represented by the following diacritics [ˍ], [˄], [◦] and [ˬ]. Thus, for instance, [1̲] and [4̲] are the highest allophones within levels /1/ and /4/.
In Trager and Smith, three phonemes of terminal juncture are added into the system. Terminal juncture phonemes represent the pitch movements at the end of the intonation phrase: fall (#), rise (||) and level (|). In this view, a pitch contour consists of a series of pitch phonemes followed by a terminal juncture phoneme. Example (18) illustrates the same sentence as (11b) described according to the parameters of the American School of intonational analysis. At the beginning and at the end of the sentence, pitch level /1/ indicates a low pitch. /3/ and /2/ stand for a high tone and a mid tone associated to the stressed syllables. Finally, the symbol # shows a final fall.

(18)



My *sis ter's *nice

1 3 2 1 #

As opposed to the British tradition, the intonational analysis proposed by the American School does not divide the pitch contours into smaller constituents (such as the nucleus or the head). Thus, the inventory of tonal categories (pitch phonemes) is the same irrespective of whether they are associated to the last accented syllable or other syllables. Later developments of the level analysis of intonation within the Autosegmental approach propose to reduce the number of pitch accents to three levels (high, mid and low) and the idea of phonemes at the terminal junctures develops into the notion of boundary tone, *i.e.* an underlying phoneme manifested phonetically by a pitch movement at an intonational boundary (Liberman 1975).

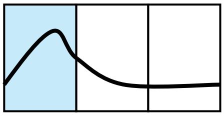
3.2.2. The latest proposals: the Autosegmental-Metrical approach and the ToBI system

One of the latest and most influential frameworks of intonational analysis within the American tradition is the *Autosegmental-Metrical (AM)* approach (Pierrehumbert 1980, Beckman and Pierrehumbert 1986, Pierrehumbert and Steele 1989, Ladd 1996, Gussenhoven 2004, among many others) whose tenets have been used to create a system for intonational annotation known as the *ToBI* (*Tone and Break Indices*) system (Beckman and Hirshberg 1994, Beckman and Alam-Eyers 1997). The AM approach to intonational analysis inherits several aspects from the early works of the American School, in particular, the level analysis of pitch (although reduced to two categories) and the notion of a boundary tone.

According to the AM framework and the ToBI system, intonation contours can be described by means of two tones, H (high) and L (low), which can be associated to stressed syllables or to the edges (breaks) of the intonation phrases. When the tones are associated with stressed syllables, they are marked with an asterisk. Thus, H* and L* indicate that a high pitch and a low pitch occur within a stressed (accented) syllable. Tones associated with the right edge of a pitch contour can be of two kinds depending on whether they signal the edge of a higher prosodic domain (the *intonation phrase*) or a lower prosodic domain (the *intermediate phrase*). Tones at the end of an intonation phrase are called *boundary tones* and are marked with %. Thus, H% and L% stand for a high pitch and a low pitch respectively at the end of an intonation phrase. Tones at the end of an intermediate phrase are called *phrase accents* and are signaled with the symbol – (*i.e.* L- and H-). In this way, H- and L- represent a high and a low pitch movement at the end of an intermediate phrase.

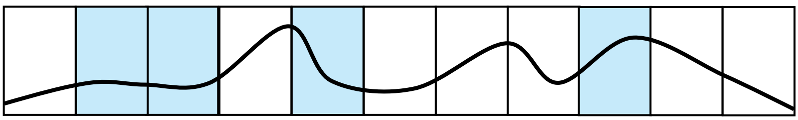
Intonation phrases contain at least one or more intermediate phrases. Whereas the intonation phrase is marked by a final boundary tone usually followed by a pause, the intermediate phrase contains a final phrase accent but no subsequent break. The example in (19) includes an intonation phrase (marked with a final L%) with only one intermediate phrase (signaled by L-). The sentence in (20) illustrates an intonation phrase with three intermediate phrases (each phrase contained in a different box). The first two intermediate phrases end with a H- phrase accent. The third intermediate phrase finishes with a L- phrase accent followed by a L% boundary tone which signals the end of the full intonation phrase.

(19)



*Mel a nie.
H* L-L%

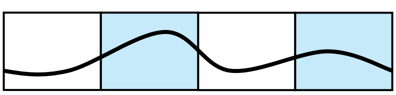
(20)



I *love *lem ons, L* L* H-	*man da rins L* H-	and *o ran ges. H* L-L%

The possibility within the AM model of analyzing pitch contours by means of two tones only (H and L) derives from the fact that this model assumes that H and L can have different realizations. For example, in a pitch contour produced with a series of H* accents, each H* accent will be *downstepped*, that is, it will be produced at a lower level than the preceding accent. Despite differences in the realizations of these H* accents, they are still interpreted as H*. The convention proposed by the AM framework to indicate a downstepped accent (*i.e.* an H* accent which is lower than the preceding one) is by means of the symbol ! (!H*). Thus, for example, the two high accents of the pitch contour in (21) will be described as H* !H*.

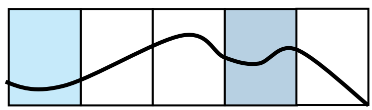
(21)



My *sis ter's *nice.
H* !H* L-L%

Apart from H* and L*, the tone inventory proposed by the AM framework also includes bitonal accents, that is, accents which are made up of two tonal targets, such as L*+H or L+H*. These tones can only be associated with stressed syllables but not with the edge of a pitch contour. In a bitonal accent, the tone with the asterisk indicates that this tone is realized within the accented syllable. The tone preceding the starred one is realized on the preceding (unstressed) syllable(s). The tone following the starred one is realized on the following (unstressed) syllable(s). For example, L*+H indicates that the stressed syllable has a low pitch followed by a rise with the peak on the post-stressed syllables. An example of an intonation phrase containing a bitonal accent (L*+H) is illustrated on the first word of the sentence **Melanie's *singing* in (22), which exhibits a low pitch on the stressed syllable (**Mel*) followed by a high pitch on the following unstressed syllables.

(22)

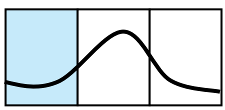


*Mel a nie's *sing ing.
L*+H H* L-L%

The latest versions of the English ToBI system (Beckman and Ayers-Elam 1997) propose a tone inventory which includes seven pitch accents, two phrase accents and three boundary tones. Pitch accents are divided into monotonal and bitonal. Monotonal pitch accents are: H* (a high peak within the accented syllable with no preceding or following valley), L* (a low pitch on the accented syllable), and !H* (a downstepped high pitch within the accented syllable). Bitonal pitch accents include: L+H* (a high pitch on the accented syllable with a preceding valley), L+!H* (a downstepped L+H*), L*+H (a low pitch on the accented syllable with a rise on the following unstressed syllable), and H+!H* (a high pitch on the unstressed syllable followed by a downstepped tone on the accented syllable).

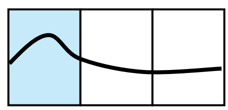
With respect to edge tones, the model proposes three phrase accents, namely, H- (high), L- (low) and !H- (downstepped high) and two *boundary tones*, H% (high) and L% (low). The model also attests an initial high boundary tone (%H) which may account for those cases where the utterance starts with a high pitch associated to initial unstressed syllables. This %H is equivalent to the concept of a high pre-head within the British tradition (see example 16b). The English ToBI proposal, describes four possible combinations of phrase accents and boundary tones at the end of a pitch contour in English, namely, L-L% (final falling as at the end of a declarative intonation), H-H% (steep rise used in a yes-no question contour), L-H% (fall-rise contour if the preceding pitch accent is H* or continuation rise if the preceding accent is L*) and H-L% (rise-fall if the preceding pitch accent is L* or final level plateaux if the preceding accent is H*). The two pitch patterns in (23) show a final rise-fall (L*H-L%), typical of an urgent statement, and a level configuration (H* H-L%), attested in a calling contour.

(23) a.



*Mel a nie
L* H-L%

b.



*Mel a nie
H* H-L%

4. Modeling the basic tunes of English intonation

In this last section, we will compare how two different approaches, namely, the British School and the AM model, describe some of the basic tunes of English intonation, such as the neutral patterns of declarative, interrogative (wh-questions, yes-no questions, and tag-questions) and imperative intonation. Some non-neutral patterns will also be covered, such as, statements showing reservations and wh-questions with a polite nuance.

The most common type of intonation for neutral statements involves a fall, that is, a pitch pattern which shows a high pitch on the last accented syllable of the contour which glides to low at the end of a sentence. If there are any accented syllables preceding this falling movement, they are usually high. This is illustrated in Figure 1 which shows the speech waveform and the spectrogram with an overlapped F0 (pitch) trace for the neutral declarative **Melanie *loves *mandarins*. The last tier includes the syllable division so as to show the association between pitch movements and the segmental string.[9] The box at the top of the Figure specifies the description of this pitch contour according to the British School and the AM model.

Neutral declarative	
British School	*AM model*
ˈMelanie ˈloves ↘mandarins.	Melanie loves mandarins. H* !H* !H* L-L%

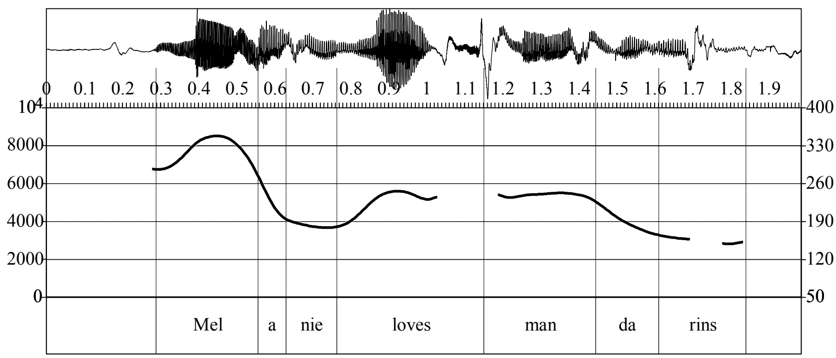


Figure 1. Speech waveform and pitch trace for the neutral declarative **Melanie *loves *mandarins.*

It is also very common to produce statements with a falling intonation which rises again at the end of the contour. The nuance of this type of tonal pattern is that the speaker has reservations about what (s)he has said. This pattern also expresses non-finality or continuity. If there are any accented syllables preceding the falling-rising movement, they tend to be falling, as illustrated in Figure 2. This utterance can have two nuances 1) Mary loves mandarins but not very much (reservations), and 2) Mary loves mandarins and also other types of fruits (non-finality).

[9] The graphic representations were obtained by means of *Praat* (Boersma and Weenink 2011).

Non-final declarative/declarative with reservations	
British School	*AM model*
ˎMelanie ˚loves ˎˊmandarins.	Melanie loves mandarins. H* !H* L-H%

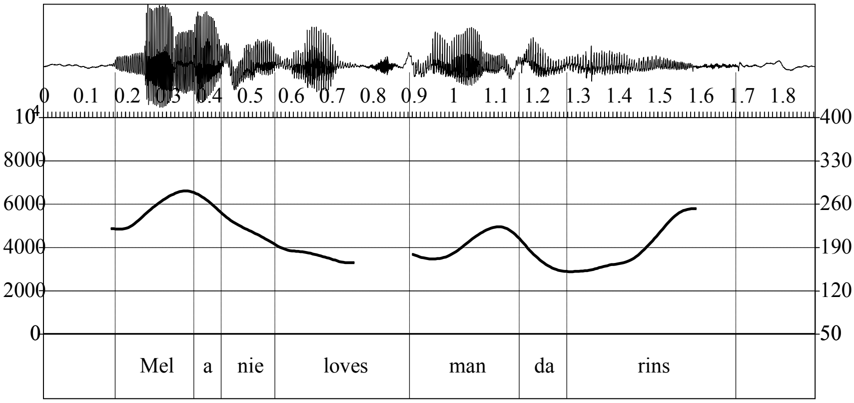


Figure 2. Speech waveform and pitch trace for the non-final declarative **Melanie *loves *mandarins.*

Wh-questions always begin with an interrogative pronoun, such as *what, where, which, why* and *how*. The most typical intonation for a wh-question is a fall. If there are any accented syllables preceding this falling movement, they are usually high. An example of a wh-question is included in Figure 3. Sometimes wh-questions are produced with a final rise. Any accented syllables preceding this rise are high. In this case it has a nuance of politeness, kindness and encouragement. This type of contour is illustrated in Figure 4.

Neutral wh-question	
British School	*AM model*
ˈWould you ˚like a ↘mandarin?	Would you like a mandarin? H* H* L-L%

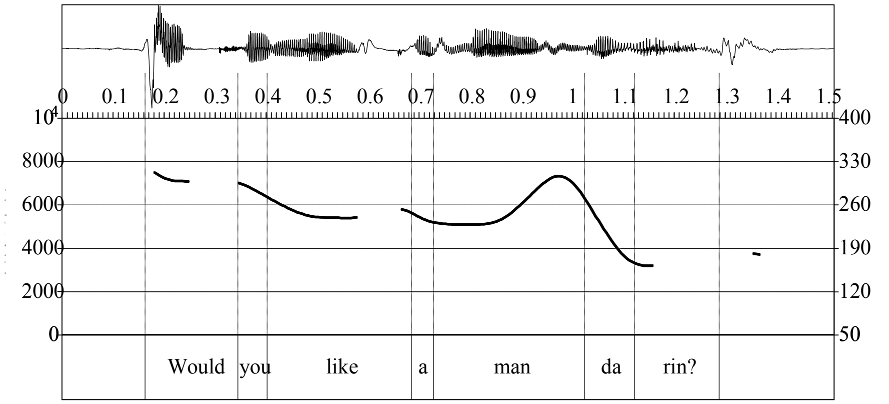


Figure 3. Speech waveform and pitch trace for the neutral wh-question **Would you *like a *mandarin?*

Polite wh-question	
British School	*AM model*
ˈWould you ˚like a ↗mandarin?	Would you like a mandarin? H* L* H-H%

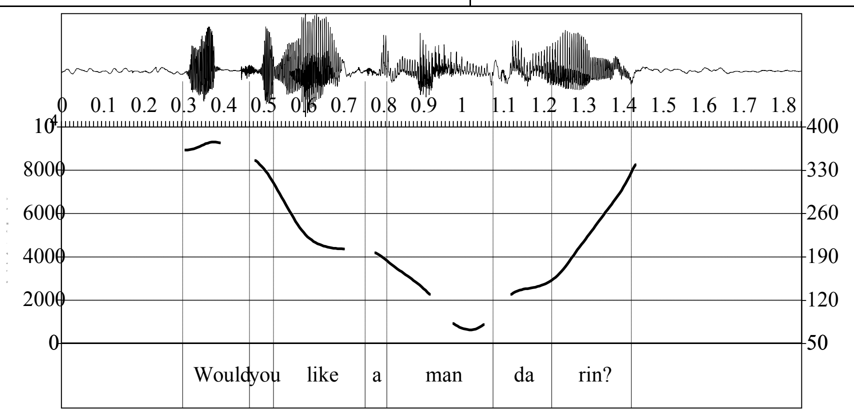


Figure 4. Speech waveform and pitch trace for the polite wh-question **Would you *like a *mandarin?*

A yes-no question is a type of question whose answer can only be “yes” or “no”. The most typical intonation for a yes-no question is a rise from a low pitch to a mid pitch. If there are any accented syllables preceding this rising movement, they are usually high. An example is included in Figure 5 for the utterance **Do you *like *mandarins?*

Neutral yes-no question	
British School	*AM model*
ˈDo you ˚like ↗ mandarins?	Do you like mandarins? H* L* H-H%

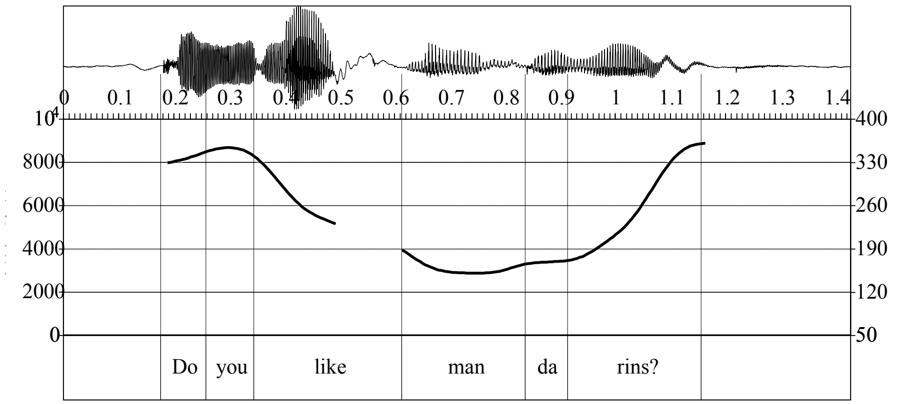


Figure 5. Speech waveform and pitch trace for the neutral yes-no question **Do you *like *mandarins?*

Tag questions, such as *(You *love* mandarins,) *don't you?* can be produced with a fall or with a rise. If they are produced with a fall, the speaker asks for *confirmation* of what (s)he has just said and (s)he only expects an agreement. This is illustrated in Figure 6. The utterance contains two phrases marked with two nuclear tones in the British tradition and with a phrase accent at the end of the first prosodic domain in the AM model.

Tag-question (asking confirmation)	
British School	*AM model*
You ˚love ↘mandarins, ↘don't you?	You love mandarins, don't you? H* L- H* L-L%

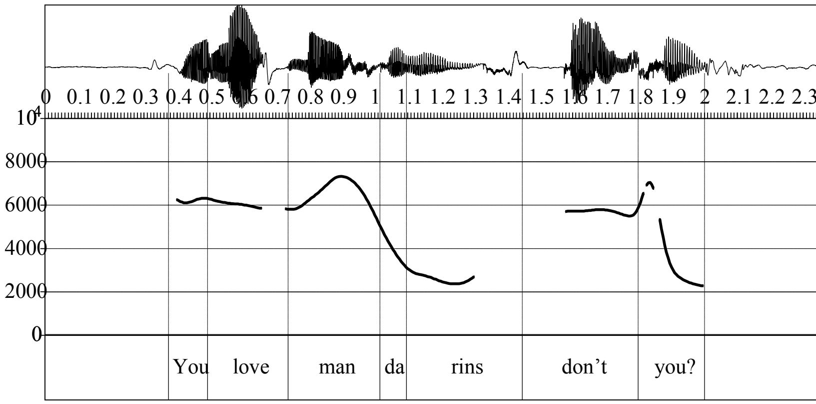


Figure 6. Speech waveform and pitch trace for the tag-question (asking confirmation) *You *love *mandarins, *don't you?*

When tag-questions are produced with a rise the speaker asks for *information* and thus (s)he expects an answer. Figure 7 includes the same tag question as in Figure 6 but with a final rise. As before, the utterance contains two phrases. In English, however, most tag questions are produced with a fall, that is, they only expect agreement from the interlocutor. Finally, the most common intonation pattern for a command is a fall from a mid pitch. Any preceding accented syllables tend to be high. The nuance conveyed by this pattern is firm and authoritative. This pattern is illustrated in Figure 8 for the sentence **Take a *mandarin*.

Tag-question (asking information)	
British School	*AM model*
You ˚love ↘mandarins, ↗don't you?	You love mandarins, don't you? H* L- L* H-H%

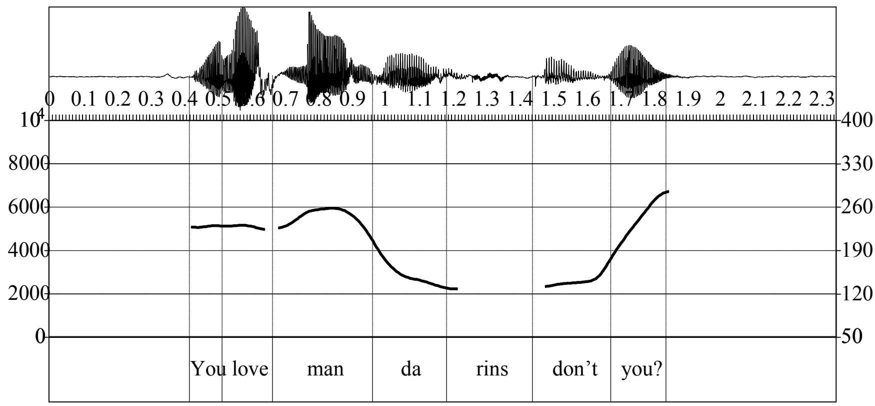


Figure 7. Speech waveform and pitch trace for the tag-question (asking information) *You *love *mandarins, *don't you?*

Neutral command	
British School	*AM model*
ˈTake a ↘mandarin	Take a mandarin H* !H* L-L%

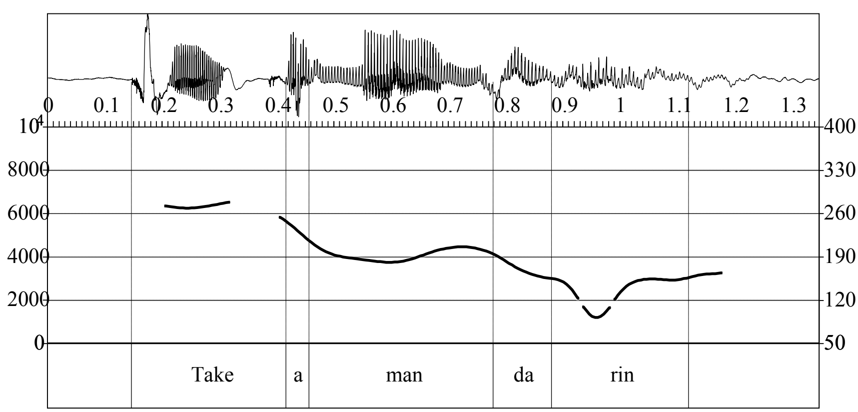


Figure 8. Speech waveform and pitch trace for the neutral command **Take a *mandarin.*

5. Conclusions

In this paper, we have presented a brief overview of two of the most famous traditions of intonational description in English, namely, the British School and the American School (with particular interest in its latest proposals, such as the AM approach and the ToBI system). The presentation of the models included in this paper is a general and broad introduction to two traditions whose insights and contents are much richer and more elaborate than the ideas offered in this summary. However, we expect that these short impressions may help students to open their interests in the field of prosody and intonational phonology.

Apart from the differences in the theoretical tenets of both traditions, the immediate applications of the two models have also been very different. Whereas the British tradition has been mainly used for pedagogical purposes, the American frameworks have been more phonologically oriented and have been applied to the description of linguistic corpora and to speech technology. Nowadays, the great impact that the AM approach and the ToBI system have had on the description of intonation is due to a number of reasons, namely, its concise tonal description tools, reduced to two tones, and the clear-cut location of such tones with respect to the segmental layer (tones are only associated to stressed syllables and to the end of phrases). These assets have been particularly beneficial for speech synthesis and speech recognition due to the explicit interaction between the segmental and the prosodic levels. Despite the several advantages of the AM model and the ToBI system, they have hardly ever been used with a pedagogical aim. When it comes to teach students the intonation of foreign languages, some of the pros, such as the reduced number of tones, may turn out to be a drawback since it is really difficult for untrained students to grasp the meaning of categories that look pretty much the same (*e.g.* H*, L+H*, L*+H, H*+L, L+!H*). Thus, despite the great impact of the AM model and the ToBI system on intonational modeling, the teaching and learning of English intonation have rarely benefitted from their insights and, for more than fifty years, most of the training materials for the study of intonation have been designed within the tenets of the British School. One of the advantages of this tradition is that it offers a detailed, perceptually-based description of the intonation contours of different types of sentences, which is very helpful for the learning of the main tunes of English. Even though the benefits of the British model for the teaching of intonation have been widely attested, this framework also presents some problems. In particular, it requires a rather long-term, face-to-face training and tuition to have a full command of it since it is difficult to grasp the alignment of pitch-trajectory tones to the segmental string.

Despite the pros and cons associated to each model, the contributions of these two schools to the analysis and description of English intonation are unquestionable. Thus, it is crucial that any student of English phonetics and phonology who wishes to have a thorough understanding of English intonation benefits from the insights of both traditions.

REFERENCES

Beckman, Mary & Ayers-Elam, Gayle (1997): “Guidelines for ToBI labeling”. <http://www.ling.ohio-state.edu/research/phonetics/E_ToBI>.

Beckman, Mary & Hirschberg, Julia (1994): “The ToBI annotation conventions”.<http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html>.

Beckman, Mary & Pierrehumbert, Janet (1986): “Intonational structure in Japanese and English”. *Phonology Yearbook* 3: 255-310.

Bloomfield, Leonard (1933): *Language.* Chicago: Chicago University Press.

Boersma, Paul & Weenink, David (2011): “Praat: doing phonetics by computer*”.* < http://www.fon.hum.uva.nl/praat/>.

Bolinger, Dwight (1951): “Intonation: levels versus configurations”. *Word* 7: 199-210.

Bolinger, Dwight (1958): “A theory of pitch accent in English. *Word* 14: 109-149.

Brazil, David (1985): *The communicative value of intonation in English.* ELR: University of Birmingham.

Brown, Gillian, Currie Karen & Kenworthy, Joanne (1980): *Questions of intonation.* London: Croom Helm.

Couper-Kuhlen, Elizabeth (1986): *An introduction to English prosody.* London: Edward Arnold.

Cruttenden, Alan (1986): *Intonation.* Cambridge: Cambridge University Press.

Crystal, David (1969): *Prosodic systems and intonation in English.* Cambridge: Cambridge University Press.

Estebas-Vilaplana, Eva (2009): *Teach yourself English pronunciation.* La Coruña: Netbiblo.

Gimson, Alfred (1980): *An introduction to the pronunciation of English.* London: Edward Arnold.

Goldsmith, John (1976): *Autosegmental phonology.* New York: Garland Press.

Gussenhoven, Carlos (2004): *The phonology of tone and intonation.* Cambridge: Cambridge University Press.

Halliday, Michael (1967): *Intonation and grammar in British English.* La Haya: Mouton.

Hockett, Charles (1958): *A course in modern linguistics.* Nueva York: Macmillan.

Kingdon, Roger (1958): *The groundwork of English intonation.* London: Longmans, Green and Co.

Ladd, Robert (1996): *Intonational phonology.* Cambridge: Cambridge University Press.

Leben, William (1976): “The tones in English intonation”. *Linguistic Analysis* 2: 69-107.
Liberman, Mark (1975): *The intonational system of English.* (PhD dissertation). MIT, Cambridge, MA.
O’Connor, Joseph & Arnold, G. F. (1973): *Intonation of colloquial English.* London: Longman.
Palmer, Harold (1922): *English intonation, with syntactic exercises.* Cambridge: Heffer.
Pierrehumbert, Janet (1980): *The phonetics and phonology of English intonation.* (PhD dissertation). MIT, Cambridge, MA.
Pierrehumbert, Janet & Steele, Shirley (1989): “Categories of tonal alignment in English”. *Phonetica* 46: 181-196.
Pike, Kenneth (1945): *The intonation of American English.* Ann Arbor: University of Michigan Press.
Schübiger, Maria (1958): *English intonation: its form and function.* Tübigen: Max Niemeyer Verlag.
Tench, Paul (1996): *The intonation systems of English.* London: Cassell.
Trager, George & Smith, Henry (1951): *An outline of English structure.* Norman, OK: Battenburg Press.
Wells, Rulon (1945): “The pitch phonemes of English”. *Language* 21: 27-40.
Wells, John (2006): *English intonation: an introduction.* Cambridge: Cambridge University Press.

CHAPTER 10. A SYSTEMIC FUNCTIONAL MODEL OF THE INTONATION OF CLAUSES IN ENGLISH

Paul Tench
Cardiff University

This chapter is intended to complement that of Eva Estebas Vilaplana's by exploring one particular model in detail as a means of producing a complete inventory of intonational forms for clauses in English spoken discourse. The basis of the description is the threefold set of systems of tonality, tonicity and tone which underlie the total integrated intonation of discourse in English at the level of the clause. Each system will be illustrated comprehensively, with examples of transcriptions of all the options that are possible at the level of the clause. All the examples are set out clearly, so that the reader can see the systems "at work" within discourse. (Features of intonation in paratones – phonological paragraphs – will not be included since they operate at a level above the clause. The best presentation of such higher levels of intonation in discourse is O'Grady, 2010.) The model is both systemic, in that it operates with sets of options; and functional, in that the set of options only exist to convey options of meanings in the minds of speakers/hearers. The model presented is based on that of Halliday's (1967, 1970), but with extensions from the work of other linguists who work within similar models, eg Tench (1990, 1996, 2005), Brazil (1997), Wells (2006).

1. Halliday's systemic-functional model of intonation

Halliday (1985) famously described intonation as the lexicogrammatical system *around* the clause in spoken discourse. Whereas other systems were described as operating either above or below the clause, intonation operates around the clause because it always works alongside it. Not only does it accompany each clause, it also provides an independent set of meanings and thus it forms an essential element of the contribution of each clause to the development of messages within a discourse.

But why focus on a systemic-functional model at all? Primarily because Halliday integrated intonation into a complete model of language in a way that was revolutionary in the 1960s. Previous studies of intonation in Britain, Europe and North America had largely presented intonation as a feature of language with little connection with grammar and discourse or even with the rest of phonology. Halliday deserves the credit that has often been accorded him for integrating disparate elements of language description. Pike (1967) had achieved something similar, but nevertheless his original full description of American intonation (Pike 1945) had been largely divorced from the rest of the language.

1.1 Intonation and phonology

Halliday did share with Pike the notion that there is a connection between intonation and other levels of phonology in the form of a hierarchy of phonological forms and units. They both developed the notion of levels or ranks within the hierarchy in a way that showed how the units of one rank functioned in the units of a higher rank, and how units were constituted of units from a lower rank. For example, different classes of phonemes had different functions within the higher rank of syllables, and syllables were shown to consist of the phonemes below them. Hence, this kind of hierarchy was dubbed “structure-function”. Different hierarchies were established for different languages to represent differences of phonological structures; in English, syllables provided the structure of stress/rhythm groups, but it was not anticipated that every language would require that particular rank. Furthermore, in English, the stress/rhythm groups supplied the elements of the structure of intonation units (called ‘tone groups’ in the original Hallidayan description). Thus intonation was formally and functionally integrated into the whole phonological system of English. (As it happens, Halliday did not recognize ranks above 'tone groups' as Pike did and others in the systemic-functional approach have since done.) The main point here is that intonation was declared to be an integral part of the phonology of English, as it would be in the phonology of any language. Previous descriptions had not recognized this in this formal way that is now widely accepted.

1.2 Intonation and grammar

Secondly, Halliday showed that intonation was intimately related to grammar, particularly the grammar of clauses. It was not, however, just the matter of the elementary connection between clause types and certain intonation patterns, eg falling intonation with declarative and imperative clauses, and rising intonation with 'polar' ('yes/no') interrogatives but falling with 'non-polar' ('*wh*-') interrogatives; this relates to the 'interpersonal' function of clauses, ie the communicative discourse functions. This latter function had been widely recognized before Halliday, of course; see Palmer (1922), Armstrong & Ward (1926), Kingdon (1958) and O'Connor & Arnold (1961), and also his contemporary, Crystal (1969); but Halliday showed also that there was no simple and direct correlation between clause type and discourse function. To take an obvious example: tagged clauses –with identical wording– take either a falling or a rising tone on the tag to indicate very different discourse contributions:

| Spain won the world cup | \didn't they | (an assertion to emphasize the expectation, perhaps, that they will win their next match)
| Spain won the world cup | /didn't they | (a query indicating the speaker's uncertainty about the proposition put forward)

Declarative clauses are most commonly accompanied by falling intonation, but might be accompanied by rising, or falling-rising intonation to indicate a different discourse function from statements. The intonational accompany-iment of all clause types is subject to a system of options for all manner of communicative functions.

However, intonational contrasts in clauses also indicate the 'ideational' function of language, presenting propositional content. Take, for instance the difference between 'defining' and 'non-defining' relative clauses:

| my friend who lives in Valencia | is married with two children ('defining' which friend, ie I have more than one friend)
| his wife | who comes from Wales | is a midwife ('non-defining', ie he does not have another wife who has to be distinguished from the other! But I am adding a piece of information that I believe to be relevant in the given discourse).

These are two relative clauses with different meanings despite identical wording, but marked by contrasting intonational forms; in this case by adding an extra intonation unit ('tone group') with the 'non-defining' type which simply inserts additional information. The upright bars (|) mark the boundaries of the intonation units in this paper, although Halliday adopted the double slash (//). On paper, the two types of relative clause may well be marked by punctuation differences, but this cannot be relied upon; in speech, the distinction is more or less consistently marked by contrastive intonation. For a full range of such contrasting syntactic functions indicated by intonation alone, see Halliday (1967), Crystal (1975), Tench (1996), Halliday & Greaves (2008).

1.3 Intonation and information structure

Thirdly, Halliday integrated intonation into a theory of information structure. Speakers stage their talk by producing a succession of pieces of information; one thing is stated first, then a second piece, followed by a third piece, etc. Each piece of information is presented in a separate unit of intonation ('tone group'), and typically each piece of information is contained within a single clause. A clause typically contains a subject and a predicate, ie something being talked about ('theme' or 'topic') and something being said about the 'theme'/'topic'. The predicate typically contains a verb with whatever complementation is required in the form of 'direct'/'indirect' object and/or adverbials. Thus, a clause expresses a representation of something happening or a state of affairs, ie a piece of information. It is thus not uncommon to find a high level of congruence between units of intonation, units of information and clauses (units of syntax), for example

| if I had the chance | I would visit Valencia | to visit my friends |

– three clauses, each with a separate piece of intonation and each accompanied by an intonation unit. This congruence of intonation, information structure and clause is usually known as 'neutral tonality' in Hallidayan terms.

However, speakers of English are not confined to neutral tonality; it is quite possible to divide a single clause into more than one intonation unit and, thereby, into more than one unit of information:

| I | would like to visit Valencia | (ie *I*, in contrast to others)

–two units within a single clause: the subject is cast as a separate piece of information from the rest of the message. There are many common cases of 'non-neutral' ('marked') tonality; see Halliday (1967), Crystal (1975), Tench (1996).

'Tonality' is the system that contains the options in the number of intonation units in relation to clauses, and thus represents speakers' management of the structure of information in their discourses. Speakers also have another system available to them, that of 'tonicity', ie the placement of the main prominence within an intonation unit. The most prominent element in an intonation unit is called the 'tonic', or the 'nucleus', and it is usually found at the end of a clause, on the final lexical item. In systemic-functional transcriptions it is usually indicated by underlining; for example:

| I would like to visit Valencia |

This positioning of the 'tonic' usually signifies that all the information in the unit is considered new information by the speaker; this is called 'neutral tonicity'. However, speakers are not confined to neutral tonicity, but can place the tonic elsewhere if they wish to identify one part as new and the rest as ‘old’ or ‘given’. For instance:

| I would like to visit Valencia |

–the tonic on *like* suggests that the speaker’s visiting Valencia is already known (‘old’/‘given’) and the only new piece of information is *like*. This ‘non-neutral/marked’ tonicity is an option available to the speaker to present information in a particular context. Similarly available is the placement of the tonic on a non-lexical item as in the following:

| I would like to visit Valencia | (the speaker’s desire to visit has already been mentioned, but now the speaker feels the need to emphasize it).

In addition to options in the systems of tonality and tonicity, the speaker has options in the system of tone to indicate the status of a piece of information. If the intonation contains a falling tone, the speaker indicates that they regard the information as being of major importance; with a rising tone, it is regarded as incomplete in itself; and with a falling-rising tone, the speaker indicates that an additional unspoken message should be understood:

| I would like to visit Va\lencia | (of major importance)
| I would like to visit Va/lencia | (incomplete without some other message following *and …*)
| I would like to visit Va\/lencia | (another message implied, eg “but not Madrid”)

The direct connection between intonation and information structure is one of Halliday’s great innovations in intonation theory. It has now become a standard feature in intonation description.

1.4 Intonation and attitude

The expression of attitude, or emotion, had long been associated with intonation, and as its primary function; this had been recognized by all linguists who had been involved in the description of intonation. Halliday recognized it too, but he added a fresh dimension to intonation theory by proposing 'neutral' forms associated with clause types. Palmer (1922), Pike (1945), Schubiger (1958), Kingdon (1958) and Crystal (1969) had all associated certain tone features with specific expressions of a range of attitudes, but none to the extent that O'Connor & Arnold (1961/1973) had; for them, each of ten intonation 'tunes', with a doubling of them all with 'emphasis', were applied to each clause type, expressing a different attitude or range of attitudes. For example, a low falling tone on the tonic of a declarative clause:

| Spain \\won |

expresses a "categoric, weighty, judicial, considered" attitude. If there had been no stressed syllable at the beginning (ie no 'head'):

| they \\won |

it would have meant "detached, cool, dispassionate, reserved, dull, possibly grim or surly" (O'Connor & Arnold 1973:113). Such a range can be quite bewildering, especially without reference to the lexical content of the clause. A high falling tone would, on the other hand, convey "a sense of involvement, light, airy", and so on for all ten 'tunes'. What O'Connor & Arnold failed to see was that their descriptors depended very much on the lexical content and the communicative function of the clause and its situational context.

Halliday developed a simpler system which involved 'secondary tones', which were variations in the pitch height of the 'primary tones' used for communicative functions and information structure. The system of secondary tones was called 'key' (not to be confused with Brazil's use of this term; see below), because it principally involved high or low versions of the primary tones. He claimed, for instance, that there is a plain, or 'neutral', version of a primary tone, which might have high or low alternatives ('secondary tones'); it is this variation in pitch which indicated 'key' and attitudinal expression. So, for example:

| Spain \won |

with a ‘neutral’ form of the falling tone in a declarative clause is a plain statement, with no attitudinal overtones, whereas

| Spain \won |

with a falling tone that started lower than the neutral starting point and accordingly finished lower was not just a plain statement, but a statement “with attitude”. Halliday frequently labelled this form as a ‘mild’ expression, but the exact attitude would depend on other factors such as lexis, communicative function and context. Uttered by an ardent Spanish football supporter, it might mean something like “Of course they won, what else would you expect?”, but uttered by a rival supporter, it might express disappointment. On the other hand, a high version – with a falling tone starting at a higher pitch – was described as ‘strong’:

| Spain \won |

Again, a declarative clause as a statement “with attitude”. Which attitude? That depends on the same factors as before: the Spanish football supporter this time might be expressing something like pride, or excitement; on the other hand, someone who expected Spain to lose might be expressing surprise. Whatever the actual emotion, it is stronger than a normal, plain, statement.

1.5 Intonation and systems

The distinctive feature of Halliday’s description of intonation was that he integrated intonation with other components of language and showed it to be as essential a component as any other in the working of language, with a specific range of meanings. The range of meanings were presented in the form of three systems – tonality, tonicity and tone – each with sets of options which were determined by their own range of meanings. Tonality indicates the speaker's management of the flow of information, with a two-item system: neutral, or non-neutral/'marked'. Tonicity indicates the speaker's perception of what constitutes new and old/'given' information; again, there is a two-item system: neutral, or non-neutral/'marked'. Tone indicates both information status and communicative function on the one hand ('primary tones') and key on the other ('secondary tones'). Halliday postulated 5 primary tones for English which were numbered as follows:

1. fall
2. high rise
3. low rise
4. fall-rise
5. rise-fall

and a combination of 1 and 3, and 5 and 3.

He postulated a system of secondary tones, which included variations of the primary ones at the tonic and in the section of the intonation unit that preceded the tonic (the 'pretonic'). There is not enough space here to set the whole system out, and in any case, the tone systems have been re-worked by others and will form the main part of this chapter.

What needs to be emphasized here is simply the observation that Halliday developed a comprehensive description of intonation in English that was totally integrated with the description of all the other components of language.

1.6 Intonation above the clause

Although Halliday noted the intonation patterns of specific sequences of clauses, he did not venture into higher reaches of intonation in discourse organization. Such higher units in language were deemed to be beyond the scope of grammar. However, other linguists did venture higher. Trim (1959) and Fox (1973) outlined the intonational features of pieces of spoken discourse above clause sequences in what became known as 'paratones', the spoken equivalent of 'para*graph*' in written discourse. (Tench (1990, 1996), in fact, referred to these features as exponents of 'phonological paragraphs'.) Pike (1967) also recognized their existence. Paratones mark larger units of information in the management of discourse, e.g. in the development of a topic. The phonetic features of paratones were identified by Crystal (1969, with his notion of 'subordination'), Lehiste (1980, 1982), Brazil (1997, who used the term 'key' to refer to them), Couper-Kuhlen (1986) and O'Grady (2010).

In the simplest terms, a paratone is marked by a relatively high level of pitch in the baseline of the first intonation unit, a progressive lowering of that baseline as the paratone proceeds and a relatively low level of pitch in its concluding intonation unit. Speakers can alter this sequence and thereby introduce rhetorical effects: a sudden rise in the baseline might indicate a highlighting of a piece of information; a maintenance of the baseline pitch might indicate that a piece of information should be interpreted as a re-stating of previous information; a sudden drop in the baseline might indicate the speaker's own comment on a piece of information. Thus, there is a system of baseline pitch levels which might either be 'neutral' as in typical progression in the development of the topic, or 'non-neutral'/'marked' where the speaker introduces rhetorical effects.

The relevance of the paratone to this chapter is that a recent innovation in the intonation systems at clause level in English has taken a paratone feature and incorporated it into the options available at clause level. This is the so called 'high rising terminal', or as it will be called in this paper, the 'raised rising tone'.

Intonation is also a characteristic of an even higher level of language, a language event itself. Recognizably different spoken language events, known as genres, have their own identifying 'prosodic composition'. News reading requires a particular style of speaking to be recognized as news reading, whereas ghost stories require a very different style. The various genres we recognize in a culture all have a distinctive sound: horse racing commentary, weddings, joke telling, stories to children, prayer, informal conversation, etc. The distinctive sound is the identifying 'prosodic composition' which includes intonation, as well as other paralinguistic features like speed of articulation, volume, timing and voice quality; see Crystal & Davy (1969). 'Prosodic composition' properly belongs to whole discourses, and it will not be considered further in this chapter except in the case of a certain type of discourse that consists of a single clause, namely, calling at a distance, eg “Dinner’s ready!”

2. Intonation systems at the level of the clause in English

We are now in a position to set out all the intonation possibilities for a single, short, simple, straightforward clause in English like *Spain won the World Cup*. It is a single clause and it is short, with only four stresses (^{1}*Spain* ^{2}*won the* ^{3}*World* ^{4}*Cup*); longer clauses with five or more stresses generally divide into two or more intonation units (eg ^{1}*Spain* ^{2}*won the* ^{3}*World* ^{4}*Cup* | *for the* ^{5}*first* ^{6}*time in their* ^{7}*history*). It is a simple clause without extensions like apposition (eg *Spain won the World Cup, the highest accolade in football*) or list items (eg *Spain won the World Cup, the European Cup, and so on*); such extensions are usually accompanied by additional intonation units, as additional pieces of information. It is also a straightforward clause in the sense it follows the default order of clause elements, eg with no marked theme (eg *The World Cup Spain won*) or clefting (eg *It was Spain who won the World Cup*); these marked structures are often also accompanied by additional intonation units.

2.1 Tone choices

We start with the range of tone choices for a clause with neutral tonality (ie one clause as one piece of information in one intonation unit), neutral tonicity (ie one piece of information all treated as new) and with a pretonic pitched at more or less the mid level of a speaker’s pitch range, but not necessarily absolutely level since there is no contrast in a slight wavering of pitch.

1 | Spain won the World \Cup |

This is the most neutral form possible; it is a plain statement with a neutral falling tone, which starts at a mid point in the speaker’s pitch range and falls;

it indicates a major piece of information, with no attitudinal marking. The 'pretonic' segment (*Spain won the World*) is typically around the mid point of pitch. This intonation pattern is Halliday's Tone 1, Brazil's *p* tone ('proclaiming'), etc. Variations of the tone indicate either a change of information status or an expression of attitude, as illustrated, for example in 2:

2 | Spain won the World \Cup |

The low version of the falling tone, starting from a mid-low point in the speaker's pitch range and falling slightly, indicates the speaker's attitude: a 'mild' tone suggesting that the information is totally expected (eg *Exactly what we predicted*). Halliday transcribes this pattern as Tone 1–.

3 | Spain won the World \Cup |

The high version, starting from a higher point in the speaker's pitch range and falling to at least the mid point or lower, indicates a different attitude: a 'strong' tone suggesting that the information might not have been expected, a sense of excitement, or surprise, etc (eg *They've done it; they've never done it before*). Halliday's transcription: Tone 1+.

4 | Spain won the World /\Cup |

This is a more intense version of the high fall; it is as if the speaker allows the voice to 'climb up' to the high pitch so as to emphasize the strength of feeling and in this way produces a rise-fall; it expresses a stronger attitude like amazement or sensing the magnitude of the information (eg *They didn't just beat Italy | they won the World /\Cup |*. Halliday's transcription: Tone 5; Brazil's *p+*.

5 | Spain won the World /\Cup |

This is an even more intense, more emotional version of the high-fall, pitched low in the speaker's pitch range, expressing what might popularly be expressed as the *wow* factor, a sense of being overwhelmed (eg *I can hardly believe it; isn't it fantastic!*). Halliday's transcription: Tone 5.

The exact phonetic range of each of these five tones will vary somewhat, but out of the whole spectrum of pitch possibilities, English speakers seem to be able to identify these five possible meanings; thus these five meanings can be treated as the phonological system of options for falling tones in declarative clauses functioning as statements as major information. Any other tone –a rise, fall-rise or a level mid– will have a different function.

A rising tone on a declarative clause will either indicate a different status of information, that is not complete in itself, or else will indicate a different communicative function, typically a question of some kind.

6 | Spain won the World /Cup |

This is how incomplete information is usually intoned; the rise, from low to about mid point in the speaker's pitch range, suggests that the speaker has not completed what they plan to say (eg …*and are champions of the world*). A different possibility is that the speaker repeats information that they believe the addressee already knows when the unit follows one with major information, eg (You probably \heard | that Spain won the World /Cup |); although the repeated information is 'given', the speaker may still want to draw attention to it as 'minor information'. Halliday's transcription in both cases is: Tone 3.

Questions are also typically produced with this very same rising tone, but usually with a polar interrogative clause:

6a | did Spain win the World /Cup |

But rising tones with declarative clauses with similar communicative functions are also possible, most notably with a high rising tone:

7 | Spain won the World /Cup |

The rise typically starts from about mid point and rises to a high point in the speaker's pitch range; the speaker appears to be questioning what they think they have heard, with strong feelings like disbelief or surprise (eg *Have I got that right?*). This is Halliday's Tone 2, which he normally associates with polar questions like 6a above, but it seems that for most questions of that kind, there is usually very little difference in the pitch movement from a simple rise; see Watt (1994).

But a speaker may express a very different attitude if the rise is a low version:

8 | Spain won the World /Cup |

This sounds like a very grudging response (eg *What more do they want?* or *I hope they don't win anything more*).

Fall-rises seem to have developed in English (Tench 2003) by combining the status of major information (the fall element) with the status of incomplete information (the rise element). If a person should say:

9 | Spain won the World \/Cup |

they would be providing major information, but with a hint that other information ought to be inferred by the addressee; it implies an unspoken extra piece of information which the speaker expects the other person to understand. If there was, for example, a dispute about the best football team in the world, the speaker who uttered 9 above might well be implying something like "That *proves* they are the best". Halliday's transcription: Tone 4; also Tone 2, which is related to focus in polar questions only; see Watt (1994) for the evidence that there is precious little difference in the pitch movements for Tones 4 and 2, and that it would be simpler to regard them as one and the same. Brazil transcribes this tone as Tone *r* ('referring').

Fall-rises, like the rise-falls, have a low pitched version:

10 | Spain won the World \/Cup |

with the same added sense of strong emotion (eg *What more evidence do you want that they are the best in the world?*). Halliday's transcription: 4.

A level mid tone is also possible:

11 | Spain won the World –Cup |

This suggests the routine listing of pieces of information that are commonly accepted in public. If someone wishes to assert that Spain has the best football team in the world, they might say, *Look,*

11a | Spain won the World –Cup | they won the –European Cup | they haven't been beaten for two –years | (*there is no better team in the world*)

For further discussion of the meanings of the mid level tone, see Tench (2003). In Tench (1996: 81) it was suggested that

> The mid-level tone may possibly establish itself as a separate contrasting tone, say in another generation's time … we may well be witnessing a change in this respect, especially as the older descriptions like Armstrong & Ward (1926) and Palmer (1922), did not include reference to a mid-level tone at all.

However, the occurrence of this tone is now so common, in a wide variety of contexts, that it is now clearly a separate option in the phonological system of current English intonation, although Halliday (1967) did not acknowledge it. And its use is not confined to misfortune as Tench (op. cit.) once suggested, as the current example shows! Brazil transcribes this tone as Tone *o* ('oblique').

Another new phenomenon in current English intonation is the 'high rising terminal'. It is a very clever device by which a speaker can provide major information and at the same time check the addressee's understanding of the importance of it.

12 | ↑Spain won the World /Cup |

If people, for example, were talking negatively about the country of Spain, a speaker who uttered 12 above might offer that piece of information and simultaneously be asking addressees if they appreciate the significance of it (eg *There are positive things to say about Spain as well*). There has been extensive literature on this relatively new phenomenon; see Tench (2003), who suggests that a more accurate term would be a ‘raised rise’ because the baseline of the intonation unit as a whole has been raised, as indicated by [↑] at the beginning of the unit , or often just the tonic itself:

12a | Spain won the World ↑/Cup |

Calling over a distance often requires a different kind of intonation pattern, as if to project the sound of the voice the appropriate distance.

13 | Spain won the World ¯–Cup |

Imagine a person wishing to provide this information in a challenging environment –distance, noisiness, etc– the tonic is first pitched level high and then level mid. A more typical scenario would be calling out

13a | ¯dinner’s –ready | or

13b | ¯Isa –belle |

In these cases, the second element, the mid level pitch, occurs with the final strong syllable, or the final weak syllable if there is no stressed syllable following the tonic.
There are thus 13 options of primary and secondary tones in an intonation unit.

2.2 Pretonic choices

The pretonic segment is the whole of an intonation unit from its beginning up to, but not including, the tonic syllable. It is often divided further into the ‘head’, which is the whole of the pretonic segment from its first stressed syllable, and the ‘pre-head’ which consists of unstressed syllables preceding the ‘head’. As indicated above, *Spain won the World* is the pretonic, which consists of a head which begins with the first stressed syllable, *Spain*; there is no pre-head in this instance. In *The national team won …*, *The* constitutes the pre-head as an unstressed syllable preceding the first stress, ^{1}*national …* (Likewise, the 'tonic segment' consists of the tonic syllable itself and any following syllables, which are called the 'tail'; for example, in *Spain won the World Cup*, there are no syllables following the tonic syllable, and so there is no 'tail' –the tonic segment consists solely of the tonic *Cup*. However, in *Spain are world champions*, the tonic segment consists of the tonic syllable, *cham-* and a tail, *-pions*.)

In English, there are a number of variations that operate as a system of secondary tone choices in the pretonic segment. Principally, they are: a low version and a high version; a falling version and a rising version, and 'stepping' and 'glissando' variations of both the falling and rising versions. All of these pretonic choices can accompany all the tones above, apart from the 'raised rise' with its fixed baseline; and so they create a vast increase in the number of intonation possibilities for a single, short, simple, straightforward clause.

The low pretonic (_) accompanies information that is deemed by the speaker to be old, ie 'given'. So, for example

14 | _Spain won the World \Cup |

would make sense in a context where *Spain won the World* was already known ('given'); for example, if someone claimed that Spain had won the World Crown, the reply might be *No* | *Spain won the World* (low pretonic) *Cup*.

The low pretonic can combine with a low falling tone, which expresses a mild attitude, 'expected' information. In a context similar to the foregoing, the reply would indicate that the speaker fully expected the addressee to have known this piece of information:

15 | _Spain won the World \Cup |

A speaker might take a stronger line, with a high fall:

16 | _Spain won the World \Cup |

They might take an even stronger line, and also one with a greater intensity of emotion:

17 | _Spain won the World /\Cup |

18 | _Spain won the World /\Cup |

The low pretonic can combine with a rise indicating incomplete or minor information; the context is the same as for 14 above, but the speaker wishes to add further relevant information:

19 | _Spain won the World /Cup | (*and …*)

It can combine with a rise indicating a question too:

19a | _did Spain win the World /Cup | (*or, was it, possibly, the World* Crown*?*)

A low pretonic can also combine with a high rise, like 7 above (*Have I go that right?*); the low pretonic seems to shift the emphasis onto *Cup*:

20 | _Spain won the World /Cup |

Halliday transcribes this as Tone −2, which is only associated with polar questions. The combination of a low pretonic with a low rise seems to suggest a lack of interest or concern:

21 | _Spain won the World /Cup |

Halliday transcribes this as Tone −3. The low pretonic can also precede a fall-rise, with the same kind of implication as 9 above, and with additional emotion, as 10 above:

22 | _Spain won the World \/Cup |

23 | _Spain won the World \/Cup |

The low pretonic regularly precedes the mid level tone and better represents the routine nature of the kind of listing of 11 above:

24 | _Spain won the World −Cup |

The low pretonic cannot precede the raised rise (12 above), but regularly accompanies the call of 13 above:

25 | _Spain won the World ¯ –Cup |

25a (| ¯dinner's –ready) | _ come and ¯get –it |

We have now presented all the possibilities with a low version of the pretonic, and most of these patterns can alternatively take a high version also. Whereas the low version suggests that the accompanying information is old ('given'), the high version draws extra attention to it, suggesting an insistence that the information is valid:

26 | ¯ Spain won the World \Cup |

The speaker adds a sense of insistence to the information, as if they were saying:

26a | ¯ I'm telling you that Spain won the World \Cup |

The speaker can choose to have the high pretonic before a low fall:

27 | ¯ Spain won the World \Cup |

This is identical to O'Connor & Arnold's 'Low Drop', which they categorized as being "categoric, weighty, judicial, considered" (op.cit.). The speaker can also choose to have the high pretonic before a high fall:

28 | ¯ Spain won the World \Cup |

This is identical to O'Connor & Arnold's 'High Drop' which they categorized as "sense of involvement, light, airy". Equally, the speaker can choose the high pretonic before the two forms of the rise-fall:

29 | ¯ Spain won the World ^Cup |
This is identical to O'Connor & Arnold's 'Jackknife', which they categorized as "impressed, awed, complacent, self-satisfied".

30 | ¯ Spain won the World ^Cup |

30 adds both the sense of insistence of the truth of the information and the intensification of the sense of awe.
The high pretonic can precede rising tones to indicate insistence of the truth of incomplete or minor information:

31 | ¯ Spain won the World /Cup |

This is identical to O'Connor & Arnold's 'Low Bounce', which they categorized as "soothing, reassuring", as if the speaker was saying *No need to worry*; this comment itself would typically be intoned in this way, as:

31a | ¯ no need to /worry |

Polar questions could be accompanied by this pattern, which suggests a more excited, more expectant way of asking the question in 6a above:

31b | ¯ did Spain win the World /Cup |

The declarative, as well as the polar interrogative clause can be accompanied by the choice of a high pretonic and a high rise to indicate the strong sense of excitement and expectancy either with a statement of disbelief or surprise (declarative) or a question (interrogative):

32 | ¯ Spain won the World /Cup |

32a | ¯ did Spain win the World /Cup |

This is identical to O' Connor & Arnold's 'High Bounce'. The high pretonic does not seem to combine with the low rise, but it could possibly combine with the mid level tone:

33 | ¯ Spain won the World –Cup |

This could suggest an excited way of stating routine lists (see 11 and 11a above). It is identical to O'Connor & Arnold's 'Terrace', which they categorized as "non-finality, without conveying any impression of expectancy"; the descriptor "non-finality" is relevant, but with the second descriptor they had in mind the kind of adverbial phrase in initial position in a clause, eg *In 2010* (*Spain won the Word Cup*):

33a | ¯ in twenty –ten | Spain won the World \Cup |

Finally, the high pretonic can combine with the 'call' tone, expressing a sense of excitement:

34 | ¯ Spain won the World ¯ –Cup |

34a (| ¯dinner's –ready) | ¯come and ¯get –it |

We have now considered the effect of both the low and high versions of the pretonic where the voice maintains a more or less level pitch, but English speakers readily react to pitch variation in the pretonic. Pretonic pitch might steadily fall or rise, or fall or rise stepwise or smoothly on each stressed syllable. We will illustrate each of these possibilities without tediously describing their application with each tone.

The effect of the gradual fall of pitch within the pretonic segment adds a sense of warmth on the part of the speaker towards the addressee, or at least a strong expectation of some kind of response (Tench 1996: 131). This gradual fall begins from a relatively high pitch on the first stressed syllable and finishes with a relatively low pitch before the tonic syllable; it is a ‘wide’ pretonic fall and is distinguishable from any slight unevenness in the pitch range described as ‘level’.

35 **|** \Spain won the World \Cup **|**

Notice the symbol (\) is ‘wide’ and is not accompanied by a tonic syllable, ie it represents a wide falling pitch associated with the pretonic. The speaker is encouraging a sceptical addressee to think positively, and does so more strongly with a high tone:

36 **|** \Spain won the World \Cup **|**

The wide falling pretonic (‘warm relationship’) does not seem to combine with a low fall (‘mild’ attitude with expected information) or with rise-falls (‘strong feeling’ would be redundant), but it readily accompanies fall-rises:

37 **|** \Spain won the World \/Cup **|**

Halliday claims in fact that the fall-rise regularly takes this wide falling pretonic and describes the combination as the usual pattern for Tone 4. The fall-rise implies an extra, unspoken, message that the addressee is expected to understand (see 9 above).

It can also combine with the rising tone for incomplete information:

38 **|** \Spain won the World /Cup **|**

and with questions:

38a | \did Spain win the World /<u>Cup</u> |

and especially with statements expressing disbelief and questions with a high rise:

39 | \Spain won the World /<u>Cup</u> |

39a | \did Spain win the World /<u>Cup</u> |

In fact, Halliday sees the pattern of 39a as the typical way of asking questions. The wide falling pretonic does not readily combine with the low rise as the meanings of each component are not compatible.
The wide rising pretonic begins with a relatively low pitch on the first stressed syllable and gradually rises to a high pitch before the tonic. It seems to suggest that the speaker is appealing to the addressee to believe the information:

40 | /Spain won the World \<u>Cup</u> |

Combined with a high fall, it suggests a protest, which is a stronger appeal to the addressee:

41 | /Spain won the World \<u>Cup</u> |

Combined with a rise-fall, it suggests an appeal to believe what the speaker strongly asserts:

42 | /Spain won the World /\<u>Cup</u> |

The wide rising pretonic can combine with the fall-rise to produce an appeal to accept the extra, unspoken, implication, (eg *so they must be the best in the world*):

43 | /Spain won the World \/<u>Cup</u> |

It can also combine with rising tones, for incomplete information:

44 | /Spain won the World /Cup |

and with the high rise particularly:

45 | /Spain won the World ˊCup |

increasing the expression of attitudes like disbelief (eg *I don't believe you!*). Again, it does not seem to combine with the low rise.
The stepping and glissando pretonic pitch forms add, respectively, emphasis and forcefulness, whether they progress downwards or upwards in pitch (Tench 1996: 132-4). In the case of stepping pretonics, a drop, or a lift, in pitch accompanies each stressed syllable in the pretonic.

46 |¯Spain −won the _World \Cup |

This is the more emphatic way of saying 35 above.
47-50 are similarly the more emphatic ways of saying 36-39 above. Rising steps are represented by:

51 | _Spain −won the ¯World \Cup |

which is the more emphatic way of saying 40 above.
52-56 are similarly the more emphatic ways of saying 41-45 above.

In the case of glissando pretonics, a wide fall, or rise, accompanies each stressed syllable in the pretonic:

57 | \Spain \won the \World \Cup |

Notice that the symbol (\) accompanies each stressed pretonic syllable as distinct from the tonic syllable (marked by underlining). Glissando pretonics are even more forceful ways of expressing information; it is a way of insistently asserting what the speaker believes to be true, typically in a contest of ideas. Example 57 represents the forceful expression of 35 above. Examples 58-61 are similarly the more forceful ways of saying 36-39. Halliday states that the pattern in

62 | \Spain \won the \World ∨Cup |

would, in fact, be the most typical pattern for the low fall-rise (his Tone 4). Rising glissando pretonics match the stepped rising pretonics in the same way with the extra degree of forcefulness:

63 | /Spain /won the /World \Cup |

Examples 64-68 represent 41-45 with the more forceful expression. Halliday maintains that the patern of 69 below – rising glissando pretonic with the low version of the rise-fall tone (his Tone 5) is the usual pattern:

69 | /Spain /won the /World ∧Cup |

We have now noted 69 ways of intoning one single, short, simple, straightforward clause, representing the enormous choice that is open to a speaker of English. The system available can be set out as follows:

```
                          mid      \_                ┐
                ┌fall     low      \ _               |
                |         high     \ _               |
                |         rise-fall high  ^\_        |
                |                   low   ∧ _        |
                |                                    |
                |         mid      /_                |
                ├rise     low      /_                |
                |         high     /_                ├ +/- pretonic choice
tone choice     ┤                                    |
                ├mid level       –_                  |
                |                                    |
                ├fall-rise mid   V_                  |
                |           low    v_                |
                |                                    |
                ├call                  ‾ –_          ┘
                |
                └ raised rise    ↑/_
```

The network shows the primary tones: fall, rise, mid level, fall-rise and call and their secondary tone variations where they exist (none for mid level or call). Each of these can enter into the pretonic choice, although not all pretonic possibilities are relevant for each tone. The ‘raised rise’ is exceptional, since the baseline is fixed. The symbols appear before a token underlining which represents the position of the tonic.
The pretonic choices appear below:

```
                   ┌mid  | (not marked)
                   ├low  | _
                   ├high | ¯
pretonic choices  ┤
                   ├falling ┬ wide        | \ ; | /
                   |         ├ stepping    | ¯ - _ ; | _ - ¯
                   └rising   ┴ glissando   | \ \ \ ; | / / /
```

The pretonic network shows the choices of mid, low and high level pretonics and the threefold choice with both falling and rising pretonics. Mid level is not marked, as being the ‘neutral’ version. The upright bar (|) shows the position of the beginning of the intonation unit. It must be remembered that not all pretonic choices occur with all tonic choices.

2.3 Tonicity choices

Neutral tonicity has the tonic syllable within the final lexical item; it usually indicates that the whole of the information within the intonation unit is being treated as new, or ‘fresh’ – the focus of information is ‘broad’, extending over the whole unit. If only the final lexical item itself is new, the rest before being old or ‘given’, then an English speaker deploys the low level pretonic to signal ‘givenness’, as already presented:

14 **|** _Spain won the World \Cup **|**

But it is possible to treat any other lexical item in the unit as new and the rest as old, e. g.

70 | Spain won the \World Cup |

This statement could only be made in a context where Spain winning a cup is already known. The focus of information in cases like 70 and 14 is said to be ‘narrow’. Likewise

71 | Spain \won the World Cup |

has ‘narrow focus’ on *won*, in the context of an exchange of messages that assume that Spain did something in the World Cup tournament. And

72 | \Spain won the World Cup |

has ‘narrow focus’ on *Spain*, in the context where winning the World Cup is already known.

It is also possible to place the tonic on a grammatical item, ie a non-lexical item. In the current example, only *the* is grammatical; it is possible to focus on it, as

73 | Spain won \the World Cup |

as if the 2010 World Cup is the only one that counts or is the most significant to date (*because Spain won it!*). Focus on a grammatical item typically indicates a contrast, such as

73a | England \has won the World Cup |

meaning that it is not the case that England never wins it! It is beyond the scope of this chapter to outline all the contrasts with grammatical items; these two examples must suffice to show that grammatical ‘narrow focus’ is possible. Whether tonicity falls on a non-final lexical item or on a non-lexical item, it is called ‘marked’ (or non-neutral’) tonicity.

It would be tedious beyond bearing now to list all the tone and pretonic choices for each of 70, 71, and 73, but, of course, they are all possible; such a list would increase the total number of possibilities by 3 times 69 (= 207). There is no pretonic in 72, and so the choices would be limited to 13. This would make a grand total of 289 (= 69+207+13) potential choices to this one single, short, simple, straightforward clause!

The network of tonicity choices is relatively simple:

```
                         ┌neutral   ┌broad focus
                         |          └narrow focus (with low pretonic)
tonicity choices        ┤
                         |          ┌non-final lexical item
                         └marked    └non-lexical item
```

2.4 Tonality choices

Neutral tonality is the congruence of one piece of information as one unit of intonation in one whole clause. Speakers manage the flow of information in their discourse by segmenting it into pieces and conveying each piece in one unit of intonation. Sometimes pieces of information get distributed into long and complicated clauses which require a succession of intonation units, a hint of which appeared in the introduction to Section 2 above. A succession of intonation units in a single clause, however short or long and complicated it may be, constitutes (non-neutral) ‘marked tonality’. All the examples given so far in Subsections 2.1 to 2.4 have maintained neutral tonality – one clause with one unit of intonation representing the speaker’s management of one piece of information. But a speaker has a choice between neutral and marked tonality even in that clause:

```
                         ┌neutral
tonality choices        ┤
                         └marked
```

Is it possible to segment a single, short, simple, straightforward clause like

Spain won the World Cup? Yes, as we did in Subsection 1.3 above; it would result in:

74 | Spain | won the World Cup |

As we explained in the introduction to Subsection 1.3, the clause is the natural unit of grammar to present an utterance about something happening, or a state of affairs. The subject of the clause typically comes first and is usually the element for presenting the 'theme' or 'topic' of a piece of information (what is being talked about), and the predicate the habitual element for presenting the 'rheme' of that piece of information (what is being said about the theme). If the subject element does indeed come first, as it has done in this example, the 'theme' is neutral; but the 'theme' becomes marked if anything but the subject comes first, eg *The World Cup Spain won*; *In 2010 Spain won the World Cup*; or *Unfortunately for Italy Spain won the World Cup*; etc. In such cases, the 'theme' being the starting point of the clause is not the subject and so it is 'marked'. 'Marked theme' almost always has an intonation unit to itself by virtue of its role as a separate piece of information, and will, therefore, almost always produce an instance of 'marked tonality', since more than one intonation unit will accompany the clause. The examples above – all single clauses – would segment as follows:

The World Cup | *Spain won*;
In 2010 | *Spain won the World Cup*;
Unfortunately for Italy | *Spain won the World Cup*:

in each case, one clause with two intonation units. (There are occasionally cases of 'marked theme' being incorporated into the one unit of intonation when the theme is very short and not strongly stressed, eg *Then/So Spain won …*)

So, how can 74 be split into two intonation units? After all, the subject comes first and acts as the theme, which is neutral. But the speaker has the choice of highlighting the theme for the sake of emphasis or contrast (Tench 1996: 83-4), and this is what happens in this case. It is a case of a neutral theme being highlighted. But when could a speaker utter the two unit version as in 74 above? Imagine someone maintaining that Brazil, for instance, is in fact the best footballing nation in the world, someone might point out *But*

74 | Spain | won the World Cup |

as two pieces of information: first, by changing the theme to *Spain*, and second, by providing the information about Spain to contradict the claim *that Brazil is, in fact, the best footballing nation in the world* . Neutral tonicity in each unit and the most neutral tone with each tonic would produce:

74a | \Spain | won the World \Cup |

But then there are all the possibilities of tone variations, eg a low fall in each unit to express a 'mild' sentiment like *as everybody knows*:

75 | \Spain | won the World \Cup |

or a 'strong' sentiment with a high fall, perhaps with a touch of anger (76), a more intense statement as showing amazement at the original claim with a rise-fall (77), or a more emotionally charged response with a low pitched rise-fall (78):

76 | \Spain | won the World \Cup |

77 | ^Spain | won the World ^Cup |

78 | ^Spain | won the World ^Cup |

In all these cases, there is a matching of tone in the two units. Although this is very common, it is not obligatory. A common sequence is a rise indicating 'incomplete' or 'minor' information, followed by a fall indicating what the speaker intends as the 'major' piece of information:

79 | /Spain | won the World \Cup |

Another common sequence is the opposite, where a fall indicates the major information followed by a rise indicating ‘minor’:

80 | \Spain | won the World /Cup |

Perhaps the most common sequence with a highlighted theme is fall-rise, which draws particular attention to the theme, followed by a fall for the major information (81), or the low variety to express greater emotional involvement (82):

81 | \/Spain | won the World \Cup |

82 | \/Spain | won the World \Cup |

Another possibility is a mid level tone for simply noting a change of theme, followed by a fall:

83 | −Spain | won the World \Cup |

A succession of rises suggests that the speaker wishes to check what they have heard (84) and more strongly (85):

84 | /Spain | won the World /Cup | (*I didn’t know that*, for example)

85 | /Spain | won the World /Cup | (*I don’t believe it!*, for example)

A succession of low rises suggests an unhappy reaction:

86 | /Spain | won the World /Cup |

The ‘raised rise’ (or, the so called ‘high rising terminal’) would also work; this would indicate that the speaker is providing two pieces of new information and asking the addressee to consider the significance of them:

87 | ↑/Spain | won the World ↑/Cup |

It is also possible to use the 'call' tone (see 13 above) as a kind of taunt (see Pike 1945) in response to the claim that *Brazil is the best footballing nation in the world*:

88 | ¯ –Spain | won the World ¯ –Cup |

All these variations mirror the choice of tone in Subsection 2.1. It would again be tedious beyond bearing to list the eight possible pretonic variations in the second unit for 74-88: low, and high, level pretonics, falling and rising pretonics that are either wide, stepping or glissando. Eight pretonic types for fifteen cases of marked tonality amount to 120 potential possibilities. However, some are more likely than others; here is a selection:

89 | \Spain | ¯ won the World \Cup | (high level pretonic, 'insistent')

90 | \Spain | _ won the World \Cup | (low level pretonic, 'given' information)

91 | \Spain | ¯ won the World \Cup | (O'Connor & Arnold's 'Low Drop'; as 27 above)

92 | \Spain | ¯ won the World \Cup | (O'Connor & Arnold's 'High Drop'; as 28 above)

93 | \Spain | ¯ won the World ^Cup | (O'Connor & Arnold's 'Jackknife'; as 29 above)

94 | \Spain | ¯ won the World /Cup | (O'Connor & Arnold's 'Low Bounce'; as 31 above)

95 | \/Spain | ¯ won the World \Cup | (fall-rise on *Spain*; otherwise as 92 above).

96 | \/Spain | \won the World \Cup | (fall-rise on *Spain*; falling pretonic; high fall)

97 | \/Spain | ¯won the −World \Cup | (fall-rise on *Spain*; stepping pretonic; high fall)

98 | \/Spain | /won the /World \Cup | (fall-rise on *Spain*; glissando pretonic; high fall)

99 | ↑/Spain | _ won the World ↑/Cup | (‘raised rises’; with low pretonic to throw emphasis on *World Cup*).

100 | −Spain | /won the World ¯ −Cup | (mid level on *Spain* notes a change of theme; rising pretonic indicates ‘appeal’; with ‘call’ as a taunt)

3. Conclusion

We have now considered exactly 100 ways of intoning a single, simple, short, straightforward clause in English, while noting that, technically, there are hundreds more. We add the 120 possibilities noted in Subsection 2.4 to the 289 noted in 2.3. If we further account for the ‘marked tonicity’ possibilities for the second unit of 74, eg *Spain* | *won the World Cup*, with the various pretonic variations, we multiply excessively the 409 already noted. The potential of the systems is vast, much vaster than many other approaches, eg Gussenhoven (2004) suggest.

We have presented intonation as an intrinsic part of the phonological hierarchy, above the level of rhythm/stress groups in English, but below the level of paratones and other, higher, segments of discourse. A unit of intonation typically embraces a whole clause, which is the grammatical unit for expressing happenings and states of affairs; and so it contains a single piece of information. That information is subject to various discourse (or ‘communicative’) functions like statements, questions, commands, etc. The tone system is used to express such discourse/communicative functions, but it also expresses the status of a piece of information and a range of attitudes. Pretonic choices express a wider range of attitudes. Tonicity is the system that a speaker uses to manage the focus of information, whether it is ‘broad’ or ‘narrow’. Tonality is the system that a speaker uses to control the flow of information by segmenting discourse into discrete units of information. These three systems of intonation represent the choices that are available to a speaker of English, clause by clause. The purpose of this chapter has been to show the vast potential of the three systems in a way never before attempted.

REFERENCES

Armstrong, L C & Ward, Ida C. (1926): *A handbook of English intonation.* 2nd ed. Cambridge: Heffer.

Brazil, David (1997): *The communicative value of intonation in English.* Cambridge: CUP.

Couper-Kuhlen, Elizabeth (1986): *An introduction to English prosody.* London: Edward Arnold

Crystal, David (1969): *Prosodic systems and intonation in English.* Cambridge: CUP.

Crystal, David (1975): *The English tone of voice.* London: Edward Arnold

Crystal, David & Davy, D (1969): *Investigating English style.* London: Longman.

Fox, Anthony (1973): "Tone-sequences in English". *Archivum Linguisticum* IV: 17-26.

Gussenhoven, Carlos (2004): *The phonology of tone and intonation*. Cambridge: CUP.

Halliday, Michael, A. K. (1967): *Intonation and grammar in British English.* The Hague: Mouton.

Halliday, Michael A. K. (1970): *A course in spoken English: intonation.* Oxford: OUP.

Halliday, Michael, A. K. (1985): *An introduction to functional grammar.* London: Edward Arnold

Halliday, M A K & Greaves W. S. (2008): *Intonation in the grammar of English*. London: Equinox.

Kingdon, Roger (1958): *The groundwork of English intonation.* London: Longman.

Lehiste, Isle (1980): *Phonetic characteristics of discourse.* Tokyo: Acoustical Society of Japan.

Lehiste, Isle (1982) "Some phonetic characteristics of discourse". *Studia Linguistica* 36/2: 117-130.

Palmer, Harold E. (1922): *English Intonation (with systematic exercises).* Cambridge: Heffer.

Pike, Kenneth L. (1945): *The intonation of American English.* Ann Arbor: University of Michigan Press.

Pike, Kenneth L. (1967): *Language in relation to a unified theory of the structure of human behavior.* 2nd ed. The Hague, NL: Mouton.

O'Connor, J. D. & Arnold, G. F. (1961/1973): *Intonation of colloquial English.* London: Longman.

O'Grady, G. (2010): *A grammar of spoken English discourse.* London: Continuum.

Schubiger, Maria (1958): *English intonation, its form and function.* Tübingen: Niemeyer.

Tench, Paul (1990): *The roles of intonation in English discourse.* Frankfurt aM: Peter Lang.

Tench, Paul (1996). *The intonation systems of English*. London: Cassell.

Tench, Paul (2003): “Processes of semogenesis in English intonation”. *Functions of Language* 10: 2
Tench, Paul (ed.) (2005): *Intonation: ways and meanings* (with CD). IATEFL: *Speak Out! 34.*
Trim, John L. M (1959): “Major and minor tone groups in English”. *Le Maître Phonétique* 112: 26-9.
Watt, D. L. E (1994): *The phonology and semology of intonation in English.* Bloomington: Indiana University Linguistics Club Publications.
Wells, John C. (2006): *English intonation: an introduction.* Cambridge: CUP.

CHAPTER 11: CONNECTED SPEECH: PRONUNCIATION OF WORDS IN CONTEXT

Sylvie Hanote
University of Poitiers

1. Introduction

Speech is a continuous stream of sounds with no clear-cut borders between them. Based on the work of linguists such as Jones (1957), Ladd (1980, 1996), Cruttenden (1997, 2001), Deschamps *et alii* (2004), Roach (2009) and Wells (2006), this chapter on connected speech concentrates on the phonetic phenomena observed "when the phonetic substance of lexical units is incorporated in a spoken continuum" (see, for instance, Deschamps *et al*. 2004: 24). These phenomena are also called phonotactic processes and they may occur within words or between words in contact.
This chapter is divided into five sections, each one dealing with one specific process: 1) rhythm and the notions of strong and weak syllables, stress-shift and vowel reduction; 2) assimilation (progressive and regressive), affecting the place of articulation, the manner of articulation or the voicing; 3) elision, especially contextual elisions affecting consonants in cases of heavy consonant clusters, or vowels leading to syllabic compression; 4) linking either by a final consonant or by an intrusive glide ([j], [w] or [r]); 5) prominence, where various acoustic correlates of prominence in speech (variations of intensity and fundamental frequency, duration of vowels and/or of initial consonants, change in the rhythm, presence of a glottal stop) are studied in relation to the preceding phenomena described.
Each process is described at the phonetic level in relation with the articulation of sounds in the speech chain and the influence contiguous sounds may have on each other. The phonetic description is visually

illustrated through spectrographic representations[1] of authentic English speech sequences recorded from the radio. Then, each process is analyzed at the phonological level in relation with the structure of the English syllable.

2. Rhythm and stress

English is often described as being a stress-timed language and English speech combines strong and weak syllables. The alternation between weak and strong syllables is even one of the most fundamental features of English phonology and it differentiates English from other languages such as French, for instance, which is said to be a syllable-timed language. The idea that stressed syllables (including pitch accent together with lexical stress) govern the rhythm of English is indeed quite common. Some authors assume that there is an equal amount of time between two stressed syllables (theory of 'isochrony') but the notion of isochrony has not been instrumentally proved and many authors claim that there is rather a "tendency" to isochrony.

In the following sentence (example (1)), it is highly probable that the beats will occur at regular intervals of time:

(1) The vast 'bulk have been 'killed by sec'tarians and 'terrorists. (Tony Blair, Interview on BBC Radio 4)

In this sentence, there is an alternation of stressed and unstressed syllables and unstressed syllables undergo reduction whether within a word (as in *terrorists* for instance) or between words in contact in connected speech (grammatical / function words are generally reduced in an utterance) as shown below in the phonemic transcription of the sentence[2] :

---

[1] The acoustic analyses are carried out with *Praat*, an open-source software tool developed by P. Boersma and D. Weenink (*The Institute of Phonetic Sciences*, Amsterdam): http://www. fon. hum.uva.nl/praat/.

[2] It should be pointed out that other phonotactic processes are likely to occur in this sentence, especially the elision of consonants (see round brackets). We shall deal with this process in section 4 of this chapter. Furthermore, it will be interesting to compare this phonemic

[ðəˌvɑːs(t)ˈbʌlkə(v)binˈkɪl(d)baɪsekˈteəriənzənˈterərɪsts]

It should be added that the more rapid the speech rate, the greater the tendency to reduction of function words. The various realizations (strong and weak forms) of some function words are presented in the following tables.

Tables 1 and 2. Full and reduced forms of some function words

		Full form	Reduced form
Prepositions	at	æt	ət
	for	fɔːʳ	fəʳ[3]
	from	frɒm	frəm
	of	ɒv	əv
	per	pɜː	pəʳ[3]
	to	tuː	tu + [V] / tə + [C]

Conjunctions	that	ðæt	ðət
	as	æz	əz
	than	ðæn	ðən
	and	ænd	ən(d)/ ən / n̩[4]
	but	bʌt	bət
	or	ɔː ʳ	ə ʳ (rare)[5]

transcription to the actual pronunciation of this sentence by the speaker recorded on BBC Radio – see section 6 of this chapter.

[3] Linking [r].

[4] The syllabic nasal will be used in collocations such as *Rock'n Roll.*

[5] The reduced form of *or* is only rarely used except in set phrases such as *two or three.*

Pronouns	me	miː	mi
	you	juː	ju + [V] / jə + [C]
	he	hiː	hi / i
	him	hɪm	hɪm / ɪm
	his	hɪz	hɪz / ɪz
	she	ʃiː	ʃi
	her	hɜː^{r}	ɜː^{r} \ hə^{r} \ ə^{r}
	we	wiː	wi
	us	ʌs	əs \ s (*let us* / *let's go* : [lets'gəʊ])
	they	ðeɪ (rare)	ðe
	them	ðem	ðəm \ əm \ ðm̩
	their	ðeə^{r}	ðər
	who	huː	hu
	whom	hʊm	hʊm
	that	ðæt	
	some	sʌm	

Determiners	his	hɪz	hɪz / ɪz
	her	hɜː^{r}	ɜː^{r} \ hə^{r} \ ə^{r}
	our	aʊə^{r}	ɑː^{r}
	your	jɔː^{r}	jə^{r}
	a, an	eɪ/	ə + [C] \ ən + [V]
	the	ðiː	ði + [V] \ ðə + [C]
	any	'eni	əni (rare)
	some	sʌm	səm / sm̩
	such	sʌʧ	səʧ

Table 3. Auxiliaries and modals

Auxiliaries		Full form	Reduced form
Be	Be	ˈbiː	bi
Present	I am / I'm	(ˈ)aɪ ˈæm	aɪm
	You are / you're	(ˈ)ju(ː) ˈɑːʳ	jɔː juə / jə[6]
	He is / he's	(ˈ)hi(ː) ˈɪz	(h)iz
	She is / she's	(ˈ)ʃi(ː) ˈɪz	ʃiz
	It is / it's	(ˈ)ɪt ˈɪz	ɪts
	We are / we're	(ˈ)wi(ː) ˈɑːʳ	wiəʳ
	They are / they're	(ˈ)ðe(ɪ) ˈɑːʳ	ðeəʳ[7]
	There is / there's	ˈðeər ˈɪz	ðeəz
	There are / there're	ˈðeər ˈɑːʳ	ðərə
Preterit	Was	wɒz	wəz
	Were	wɜːʳ	wəʳ
Have	Have	hæv	həv /əv / v[8]
Present	Has	hæz	həz / əz / z
	I have / I've	(ˈ)aɪ ˈhæv	aɪv
	He has / He's	(ˈ)hi(ː) ˈhæz	hiz
	She has / She's	(ˈ)ʃi(ː) ˈhæz	ʃiz
	It has / It's	(ˈ)ɪt ˈhæz	ɪts
	We have / we've	(ˈ)wi(ː) ˈhæv	wiv
	You have / you've	(ˈ)ju(ː) ˈhæv	juv
	They have / they've	(ˈ)ðe(ɪ) ˈhæv	ðeɪv
Preterit			
	Had	ˈhæd	həd / əd / d[9]
	I had / I'd	(ˈ)aɪ ˈhæd	aɪd
	He had / he'd	(ˈ)hi(ː) ˈhæd	hid

[6] The reduced form is fairly unusual in Received Pronunciation (see Wells, 2000: 866).

[7] [ðer] in GenAm. In GenAm., there is also a weak form [ðᵊr], but not in RP (Wells, 2000: 776).

[8] Used only after a vowel.

[9] The contracted weak form [d] is used mainly after a vowel (and is often written *'d*).

Modals	can shall could should will would / 'd cannot / can't shall not / shan't could not / couldn't would not / wouldn't will not /won't do not / don't	kæn ʃæl kʊd ʃʊd wɪl wʊd 'kæ(n)'nɒt 'ʃæl'nɒt 'kʊd 'nɒt 'wʊd 'nɒt 'wil 'nɒt 'du: 'nɒt	kən / kn̩ ʃl̩ kəd (only occasional) ʃəd (only occasional) wl̩ / l̩ ('ll) wəd(occasionally) / d (kə'nɒt) / 'kɑ:nt 'ʃɑ:nt 'kʊdn̩t 'wʊdn̩(t) wəʊn(t) dəʊn(t)

Furthermore, the speech rate and the reduction of forms greatly depend on the speaker and his/her purpose. If the context of the previous example is broadened (see example (2)), one can see that complete isochrony is probably not respected: some function words are emphasized[10] (*of* in the prepositional phrase *of those* and the auxiliary *have* in *have died*) and there are parentheses[11] (*this is based on the international Red Cross figures* ; *and there's far too many of course*) which are likely to be said faster and at lower pitch than in the rest of the utterance:

(1) One of the things I do in the book is to correct some of the… the… the… the really very exaggerated figures as to what's happened in the conflict and draw attention –this is based on the international Red Cross figures– um draw attention to the fact that *of* those that *have* died – and there's far too many of course –the vast bulk have been killed by sectarians and terrorists. (Tony Blair, Interview on BBC Radio 4).

Thus, although there is clearly a tendency to isochrony in English and to a reduction of function words in classic normal speech, isochrony is not an exact rule and it largely depends on the speaker and his/her purpose.

[10] See italicized words. The notion of prominence and its acoustic correlates will be dealt with in detail in section 6 of this chapter.

[11] Between dashes in the example.

The alternation of strong and weak syllables also sometimes justifies the displacement of primary stress in a word. This process is called 'stress-shift'. For instance *ˌmiddle-ˈclass* (adj.) when isolated follows the /201/ stress-pattern but when used as a compound adjective in *a middle-class citizen* / *a middle-class life* for instance, it undergoes stress-shift and is stressed as follows: *a ˌmiddle-class ˈcitizen* / *a ˌmiddle-class ˈlife*. Similarly, in the Noun Phrase *a Chinese restaurant* where the adjective *ˌChiˈnese* is followed by the noun *ˈrestaurant* stressed on its first syllable, the adjective will undergo stress-shift and will be stressed as follows: *a ˌChinese ˈrestaurant*.

3. Assimilation

Assimilation is one of the most common processes in the spoken chain. In natural connected speech, especially casual speech, the pronunciation of one phoneme may influence the pronunciation of other phonemes that are directly in contact with it, whether in the same syllable or in two contiguous syllables. As Cruttenden states (2001: 281), speech must be seen as an articulatory continuum rather than as the juxtaposition of discrete units independent from one another, and there is mutual influence of phonemes in contact in the speech chain, sometimes leading to the merging of certain qualities of the phonemes[12].

The process of assimilation consists in the spreading of a phonetic feature to the right (progressive assimilation) or, more often, to the left (regressive assimilation) of its departure point (see also Deschamps *et alii*, 2004: 27). Sometimes, a fusion (or coalescence) of phonemes may also take place. The feature spreading may affect the place of articulation, the manner of articulation or the voicing.

3.1 Regressive assimilation

Regressive assimilation is also called anticipatory assimilation as the features of one phoneme are anticipated in the articulation of the preceding

[12] Some authors consider assimilation a case of co-articulation, others clearly distinguish the two processes. For further detail on this debate, see Pavlik (2009: 12).

phoneme. For example, given two segments (A and B), segment A (the first one of the two) is "affected" by the following segment (segment B). The direction of the process is thus backwards: A <– B. That is why it is also called 'backward assimilation'.

Regressive assimilation is the most common case of assimilation; it is very frequent in all languages and it may affect the three articulatory features of place, manner and voicing but only one feature can be affected at a time.

In *question mark*, for instance, the place of articulation is affected: the final nasal [n] in *question* assimilates to the following consonant ([m]) and tends to be labialized: [n] is thus changed into [m] and the compound is pronounced ['kwesʧə(m)mɑːk][13]. The same kind of process occurs in *Great Britain*, being pronounced [ˌgreɪp'brɪtn̩] as the apico-alveolar [t] of *Great* is likely to be changed into the bi-labial [p] because of the following [b]; for the same reasons, *football* will be pronounced ['fʊpbɔːl], *shortcut* ['ʃɔː(k)kʌt], *Commonwealth* ['kɒməmwelθ], or *rising prosperity* ['raɪzɪmprɒ'sperəti]. The process of assimilation is particularly frequent in compounds or in collocations but it is also frequently found between any words in contact in normal colloquial speech.

It should be noted that, sometimes, more than one segment may be affected by the process of assimilation as in *can't be* ['kɑːm(p)biː] where the final apico-alveolar plosive [t] in *can't* turns into the bilabial plosive [p] before the bilabial plosive [b], which in turn has an influence on the preceding apico-alveolar nasal [n] which is likely to change into the bilabial nasal [m]. The same kind of "double" process occurs in the compound *handbag* or in the phrase *won't go* leading to the following pronunciations: *handbag* ['hæm(b)-bæg] and *won't go* [wəʊŋ(k)'gəʊ].

Another feature that may be affected by regressive assimilation is the manner of articulation as in *good night* where the apico-alveolar plosive [d] changes into the apico-alveolar [n] because of the following nasal. *Good night* will then be pronounced [(ˌ)gʊ(n)'naɪt].

---

[13] The round brackets indicate that there might be another phonotactic process taking place here –when assimilation leads to having two identical consonants one next to the other –called 'elision'. We shall deal with that process further below.

Then, the last feature that may be affected in cases of assimilation is the voicing feature as in *have to*, for instance, where the voiced last consonant of *have* turns into a voiceless consonant because of the influence of the following unvoiced apico-alveolar plosive, thus leading to the pronunciation: ['hæftə].

3.2 Progressive assimilation

Progressive assimilation (also called perseverative assimilation) occurs when the feature spreads from left to right. If you have two segments (A and B) and the second one is affected by the previous one, then it is progressive assimilation (A –> B). Progressive assimilations are less frequent than regressive ones, the usual examples of this phonotactic process being the particular cases of the pronunciations of the grammatical morphemes <-s> or <–ed>.

As mentioned by Deschamps *et alii* (2004: 38-39), for instance, the morpheme <-s> in English (plural, genitive, weak form of auxiliaries) may be realized in three different ways:

- [ɪz] after sibilants (*crashes*, *bushes*, *judges*, *roses*, *glasses*, *catches*);
- [s] after voiceless obstruents except sibilants (*cuts*, *cooks*, *cups*…);
- [z] in all other cases[14] (*boys*, *cars*, *sins*).

Thus, in normal rapid speech, *Jack is here* will be pronounced ['ʤæks'hɪə] but *Jane's had a baby* will be pronounced ['ʤeɪnz'hædə'beɪbi].

Similarly, the morpheme <–ed> in English (preterit, past participle, adjectives) may be realized in three different ways:

- [ɪd] after the two apico-alveolar plosives [t] and [d] (*pleaded*, *scented*, *added*);
- [t] after voiceless obstruents except [t] (*crashed*, *stopped*, *kicked*);
- [d] in all other cases (*loved*, *died*, *plugged*).

[14] After a voiced consonant or a vowel.

The spectrograms below show the difference in pronunciation of the morpheme <–ed> in the two following words (*concluded* and *cloned*) in the following sentence:

> Government advisers have concluded that meat and milk from cloned animals is safe to eat. (BBC Radio 4)

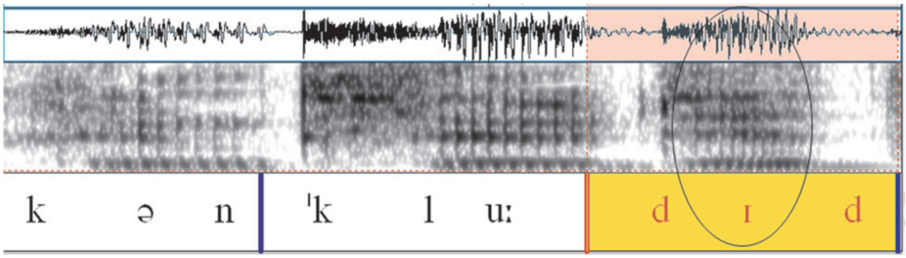


Figure 1

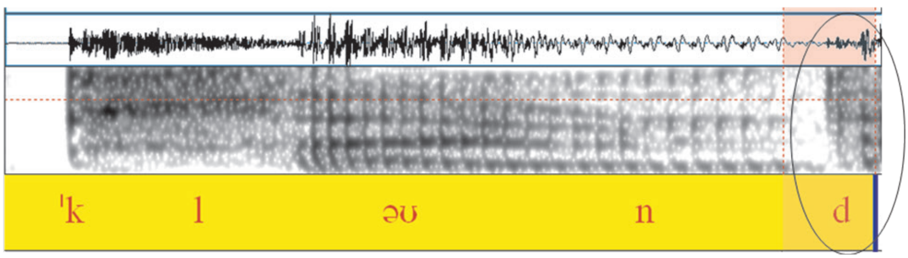


Figure 2

In the figures above, the signal is shown on the top line, then the spectrogram, and beneath the spectrogram the phonetic transcription aligned at the phoneme level. In figure 1, one can see that there is a vocalic element ([ɪ]) after the apico-alveolar plosive [d] whereas in figure 2, the apico-alveolar plosive [d] directly follows the nasal [n].

Other cases of progressive assimilation are sometimes found in rapid speech, especially within words such as *opened* pronounced ['əʊpm̩d] or *bacon* pronounced ['beɪkŋ][15] and it may also occur between words in collocations such as *up and down* ['ʌpm̩'daʊn], for instance.

[15] Examples also mentioned in Deschamps *et alii* (2004: 27)

Assimilation is not only a process that takes place in connected speech in contemporary English but there are also diachronic (or historical) assimilation forms originating in Latin etymology.

The Latin pre-verbs of prefixed verbs have been submitted to the process of assimilation. The English language, which has massively borrowed these verbs, has kept a trace of these cases of assimilation. For instance, the prefix *ex-* is unchanged before a plosive in *expose*, *expire*, *extoll*, *excuse* and before a vowel in *exact* and *examine* but it has a reduced form <e-> before the sonorants [l] and [r] as in *erect* and *elect* and an assimilated form before a fricative consonant (*effect*, *efface*), both processes leading to the opacification of the word formation[16].

One might add that in French, the spelling has fossilized ancient assimilations that cannot be heard anymore today because the former nasal consonant is mute as in *lampe* (which is spelt with an <m> before the bilabial consonant [p]) and *lande* (with an <n> before the apico-alveolar consonant [d]). Now, the graphic nasal consonant is there as the trace of the nasality of the vowel: [ɑ̃].

4. Palatalization

Palatalization is a special case of assimilation: in *partial,* for example, borrowed from the Old French *parcial* from Medieval Latin[17], regressive assimilation occurred based on the place of articulation; in contact with the palatal glide [j], the sibilant [s] became palatal and the palatal glide [j] lost its voicing. The merging of the two produced [ʃ]: this process is called palatalization.

Palatalization is also sometimes called ‘yod coalescence’ or ‘coalescent assimilation’ because of the merging of two segments into one forming a new qualitative sound; it may occur within the boundaries of a word or between words in contact in the speech chain and it may be of two types:

---

[16] The same process of assimilation has occurred with many other prefixes in English such as *ad-* in *accompany*, *appeal* or *affront* and it is now difficult synchronically to determine if *abbreviation*, for instance, is formed with the prefix *ab-* or with the prefix *ad-*, the final <d> being changed to <b> because of the following <b>.

[17] See *Chambers Dictionary of Etymology* (2005: 759).

- The first one is when the new sound that is formed shares a feature with one of the original segments as in *horseshoe* where the combination of [s] + [ʃ] turns into [ʃ] or in *Is she?* where the combination of [z] + [ʃ] turns into [ʃ].

- The second one occurs when the fusion of the two initial segments produces a new sound, for instance when the combination of [t] + [j] is changed into [tʃ] in words like *question* or *digestion*, or in phrases like *the ball that you brought* [ðə'bɔːlðəʧu'brɔːt]).

4.1 Palatalization within word boundaries

Palatalization usually occurs after the stressed syllable in pre-final position :

- [s] + [j] becomes [ʃ] regardless of the spelling: <ss> in *confession*, <c> in *delicious*, or <t> in *portion* or *partial*[18] ;
- [k] + [s] + [j] becomes [kʃ] in *action* ;
- [z] + [j] becomes [ʒ] in *television* ;
- etc.

Other words may fluctuate between [dj] and [dʒ] as *educate* or *soldier*. *Issue* may also fluctuate between ['ɪʃuː] and ['ɪsjuː][19] as shown in the spectrograms below:

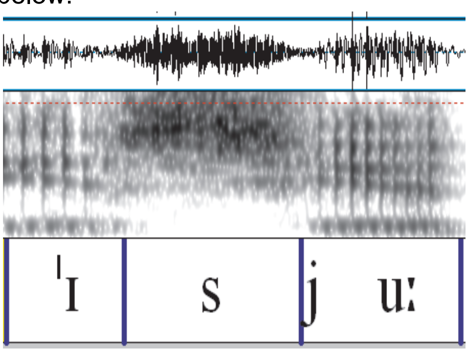


Figure 3

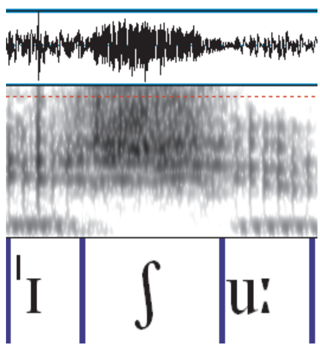


Figure 4

[18] Already mentioned above.

[19] BrE 1998 poll panel preference: ['ɪʃuː] 49%, ['ɪsjuː] 30%, ['ɪʃjuː] 21 %; in AmE always ['ɪʃuː] (See Wells 2008: 404).

In a stressed syllable, palatalization does not usually occur: for instance, the combination [s] + [j] is pronounced [sj] in *con'sumers*[20] (see spectrogram below):

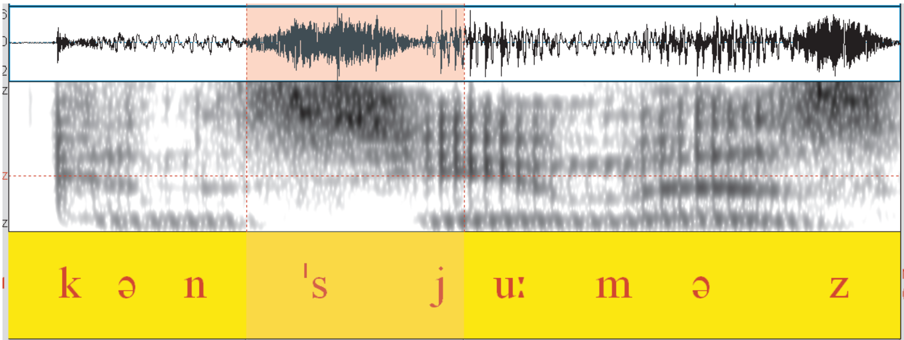


Figure 5

4.2. Palatalization between words in contact

As mentioned before, the palatalization process described above may also occur between word boundaries as in [bə'kʊdʒu'telmi'waɪ] (*But could you tell me why?*) where the reduced form of *could you* undergoes palatalization: [d] + [j] -> [dʒ] as shown in figure 6 below:

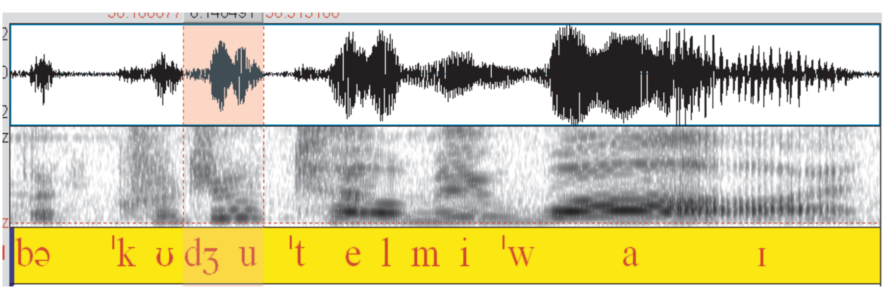


Figure 6

[20] With a tendency to drop the palatal sonorant in this context [kən'su:məz].

5. Elision

This process may affect consonants (especially when heavy obstruent clusters are present) as well as vowels, leading to syllabic compression. Elision either is a long established internal process, or occurs in the speech chain, leading to contextual elisions, especially in rapid speech.

5.1 Elision of consonants

Historically, the elision of plosives in the middle of a cluster of three consonants has led to a mute consonant in the present-day pronunciation of words like *castle*, *whistle*, *rustle*, *glisten*, *fasten*, *hasten*, *often*[21], *soften*, *muscle*. The historical elision of consonants at the beginning of a word or at its ending has also been clearly established. It is the case of the initial [w] in *write, wrong* or *wraith* for instance, the initial [p] in *psychology* and *psalm* or the initial [g] in *gnash, gnarl* or *gnaw*. At the end of the word, it is the case of the final [b] in *lamb* or of the final [m] in *hymn*, making the latter a homophone of the personal pronoun *him*.

In present day English, the same process of elision within a word usually occurs in words like *exactly* [ɪg'zækli], *handbag* ['hæmbæg], *postpone* ['pəʊspəʊn] and it often occurs in conjunction with assimilation. It is therefore interesting to contrast the elision of [t] in words like *events* to the opposite tendency to insert an epenthetic [t] in words like *fence*, *dance*: *tense* and *tents* might therefore sound the same, either [tens] or [tents].

In the speech chain, the same process of elision is frequent, especially in rapid and colloquial speech. There again, it usually occurs in conjunction with assimilation.

It may occur when heavy sequences of obstruents are found in contact, as in *the fact that* [ðə'fækðət], for instance, where the final [t] of *fact* is omitted but this is not systematic, as shown through the following example, where there is no elision of the apico-alveolar [t] at the end of *next* even though it is part of a heavy consonant cluster:

---

[21] Sometimes also pronounced ['ɒftən] in British English.

(3) Government advisers have concluded that meat and milk from cloned animals is safe to eat. The independent experts have been examining the issue following reports over the summer that meat from the offspring of a cloned animal was sold to consumers without the required authorization. The Food Standards Agency will consider the issue next month before advising ministers. (BBC Radio 4).

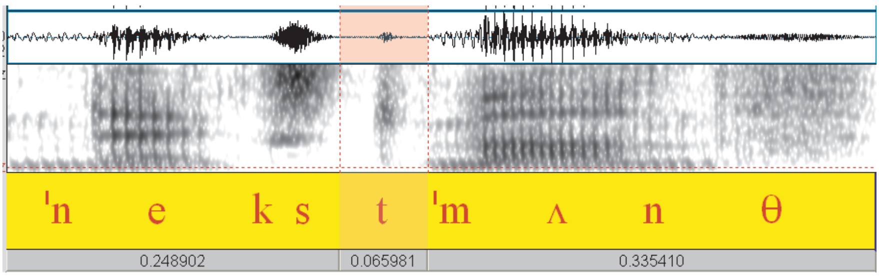


Figure 7

In Figure 7, the plosive is visible (and is affricated); it is not elided. Elision occurs very frequently when two consecutive consonants are identical. In the following sequence drawn from example (3), one of the two fricatives in *was sold* is elided (the length of the fricative [s] is identical to the length of the same fricative in *consumers* for instance). Similarly, the final [d] in *sold* is not pronounced as it is followed by the voiceless apico-alveolar plosive [t].

(3a) meat from the offspring of a cloned animal was sold to consumers

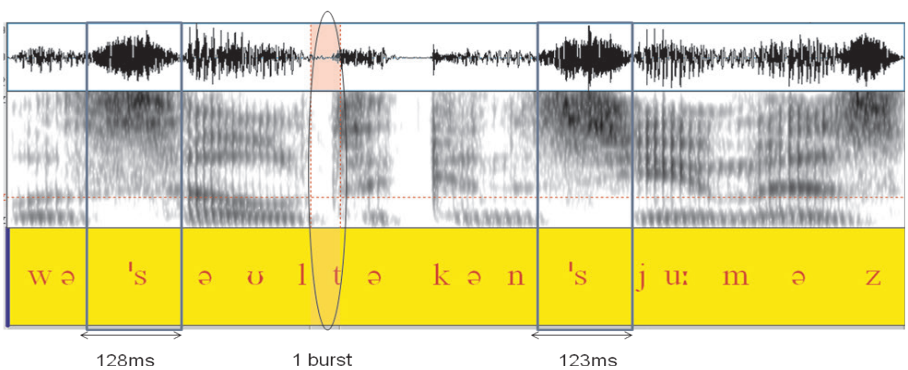


Figure 8

By contrast, in the following sentence:

(3b) Government advisers have concluded that meat and milk from cloned animals is safe to eat.

two fricative sounds are pronounced, one of them being voiced (see the voicing line on the first fricative sound):

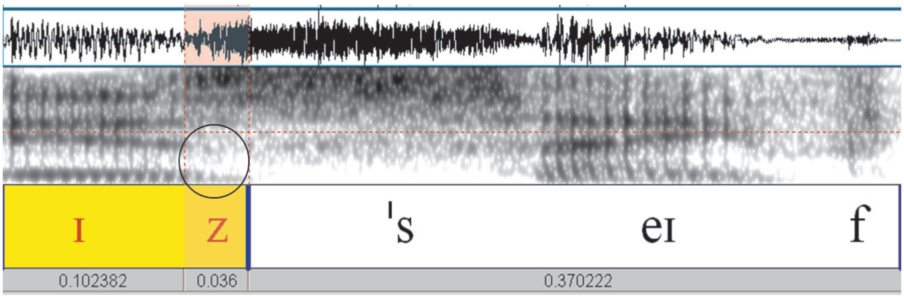


Figure 9

5.2 Reduction, assimilation and elision of grammatical markers

The reduction, assimilation and elision of grammatical markers are common in English. These speech chain phenomena may lead to semantic ambiguities as in the following: [hu'spɔɪldɪt] (*Who spoiled it? / Who has spoiled it?*); [aɪ'drɒpðə'bʊk] (*I drop the book / I dropped the book / I had dropped the book*). These ambiguities are generally solved thanks to the context.

5.3 Elision of vowels and syllabic compression

As shown in the section on rhythm, weak syllables in English undergo a process of reduction. This process may lead to the elision of vowels and thus to the loss of one syllable in the word.

There is historical evidence of the process through words like *chapter* or *evening*, which are the result of the compression of trisyllabic words (†*chapiter*, *even + –ing*) into disyllabic ones.

In present day English, whole syllables may be elided in rapid speech after primary stress, especially in the vicinity of [r]: *history*, *library, missionary, different, interest*, *general,* or in the vicinity of [l] in *catholic, awfully, chocolate, family.*

This tendency may also be observed in words like *Mexican,* pronounced in two syllables in colloquial speech ['mekskən], *medicine* ['medsn̩], or *comfortable* which is pronounced in three syllables [ˈkʌmftəbl̩] in rapid speech. Some preprimary stress syllabic compressions are also possible in words like *police*, leading first to a pronunciation with a syllabic [l̩]: [pl̩iːs] then tending to the loss of the syllabicity of the sonorant [l] and thus to the pronunciation of the word in only one syllable. This pronunciation of the word *police* is not restricted to colloquial speech; it is often heard on the radio (especially on the *BBC*) but it must be pointed out that it is restricted to rapid speech.

Acoustic evidence of that loss of one syllable is shown through the following spectrograms (figures 10 and 11). In figure 10, the audio file analysed is the word *government* as pronounced in isolation in the *Longman Dictionary of Contemporary English* (2003) and one can see that there is a vocalic element between the fricative [v] that is the onset of the second syllable and the nasal [n] which is the coda of that same syllable, whereas in figure 11, the spectrogram of the same word pronounced in normal speech (drawn from the radio) shows that the weak syllable is elided as there is no vocalic element between the fricative and the nasal. It must be added that the fricative [v] is short and almost inaudible.

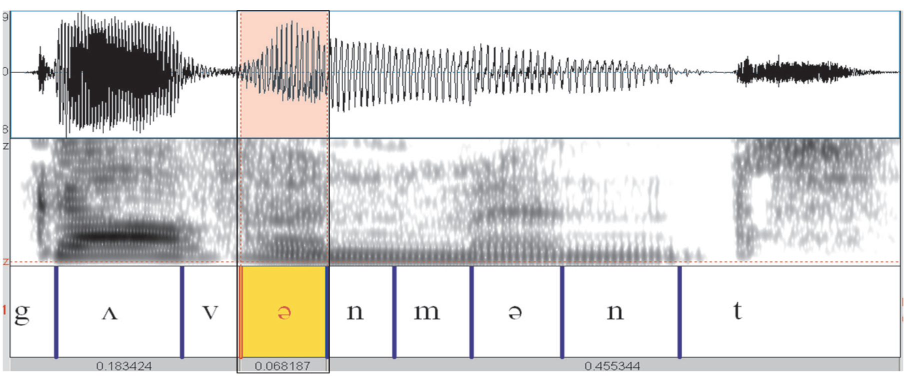


Figure 10. Pronunciation of *government* in the Longman Dictionary (2003)

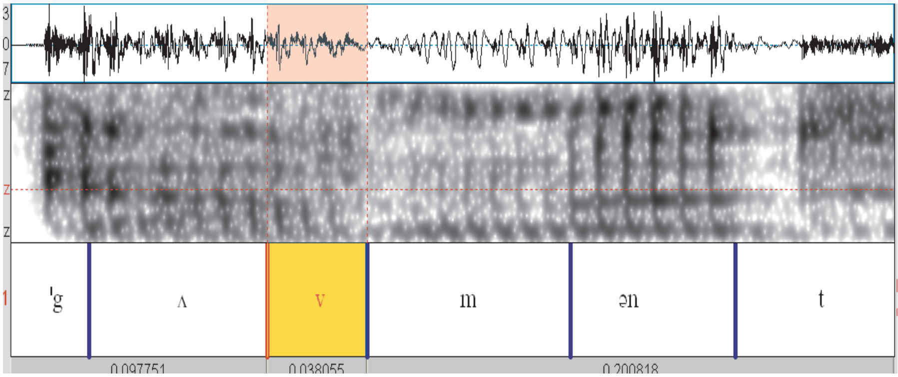


Figure. 11 Pronunciation of *government* in normal speech, Radio extract: BBC R4.

6. Diaeresis, synaeresis and smoothing

Below is a verse from Racine's play *Esther* (1689), Act 1, Scene 4:

> Hélas ! Ce peuple ingrat a méprisé ta loi
> La nation chérie a violé sa foi ;
> Elle a répudié son époux et son père,
> Pour rendre à d'autres dieux un honneur adultère.

where the French words *nation* and *violé* have to be pronounced in three syllables for the line to be metrically complete. For the same reason, the word *répudié* has to be pronounced in four syllables. These three words illustrate the process of synaeresis which has taken place since then. Indeed, in spoken French today, the words *violé* and *nation* are pronounced with only two syllables and the word *répudié* in three syllables.

The phenomena of diaeresis and synaeresis also occur in English. For instance, if *nation* in English is pronounced in two syllables because it is subject to the process of synaeresis (and has undergone palatalization), *violate* and *repudiate* are pronounced with three and four syllables, respectively, because the two contiguous vowels (<i> + <o> in one case and <i> + <a> in the other) are the nuclei of two different syllables: they are subject to diaeresis.

6.2 Smoothing

In rapid speech, when a closing diphthong is followed by [ə], a special case of elision may occur. It is called 'smoothing' and it consists in the elision of the second element of the closing diphthong. For instance, words like *hour* (or *our*), *diagram* and *nowadays* will be smoothed into [aə] (['daəgræm], ['naədeɪz]) and sometimes to [ɑː]. In the following excerpt from *BBC Radio 4*, the process of smoothing occurs in the word *fire*:

(4) Well then she was a fire-watcher and she learnt exactly what to do and she gave up two nights a week to fire-watching in Curson Street. Then somebody said would she lecture to new fire-watchers, so she said er yes she would, and when she'd done it about three times, the lady who ran it all asked to see her and said, "Would you mind giving up the lectures?" and she said, "No, not the least, but could you tell me why?" "Well", she said, "it's your voice, your accent irritates people so much they'd like to put you on the fire!" (Lady Mosley, Interview, BBC Radio 4)

If one compares the way the word is pronounced in the audio file (figure 12) to the same word recorded in the Longman Dictionary (figure 13):

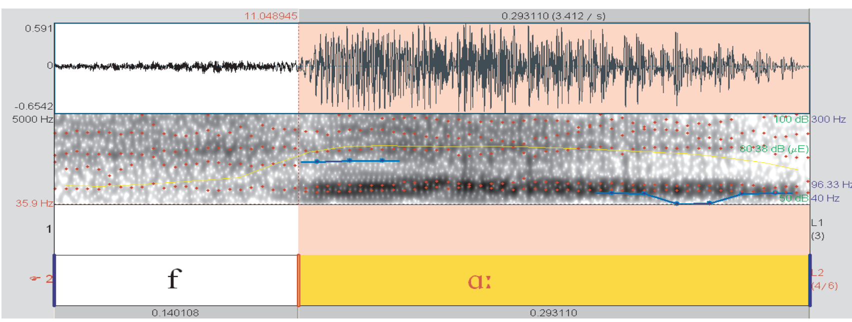


Figure 12

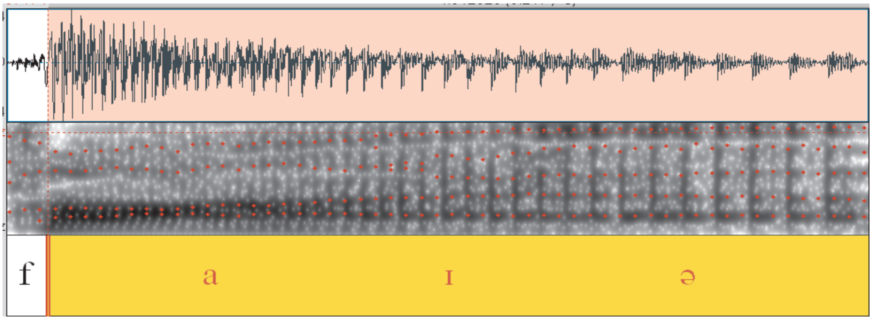


Figure 13

one can see that there is only one phoneme in the first audio file ([ɑː]) whereas there is an evolution in the second one (from [aɪ] to [ə]) and a linking process occurs between the two.

The smoothed pronunciation of *fire* could lead to ambiguity, as shown through the following anecdote, well-known to phoneticians:

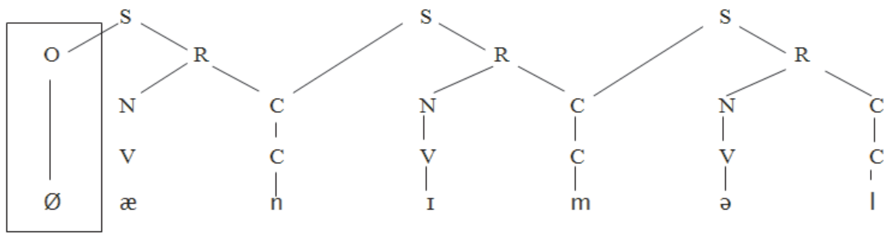
(5) In a small southern town there was a "Nativity Scene" that showed that great skill and talent had gone into creating it. One small feature bothered me. The three wise men were wearing firemen's helmets. Totally unable to come up with a reason or explanation, I asked the lady behind the counter about the helmets. She exploded into a rage, yelling at me, "You darn Yankees never do read the Bible!" I assured her that I did, but simply couldn't recall anything about firemen in the Bible. She jerked her Bible from behind the counter and riffled thru some pages, and finally jabbed her finger at a passage. Sticking it in my face she said "See, it says right here, 'The three wise men came from afar.'" (Firemen, Yankees, and Nativity Scenes, an anecdote)

7. Linking

As mentioned in the introduction of this chapter, in real connected speech, words are not uttered separately but are linked to one another and pronounced in a continuous stream of sounds.

While the use of linking consonants in French is closely related to syntactic structure and is subject to specific rules (compulsory in *un grand‿homme*, it is optional in *Nous partons(‿)à Paris* but forbidden in **Ils sont partis‿au zoo*), in English, the linking process appears to be more strictly phonetic and

is closely related to the structure of the syllable, usually described as a CVC structure: the initial consonant(s) form(s) the onset (O) and the rest of the syllable is the rhyme (R) which is itself composed of a nucleus (N), which is a vowel, and a coda (C), consisting of one or several consonants.
Words beginning with a vowel have no consonant onset:



Thus, they “borrow” the onset they require from the coda of the preceding syllable, if there is one.
It should be noted here that as Deschamps *et alii* (2004: 25) point out, there is historical evidence of this phenomenon in the double phonological form of the indefinite determiner of English (*a/an*), for instance. Indeed, the etymological [n] has been retained only when it can serve as the onset of the initial syllable of the word following the determiner:

a + consonant-initial word: *a man* [ə'mæn]

an + vowel-initial word: *an animal* [ə'nænɪməl]

In the following example drawn from the radio and already mentioned in this chapter,

(3) Government advisers have concluded that meat and milk from cloned‿animals‿is safe to eat. The independent experts have been examining the issue following reports over the summer that meat from the offspring of a cloned‿animal was sold to consumers without the required authorization. The Food Standards Agency will consider the issue next month before advising ministers. (BBC Radio 4)

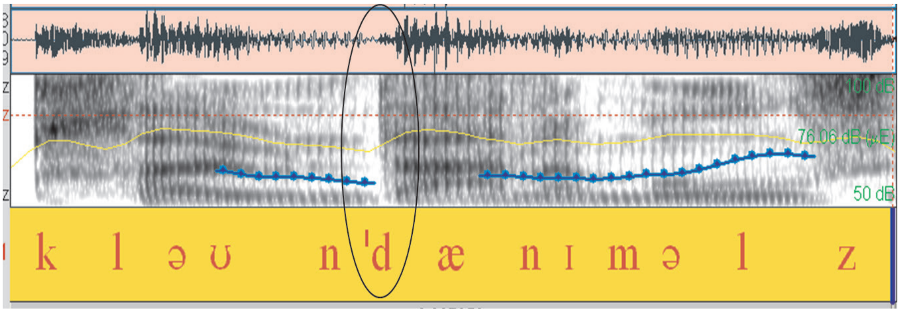


Figure 14

the word *animal* is said twice, preceded by the adjective *cloned*. As shown through the spectrogram in figure 14, there is no pause between the two words and the final consonant of the adjective is used as the onset of the next syllable, leading to the following pronunciation: [frəm'kləʊn'dænɪməlz]. The same process occurs in the following example:

(6) So, you've not exactly had a… a privileged‿upbringing. I mean you… you… you're not exactly part of the sort of middle-class, are you? (BBC Radio 4)

where there is no break between the last syllable of the adjective *privileged* and the first syllable of the noun *upbringing*:

['səʊjuv'nɒtɪg'zækli'hædə'prɪvɪlədʒ'dʌpbrɪŋɪŋ]

This absence of clear-cut borders between words may lead to problems of ambiguity as in the following pair: *an aim* / *a name*. Only the context helps to distinguish between these words. Here is another example of possible ambiguity: ['gɪvmisəm'mʌniɔː₍ˌ₎raɪ'skriːm] may correspond to the following sentences: *Give me some money or I scream* / *Give me some money or ice-cream*.

There are also well-known examples of misdivisions that occurred at some point of the word's history: for instance, *apron* evolved by misdivision of a ‘napron’ (from OF ‘naperon’) and *nickname* comes from ME ‘neke-name’ (from ‘an + ekename[22]’).

But when the rhyme of the syllable is a vowel followed by an empty coda, there is no consonant available for the linking process. Thus, an excrescent sonorant ([j], [w] or [r]) is introduced to realize the linking and it depends on the type of vowel forming the nucleus of the previous syllable:

- the palatal glide [j] is used after front closing tense vowels [i(ː), eɪ, aɪ, ɔɪ]: *Be on time!* ['biːjɒn'taɪm]; *a toy animal* [əˌtɔɪ'jænɪməl]; *Joy and sorrow* ['ʤɔɪj̱ən'sɒrəʊ]; *Stay at home* ['steɪj̱ət'həʊm];
- the bilabial glide [w] arises after back closing tense vowels [u(ː), əʊ, aʊ]: *Go away!* ['gəʊwə'weɪ]; *How interesting!* ['haʊ'wɪntrəstɪŋ]; *now and then* ['naʊwən'ðen];
- and the apico-alveolar sonorant [r] is used after the central vowel [ə] or after the centring tense vowels [ɪə, eə, ʊə, ɔː, ɑː, ɜː]: *far away* ['fɑːrə'weɪ]; *here and there* ['hɪərən'ðeə]; *rather exciting* ['rɑːðərɪk 'saɪtɪŋ].

It should be noted that the linking [r] occurs in contemporary British English each time a centring diphthong or a schwa is followed by an initial vowel, whether or not < r > is present in the spelling[23] :

the Shah of Persia [ðə'ʃɑːrəf'pɜːʃə]
Law and Order ['lɔːrə'nɔːdə]

but this ‘intrusive [r]’ is sometimes considered as substandard in morphological derivation if there is no < r > in the spelling. For that reason, Wells (2000: 629) chose to show the linking [r] within a word in “ordinary type” (because it is “compulsory”) whereas the intrusive [r] is shown in “raised type” (because it is “optional and disapproved of by some”). The examples used by Wells are *storing*: ['stɔːrɪŋ] vs. *thawing* ['θɔːʳɪŋ][24].

[22] Meaning ‘additional name’.
[23] When <r> is not present in the spelling, it is called ‘intrusive [r]’.
[24] Probably also pronounced ['θɔːwɪŋ] by some native speakers.

Acoustic evidence of the process of linking is shown in the following figures where there is no blank in the signal but a glide realizing the link between the vowels.

(2) Government advisers have concluded that meat and milk from cloned animals is safe to‿eat. The independent experts have been examining the issue following reports over the summer that meat from the‿offspring of a cloned animal was sold to consumers without the required authorization. The Food Standards Agency will consider the issue next month before advising ministers. (BBC Radio 4)

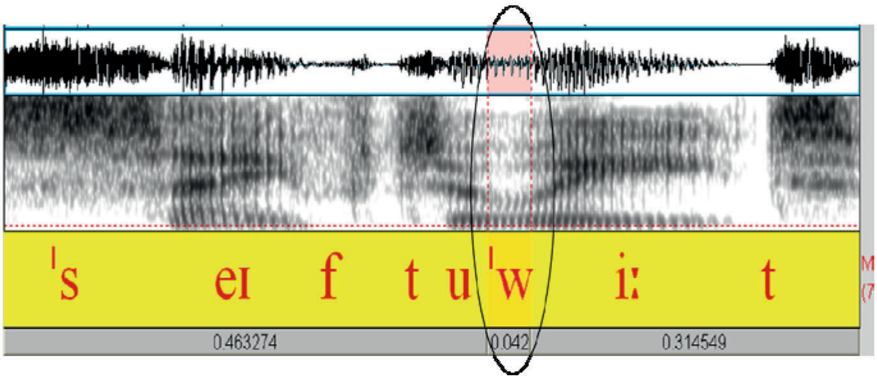


Figure 15. 'safe to‿eat'

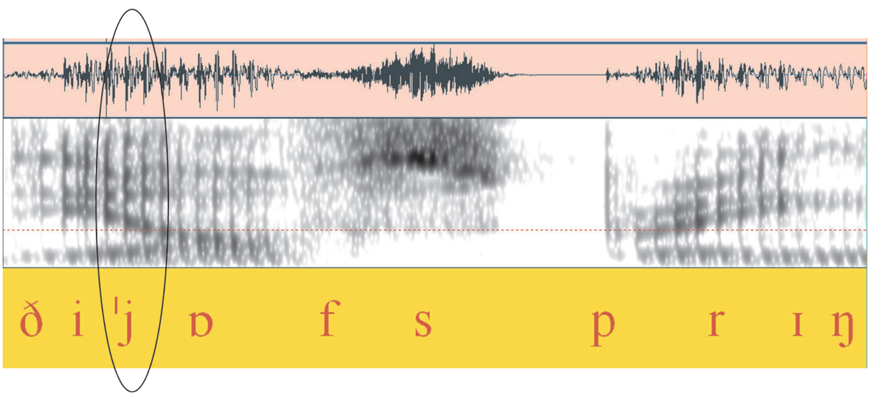


Figure 16 : 'the‿offspring'

The same process occurs in the following examples :

(7) Have you ever‿allowed a child to have a drop of wine with the meal at home? Well, if you have and if you keep doing it, you could find yourself in trouble with the law! (BBC Radio 4)

(8) I don't think that's the right way to go really because um, people will just go out and get wasted on vodka‿and stuff. (BBC Radio 4)

(9) Good morning. Doctor Callaghan, do you think this document will help doctors deal with the... the sort of cases that we've just been hearing about?
I think it will, I think it clarifies the law‿around advanced directives... (BBC Radio 4)

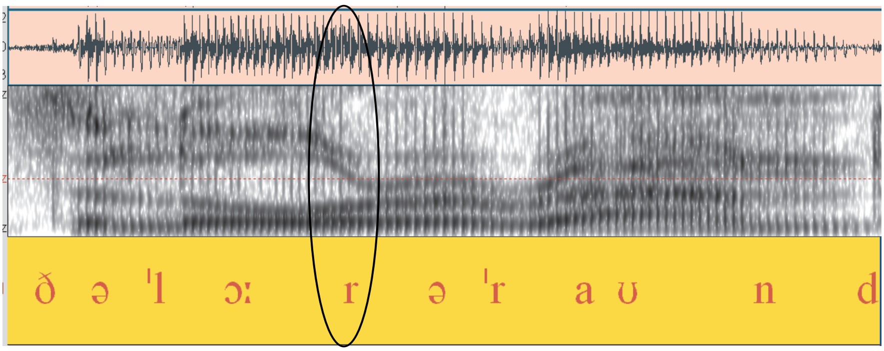


Figure 17 : 'the law‿around'

8. The glottal stop

For various reasons (mainly emphasis), the speaker may sometimes use a glottal stop to avoid the linking process between two contiguous vowels. The glottal stop, symbolized as [ʔ], is a plosive pronounced in the glottis. Like other plosives, its articulation involves two stages: first an interruption in the air flow (blank on the signal), then a burst.

(3) Government advisers have concluded that meat and milk from cloned animals is safe to eat. The independent experts have been examining the issue following reports over the summer that meat from the offspring of a cloned animal was sold to consumers without the required authorization. The Food Standards Agency will consider the issue next month before advising ministers. (BBC Radio 4)

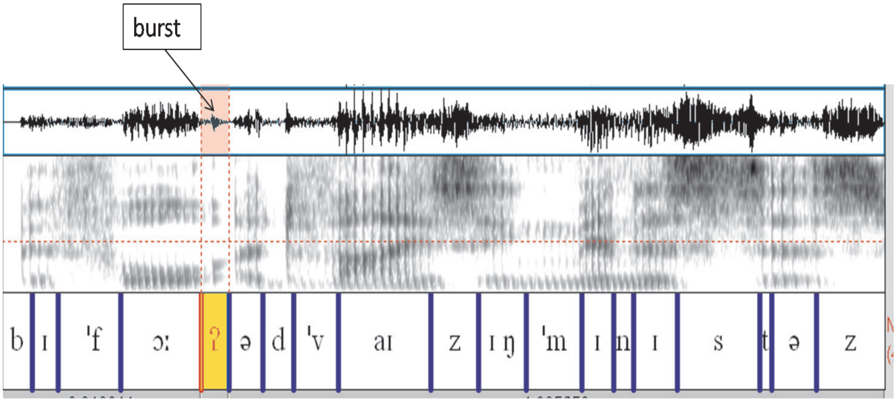


Figure 18. ‘before ʔ advising ministers’

(11) Schoolchildren in Southern Sweden are being terrorised by <u>a drunken elk</u>. The creature, which weighs more than five hundred kilos, has been eating large numbers of fermented apples from a tree. Police say its behaviour has become erratic, and mad. (BBC Radio 4)

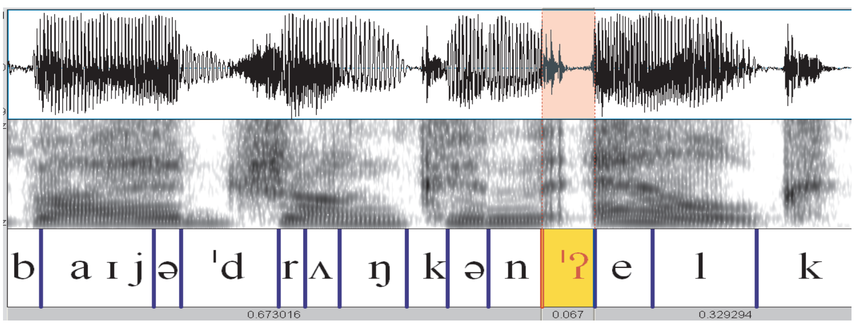


Figure 19. ‘by‿a drunken ʔelk’

In these two examples, linking is thus avoided by the intrusion of a glottal stop[25], which is sometimes used as an expressive means of emphasis before a word-initial stressed vowel. The various means of marking

[25] See also the phenomenon of ‘creaky voice’ occurring between *drunken* and *elk* in figure 19.

emphasis will be the subject of the last section of this chapter on connected speech.

9. Prominence in speech

There are several factors, segmental and suprasegmental, that make a syllable sound more prominent than the others in normal speech. Many of these factors have already been studied –see Cruttenden 1997, Gussenhoven 2004, Herment 2001, Hirst & Di Cristo 1998, Nicaise & Gray 1998, Roach 2009, among others.
At the suprasegmental level, the markers that are usually mentioned are:

- a change in pitch level and in pitch contour,
- generally associated with a change in loudness (prominent syllables are often assumed to be pronounced louder than the others),
- sometimes a pause being inserted before the element designed to be prominent,
- a change in rhythm (either an acceleration or a deceleration in speech rhythm),
- and it would be useful to add to these well-known factors a further one, namely the absence of the phonotactic processes described in the previous sections (linking, assimilation and elision processes for instance), sometimes due to the insertion of a glottal stop.

As far as segmental factors are concerned, one can include:

- vowel quantity and vowel quality,
- the length of the consonant(s) at the onset of the prominent syllable.

In the following example from BBC Radio 4, in which the linguist John Honey is being interviewed, some words are particularly salient, the most prominent being the adjective *bad*, italicized in the example:

(11) Many linguists have argued for thirty years that one of the bad things about teaching Standard English, white, Anglo-Saxon, middle-class speech, is that it is exclusive, it serves to exclude others. “Quite wrong”,

says my book, Standard English is the only one that is not exclusive all the others're exclusive, you and I can't learn Black English, there are no text-books [...] (J. Honey « Mother Tongue », Bill Bryson, BBC R4)

The following spectrogram shows that prominence is not due to one isolated acoustic marker but to a combination of markers:

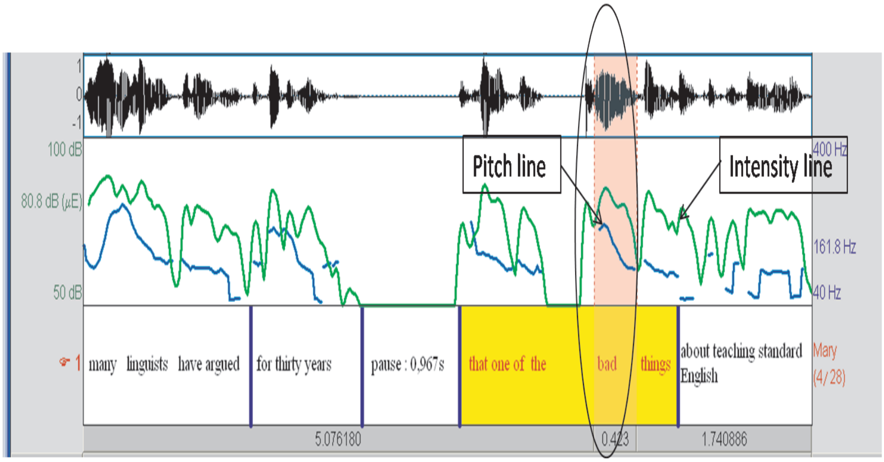


Figure 20

- The pitch line is high on the adjective[26] with a large movement on it, the rest of the sentence '*things about teaching standard English*' being uttered at low level.
- There is a deceleration on the adjective: '*one of the*': 66ms; '*bad*': 92ms; '*things about teaching Standard English*': 180ms. This deceleration is due to segmental factors (vowel quantity and consonant length). First, it is important to notice here the long period of silence preceding the release phase in the pronunciation of the initial plosive [b] (= 46ms) in *bad*, see figure 21. This long period of silence gives the hearer the impression that there is a pause before the adjective, making it particularly prominent in Honey's discourse. Furthermore, as the release of the plosive closure is not made rapidly, a fricative sound follows the burst. As described by Cruttenden (2001: 160), "plosives

[26] But no higher than on other syllables of other words in the context, see for instance the first syllable of *Many* or the first syllable of *thirty* in the prepositional phrase '*for many years*', which are also important words in Honey's discourse.

made with this slow, fricative release, are said to be affricated"; the consonant is then devoiced, turning from lenis (weak) to fortis (strong).

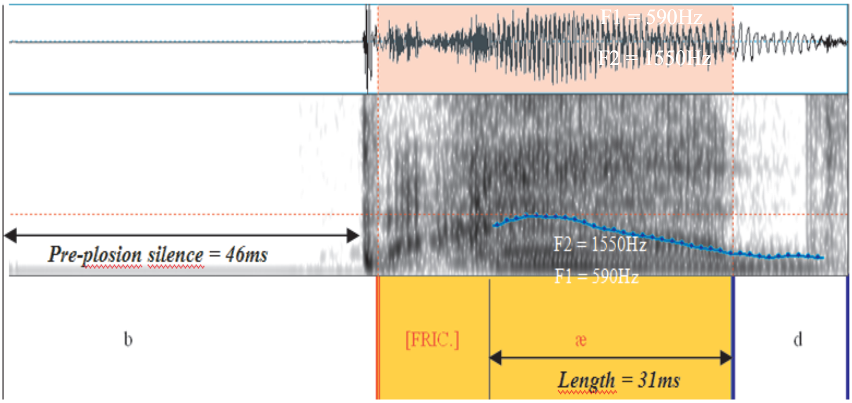


Figure 21. *'bad'*

- Other factors are vowel quantity and vowel quality. As far as quantity is concerned, one can see in figure 21 that the lax vowel [æ] is relatively long in *bad* (the length of the vowel is 31ms) whereas it is shorter when pronounced in other words from the same extract (21ms in *Saxon* for instance and even 12ms in *Black* (*Black English*)). But it should be noted that the pronunciation of vowels depends on the phonemes that directly precede or directly follow them. For instance, in this example, the vowel is followed by the voiced apico-alveolar plosive [d]. This consonantal environment has an impact on the length of the vowel, as shown by Cruttenden (2001: 111): "this traditionally short vowel [æ] is now generally longer in RP than the other short vowels […]. Such lengthening is particularly apparent before voiced consonants e.g. in *cab*, *bad*, *bag*, *badge*, *man*; /æ/ in these contexts is almost equivalent to the long vowels". As far as vowel quality is concerned, there is no instrumental evidence to prove a particularly high degree of openness of the vowel, as its first two formants[27] are quite typical of the data usually given for [æ].

[27] F1 = 590Hz; F2 = 1550Hz.

To the markers described above, one can add the insertion of a glottal stop before a word-initial stressed vowel as shown below:

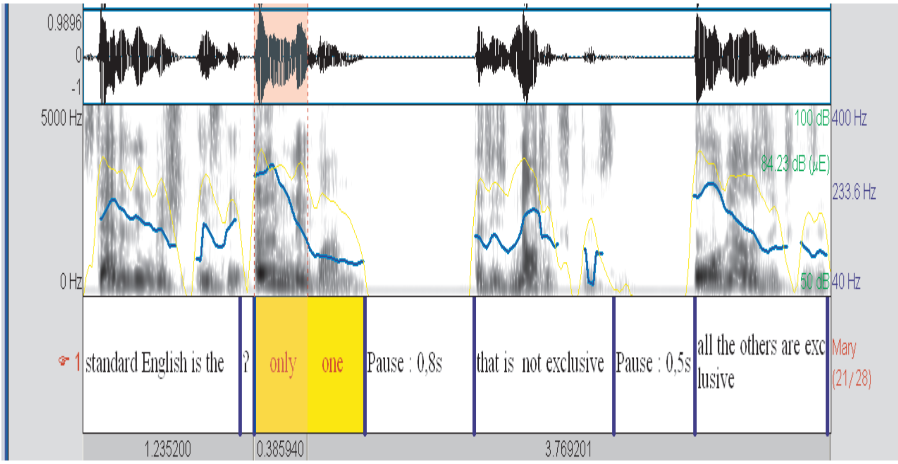


Figure 22

There is a very short discontinuity in the signal between the two vowels in contact –[i] from the determiner *the* and [əʊ] from the adjective *only*– which reflects the insertion of a glottal stop between these two words. Associated to high pitch and high intensity on the first syllable of the adjective, the glottal stop is used as a means of emphasis. This phenomenon, described by Wells in his dictionary (2008: 327), is also called hard attack: "ʔ is optionally used as a way of adding emphasis to a syllable that begins with a vowel sound".

Thus, the acoustic markers of prominence in oral speech are:

- a significant rise in pitch,
- a rise in intensity (loudness),
- a change in rhythm (acceleration or deceleration),
- a change in vowel and consonant quantities and qualities (segmental factors),
- sometimes the presence of a pause or of a pre-vocalic glottal stop.

It should be added that prominence occurs not only on lexical words but also on function words.

9.1 Prominence of function words

Function words are usually unstressed and reduced in normal speech but they may also be made prominent in discourse. It can be the case of prepositions or auxiliaries as in the next example already mentioned at the beginning of this chapter:

(3) T. Blair : One of the things I do in the book is to correct some of the… the… the… the really very exaggerated figures as to what's happened in the conflict and draw attention –this is based on the international Red Cross figures– um draw attention to the fact that *of* those that *have* died – and there's far too many of course– the vast bulk have been killed by sectarians and terrorists. (Tony Blair, Interview on BBC Radio 4)

Two function words are prominent here: the *of* and *have*.

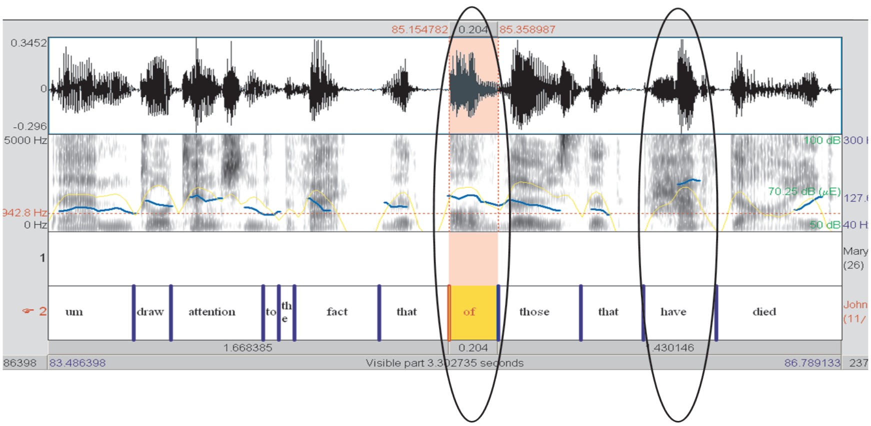


Figure 23

As in the previous examples, it is a combination of markers that makes the preposition and the auxiliary sound prominent in the speaker's speech:

- discontinuities in the signal[28],
- a sudden rise in the intensity and in the pitch lines,
- the full forms of these two function words pronounced [ɒv][29] (see figure 24) and [hæv][30] (see figure 25) when the reduced forms would

---

[28] See the short breaks of 130ms between *that* and *of* and of 90 ms between *that* and *have*.

[29] Length of vowel = 120 ms; F_1 = 570 Hz and F_2 = 800 Hz.

[30] Length of vowel = 80 ms; F_1 = 850 Hz and F_2 = 1578 Hz.

- be more common and “neutral” (see figure 26 where the auxiliary *have* is reduced to [ə]).

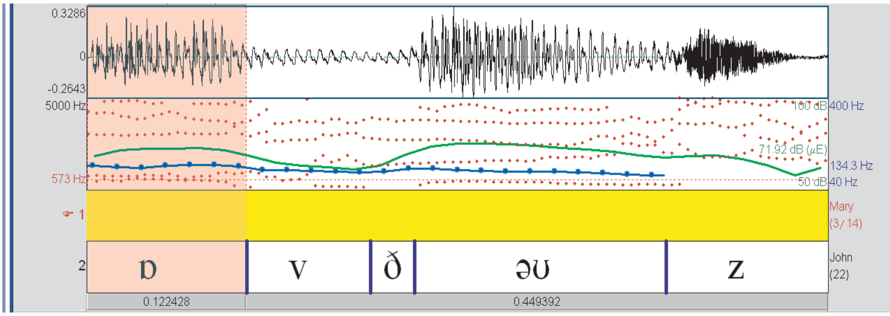


Figure 24

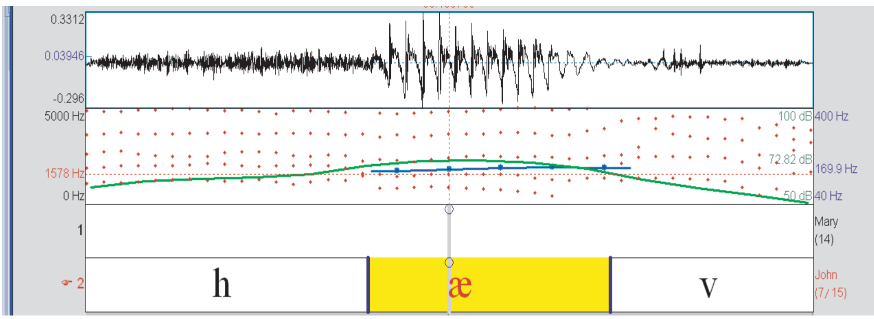


Figure 25

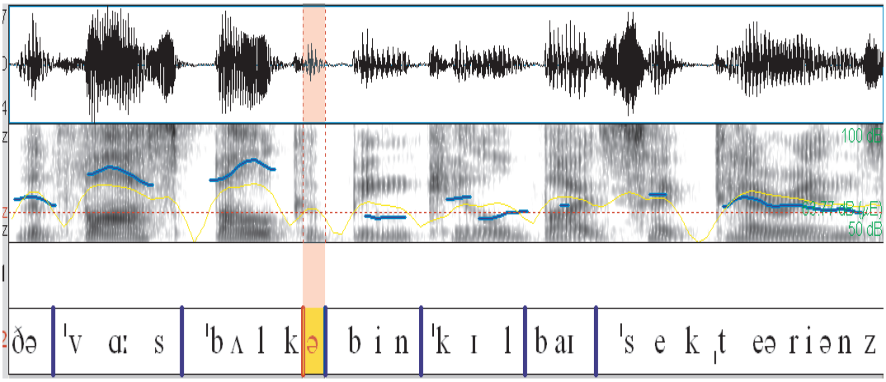


Figure 26[31]

In the following example drawn from the same document as example (2) but uttered by the journalist, other function words are highlighted (the auxiliaries *do* and *were*) through the same acoustic markers as before (rise in the intensity and pitch lines and full forms of the grammatical markers – see figures 27 and 28):

(12) Journalist: We *do* know that Al Qaida increased their activities in Iraq after the war so those people may have died as a result of terrorism but that terrorism *was* there as an *in*direct result of your decisions. (Tony Blair, Interview on BBC Radio 4)

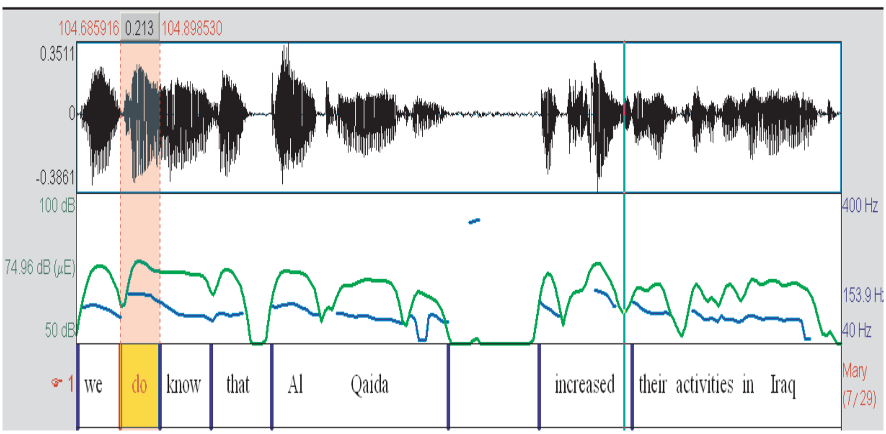


Figure 27

---

[31] It should be noted here that the pitch line being slightly higher on the first syllable of *sectarians* than on the others, it is noted as bearing primary stress even if the rules governing lexical stress should impose the stress on the second syllable. It is an "expressive" accent.

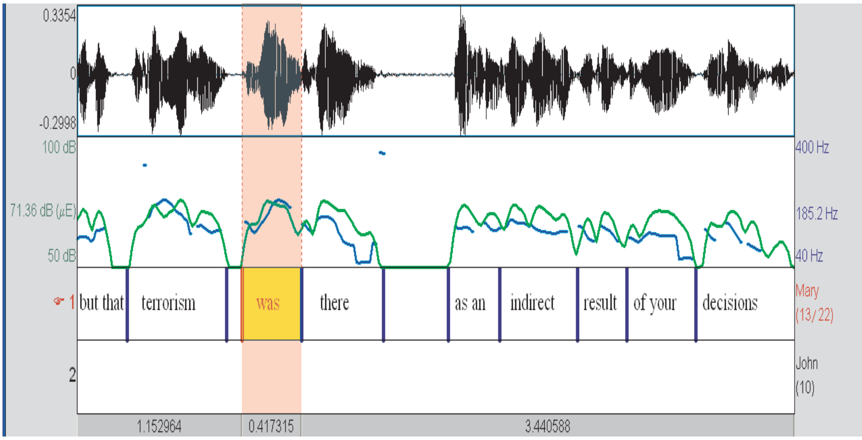


Figure 28

It should be noted that yet another element is prominent in this example (the prefix *in-* in the adjective *indirect*: *as an **in**direct result of your decision*); the prominence of prefixes in speech will be dealt with below in the final section of this chapter.

9.2 Prominence and morphophonology: prefixed words

Numerous studies on prefixed words have been carried out by lexicologists (see Aronoff 1976, Bauer 1983, Tournier 1985, Paillard 2000, among others) and by morphophonologists (see for instance Guierre 1979, 1984, Guyot-Talbot 2003, Trevian 2003, 2010). A clear-cut distinction is generally made between separable and inseparable prefixes in relation with the transparency of the derivation: for instance, *de-* is a separable prefix in *de-ice* (= to remove ice from) but is inseparable in *deceive* where the prefix is recognizable only because other words built with the same root exist in contemporary English (*receive*, *conceive*, *perceive*, etc.)[32].
Prefixes, even separable ones, are bound morphemes and do not usually bear lexical primary stress but in normal speech, they may be prominent especially in cases of contrast:

---

[32] The prefix is etymological (Latin roots) and opacified.

> [Prefixes] may receive nuclear stress (the most important stress in a sentence) if they are contrastively stressed; the following dialogue exemplifies this:
> A It's probable that Peter will win.
> B Really? I'd have thought that was very improbable [ˋɪmˌprɒbəbl̩]
> (Fudge, 1984: 164)

The same kind of process as described by Fudge occurs in the following authentic example from the American National Public Radio (NPR) where the separable prefix *mis-* is prominent in *misperception*, reinforcing the contrast with *perception* in the context.

(13) Exercise can be hard on the joints at least that's a common perception – make that *mis*perception. (Exercise Studies Find Good News For The Knees in All things Considered, NPR, 04/09/2009)

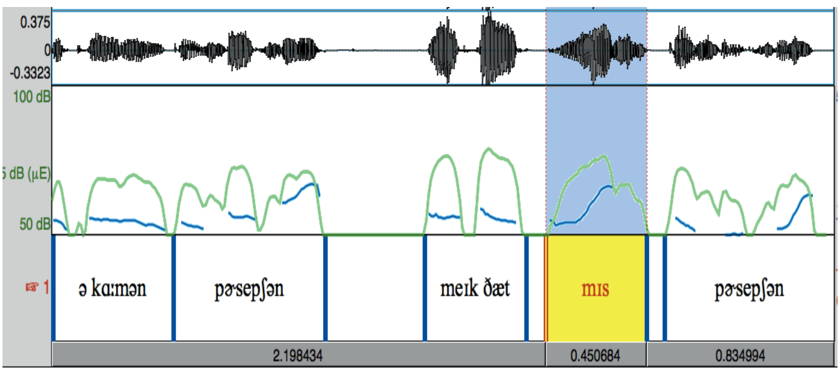


Figure 29

As shown in figure 29, there is high intensity on the prefix, associated to a rise-fall movement of the pitch, both being the markers of a stress-shift on the prefix which not only bears primary stress but is also the nuclear element of the intonation phrase. Short pauses before and after the prefix also reinforce its prominence and emphasize the contrast between the two antonyms[33].

Prominence in spoken discourse does not occur only for contrastive purposes but is also possible in other contexts as shown in example (13), already mentioned in the preceding section:

(12) We do know that Al Qaida increased their activities in Iraq after the war so those people may have died as a result of terrorism but that terrorism was there as an *in*direct result of your decisions (Tony Blair, Interview on BBC Radio 4)

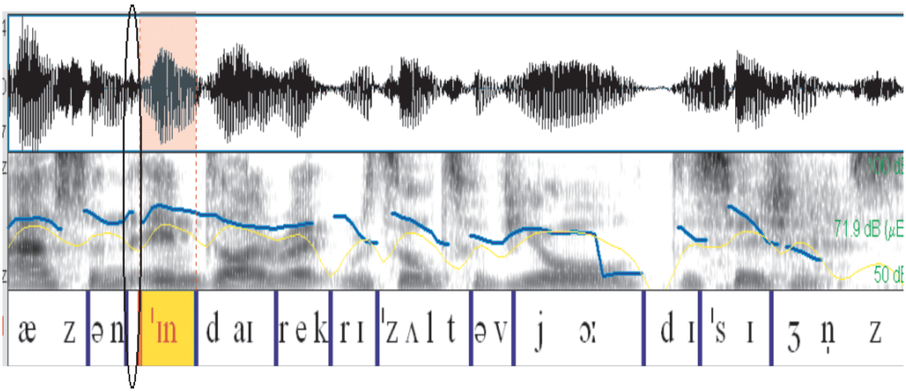


Figure 30

The prefix is made prominent through high pitch and through the presence of a glottal stop which prevents the linking process between the determiner and the adjective from taking place. There is no particular explicit contrast with an antonym in the context here (*direct* vs. *indirect results*); the prominence of the prefix is due to the strategy of the journalist who wants to be cautious in his way of addressing the former British Prime Minister; it can be characterized as 'pragmatic' prominence.

---

[33] The prominence of prefixes in speech occurs not only on separable prefixes but is also possible when the prefix is inseparable in pairs such as *increase / decrease* or *accelerate / decelerate* as in the following example (See Videau & Hanote, 2012):

> It would be very difficult to claim that Iraq was stepping up its unconventional weapons programs cos in fact Sadam Hussein was *de*celerating that effort rather than *ac*celerating it. (*Downing St. Memos Hint at Early U.S. Plans for Iraq* in *All Things Considered*, NPR, 15/01/2005)

10. Conclusion

The phonotactic processes described in this chapter are related to speech register: the more formal the discourse, the less frequently the phonotactic phenomena occur because there is a tendency to stay closer to the form of isolated words. On the contrary when the speech is rapid and colloquial, most of the phenomena described above (assimilation, elision, compression and linking) are expected to take place. As far as prominence is concerned, it was shown in this chapter that it is not the result of one particular marker but of a combination of markers and that it may affect any element of speech, whether lexical, grammatical or morphological. As Wennerstrom (1993: 313-4) states, “any syllable can be made prominent for metalinguistic purposes such as repair, clarification for the hearer or a stylistic effect”.

REFERENCES

Aronoff, Mark (1976): *Word Formation in Generative Grammar*. Cambridge & Massachussets: MIT Press.

Bauer, Laurie (1983): *English Word-Formation*. Cambridge: Cambridge University Press.

Boersma, Paul P.G. & Weenink, David (1992-2010): *Praat: doing phonetics by computer*. Amsterdam: University of Amsterdam.

Cruttenden, Alan (1997): *Intonation*. Cambridge: Cambridge University Press. (2nd ed.)

Cruttenden, Alan (2001): *Gimson's Pronunciation of English.* Arnold Publishers. (6th ed.).

Deschamps, Alain *et alii* (2004): *English Phonology and Grapho-Phonemics*. Paris: Orphys.

Fudge, Erik (1984): *English Word-Stress*. London: George Allen & Unwin Ltd.

Guierre, Lionel (1979): *Essai sur l'accentuation en anglais contemporain*. Thèse d'état, Université Paris VII.

Guierre, Lionel (1984): *Drills in English Stress-Patterns*. Paris: Armand Colin – Longman, 4th edition.

Gussenhoven, Carlos (2004): *The Phonology of Tone and Intonation.* Cambridge: Cambridge University Press.

Guyot-Talbot, Anne (2003): *Les préfixes et la reconnaissance morphologique*. Thèse de doctorat non publiée, Université Paris VII.

Herment, Sophie (2001): *L'emphase dans le discours spontané anglais: corrélats acoustiques et prosodiques*. Thèse de doctorat sous la direction de Daniel Hirst, Université d'Aix-Marseille I.

Hirst, Daniel & Di Cristo, Albert (1998): *Intonation Systems : a Survey of Twenty Languages.* Cambridge: Cambridge University Press.

Jones, Daniel (1957) [1918]: *An Outline of English Phonetics*. Cambridge: Heffer & Sons LTD. (8th ed.)

Ladd, Dwight Robert (1980): *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.

Ladd, Dwight Robert (1996): *Intonational Phonology*. Cambridge: Cambridge University Press.

Nicaise, Alain & Gray, Mark (2005) [1998]: *L'intonation de l'anglais*. Armand Colin: Editions Nathan.

Paillard, Michel (2000): *Lexicologie contrastive anglais-français. Formation des mots et construction du sens*. Gap & Paris : Ophrys.

Pavlík, Radoslav (2009): "A typology of assimilations". *SKASE Journal of Theoretical Linguistics, Vol. 6, no. 1*, 2-26. <http://www.skase.sk/Volumes/JTL13/index.html> [downloaded 20/03/2013]

Roach, Peter (2009) [1983]: *English Phonetics and Phonology*. Cambridge: Cambridge University Press. (4th ed.)
Tournier, Jean (1985): *Introduction descriptive à la lexicogénétique de l'anglais contemporain*. Paris – Genève: Champion – Slatkine.
Trevian, Ives (2003): *Morphoaccentologie et processus d'affixation de l'anglais*. Bern : Peter Lang.
Trevian, Ives (2010): *Les affixes anglais, productivité, formation de néologismes et contraintes combinatoires : De la diachronie à la synchronie*. Bern: Peter Lang.
Videau, Nicolas & Hanote, Sylvie: "Préfixation et focalisation", *In:* Cappeau, Paul & Hanote Sylvie (eds.). *Focalisation(s). Saillance dans les langues : lexique, syntaxe, prosodie*. Rennes: Presses Universitaires, 35-58.
Wells, John, Christopher (2006): *English Intonation: an Introduction*. Cambridge: Cambridge University Press.
Wennerstrom, Ann (1993): "Focus on the prefix: evidence for word-internal prosodic words". *Phonology Vol. 10 n°2*. Cambridge: University Press, 309-324.

Dictionaries:

Chambers Dictionary of Etymology (2005). Edinburgh: Chambers Harrap Publishers Ltd.
Longman Dictionary of Contemporary English (2003). London: Pearson Longman. (4th ed.)
Wells, John Christopher (2000): *Longman Pronunciation Dictionary*. London: Pearson Education Longman. (2nd ed.)
Wells, John Christopher (2008): *Longman Pronunciation Dictionary*. London: Pearson Education Longman. (3rd ed. With CD-Rom)

Audio Corpus:

BBC Radio 4: http://www.bbc.co.uk/radio4/
National Public Radio: http://www.npr.org/

PART III

New Developments in English

CHAPTER 12. BRITISH OR AMERICAN ENGLISH? A STUDY OF FOUR PHONOLOGICAL VARIABLES IN CLIFF RICHARD'S SONGS AND SPEECH OVER A PERIOD OF 50 YEARS[1]

Erika A. Larsen & Inger M. Mees
Copenhagen Business School

1. Introduction

Most sociolinguistic studies have concentrated on linguistic variation in speech, whilst the amount of research dealing with variation in singing is more limited. However, following the appearance of Trudgill's (1983) inspirational paper "Acts of conflicting identity: the sociolinguistics of British pop-song pronunciation", a number of researchers have investigated the pronunciation of popular singers.

Since the 1950s, British singers have modified their accents when singing, employing features that are characteristic of American accents. The motivation for using US pronunciations was that a large portion of pop music in the western world had American roots and that this American dominance led to imitation (Trudgill, 1983: 144). A number of theories have been proposed both by Trudgill and other scholars (e.g. Simpson, 1999; Beal, 2009; Gibson, 2010) to account for this variation. Some have seen Americanisms as an attempt to assert an American identity while others believe that the themes of songs (Le Page, 1978; Simpson, 1999), or the audience (Bell, 1984), are factors that determine the choice of accent.

[1] We wish to thank Beverley Collins (Leiden University Centre for Linguistics), Catherine Ross (Copenhagen Business School) and Christina Høøck Osorno for many insightful comments and suggestions on earlier versions of this paper.

Trudgill (1983: 141-2) mentions six features in British artists' pronunciation in songs which are different from those employed in their speech. These are the following:

- the pronunciation of intervocalic /t/ as a type of /d/ in words like *matter* (termed T Voicing, Wells, 1982: 248-51);
- the choice of the TRAP vowel in BATH words (Wells, 1982: 133-5);
- the use of non-prevocalic /r/ (i.e. rhoticity, Wells, 1982: 75-6);
- monophthongisation of the PRICE vowel /aɪ/;
- the replacement of the LOT vowel by PALM (e.g. *rock*: British English /rɒk/ vs. American /rɑk/);
- centralisation of the STRUT vowel /ʌ/.

Although most of these variants can also be found in many British English accents, there is no single variety that contains all the features and "the vast majority of singers who use these forms when singing do not do so when speaking" (Trudgill, 1983: 143). Trudgill concludes (1983: 142-3) that there can be no doubt that British singers are indeed trying to modify their linguistic behaviour when they sing.

However, later research has generally not confirmed that the use of such Americanisms is a conscious shift away from speech, e.g. Drager's (2003) study (cited in Gibson, 2010: 48). Similarly, Gibson's work on the production and perception of vowels in New Zealand popular music led him to conclude (2010: ix) that although "early research on singing pronunciation in popular music […] described the use of American pronunciation in pop music as an act of identity which involved effort and awareness (Trudgill, 1983)", his own findings "support the claims of more recent studies which suggest, conversely, that it is the use of non-American accent features which requires a wilful act of identity (Beal, 2009; O'Hanlon, 2006)".

Trudgill found that "from around 1964 on, British singers generally began trying less hard to sound like Americans" (1983: 153). It is now thirty years ago since his article appeared, and as we were interested in discovering whether this change has continued in the intervening years, it was decided to carry out a case study. As the time dimension was important, it was necessary to find a singer who had remained popular since the advent of modern pop music in the middle of the 20th century. This restricted the number of possibilities, all the more so because we wished to select an artist whose normal speech was a type of standard British English (rather than a regional variety). The singer we chose was Cliff Richard (CR), who has been highly successful in the realm of popular music since the 1950s, and for whom it is easy to obtain recordings not only of his singing but also, importantly, of his speech. On the basis of a pilot study, the first four of Trudgill's six variables were selected for analysis (see Section 2 for a more detailed description). Three issues were addressed. First, it was investigated whether or not the trend, noted by Trudgill, for Americanisms to decrease over time is ongoing. We considered a period spanning the late 1950s to the mid-2000s. Secondly, the songs were categorised according to genre – based on such factors as theme, arrangement, tempo, and phrases and idioms (American vs. British)– to discover to what extent this has an impact on the choice of variant. Finally, we examined a number of interviews with CR to find out whether it is indeed true that he modifies his pronunciation when he sings.

2. British and American English

The British prestige accent is variously known as Received Pronunciation (RP), General British (GB), BBC English, Non-regional Pronunciation (NRP), and Standard Southern British English (SSBE). It is often defined as non-localisable (e.g. Wells (1982), Roach (2004, 2009), Collins and Mees (2013), and Cruttenden (2014) and has been thoroughly described firstly by Jones (1918) and later by, amongst many others, Collins and Mees (2013), Cruttenden (2008), Roach (2004, 2009), and Wells (1982). The term RP has been in use since Jones introduced it in the revised introduction to the reprint of the second edition (later redesignated third edition) of his *English Pronouncing Dictionary* (Jones, 1926). But the suitability of the phrase is increasingly being questioned, as the accent is nowadays taken to be more encompassing than its original somewhat narrower definition, which included mainly "persons who had been educated at one of what in Britain are called public schools (actually very expensive private schools), the most famous being perhaps Eton, Harrow and Winchester (Collins & Mees, 2013: 4). To take account of the variability within RP, Wells (1982: 279-301) distinguished four types: mainstream RP (the speech of the upper middle class), upper-crust RP (the speech of the upper class), adoptive RP ("spoken by adults who did not speak RP as children", Wells, 1982: 283) and Near-RP (referring to "any accent which, while not falling within the definition of RP, nevertheless includes very little in the way of regionalisms which would enable the provenance of the speaker to be localized within England", Wells, 1982: 297). Despite this fine-tuning of the concept, there is still widespread dissatisfaction with the term RP, and from Section 3 onwards in this paper we shall therefore use the term NRP, taking it to refer to a neutral type of modern British English which lacks regional accent features.

It is difficult to say precisely which of Wells's types describes Cliff Richard's speech best. We have been unable to find recordings of his speech as a child and adolescent, but since he lived in India until the age of eight (see Section 4), he is most likely originally to have spoken a type of educated Anglo-Indian,[2] which was later modified when he and his family settled in the south of England. This would make him an adoptive RP speaker. Later recordings (from age 20) indicate that CR's pronunciation is in many respects very similar to mainstream RP; it is only his frequent voicing of intervocalic /t/ and glottalised realisations of pre-vocalic /t/ across word boundaries that would perhaps have put him in the category of Near-RP (Wells, 1982: 299).

We would *not* classify CR as a speaker of Estuary English. For the past 30 years there has been a lively debate on this new variety of British English which was first brought to the attention of the general public in an article by David Rosewarne in the *Times Educational Supplement* (Rosewarne, 1984). Estuary English refers to the variety spoken in the south east of England

---

[2] Interestingly, in a Channel 4 film documentary entitled "The Real Cliff Richard" (accessed 28 May 2012 at http://www.youtube.com/watch?v=qcYd_pE7d4c), there is no reference to Cliff Richard's accent. However, a recent newspaper article (9 February 2012) mentions Cliff Richard "being bullied at his Surrey primary school for his deep tan and Indian accent". Accessed 8 March 2012 at http://www.express.co.uk/posts/view/ 301018/Hungry-youngies-to-golden-oldies.

along the River Thames and its estuary. Though scholars have found it difficult to pinpoint the exact features of this accent, it is generally defined as being intermediate between RP and popular London speech. Estuary English is often stated (see, for instance, Altendorf 1999; Wells, 1994a) to be characterised by vocalisation of pre-consonantal and pre-pausal /l/, e.g. *silk* [sɪok], and extensive use of glottal stop, e.g. *move it over* ['muːv ɪʔ 'əʊvə]). In addition, there is a tendency towards "diphthong shift" –a term denoting changes in the starting points of the vowels, whereby front closing diphthongs shift anti-clockwise whilst back closing diphthongs move clockwise, as diagrammed in Fig. 1 (from Wells, 1982: 256). For more detail on diphthong shift, see Wells (1982: 256-7 and 306-10).

iː			uː
↓			↓
eɪ	oɪ		əʊ
↓	↑		↓
aɪ →	ɔɪ	æʊ ←	aʊ

Figure 1. Diphthong shift shown in schematic and simplified form (Wells, 1982: 256)

Wells (1994a) states that whilst Estuary English is characterised by the London features mentioned above, it does not contain any of the basilectal features of this accent, e.g. H Dropping (*head* [ed]) or TH Fronting (replacement of the dental fricatives /θ ð/ by labio-dental /f v/, e.g. *three months* ['friː 'mʌnfs], *brother* ['brʌvə]). Cliff Richard's speech displays neither of these low-status Cockney features, nor does it exhibit l-vocalisation or traces of diphthong shift (see also Section 4 below).

In the United States, the standard variety is often referred to as "General American" (GA) or "Network English" (Wells, 1982: 470). This is not a uniform accent but represents a number of different American accents without regional features and is spoken by the educated population of the Midwest (Tottie, 2002: 13). Two other major American varieties are constituted by the eastern and southern accents of the USA, which, amongst other features, both differ from GA in the absence of /r/ before consonant and pause (Wells, 1982: 469-70). However, for the present study, the most important variety of US English is African American Vernacular English (AAVE), which is heard in much of the music that has been an important source of inspiration for artists both in Britain and the United States.[3]

As stated above, four variables were investigated in our study: T Voicing; the use of /æ/ in BATH words; pronunciation of non-prevocalic /r/; and monophthongisation of the PRICE vowel. These are briefly discussed below.

British singers tend to voice /t/ (symbolised as [t̬]) in intervocalic contexts (e.g. *waiting* [ˈweɪt̬ɪŋ]) –a feature which is characteristic of American varieties (Wells, 1982: 248-50)– rather than pronouncing a voiceless [t] or glottal stop [ʔ], which are the allophones used by many British speakers. To non-Americans, voiced /t/ sounds very similar to /d/ (Wells, 1982: 248). In GA, T Voicing is heard not only between vowels, but also between a vowel and syllabic /l, r/ (e.g. *little* [ˈlɪt̬l], *letter* [ˈlet̬r̩]), and between /r/ and a vowel (e.g. *hurting* [ˈhɜːrt̬ɪŋ]). Furthermore, it occurs both word-internally (e.g. *Betty* [ˈbet̬i]) and across word boundaries (*bet it* [ˈbet̬ ɪt]). T Voicing is also found in RP, but its occurrence is far more restricted. However, it can be heard regularly in high-frequency words and phrases, particularly across word boundaries, e.g. *pretty, better, but I, that I, not a, what a, lot of* (Collins & Mees, 2013: 90). It is also a characteristic feature of London speech (Wells, 1982: 324-5).

When pronouncing the BATH words –a set of words in which orthographic <a> is followed by a voiceless fricative (e.g. *laugh, bath, pass*) or a nasal cluster (e.g. *chance, example*, *advantage*)– British singers do not select the PALM vowel, /ɑː/, which is the norm in RP and the south-east of England, but TRAP, /æ/, which is heard in the United States and Canada (Trudgill & Hannah, 2002: 12, 38). Thus a word like *dance* is said as /dæns/ instead of /dɑːns/. In Britain, the TRAP vowel is also often selected for the BATH words in northern, West Country and Welsh English accents –though not in all items belonging to this set. For instance, in the north of England words like *can't* and *half* take the PALM vowel while other words (e.g. *dance* and *laugh*) take

[3] See, for instance, the *Encyclopaedia Britannica* articles on African Americans, popular music and rock.

TRAP (Trudgill, 1983: 142). For more detail on the BATH words, see Wells (1982: 78, 133-5, 232-4) and Trudgill & Hannah (2002: 12).

There is a tendency for British singers to pronounce /r/ in pre-consonantal and pre-pausal position (e.g. *girl*, *car*) even if their own speech follows the pattern of non-rhotic varieties –thus only pronouncing /r/ when it occurs before a vowel (e.g. *red, very, far away*). GA is rhotic, whilst RP and most varieties found in England (except the West Country and parts of Lancashire) are non-rhotic (Trudgill & Hannah, 2002: 14). It is interesting that for non-rhotic composers and singers *data* ['deɪtə] would therefore rhyme with *later* ['leɪtə], and *Anna* ['ænə] with *manner* ['mænə], whereas pairs like these do not rhyme for users of rhotic varieties ['deɪt̬ə -'leɪt̬r̬], ['ænə – 'mænər] (cf. Trudgill, 1983: 153). As stated above, southern varieties of the USA are non-rhotic and the same holds true for AAVE (Wells, 1982: 557).

The PRICE vowel, which in most British accents is realised as a diphthong, is often pronounced as [aː] in pop and rock songs. Thus *high* is said as [haː], whereas British English normally has [haɪ] (Trudgill & Hannah, 2002: 41). The three variables previously discussed are characteristic of most accents of American, including the standard variety, but the fourth –monophthongisation of PRICE– is not heard in GA. However, it is widespread in the American South and in AAVE (Edwards, 1997: 204; Wells, 1982: 557) and, in fact, it has even been termed "the confederate vowel" (Underwood, 1988: 421). In some regions,[4] [aː] is employed only word-finally and before voiced consonants (e.g. *lie, cry, eyes, time*) whilst a diphthong is used before voiceless consonants, e.g. *night, life* (Trudgill & Hannah, 2002: 41). In these accents, the expression *At my time of life* would consequently be rendered as [ət 'maː 'taːm əv 'laɪf].

3. Song vs. speech

Song and speech are two different modalities that are not immediately comparable. Morrissey (2008) is one of the few who has concentrated on this aspect: "As any vocalist appreciates, there are certain speech sounds

[4] No precise indication of the regions is given by Trudgill & Hannah.

that lend themselves better to singing than others, because, to put it simply, they ‘carry’ the tune” (2008: 211-12). Morrissey’s point of departure is Trudgill’s (1983) concluding remark that the theories proposed in his paper still cannot account for why certain American features are retained whilst others are replaced by British variants. He quotes Trudgill’s example of the Clash, who sing [kɑːnt gɛt̬ ə'hɛd] instead of having either [kɑːnt] and [gɛʔ ə'hɛd] (British English) or [kænt] and [gɛt̬ ə'hɛd] (American English) (Trudgill, 1983: 159-60). Morrissey (2008: 211) reminds us that the data investigated are actual performances of songs, which are subject to constraints other than sociolinguistic factors, e.g. the lyrics, tempo and rhythm.

Some speech sounds are more sonorous than others, which means that they are perceived as louder than other sounds with the same length, stress, and pitch (Ladefoged, 1982: 221, 284). They have more carrying power and require less energy to produce (Hawkins, 1984: 98) and are therefore more suited to singing as they make it easier for the singer to be heard above the accompaniment (Morrissey, 2008: 212). The sonority of the sounds is related to the degree of openness of the oral, nasal and pharyngeal cavities. The list below shows the degree of sonority of the English phonemes, ranked in order from most to least sonorous (after Hawkins, 1984: 99, and Burquest & Payne, 1993: 101).

1) Open vowels /æ ɑ ɒ ʌ ɔ/
2) Close vowels /ɪ e u ʊ ə/
3) Semi-vowels /w j/
4) Liquids /l r/
5) Nasals /m n ŋ/
6) Voiced fricatives /z v ð/
7) Voiceless fricatives /s f θ/
8) Voiced stops /b d g/
9) Voiceless stops /p t k/

It can be seen from the list that one of the four variables studied here (the choice of TRAP or PALM in BATH words) is equally sonorous in both British and American varieties (both alternatives being open vowels). Two of the

remaining features are more sonorous when sung with a pronunciation inspired by GA/AAVE than with NRP: /aɪ/ reduced to [aː] (more sonorous in AAVE),[5] and intervocalic /t/ realised as [t̬] (more sonorous in both GA and AAVE). With respect to the latter, Morrissey (2008: 213) concludes that the American variant "is relatively sonorous and thus more singable, which is why we can take it as the unmarked form, particularly in *legato* phrasings".
As regards our final variable (r), Morrissey (2008: 213) observes that

> Singability may also have an impact on rhoticity [...]. A postvocalic /r/ represents a closure in comparison to the position of tongue and mouth during the pronunciation of the vowel. For the singer this creates two problems: firstly, the closure reduces the opening for the air flow, thus reducing sonority, and, secondly, it requires a decision as to when the tongue should begin to move from the vowel constellation towards the alveolar approximation of the [ɹ]. For a singer, non-rhoticity may therefore be preferable, particularly in sustained notes […]. [These two problems] may be a reason why non-rhotic realisation is relatively widespread, especially in monosyllabic words, at the end of a line in the lyrics, or generally in a sustained note.

Apart from the issue of sonority, it should be noted that non-rhoticity is not necessarily a deliberate choice to sound British, as the American variant in this case is more difficult to imitate for a person with a non-rhotic accent. In a discussion of Le Page's theory of linguistic behaviour (Le Page, 1978, cited in Trudgill, 1983), Trudgill quotes and elaborates on four factors which are assumed to have an impact on the degree to which singers are actually able to modify their linguistic behaviour. These constraints can to an extent explain why artists sometimes fail to hit the target, particularly with respect to the distribution of /r/. The degree of success is governed by the singers' ability to (1) identify the group; are they aiming at a rhotic (GA) or non-rhotic variety (AAVE)?; (2) access to the groups and ability to analyse their speech patterns (failure to do so may lead to hypercorrect /r/; see below); (3) conflicting motivations towards different models (Trudgill showed that the

---

[5] Diphthongs are not mentioned in the list, but it can be concluded that the PRICE vowel is less sonorous if the final element is realised with a closer vowel like [i], [ɪ] or [ə] than if pronounced as a steady-state open vowel. Furthermore, as Morrissey (2010: 214) observes, "notes sustained on /iː/ create a screechy, rather unpleasant sound and are therefore avoided".

strength of the motivation towards the American model was reduced round about 1964); (4) their ability to modify their linguistic behaviour, which perhaps decreases with age (Trudgill, 1983: 145-54).

As Trudgill states, "[we] can assume […] that many singers would pronounce all non-prevocalic /r/s if they could [but] in the flow of the song, they are not consistently able to do so" (Trudgill, 1983: 149-150). Even if the singers identify the accent they wish to imitate as "American" and therefore sound more /r/s than in their normal speech, they may fail to analyse the model correctly, and thus their usage of /r/ occasionally overshoots the mark, so that /r/ is also pronounced before consonant and pause in cases where it does not occur in the spelling. It is, for instance, heard in Cliff Richard's "Bachelor Boy", where *a bachelor boy* is sung as /ə**r** 'bætʃələr bɔɪ/ (Trudgill, 1983: 148-9).

As the American variants are on the whole easier to pronounce from the point of view of singing, they can be regarded as the unmarked forms in pop songs (Morrissey, 2008: 213). So if an artist sings, for instance, a voiceless [t] or a glottal stop, it is likely to be a conscious choice to use a British pronunciation (i.e. a marked form).

Although it would be overstating matters to say that the factor of sonority can explain everything, it does provide another possible reason for some of the pronunciation inconsistencies pointed out by Trudgill (1983) and Simpson (1999).

4. Cliff Richard: biographical details

Sir Cliff Richard (real name Harry Webb) was born at Lucknow, India, in 1940, of British parentage. Until 1948 the family lived in India, where his father worked for a catering firm, before they moved to Cheshunt in Hertfordshire, to the north of London. In his autobiography, CR explains that they went from great wealth (servants, very comfortable circumstances) to total poverty. They slept on the floor in one room for a year, and only survived through the help of friends and relatives. When they eventually acquired a small government-subsidised council house, they regarded this as the height of luxury (Richard, 2008). CR left school at the age of 16, after which he was employed by the same factory in which his parents had found work. His first record, *Move it*, appeared in 1958. Ever since he has been active in the music industry and has recorded material from many different genres. For the first ten years of his career, he worked with his band the Drifters (later renamed the Shadows), and mainly recorded rock and roll. In the late 1960s and early 1970s he moved on to folk songs and religious material before returning to classic rock in 1975 (Turner, 2008).

According to one biographer, Steve Turner, Cliff Richard, remarkably, did not adopt the North London accent of his peers when he moved to England. By 1960, when he was 20, his accent “was becoming more well-rounded” and he “was losing the Anglo-Indian lilt that he had brought to Britain” (Turner, 2008: 160). This was probably the result of a change of image. As early as 1959, when he produced his second album, he had moved from rock and roll to more mainstream songs in order to become an “all-round entertainer” and so have a better chance of enjoying a long-lasting career (Turner, 2008: 159).

Given this information, one would not expect CR’s speech to manifest Indian English features such as retroflex consonants, /θ ð/ replaced by /t d/ (a process called TH Stopping, Wells, 1982), and confusion of /v/ and /w/ (Collins & Mees, 2013: 195). Nor would he be likely to produce regional London variants (Wells, 1982: 301-34), e.g. H Dropping (deletion of /h/, e.g. *help* pronounced as [elp]), TH Fronting (*three = free; further = fervour*), intervocalic /t/ pronounced as [ʔ] (*water* [ˈwɔːʔə]), or diphthong shift. A cursory examination of his speech from those days revealed that features such as the above were indeed absent or, if they had been part of his idiolect, he had removed them by the end of his teens. In one of the earliest recordings of CR’s speech we have been able to find (a TV interview from 1961),[6] ***h**appens* and ***h**ad* are pronounced with /h/, *chari**t**ies* with voiceless [t], *Por**t** Elizabeth* and *forge**t** it’s* with [t̬], and ***th**ink, **th**ings, **th**ree* with /θ/; the pronunciation of the diphthongs FACE, GOAT, PRICE, MOUTH /eɪ, əʊ, aɪ, aʊ/ in words like *w**a**ve, sunb**a**thing*, *m**ai**nly; sh**o**ws, m**o**st; wh**i**tes, t**i**me*; *(Cape)*

[6] The Newsreels of Cliff Richard 1961-1978, on *His Land* [DVD] (2003).

Town, found is fully NRP. The only feature characteristic of Indian English is the relatively unaspirated nature of stressed syllable-initial /p t k/ (and corresponding relatively voiced /b d g/; see Collins & Mees, 2013: 195). The interview also showed that CR pronounced the variables studied here in accordance with an NRP model and not as GA/AAVE. He has the PALM vowel in BATH words (*ask, dance, can't),* does not sound non-prevocalic /r/ (*forget, normal, fair)*, and realises the PRICE vowel as a diphthong [aɪ] in *time* and *whites*. Word-internal intervocalic /t/ is voiceless (as in NRP), although word-final intervocalic /t/ is voiced (NRP or Near-RP); see also Sections 2 and 6.3.

As regards his music, there can be no doubt that at the beginning of his career CR belonged to the group of singers who had an American pronunciation as their model. He has said in countless interviews that American singers like Buddy Holly, Ricky Nelson, and especially Elvis Presley, inspired him to start a career of his own and to imitate them.

5. Methodology

5.1 Research questions

As stated in our introduction, the study was guided by three research questions:

1) Has the use of American pronunciations in CR's songs decreased over time? (This was investigated to discover if Trudgill's 1983 finding that British singers generally began sounding less like Americans from the mid 1960s onwards could be confirmed).

2) Alternatively, does genre –based on such factors as theme, arrangement, tempo, and phrases and idioms– affect CR's choice of a British or American variant in his songs?

3) Can the use of American variants found in CR's songs also be identified in his free speech?

5.2 The variables and their calculations

A linguistic variable (enclosed in parentheses) can be defined as “a linguistic item with two or more variants involved in covariation with other social and/or linguistic variables” (Chambers & Trudgill, 1998: 50), e.g. (t) and (r). Phonological variables which concern consonants tend to have discrete variants while vowels are often continuous, but they can be categorised as discrete variants to enable calculations (Milroy & Gordon, 2003: 138-9). All the variables in this study –(r), (t), (æ) and (ai)– were discrete with two possibilities only. For (r), it was a question of whether **<r>** was present or absent, that is, [r] or 0 (i.e. zero), while the other variables all had two possible forms: (t): [t] or [t̬] ; (æ): [ɑː] or [æ]; and (ai): [aɪ] or [aː].

The calculations were performed as follows. First, all the occurrences of a particular variable were registered; for example, in the case of the variable (ai), it was necessary to find all words that contained a PRICE vowel. We listened to the realisations, and subsequently each token was assigned the value 0 for a British pronunciation and the value 1 for an American pronunciation. Thus the pronunciation [aɪ] (the NRP variant) was given the value 0, while [aː] (the AAVE variant) was assigned the value 1. Subsequently, the percentage of American realisations was calculated for the songs in each set. For example, in Set 1 there were 44 words containing (ai), 14 of which were pronounced with the AAVE variant [aː]: 14/44*100 = 31.8 per cent. The other variables were computed in the same manner.

5.3 Data

The empirical foundation for this study is formed by 25 songs by Cliff Richard representing different genres and periods (see Larsen, 2011) as well as interviews from various decades. As a first step, CR’s recordings were divided into five consecutive periods; subsequently, one album was chosen from each period and from each album five songs were selected for analysis. The criterion for selection was for the songs to have been released as singles, as these often reflect the image the singer wishes to portray. However, not all albums contained five singles, so the remaining songs were chosen by lot. The most interesting variables were identified in a pilot study and, as stated above, four were chosen which have also been investigated by Trudgill and others.

The first set was recorded in 1958-59, Set 2 in 1968-69, Set 3 in 1975-76, Set 4 in 1995, and Set 5 in 2004. Set 1, unlike the remaining sets, does not contain tracks from albums, as songs recorded in this era were typically released as singles. The criterion for selection was whether the songs ranked among the top ten on the British hit lists. The first three songs are rock and roll; the last two mark a style shift towards slow ballads. The American rock influence is clearly manifest, and the lyrics contain many American slang expressions of the period (*baby, groove, shake, cat* 'player or devotee of jazz').

The second set is taken from the album *Sincerely*, which was the first album CR recorded without his backing band the Shadows. In the middle of the 1960s, he had become a Christian and devoted much of his time to religious activities, even considering giving up his career as a singer. The genre is still pop, but no longer rock 'n' roll, but rather "folk-rock-gospel-orientated" (Lewry & Goodall, 2004: 6). The lyrics no longer contain American expressions. One of the songs, *London's Not Too Far*, is clearly directed at a more British audience.

The third set of songs comes from the album *I'm Nearly Famous*. At the beginning of the 1970s, CR was uncertain as to what direction his career should take. His manager worried about his falling record sales and therefore attempted to find a new producer. He chose Bruce Welch, a former member of the Shadows, who had collected different songs, notably in the USA. The songs are more rock than those he had recorded earlier. The album sold well, both in Britain and the United States, and was by several reviewers called Cliff's renaissance (Lewry & Goodall, 2001: 9). The influence from the 1950s is clear.

The songs in Set 4 are taken from the album *Songs from Heathcliff*. They were part of CR's musical *Heathcliff*, based on Emily Brontë's book *Wuthering Heights* from 1847. The action is set in eighteenth-century Yorkshire, and CR changed his pronunciation on stage in the direction of a northern English accent although, as one reviewer wrote, it sounded like a mixture of Weybridge (his home, approximately 25 kilometres south west of London), New York and Macclesfield (Bailey, 1996). The songs belong to a mainstream pop genre. In terms of content, they are different from the other material chosen for this study, as they stem from a musical and have to relate to the plot. There are no American expressions, and in certain places the language is outdated as some of the lines have been borrowed directly from the novel. Here we are dealing with two conflicting models, namely pop song vs. north of England.

The songs in Set 5 are from the album *Something's Goin' On* from 2004. The album was largely recorded in Nashville, USA, and of the approximately 20 songwriters only one had a British background. CR said that he wanted an album in the style of *Shania Twain meets ZZ Top* (Turner, 2008: 353), a mixture of pop and country with blues and rock. Here he clearly returns to the roots of rock and roll music. American places and expressions are mentioned (*Albuquerque, yada* 'boring or empty talk'*, cup of joe* 'coffee'). In summary, Sets 1, 3 and 5 are more explicitly American than Sets 2 and 4; notably Set 4 is clearly British inspired.

6. Results

Our first research question aimed at determining if the use of American forms by CR decreased at the same time as a British tradition for pop music was becoming more firmly established. If this was indeed the case, there should be a gradually declining number of American features from Sets 1 to 5 (see Section 6.1 for results). Alternatively, if there is a relationship between genre and accent –irrespective of the period in which the songs were recorded– it can be assumed that songs relating to the USA will have more American features than will other genres. Here one would expect Sets 1, 3 and 5 to display a larger number of American features than Sets 2 and 4 (Section 6.2 below). Finally, we also looked at the correspondences between

CR's pronunciation in interviews and songs to discover to what extent the American features in his singing were carried over from his natural speech (Section 6.3).

6.1 The time dimension

As stated above, Trudgill observed that from the middle of the 1960s, the motivation to sound American decreased, even with artists who made their breakthrough in this decade. He analysed 14 albums recorded between 1963 and 1970: ten by the Beatles and four by the Rolling Stones. For example, the Beatles exhibited a large decrease in the use of /r/ from 47 per cent in 1963 to a mere 3 per cent in 1970. The Rolling Stones showed a drop from virtually 50 per cent in 1964 to less than 20 per cent in 1967 (Trudgill, 1983: 151-3).

Table 1 presents the percentage of GA/AAVE realisations in the five sets chosen from Cliff Richard's repertoire.

Table 1. Percentage of American realisations. Songs categorised in periods of ten years.[7] In the table, (r) represents the percentage of non-prevocalic /r/; (t) the percentage of voiced /t/; (ai) the percentage of [aː]; and (æ) the percentage of TRAP in BATH words.

	(r)	(t)	(ai)	(æ)
Set 1 (mid 50s)	11.9%	100%	31.8%	100%
Set 2 (late 60s)	2.6%	100%	28.3%	100%
Set 3 (mid 70s)	4.8%	100%	27.3%	100%
Set 4 (mid 90s)	24%	87.5%	5.8%	100%
Set 5 (mid 00s)	10.4%	100%	11.1%	100%
Mean	10.7%	97.5%	20.9%	100%

[7] No songs from the 1980s were included. From the 1970s to the mid 1990s Cliff Richard followed a normal pop career, recording songs he thought would make the hit lists. Consequently, there were no dramatic changes in style in this period.

It can be seen that CR's use of the TRAP vowel in BATH words remains at 100 per cent throughout his singing career. This is remarkable as he consistently employs a PALM vowel in his speech (see Section 6.3). Similarly, the pronunciation of /t/ is virtually completely stable, being voiced in all sets except for a negligible drop in Set 4. (Interestingly, CR's spoken language exhibits both voiceless, voiced and glottalised allophones for intervocalic /t/, see Section 6.3.). Thus for these two variables there is no decrease in the use of American variants over time.

Let us now consider the development of the two other variables. Remarkably, at no stage over a period of 50 years does /r/ in non-prevocalic position occur more than a quarter of the time (i.e., in the 1990s). It is notable that in the era when pop singers were most intent on imitating American (in the 1950s and early 1960s), CR uses post-vocalic /r/ in no more than 12 per cent of the possible cases. It should be borne in mind that even though General American is rhotic, AAVE is non-rhotic (and thus similar to NRP in this respect), and it is conceivable that CR was imitating AAVE rather than GA, in which case one would expect a low percentage of non-prevocalic /r/. Nevertheless, this does not explain why the highest percentage of non-prevocalic /r/ is found in Set 4, which was the most British set in the corpus; see also Section 6.2, where the songs have been categorised according to features of genre (British-inspired or American-inspired).

For (ai), it is important to note that the pronunciation of the diphthong is the same in both NRP and GA. So here it is only possible to measure whether there has been any difference in the amount of AAVE pronunciation. The analyses revealed that this is the only variable for which there is a change over time. For whatever reason, there appears to be a marked drop in the use of the AAVE variant after the 1970s. In this case, it is clearly not the genre that elicits a higher number of AAVE variants, as the lowest percentage of AAVE [aː] is found in both Sets 4 (British) and 5 (American-inspired). While the British set contains virtually only diphthongal realisations, which is in accordance with what one would predict, this is surprisingly also the case for the American set, where one would have expected a higher number of monophthongal variants, but in fact only 11 per cent of the PRICE words are actually realised with [aː] (see also Section 6.2 below).

Altogether there is no clear evidence that American pronunciations have decreased in the course of time. Two variables (T Voicing and BATH words) remain almost totally American whatever era the songs were recorded in, and one (non-prevocalic /r/) reveals inconsistent usage. Only (ai) shows a decrease in AAVE pronunciation over time.

One explanation for the stability of American pronunciations across time has been proposed by Beal (2009). She suggests that whilst GA/AAVE features from the 1960s were copied by British artists because they were associated with the USA, they have in the meantime become a genre convention which is associated with "mainstream pop", so that "any British artist (or indeed amateur performer) singing a mainstream pop song will employ these features as a matter of course" without involving any conscious act of identity (Beal, 2009: 229). This means that the American pronunciation has become the unmarked variant and that any other accent will be perceived as marked or deviating from the convention, and consequently used only with a specific purpose in mind. We noted earlier (Section 3) that Morrissey also concluded that American variants were the "unmarked" forms though he used the label in connection with the issue of sonority rather than the question of an unconscious manifestation of identity.

6.2 Songs re-categorised according to genre

Having established that the use of US variants on the whole does not decrease over time, we now proceed to explore another possibility, namely that the pronunciations opted for by British singers are determined by the genre of the material. One of the communication theories Trudgill proposes as an explanation for the reason singers modify their accents is the sociolinguistic concept of "appropriateness" (similar to Halliday, McIntosh & Strevens's 1964 "register analysis"; see Simpson, 1999). As Trudgill (1983: 143) observes:

> As is well known, different situations, different topics, different genres require different linguistic styles and registers. The singing of pop music in

> this way, it could be argued, is no different from vicars preaching in the register appropriate to Church of England sermons, or BBC newsreaders employing the variety appropriate for the reading of the news.

To test the model of appropriateness, the songs were re-grouped taking into account factors such as genre (e.g. folk song, country, rock), location and subject matter, phrases and idioms, arrangement, and tempo. The following categorisation was arrived at (the grounds on which the classification is founded are shown in brackets after the song title).

GB
You'll want me (1969) (arrangement, GB writers)
London's not too far (1969) (arrangement, GB folk song, GB writer, GB place)
I'm not getting married (1969) (GB writers)
The following five songs are all from the musical *Heathcliff* (written by Cliff Richard himself with the aid of British theatre director Frank Dunlop, lyrics by Tim Rice, music composed by John Farrar), based on the well-known British novel, action takes place in Yorkshire, England.
Misunderstood man (1995)
Had to be (1995)
I do not love you Isabella (1995)
Marked with death (1995)
Be with me always (1995)

US
Move it (1958) (rock)
High class baby (1958) (rock)
Mean streak (1959) (rock)
Living doll (1959) (country-ish)
Travellin' light (1959) (country-ish), US writers
All the above were composed by GB songwriters (except *Travellin' light*), but show a clear US influence (themes, arrangement, idioms)
What car (2004) (US written, recorded, US r'n'r themes)

A thousand miles to go (2004) (US written and recorded, r'n'r theme, mentions US places).

There was insufficient information to classify the remaining songs, and these were therefore excluded from the new calculations.

We saw in Table 1 that the choice of the TRAP vowel for BATH words was 100 per cent throughout, so the re-categorisation naturally does not affect this score. The percentages for T Voicing are 93 per cent and 100 per cent for the British and American sets respectively, but note that the British score is reduced owing to a single occurrence of voiceless [t] in one song ("I do not love you Isabella"), and if this instance were disregarded, both US and GB results would amount to 100 per cent.

Table 2. Number and percentage of American realisations. Songs categorised according to genre (GB or US)

GB	(r)	(t)	(ai)	(æ)
You'll want me	0/13	-	0/2	-
London's not too far	0/26	-	2/10	-
I'm not getting married	0/14	6/6	3/14	1/1
Had to Be	5/18	1/1	2/7	-
Misunderstood man	9/37	6/6	1/8	-
I do not love you Isabella	4/18	0/1	0/8	1/1
Marked with death	4/47	-	0/15	-
Be with me always	14/30	-	0/14	1/1
Total	36/203	13/14	8/78	3/3
Total (in per cent)	17.7%	92.9%	10.3%	100%

US	(r)	(t)	(ai)	(æ)
Move it	0/4	2/2	2/4	-
High class baby	0/11	7/7	12/12	9/9
Mean streak	0/19	4/4	0/4	2/2
Living doll	5/7	3/3	0/12	-
Travellin' light	0/1	2/2	0/12	4/4
What car	5/29	3/3	0/2	4/4
A thousand miles to go	6/36	6/6	5/20	7/7
Total	16/107	27/27	19/66	26/26
Total (in per cent)	15%	100%	28.8%	100%

From Table 2 it can be seen that the degree of rhoticity is very similar in the two sets. Since it was also impossible to discern a distinct pattern when the songs were classified chronologically, it can be concluded that for the variable (r) neither the time dimension nor the genre can be adduced as an explanation for the distribution found. The most notable finding is perhaps that the occurrence of non-prevocalic /r/ is so limited. Trudgill's (1983) figures for this variable in the songs performed by the Beatles and the Rolling Stones are much higher than those for CR.

The only clear difference is again to be found in the (ai) variable, the US-inspired songs having close to 30 per cent monophthongal (AAVE) realisations as compared with the British songs, which only attain a score of 10 per cent. Here it should be noted that five of the seven songs constituting the American set are from the earliest decade (1950s), which was where the number of [aː] realisations was highest in our chronological classification (31.8 per cent see Table 1), and therefore it is difficult to say whether the variation in the use of [aː] is a feature that diminishes over time or whether it is related to the genre of the songs, or both.

6.3 Comparison of Cliff Richard's pronunciation in songs and interviews

In order to establish to what extent Cliff Richard's use of American variants is restricted to his singing, we listened to some samples of his free speech. A number of interviews from different periods of his career were selected and analysed to determine whether the pronunciation patterns found in his songs could also be identified in his speech. The earliest recording we have been able to trace dates from 1960; in addition, we listened to interviews from 1969, the early 1970s, 1983, 1996, 2003 and 2009.[8]

---

[8] The interviews can be accessed at http://www.amazon.co.uk/Cliffs-Personal-Message-to-You/dp/B004RLT65S (1960); http://pd.nrk.no/h264/Historiefortelling/2009-11-17/Cliff_Richard_og_Einar_Lunde.mp4 (1969); http://www.hadland.me.uk/interviews.htm (early 1970s); http://www.youtube.com/watch?v=hXW997Iv2sI (1983) http://www.youtube.com/watch?v=P2TYwQ6TD0Q (1996); http://www.mattblank.com/interviews/cliff-richard/ (2003); http://www.youtube.com/watch?v=akx74_WLzBg (2009).

In his spontaneous speech, CR systematically chooses the PALM vowel for BATH words throughout his career. In none of the interviews did we hear a TRAP vowel. In the 1960 recording, three BATH words were found, *can't, glass* and *task*; all were pronounced with PALM. The 1969 interview contained the words *rather, ask,* and *answer*, all of which were again said with the PALM vowel. No BATH words occurred in the interview from the early 1970s. In 1983 there were another three instances of BATH words (*glass, after,* and *can't*). Again CR selected the PALM vowel in all cases. Similarly, in 1996, CR opted for PALM in all instances (*answer, rather, halfway,* and *dance*). Finally, the interviews from 2003 and 2009 feature the words *can't, autograph, example, after* and *dancers, danced, can't*, respectively, all of which are again pronounced with the PALM vowel. Since CR uses TRAP systematically in his songs, there can be no doubt that this, therefore, is a conscious choice.

The pronunciation of pre-consonantal and pre-pausal /r/ is absent from CR's free speech. In other words, his pronunciation is consistently non-rhotic. Note that he does employ both linking and intrusive *r* (e.g. *never ever* ['nevər 'evə], *Calcutta, I (think)* [kæl'kʌtər aɪ] in the 1983 interview), both of which are characteristic of non-rhotic rather than rhotic accents (Trudgill & Hannah, 2002: 14-15; Wells, 1982: 223). Even though CR does not adopt non-prevocalic /r/ systematically in his songs, we found that it occurred up to a maximum of 25 per cent of the time, which –like the use of TRAP for BATH words– would appear to indicate that it is a deliberate (or perhaps semi-deliberate) choice even if it is neither employed consistently nor always correctly.

Monophthongal realisations of PRICE appear to be absent from CR's free speech except in a few cases where we are dealing with a high-frequency item in unstressed position (e.g. the personal pronoun *I,* which even by NRP speakers is often pronounced as [a]; Wells, 2008: 397). Otherwise, our impressionistic observations suggest that this AAVE variant is not to be heard in his normal speech.

We can conclude that these three features have been adopted –albeit in some cases to a limited extent– to sound more GA/AAVE. However, the use of voiced intervocalic /t/ in the free speech material from 1960 to 2009 revealed a more complex and diverse picture. CR uses three different allophones for this consonant: [t], [t̬] and [ʔ]. The last of these, generally referred to as “T Glottalling”, is the replacement of an alveolar articulation by a glottal articulation. The phenomenon occurs in a range of environments. Use of T Glottalling in syllable-final pre-consonantal position has now become widespread in present-day NRP whilst the adoption in other contexts (pre-pausal, pre-vocalic) is gradually gaining prestige, at least if employed in limited quantities. As Wells (1994b: 201) puts it:

> [Glottalling] is by now very firmly established in casual RP before obstruents (*football* [ˈfʊʔbɔːl], *it's quite good* [ɪʔs kwaɪʔ gʊd]) and is increasingly heard before other consonants (*atmosphere* [ˈæʔməsfɪə], *partly* [ˈpɑːʔli], *Gatwick* [ˈgæʔwɪk]). Among younger RP-speakers it can even be heard finally before vowels (*pick it up* [pɪk ɪʔ ʌp]) or in absolute final position (*Let's start!* [leʔs stɑːʔ]). Intervocalically within a word, it remains firmly excluded from RP (cf. Cockney *city* [ˈsɪʔi]). Nevertheless, the increased use of glottal stops within RP may reasonably be attributed to influence from Cockney and other working-class urban speech. What started as a vulgarism is becoming respectable.

Wells’s observations gain empirical support from a study conducted by Fabricius (2000) on the use of T Glottalling amongst ex-public school university students at Cambridge University. She concluded that it “has to some extent lost its stigma, *but not yet acquired prestige*, in word-final pre-pausal and pre-vocalic environments” (Fabricius, 2000: 145).

Our earliest free speech recording of Cliff Richard (1960) contains the following instances of intervocalic /t/: *right away*, *pretty* (which occurred twice), *shatter*. In all cases, CR pronounces a voiced /t/.[9] Thus rather than introducing voiced /t/ as an Americanism in his songs, we would posit that it is likely that CR transfers a Near-RP or London feature (Wells, 1982: 299, 324-5) that is already present in his own speech. Whilst Cockney (the most basilectal variety of London speech) would opt for a glottal stop in this environment, T Voicing is a feature to be heard from London speakers somewhat higher up the social scale (Sivertsen, 1960: 119) – though not at the very top, where voiceless variants are found, as in NRP. Sivertsen stated

[9] A final instance of intervocalic /t/ is found in *sort of*, but here /t/ was elided.

that her informants from the Bethnal Green district of the East End of London considered [ˈbetˢə] (i.e. /t/ realised with affrication) for *better* "posh", [ˈbeṱə] "normal", and [ˈbeʔə] "rough". That Cliff Richard's use of voiced /t/ is not a fad from the 1960s but is a feature that remains part of his natural speech throughout his life can be confirmed if one listens to a relatively recent interview (recorded over 40 years later in 2003), from which it is evident that he still employs a large amount of T Voicing both word-internally (*better, greater, British, Peter, autograph, matter, getting*), and word-finally across word-boundaries (*get a, get it, but I, bit of, about it*).

The interview from 1983 is slightly different from those in the other decades in that there are more glottalised realisations of word-final intervocalic /t/ –in fact, approximately a third of such tokens are glottalised. Note that CR's use of this variant is indeed restricted to word-final contexts and never occurs in word-internal position, which means that in this respect his speech falls within the range of NRP or Near-RP, and so never approaches Cockney. Approximately ten per cent of the occurrences are voiceless (these primarily being found word-internally, e.g. *uprooted, waiting*). Nevertheless, despite the higher number of glottalised and voiceless variants, the overwhelming majority of occurrences remain voiced. We can therefore conclude that the use of T Voicing in CR's songs is perhaps not so much a deliberate choice to sound American as simply a feature carried over from his normal speech.

7. Conclusion

In this chapter, we have analysed stylistic variation in Cliff Richard's songs from five different decades to study to what extent American features are employed throughout his long career. In addition, we have considered the influence of genre on the choice of variant. Finally, we have attempted to discover which, if any, of these features can also be found in his free speech.

In the songs, the vowel in the BATH words was consistently realised as /æ/, and intervocalic /t/s were voiced in all cases except one. On the other hand, /aɪ/-monophthongisation showed considerable variation, diminishing markedly in the most recent years. Non-prevocalic /r/ was pronounced in approximately 10 per cent of the tokens in the first and last sets with both lower and higher percentages in the intervening sets, though it was never higher than 25 per cent. Thus it has been shown that there is no clear reduction of GA/AAVE features in CR's songs. The only variable where the GA/AAVE realisations have decreased is (ai). We can conclude that the time dimension does not appear to have a major effect on the degree to which Americanisms are used.

Nor can conclusive support be found for a relationship between genre and accent when the songs are re-categorised as "British" or "American" in terms of theme, arrangement, place, and vocabulary (which in the present analysis meant that a more British-inspired content was expected to result in fewer GA/AAVE features). The American rhotic pattern is in fact higher in one of the British sets, while the percentage of TRAP in BATH words and voiced /t/ is to all intents and purposes identical in both the British and American sets, and thus the only feature that stands out is the monophthongal AAVE realisation of (ai) in the more American sets. This, however, was also the only variable that exhibited a systematic pattern according to chronological time and consequently neither of the classifications yields clear results.

The data to be found in Section 6 seem to confirm that singing is different to speaking. This may be in part because of sonority, which places another constraint on the singer's ability to produce the desired pronunciation, and as demonstrated earlier, the GA/AAVE-inspired pronunciations of two of the four variables examined in this paper –[t̬] and [aɪ]– are more sonorous than their NRP counterparts and one is equally sonorous in both varieties (the choice of TRAP for BATH words). Thus sonority may well also have a degree of influence on Cliff Richard's singing accent. However, its importance should not be overemphasised. Gibson, who explored this issue and indeed found some evidence for a sonority-related effect, concluded that "singing inherent effects like this can only explain a portion of the variability between singing and speaking. Most of the differences between singing and speech appear to be caused by social and stylistic motivations" (Gibson, 2010: ix).

Our findings indicate that Beal's observation that an American pronunciation has become a genre convention for pop/rock songs is most likely to be correct, and therefore the American model is still relatively prominent in the newest material and in the more British-orientated sets. Notably /æ/ in BATH words is a consistent feature. Consequently, it appears that pop singers automatically sing with certain GA/AAVE features because non-American features in singing are perceived both by performers and by their audiences as an expression of the singer wanting to be seen as something different from a commercial pop artist. When British singers in the 1950s imitated American features, it could be regarded as an intentional shift, whereas shifts today happen in a sense automatically and not as a consequence of a conscious wish to identify with Americans, but because the listeners expect it, or as Gibson (2010: 128) puts it: “The use of AmE features […] in the singing of pop music is therefore an unmarked style that is determined by the situation, rather than by a desire to change the situation.”

We may conclude that Cliff Richard’s use of American features is probably not based on a conscious awareness of the linguistic rules, but is related to a complex of factors. These include the origin of the song, the requirements of the genre, his native accent, the degree of awareness of the variable, the ability to modify his accent, the context in which the variable occurs, sonority, and whether the sound is sung on a sustained note. Altogether, CR’s use of American features is perhaps more haphazard than might at first be thought. Rather than being based on explicit knowledge of this variety (the only area where he achieves complete consistency being constituted by the BATH words), it is more likely that he is making a subconscious gesture towards American English, which is achieved partly by adopting a few of the characteristics the British perceive to be American, and partly by retaining one feature of his own idiolect (voiced /t/). This is perhaps also why he is not totally successful from a linguistic point of view (e.g. the use of hypercorrect /r/).[10] However, even the rather limited use of Americanisms is enough to give the flavour –which is perhaps, both for the singer and for his audience, all which is needed.

---

[10] The fact that the use of non-prevocalic /r/ –contrary to our expectations– is at its highest in the Heathcliff musical (requiring non-rhotic northern English) is perhaps an indication that Cliff Richard in some cases simply chooses a pronunciation which is deviant from his own when wishing to hint at a different accent.

REFERENCES

Altendorf Ulrike (1999): “Approaching the notion of ‘Estuary English’”. In: Paradis, Carita (Ed.). *Recent trends in the pronunciation of English: Social, regional and attitudinal aspects*. Stockholm: Almqvist & Wiksell, 15-31.

Bailey, Eric (1996): “Passions of the bachelor boy”. *The Telegraph*. <http://www.telegraph.co.uk/culture/4706060/Passions-of-the-bachelor-boy.html> [downloaded 22/01/2012].

Beal, Joan C. (2009): “You’re not from New York City, you’re from Rotherham: Dialect and identity in British Indie Music”. *Journal of English Linguistics* 37/3: 223-240. Accessed 7 February 2012 via the data base Sage Journals Online.

Bell, Allan (1984): “Language style as audience design”.*Language in Society* 13/2: 145-204.

Burquest, Donald A & Payne, David L. (1993): *Phonological analysis: A functional approach*. Dallas: Summer Institute of Linguistics.

Chambers, Jack K & Trudgill, Peter (1998): *Dialectology* (2nd ed.). Cambridge: Cambridge University Press.

Collins, Beverley & Mees, Inger M. (2013): *Practical phonetics and phonology* (3rd ed.). Abingdon: Routledge.

Cruttenden, Alan (2008): *Gimson’s pronunciation of English* (7th ed.). Abingdon: Routledge.

Drager, Katie (2003): *When I sing I like to sound American*. (Unpublished honours dissertation) University of Canterbury.

Edwards, Harold T. (1997): *Applied phonetics: The Sounds of American English* (2nd ed.). San Diego: Singular.

Fabricius, Anne H (2000): *T-Glottalling between stigma and prestige: A sociolinguistic study of modern RP*. (Unpublished PhD. dissertation). Copenhagen: Copenhagen Business School.

http://www.akira.ruc.dk/~fabri/Fabricius-2000-PhD-thesis.pdf

Gibson, Andy (2010): *Production and perception of vowels in New Zealand popular music.* (MPhil thesis) Auckland University of Technology. <http://aut.researchgateway.ac.nz/handle/10292/962> [downloaded 22/01/ 2012].

Halliday, Michael A. K, McIntosh, Angus & Strevens, Peter D. (1964): *The linguistic sciences and language teaching*. London: Longmans.

Hawkins, Peter (1984): *Introducing Phonology*. London; Dover, N.H.: Hutchinson.

Jones, Daniel (1918): *An Outline of English Phonetics*. Leipzig: Teubner. Rpt. In: Collins, Beverley and Mees, Inger M (Eds.) (2003). *Daniel Jones: Selected Works, Vol. I: Phonetics of English*. London: Routledge.

Jones, Daniel (1926): *An English Pronouncing Dictionary* (3rd ed.). London: Dent. First published 1917.

Ladefoged, Peter (1982): *A Course in Phonetics* (2nd ed.). San Diego: Harcourt Brace Jovanovich.

Larsen, Erika (2011): *Stilistisk variation i popsange: et casestudie*. [Stylistic variation in pop songs: a case study] (Unpublished Master's thesis) Copenhagen: Copenhagen Business School.
<http://studenttheses.cbs.dk/bitstream/handle/10417/2044/erika_anne_marie_larsen.pdf?sequence=1>

Le Page, Robert Brock (1978): "Projection, focussing, diffusion". *Society for Caribbean Linguistics Occasional Paper,* no. 9.

Lewry, Peter & Goodall, Nigel (2001): [Liner notes]. *I'm nearly famous* [CD]. London: EMI Records.

Lewry, Peter & Goodall, Nigel (2004): [Liner notes]. *Sincerely* [CD]. London: EMI Records.

Milroy, Lesley & Gordon, Matthew J. (2003): *Sociolinguistics: Method and interpretation*. Malden, MA: Blackwell.

Morrissey, Franz Andres (2008): "Liverpool to Louisiana in one lyrical line: style choice in British rock, pop and folk singing". In: Locher, Miriam A & Strässler, Jürg (Eds.). *Standards and norms in the English language*. Berlin: Mouton de Gruyter, 195-218. [accessed 10/02/2012 via the database ebrary.

O'Hanlon, Renae (2006): "Australian hip hop: A sociolinguistic investigation". *Australian Journal of Linguistics* 26/2: 193-209.

Richard, Cliff (2008): *My life, my way*. London: Headline.

Roach, Peter (2004): "Illustrations of the IPA. British English: Received Pronunciation". *Journal of the International Phonetic Association* 34/2: 239-245.

Roach, Peter (2009): *English Phonetics and Phonology* (4th ed.). Cambridge: Cambridge University Press.

Rosewarne, David (1984): Estuary English. *Times Educational Supplement*, 19 October 1984.

Simpson, Paul (1999): "Language, culture and identity: with (another) look at accents in pop and rock singing". *Multilingua* 18/4: 343-367.

Sivertsen, Eva (1960): *Cockney phonology*. Oslo: Oslo University Press.

Tottie, Gunnel (2002): *An Introduction to American English*. Oxford: Blackwell.

Trudgill, Peter (1983): "Acts of conflicting identity: The sociolinguistics of British pop-song pronunciation". In: Trudgill, Peter (Ed.) *On dialect: Social and geographical perspectives*. Oxford: Blackwell.

Trudgill, Peter & Hannah, Jean (2002): *International English: A guide to varieties of Standard English*. (4th ed.). London: Arnold.

Turner, Steve (2008): *Cliff Richard: The biography.* Oxford: Lion.

Underwood, Gary N. (1988): “Accent and identity”. In: Thomas, Alan (Ed.). *Methods in dialectology.* Clevedon: Multilingual Matters, 406-428. <http://books.google.com/books?id=Ov4xywn8r-EC&printsec=frontcover&hl=da#v=onepage&q&f=false> [accessed 07/02/2012 via Google Books].

Wells, John C. (1982): *Accents of English* (3 vols). Cambridge: Cambridge University Press.

Wells, John C. (1994a): “Transcribing Estuary English: A discussion document”. *Speech, Hearing and Language: UCL Work in Progress* 8: 259-267.<http://www.phon.ucl.ac.uk/home/estuary/transcree.htm> [downloaded 20/02/2012].

Wells, John C. (1994b): The Cockneyfication of R.P.? In: Melchers, Gunnel & Johannesson, Nils-Lennart (Eds.), *Nonstandard Varieties of Language. Papers from the Stockholm Symposium 11-13 April, 1991*, Stockholm: Almqvist & Wiksell, 198-205.

Wells, John C. (2008): *Longman Pronunciation Dictionary* (3rd ed.). Harlow: Pearson Education.

DISCOGRAPHY

Move it (1958)
High class baby (1958)
Mean streak (1959)
Living doll (1959)
Travellin' light (1959)
From Cliff Richard, *The Singles Collection* [CD 1], EMI Records Ltd. (2002). 7243 5 37551 2 6.

In the past (1969)
Will you love me tomorrow? (1969)
You'll want me (1969)
I'm not getting married (1969)
London's not too far (1969)
From Cliff Richard, *Sincerely* [CD], EMI Records Ltd. (2004). 7243 4 73402 2 2.

Miss you nights (1975)
Devil woman (1976)
I can't ask for anymore than you (1976)
You've got to give me all your lovin' (1976)
Such is the mystery (1976)
From Cliff Richard, *I'm Nearly Famous* [CD], EMI Records Ltd. (2001). 7243 5 33114 2 1.

Misunderstood man (1995)
Had to be (1995)
I do not love you Isabella (1995)
Marked with death (1995)
Be with me always (1995)
From Cliff Richard, *Songs From Heathcliff* [CD], EMI Records Ltd. (1995). 7243 8 35762 2 7.

Somethin' is goin' on (2004)
A thousand miles to go (2004)
I will not be a mistake (2004)
I cannot give you my love (2004)
What car (2004)
From Cliff Richard, *Something's Goin' On* [CD], Decca Music Group Limited. (2004). 475 6408.

Heathcliff dialogue (1996)
From Cliff Richard, *Heathcliff Live* [2-CD], EMI Records Ltd. (1996). 7243 8 54768 2 2.

CHAPTER 13. RECENT CHANGES IN ENGLISH PHONETICS AND PHONOLOGY AND THEIR REPRESENTATION IN PHONETIC NOTATION

Brian Mott
University of Barcelona

1. Introduction

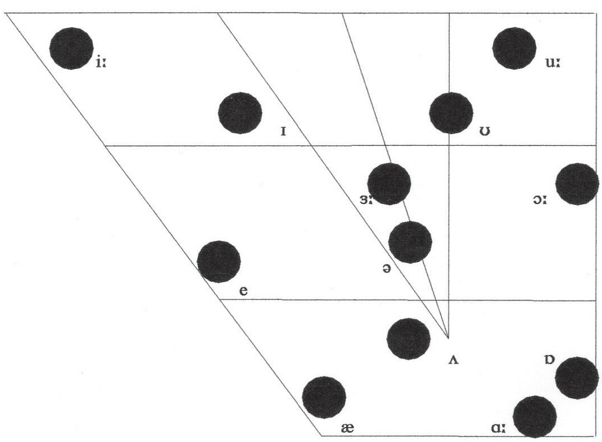
Linguistic variation and change has been at the forefront of research into language for some time now. William Labov has been the protagonist in so called "secular linguistics" ever since the publication of his 1966 study entitled *The Social Stratification of English in New York City*, one of his more recent contributions being concerned with the spread of the New York City vowel system to other American cities (2007). In particular, the author reports on the NYC pattern of tensing short ‹a› in four other American communities, namely northern New Jersey, Albany, Cincinnati, and New Orleans. In the antipodes, Bauer & Warren (2004) and Bauer et alii (2007) have summarized the present-day phonology of New Zealand English, while Gordon & Maclagan (2004) have looked at regional and social differences in this variety. Meyerhoff & Niedzielski (2003) report on a recent shift in New Zealand from British-like features to more American-like ones. Likewise, Horvath (2004) outlines the phonology of Australian English, while Bradley (2004) looks at its regional characteristics. Cheshire *et alii* (2011) address issues arising from the emergence of Multicultural London English, such as what features identify it and at what age they are acquired.

In the present article, which is intentionally a critical survey rather than an attempt to present new data, I examine the principal changes that have taken place in English (RP or SSB) over the past fifty years or so, adding comment on the accounts provided of them where appropriate, and discuss how they are represented notationally in the leading pronunciation dictionaries, in particular the *Longman Pronunciation Dictionary* (LPD), the *Cambridge English Pronouncing Dictionary* (CEPD), and *The Oxford Dictionary of Pronunciation for Current English* (ODP).

2. Vowel shifts

2.1. Stressed vowels

Hawkins and Midgley (2005) report on an anti-clockwise movement in the DRESS, TRAP, FOOT and GOOSE vowels between older and younger RP speakers, so that the first two vowels are now more open, and the latter two are fronter. Accordingly, I have made adjustments to the vowel diagram that I use with students and which appears in Mott (2011a: 69), both with respect to the above-mentioned four vowels and the other vowels in the present-day RP system (see fig. 1) after comparing descriptions of them in several of the major works on English phonetics and phonology (Collins & Mees 2008, Cruttenden 2008, Roach 2009). It goes without saying that the tongue positions shown can only be approximate owing to the fact that the degree of closeness and centralization of a vowel depends on whether it is stressed or unstressed, its position in the word and whether this word is said in its citation form or used in connected speech. Therefore, the dots are intended to give some idea of the latitude that we find in the articulation of vowels rather than suggest that we can pinpoint their articulation.



1. /iː/ FLEECE	7. /ɔː/ THOUGHT, NORTH, FORCE
2. /ɪ/ KIT	8. /ʊ/ FOOT
3. /e/ DRESS	9. /uː/ GOOSE
4. /æ/ TRAP	10. /ʌ/ STRUT
5. /ɑː/ START, PALM, BATH	11. /ɜː/ NURSE
6. /ɒ/ LOT, CLOTH	12. /ə/ LETTER, COMMA

Figure 1. The twelve English vowel phonemes (based on Mott 2011a: 69). Below them are John Wells' keywords (1982: xviii-xix)

As far as the DRESS vowel is concerned, although Gimson, some years ago (e.g. 1970: 104), said that the "general RP variety" of this vowel was closer to Cardinal Vowel 2, up-to-date descriptions recognize a more open quality: for example, Collins & Mees (2008: 98) mention that the vowel has an open allophone, which they place in the Cardinal Vowel 3 area, before dark [l], while they situate the unmarked variety just above it (see Fig. 2).

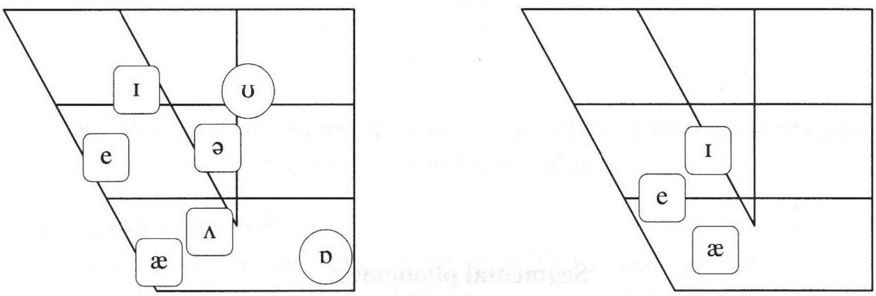


Figure 2. English checked vowels and /ə/ (left) and English front vowels before dark [l] (right). Collins & Mees 2008: 98 (for the variety that they call "Non-Regional Pronunciation"). Reproduced by kind permission of the authors

As regards phonetic notation, Schmitt (2007) argues cogently for the use of epsilon to represent the DRESS vowel, presenting articulatory and perceptual evidence for its proximity to Cardinal Vowel number 3. Furthermore, Schmitt (2007: 325) points out that many languages have both [e] and [ɛ] in their phonological systems, and that using the symbol which is normally used for the closer vowel to represent the DRESS vowel in English is confusing and misleading for the speakers of these languages. German, for example, has [eː] in the first syllable of *Lethe* 'Lethe' (mythology: a river in Hades that caused forgetfulness) and [ɛ] in the first syllable of *Letter* 'character', and there is little difference between the stressed vowel in German *Letter*, which German phoneticians transcribe with [ɛ] (/ˈlɛtɐ/), and that of English *letter*, which English phoneticians usually transcribe with [e] (/ˈletə/), though Upton *et alii* (2001) use the symbol [ɛ] in ODP.

Certainly, there is a significant advantage to be gained from using epsilon for SSB DRESS, both from the point of view of phonetic accuracy and from the contrastive or variationist viewpoint, since, as Schmitt (2007: 235) points out, many varieties of English outside that of southern England (notably South African, Australian, New Zealand, Scottish) have a DRESS vowel that is closer than the southern English one. And even in southern England itself one finds closer varieties. In a recent contribution (Mott 2012a: 9-10), I report on the speech of three elderly male Londoners, who gave an average F1 of 499 Hz for the DRESS vowel as against Cruttenden (2008: 99), where an average of 560 Hz is given for RP male speakers according to figures taken from Deterding 1997. (These figures refer to words in their citation form; for figures acquired from words in context, see Mott 2011b.)

As regards the TRAP vowel, Cruttenden (2008: 112) says "Only tradition justifies the continuing use of the symbol 'æ' for this phoneme", and in fact in the fifth edition of this text the vowel was resited (i.e. lowered) with respect to Gimson's earlier positions (Windsor Lewis 2003: 146). Weiner & Upton (2000) refer to the fact that Oxford dictionaries for native speakers from the early 1990's onwards use [a] not [æ], which Upton *et alii* (2001) accordingly do, and this is consistent with the present-day pronunciation of the vowel in SSB, though Windsor Lewis (2003: 146) is unenthusiastic about the change of symbol on the grounds that the lowering has not taken place in American and Australian accents. Incidentally, as [æ] was originally backer in the history of English and became fronter through Anglo-Frisian Brightening or First Fronting (Lass 1994: 42-44), and is now lowering and backing once again, it constitutes a classic example of the ebb and flow phenomenon that may affect sound segments over long periods of time.

Weiner & Upton (2000) also favour using [ɛː] for [ɛə] or [eə], a diphthong which has undergone smoothing in the pronunciation of some people, as happened also to [ɔə] when it became replaced by monophthongal [ɔː]. The /ɔː/ – /ɔə/ merger is now complete in RP, though not in London, where [ɔə] is used in place of /ɔː/ in open final position (cf. *board* /bɔːd/ and *bore+d* /bɔə+d/). Like Windsor Lewis (2003: 148), I believe that diphthongal values in words like *where* and *chair* are by no means yet a thing of the past, even though the diphthongal glide is often only slight, so I see no reason to transcribe regularly with the monophthong /ɛː/ for the moment. As far as ELT practice and training is concerned, certainly, foreign learners of English will benefit from attempting to produce articulations in the above instances with little diphthongization since, if they do diphthongize, in the worst of cases, it may lead to the production of two syllables instead of one.

The [ɪə] diphthong seems to be suffering the same fate as [ɛə] or [eə], so that many young speakers in southern England now omit the glide towards a central tongue position, thus producing a sound more like [ɪː]. This may occur in words like *Ian* in the variety that Tench (2011: 31) labels *SESP* (*Southern England Standard Pronunciation*). It is also noticeable that [ʊə] may sometimes be levelled to [ʊː], a development independent of the merger of the CURE set (*poor*, *sure*, *tour*, *jury*) with /ɔː/, particularly in common words like *poor* and *sure*.

Fronting of the FOOT and GOOSE VOWELS is now well established, and the latter has received more attention than the former, probably because of its salience through a more noticeable degree of fronting. It has also been fronting for longer (/uː/ probably since the 1960s, but /ʊ/ probably only since the 1990s), a fact corroborated by Harrington *et alii* (2011: 137) when they say that "… the diachronic shift in /ʊ/ is likely to be a more recent innovation than that of its tense counterpart". One sometimes comes across humoristic spellings, such as "kewl" for "cool", which show speaker awareness of this change in articulation in the GOOSE vowel.

GOOSE-fronting probably arose through consonant-on-vowel coarticulation in coronal contexts, leading to "phonologization of the variants in a fronting context and a consequential realignment in perception of the back variants towards the front" (Harrington 2012: 103). For example, an alveolar context would produce a high F2, which could be interpreted by the listener as part of the vowel instead of part of the context. Note in particular that /uː/ is preceded by yod in 70% of cases, which provides a potent fronting force. Perhaps the young compensate perceptually to a lesser extent for context, so that they perceive /uː/ after yod as front and consequently reproduce it that way. Moreover, it is possible that the consonant distribution of the world's languages favours this fronting, i.e. there is a relatively high proportion of acute consonants that would be likely to have a fronting effect on adjacent vowels.

Another reason for GOOSE-fronting could be an articulatory one: rounded back vowels require greater effort, so there is a tendency to avoid full backing (and rounding). Japanese lends support to this hypothesis, as it has a high back vowel which is not as peripheral as the other vowels in its system. Similarly, although Standard German has a high back vowel which is backer than in English, "It is often palatalized and centralized in East Central dialects such as Thuringian and Saxon" (Russ 2010: 45).

On the question of whether fronting of FOOT and GOOSE has also been accompanied by unrounding of the lips, Harrington *et alii* (2011) confirm that the shift has involved a realignment of the tongue but not of the lips.

Although GOOSE-fronting is a global phenomenon diffusing to more and more English-speaking communities and its spread is virtually complete today (Cheshire *et alii* 2011: 156), phoneticians are reluctant to make any concessions in its transcription. For example, Tench (2011: 9, 13), while accepting the use of epsilon and admitting /a/ alongside traditional /æ/, still proposes /uː/ for the GOOSE vowel. Presumably, there are still today relatively back realizations of the vowel. It is for this reason that I have assigned a cautious, conservative position to it on my own chart (fig. 1), but surely adoption of barred lower-case Roman ‹u› in the transcription of present-day English, i.e. [ʉ], which is used by the IPA for a high central rounded vowel (Pullum & Ladusaw, 1996: 183), would not be amiss and, in any case, it would be more practical and less complex than the alternative solution of the [u] symbol with a subscript diacritic to show fronting: [u̟].

Other changes have also been taking place in the English vowel system, but these are not all as clear as the ones just described. Wells (1982: 281) says

that the STRUT vowel is a back articulation in upper-crust RP (U-RP), as it was in Daniel Jones' day, but that mainstream RP STRUT is fronter and very close to the lowered variants of TRAP (1982: 291-292), so that words like *match* and *much* have become more similar. On the other hand, Fabricius (2007: 310) says that STRUT is now backing and rising among middle-class speakers, while Harrington *et alii* (2000), in their study of the Queen's speech in her Christmas messages, claim that for her it has backed and lowered. Ignoring the obvious time lapse between Wells' observations and those recorded in the other, more recent publications, perhaps we nevertheless need to take into account the fact that, in the case of elderly speakers, like the Queen, their speech organs may have undergone change if we make them the object of any study that attempts to chart modifications over a period of time, and this is an important factor that needs to be controlled in scientific analyses.
For the moment, it would probably be wise for teachers to produce pronunciation models for their students that keep TRAP and STRUT separate, i.e. front v. front-central. Fortunately, the problem of pinpointing distinctions between these vowels for foreign learners of English is partly solved for us by the fact that TRAP, though traditionally grouped with the short vowels, is often noticeably long nowadays and is probably undergoing a change in phonological length in order to remain distinct from STRUT. This increased duration is a fact that was observed by Jones (1960: 235) in certain words like *bad*, *bag*, *glad*, *jam*, *man* and *sad*. Jones makes particular reference to "adjectives ending in -*ad*" and to the fact that "Long [æ:] is most frequently found before voiced consonants, but is not confined to these situations", so we can assume that the voiced environment would instigate the change, particularly in the case of adjectives that might have some kind of emotive content and therefore be liable to receive emphasis.

2.2. Unstressed vowels

I deal with the unstressed vowel system of English in some detail in Mott (2009). However, I would like to reiterate my mention of the use of schwa in Modern English in some contexts which originally had the KIT vowel. Schwa has taken over, or is taking over, from /ɪ/ in many words like *family* /ˈfæməli/ ~ /ˈfæmɪli/ and *mistake* /mɪˈsteɪk/ ~ /məˈsteɪk/, in which there is a letter ‹i› in the spelling, and in the suffixes *-ible* and *-ity*, among others, also spelt with ‹i›: *possible* /ˈpɒsəbl/ ~ /ˈpɒsɪbl/, *terrible* /ˈterəbl/ ~ /ˈterɪbl/, *quality* /ˈkwɒləti/ ~ /ˈkwɒlɪti/, *ability* /əˈbɪləti/ ~ /əˈbɪlɪti/. To reflect this alternation, Upton *et alii* (2001) wisely use the composite symbol [ɨ]; thus: /ˈpɒsɨbl/, etc. On the other hand, García Lecumberri & Maidment (2000: 24, for example), I believe, are jumping the gun when they opt for [əd] instead of [ɪd], and [əz] instead of [ɪz], to represent the RP pronunciation of the regular past tense morpheme <ed> and the present tense/plural/genitive morpheme <es>, respectively (e.g. /ˈendəd/ *ended*, /ˈweɪstəd/ *wasted*, /ˈkɔːsəz/ *courses*), since my impression is that in these morphs the KIT vowel still predominates, perhaps through pressure from minimal pairs like *seizes* /ˈsiːzɪz/ v. *Caesar's* /ˈsiːzəz/, *offices* /ˈɒfɪsɪz/ v. *officers* /ˈɒfɪsəz/, and *chatted* /ˈtʃætɪd/ v. *chattered* /ˈtʃætəd/ (see Cruttenden 2008: 109; Upton *et alii* 2001: xiii).

Upton *et alii* (2001) also use their composite symbol in the verb/noun suffixes ‹-ed› and ‹-es›, which have just been mentioned, as in *noted* /ˈnəʊtɨd/ and *houses* /ˈhaʊzɨz/, and rightly add comment to the effect that American English pronunciations are normally [əd] and [əz], but British pronunciations are normally [ɪd] and [ɪz] (2001: xiii).

Another interesting phenomenon in the unstressed vowel system is the tensing of the final vowel of words like *happy* /ˈhæpi/ or the prevocalic segment in the second syllable of forms like *devious* /ˈdiːviəs/ (Wells 1982: 294), which in RP is now identified more often than not with the FLEECE vowel rather than the KIT vowel, as shown by use of the symbol [i]. Theoretically, at least, this symbol is intended to represent the neutralization of the /iː/ v. /ɪ/ opposition, and the fact that a new phonemic alignment of the KIT vowel with the FLEECE vowel in these positions is assumed, as well as to reflect the tendency of younger speakers to produce a tenser vowel, even in inflected forms, where the vowel is blocked by a consonant: *carry*, *carries*, *carried* /ˈkæri, ˈkæriz, ˈkærid/; *ferry*, *ferries* /ˈferi, ˈferiz/. However, in practice, the change appears to have been applied inconsistently in LPD in some cases. To take just one example, compare *archetype* /ˈɑːkitaɪp/ with *architect* /ˈɑːkɪtekt/. Both contain a prefix from Greek *arkhi-* 'chief, principal', but the first is transcribed with [i], while the second is transcibed with [ɪ]. There can be no doubt that Professor Wells gave careful consideration to the transcription of each and every single word in his dictionary, but the result is none the less confusing for the average user of the reference work. CEPD, in contrast, uses [ɪ] in both words.

In the third edition of LPD, Wells extends the use of [i] to the following prefixes: *be-, de-, e-, pre-, re-* (*begin* /biˈgɪn/, *decide* /diˈsaɪd/, *elect* /iˈlekt/, *prevent* /priˈvent/, *receive* /riˈsiːv/), though, once again, not consistently: witness *edition* /iˈdɪʃᵊn/ v. *event* /ɪˈvent/. With respect to this innovation, Windsor Lewis (2009: 238) comments: "The choice of [i] rather than [ɪ] or [ə] in some common words, e.g. *believe* and *remind*, may not meet with universal assent." ODP adopts the simpler, more practical solution of transcribing all of these prefixes with the composite symbol [ɨ]: /ɨˈdɪʃn/, /ɨˈvɛnt/, /prɨˈvent/, etc.

In his foreword to the third edition of the LPD, Wells (2008: xiii) also includes *se-* among the group of prefixes referred to above, even though it is shown to be variable in the dictionary: *seduce* /sɪˈdjuːs, səˈdjuːs/, *semester* /səˈmestə, sɪˈmestə/, but *select* /səˈlekt/, *serene* /səˈriːn/, presumably in keeping with the fact that schwa is more usual before a liquid. No doubt, this oversight will be corrected in future editions.

In similar fashion to [i], the symbol [u], a shortened variant of the GOOSE vowel, is now used in LPD to represent the neutralization of /uː/ and /ʊ/, showing alternation with schwa in some cases, like *regulate* /ˈregjuleɪt/ ~ /ˈregjəleɪt/. If [u] is truly to be taken as an archiphoneme in such cases, Wells' use of [ʊ] before syllables with a weak vowel (e.g. *regular* /ˈregjʊlə/ ~ /ˈregjələ/) would seem to be superfluous, and the ODP adoption of the composite symbol [ʉ] is a welcome practical solution in the transcription of all such words: /ˈrɛgjʉlə(r)/, /ˈrɛgjʉleɪt/, etc., parallel to that provided by the use of [ɨ] for words that show alternation in the use of [ɪ] and [ə] (see above).

To conclude these observations on the changes that have affected or are in the process of affecting English vowels, we might note the extended use of schwa in words apart from those mentioned above, in which it alternates with [ɪ], [u] or [ʊ]. The diversification of contexts with schwa should come as no surprise since centralization of unstressed vowels has been a feature of

English for over a thousand years. But there are many new additions to the class like *torrential* /tə'renʃl/ < /tɒ'renʃl/ and *nobody* /'nəʊbədi/ < /'nəʊbɒdi/, and even cases of reduction in former heavy syllables, as in *September* /səp'tembə/ < /sep'tembə/ and *November* /nə'vembə/ < /nəʊ'vembə/ (for further examples, see Mott 2011a: 121). Reduction of unstressed vowels is more likely in common words, or in words especially familiar to the speaker, than in less common words. Thus, New Yorkers pronounce *Manhattan* as /mən'hætn/ < /mæn'hætn/, and trombonists may refer to their instrument as the /trəm'bəʊn/ instead of the /trɒm'bəʊn/ (Aitchison 2013: 89).
Needless to say, vowel reduction can also affect grammatical words when they are unstressed. The many frequently used weak forms of spoken English are well known to students of phonetics, but consider also the following less common cases, which are generally given little or no attention in English phonetics courses:

more or less /'mɔːr ə 'les/
not for love nor money /'nɒt fə 'lʌv nə 'mʌni/
not so bad /'nɒt sə 'bæd/
I haven't any left /aɪ 'hævnt əni 'left/
How many do you need? /'haʊ məni dju 'niːd/

3. Consonant changes

The changes in the articulation of some of the English consonants is a topic which is much easier to deal with than the questions relating to the vowels that we have considered above. Contact of the speech organs is simpler to trace and locate than tongue height and retraction.
Palatalizing sequences of [t] + [j] and [d] + [j] (YOD-coalescence) in English seems to be the norm nowadays among young RP speakers in words such as *tune*, *Tuesday*, *perpetual*, *reduce* and *education*. Wells' preference poll for *tune* (LPD 2008: 58), for example, shows that 54% of British English speakers prefer the pronunciation /tʃuːn/ to /tjuːn/. The figure rises to 68% if only those speakers born since 1981 are counted. Naturally, there are

regional differences, too, with Scottish speakers preferring a palatal in *soldier* /ˈsɒdʒə/ (with ʟ-vocalization, too), like traditional RP /ˈsəʊldʒə/, but with South Wales not palatalizing: /ˈsəʊldjə/.

Use of /ɪn/ for /ɪŋ/ in the gerund (but not other word classes –witness the nouns *ceiling* /ˈsiːlɪŋ/ and *Waring* /ˈweərɪŋ/, etc.) is likely to raise even fewer eyebrows than palatalization of [t] and [d], owing to its lack of salience. In fact, it was only ousted as the fashionable pronunciation in the twentieth century; in the eighteenth and nineteenth centuries, it was in vogue (Cruttenden 2008: 213). Use of one or the other form now depends on social and stylistic factors. Additionally, confusion may sometimes arise through the coexistence of forms like *Hawking* and *Hawkins* with the gerund and common noun *hawking*, so that, for example, people may not be certain as to whether the famous British physicist is *Hawking* or *Hawkin*, and they may sporadically pronounce other nouns ending in /ɪŋ/ with an alveolar nasal, too.

Glottalling and tapping of /t/, as in *sitting pretty* [ˈsɪʔɪŋ ˈprɪʔi] ~ [ˈsɪt̬ɪŋ ˈprɪt̬i], are slightly more complex. Glottalling before another consonant (*what time* [ˈwɒʔ ˈtaɪm], *footpath* [ˈfʊʔpɑːθ], *Luton* [ˈluːʔn], *Gatwick* [ˈgæʔwɪk]) has been normal in RP for a long time, but glottalling between vowels and in absolute final position (*water* [ˈwɔːʔə], *what* [wɒʔ]), though very common now all over Britain, is still stigmatized, so it is best for foreign learners not to reproduce it. T-tapping, though typical of American English, is now common in SSB, and much less frowned upon than glottalling, especially in fast speech and in certain common words and phrases, such as *better* [ˈbet̬ə], *pretty good* [ˈprɪt̬i ˈgʊd], *what does it matter?* [ˈwɒʔ dəz ɪʔ ˈmæt̬ə], *as a matter of fact* [əz ə ˈmæt̬ər ə ˈfækt], *what I think is* … [wɒt̬ ˈaɪ θɪŋk ɪz], *but I think* … [bət̬ ˈaɪ θɪŋk], *a bit of jazz* [ə ˈbɪt̬ əv ˈdʒæz], etc.

To avoid use of the diacritic, [t̬] is sometimes transcribed as [ɾ], a symbol preferably reserved for the alveolar flap, as found in Spanish. But there can be no real objection to this practice as long as the phonetic description of the segment and its phonological alignment with the /t/ phoneme are made clear, which is unfortunately not the case with ODP's adoption of the symbol [d] to represent the tap in American English, e.g. *city* /ˈsɪdi/, *latter* /ˈlædər/, *potato* /pəˈteɪdoʊ/, etc. In fact, this transcription suggests that pairs like *city* and *Syddy*, *latter* and *ladder*, and *atom* and *Adam*, are always identical in pronunciation in American English, which may not be true. To mention just one detail of this very complex phonetic issue, some speakers of American English tap both intervocalic [t] and [d], while others only tap the [t].

L-vocalization is not just found in English, but is widespread among the world's languages. For example, Brazilian Portuguese has vocalized the [l] in *Brazil,* whereas in Peninsular Portuguese it is retained. From the Serbian name for the capital *Beograd* 'White City', we can see that an earlier [l] has vocalized when we compare it with the English version *Belgrade*. And northern Romanian forms like *aub* < *alb* 'white' and *a ascuta* < *a asculta* 'to listen' reveal local vocalization as compared to Standard Romanian. L-vocalization has not phonologized yet in English, as many instances of vocalization have in Danish, so it is still a sociolinguistic variable, but as Wells (1982: 259) says: "From its putative origins in London and the surrounding counties, L-vocalization is now beginning to seep into RP. It seems likely that it will become entirely standard in English over the course of the next century". Since Wells' prediction in 1982, it would appear that L-Vocalization has indeed increased in popularity, only curbed to some extent by the influence of spelling. Foreign learners should be encouraged to adopt it, at least in informal situations, to avoid producing an [l] that is too clear. The word *children* sounds much more native if pronounced as ['tʃɪodrən] or ['tʃɪʊdrən] (or even without [ɪ] –see Wells 2008: 142) rather than with some kind of [l].

As regards narrow transcription of the vocoid [l], a symbol representing a mid-high or high back vowel ([o] or [ʊ]) is adequate, even though the segment may not be fully rounded and thus more like [ɤ].

Intrusive [r] is "now widely heard and accepted" (Tench 2011: 102) and, as it is often likely to pass unnoticed (except to Scottish speakers, who hate it!), it is not a salient feature and can even be encouraged in foreign learners of English. It was controversial in the 1960s and 1970s, but is now so usual that LPD3 admits intrusive [r]s without comment (e.g. *withdrawal* /wɪð'drɔːʳəl/, *sawing* /'sɔːʳɪŋ/), as does ODP, which remarks: "Long condemned by teachers of pronunciation, this is nevertheless a firmly established feature of today's mainstream RP" (Upton *et alii* 2001: xii). CEPD simply omits such cases. As regards advocating its use among non-

native learners of English, I, personally, can only add that I would find it hard not to insert an [r] in phrases like *vodka-r and orange*, and I certainly can't imagine pronouncing the adjective *banana-r-y* without an [r]. One advantage for the speaker is that it solves the problem of hiatus and reinforces the natural, primaeval CV tendency of language. Hundreds of British tourists flock to Spain's coasts every summer and sing *Que viva-r España*!
As a closing remark on the changes in the consonant system of British English, we might add that it has recently been observed that the [s] in the cluster ‹str› appears to be increasingly undergoing palatalization (Rutter 2011). While this has not been unknown outside RP (see, for example, Mott 2012b: 85), its adoption as a feature on the increase in the standard may come as a surprise.

4. Conclusions

The preceding paragraphs have reported on the fact that, in recent decades, English has been undergoing significant and relatively rapid changes in pronunciation which need to be taken into account in our descriptions, phonetic notation and teaching of the language. The vowel shifts are more complex than the consonantal changes, which are easier to account for in phonetic terms. If in doubt as to the question of whether any steps should be taken to reflect such developments in our systems of transcription, we might recall the words of Weiner & Upton (2000: 44-45): “Sounds, like the other components of language, do not remain fixed forever”. Therefore, it is necessary to periodically review our methods of committing them to paper. Any drastic, precipitate amendments to well established tradition should be avoided but, in a language like English, in which the vowels in particular are susceptible to substantial change, it is occasionally necessary to reconsider what has become orthodox practice lest it come to be seen as immutable, and, not least, eventually turn out to be an inaccurate, out-of-date representation of present-day reality. To what degree our phonetic symbols should be modified will depend on just how salient the shifts in pronunciation have become, and to what extent they are now considered to be the norm.

REFERENCES

Aitchison, Jean (2013): *Language Change: Progress or Decay?* 4th ed. Cambridge: CUP.

Bauer, Laurie & Warren, Paul (2004): “New Zealand English: phonology”. In: Schneider, Edgar W. *et alii* (eds.). *A Handbook of Varieties of English* 1. Berlin: Mouton de Gruyter, 580-602.

Bauer, Laurie, Warren, Paul, Bardsley, Dianne, Kennedy, Marianna & Major, George (2007): “New Zealand English”. *Journal of the International Phonetic Association* 37/1: 97-102.

Bradley, David (2004): “Regional characteristics of Australiin English: phonology”. In: Schneider, Edgar W. *et alii* (eds.). *A Handbook of Varieties of English* 1. Berlin: Mouton de Gruyter, 645-655.

Cheshire, J., Kerswill, P., Fox, S. & Torgersenc, E. (2011): “Contact, the feature pool and the speech community: The emergence of Multicultural London English”. *Journal of Sociolinguistics* 15/2: 151-196.

Collins, Beverley & Mees, Inger M. (2008): *Practical Phonetics and Phonology*. 2nd ed. London: Routledge.

Cruttenden, Alan (2008): *Gimson’s Pronunciation of English*. 7th ed. London: Arnold.

Deterding, David (1997): “The formants of monophthong vowels in standard southern British English pronunciation”. *Journal of the International Phonetic Association* 27: 47-55.

Fabricius, Anne. H. (2007). “Variation and change in the trap and strut vowels of RP: a real time comparison of five acoustic data sets”. *Journal of the International Phonetic Association* 37/3: 293-320.

García Lecumberri, María Luisa & Maidment, John A. (2000): *English Transcription Course*. London: Arnold.

Gimson, Alfred C. (1970): *An Introduction to the Pronunciation of English*. 2nd ed. London: Edward Arnold.

Gordon, Elizabeth & Maclagan, Margaret (2004): “Regional and social differences in New Zealand: phonology”. In: Schneider, Edgar W. *et alii* (eds.). *A Handbook of Varieties of English* 1. Berlin: Mouton de Gruyter, 603-613.

Harrington, Jonathan (2012): “The coarticulatory basis of diachronic high back vowel fronting”. In: Solé, Maria-Josep & Recasens, Daniel (eds.). *The initiation of sound change: Perception, production, and social factors*. Amsterdam & Philadelphia: John Benjamins.

Harrington, Jonathan, Kleber, Felicitas & Reubold, Ulrich (2008): “Compensation for coarticulation, /u/-fronting and sound change in Standard Southern British”. *Journal of the Acoustical Society of America* 123: 2825-2835.

Harrington, Jonathan, Kleber, Felicitas & Reubold, Ulrich (2011): "The contribution of the lips and the tongue to the diachronic fronting of high back vowels in Standard Southern British English". *Journal of the International Phonetic Association* 41/2: 137-156.
Harrington, Jonathan, Palethorpe, Sally Anne & Watson, Catherine I. (2000): "Monophthongal vowel changes in Received Pronunciation: an acoustic analysis of the Queen's Christmas broadcasts". *Journal of the International Phonetic Association* 30: 63-78.
Hawkins, Sarah & Midgley, Jonathan (2005): "Formant frequencies of RP monophthongs in four age groups of speakers". *Journal of the International Phonetic Association* 35/2: 183-199.
Horvath, Barbara M. (2004): "Australian English: phonology". In: Schneider, E. W. *et alii* (eds.). *A Handbook of Varieties of English* 1. Berlin: Mouton de Gruyter, 625-644.
Jones, Daniel (1960): *An Outline of English Phonetics*. 9th ed. Cambridge: Heffer.
Jones, Daniel (2006): *Cambridge English Pronouncing Dictionary* (CEPD). 17th ed. edited by Peter Roach, James Hartman & Jane Setter. Cambridge: CUP. (Original title: *English Pronouncing Dictionary*)
Labov, William (1966): *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
Labov, William (2007): "Transmission and diffusion". *Language* 83: 344–387.
Lass, Roger (1994): *Old English. A Historical Linguistic Companion*. Cambridge: CUP.
Meyerhoff, Miriam & Niedzielski, Nancy (2003): "The globalization of vernacular variation". *Journal of Sociolinguistics* 7/4: 534-555.
Mott, Brian (2009): "Practical advice on the transcription of the unstressed vowel system for non-native students of English". In: Čubrović, Biljana & Paunović, Tatjana (eds.). *Ta(l)king English Phonetics Across Frontiers*. Newcastle: Cambridge Scholars Publishing, 33-41.
Mott, Brian (2011a): *English Phonetics and Phonology for Spanish Speakers*. 2nd ed. Barcelona: Publicacions i Edicions UB.
Mott, Brian (2011b): "The Monophthongs of Traditional Cockney and Popular London Speech in Context". *Philologia* 9. Belgrade: University of Belgrade, 1-8.
Mott, Brian (2012a): "Vowel Frequencies in Traditional Cockney and Popular London Speech". In: Čubrović, Biljana & Paunović, Tatjana (eds.). *Exploring* Newcastle: Cambridge Scholars Publishing, 3-11.
Mott, Brian (2012b): "Traditional Cockney and Popular London Speech." *Dialectologia* 9: 69-94 (Revista electrònica). Barcelona: Universitat de Barcelona.
Pullum, Geoffrey K. & Ladusaw, William A. (1996): *Phonetic Symbol Guide*. 2nd ed. Chicago: University of Chicago Press.
Roach, Peter (2009): *English Phonetics and Phonology: A practical course*. Cambridge: CUP.
Russ, Charles V. J. (2010): *The Sounds of German*. Cambridge: CUP.

Rutter, Ben (2011): "Acoustic analysis of a sound change in progress: The consonant cluster /stɹ/ in English". *Journal of the International Phonetic Association* 41/1: 27-40.
Schmitt, Holger (2007): "The case for the epsilon symbol (ɛ) in RP dress". *Journal of the International Phonetic Association* 37/3: 321-328.
Stibbard, Richard (2000): Review of María Luisa García Lecumberri & John A. Maidment, *English Transcription Course*. *Journal of the International Phonetic Association* 30(1/2): 79-86.
Tench, Paul (2011): *Transcribing the Sound of English*. Cambridge: CUP.
Upton, Clive (2004): "Received Pronunciation". In: Schneider, Edgar W. *et alii* (eds.). *A Handbook of Varieties of English* 1. Berlin: Mouton de Gruyter, 217-230.
Upton, Clive, Kretzschmar, William A. & Konopka, Rafal (2001): *The Oxford Dictionary of Pronunciation for Current English* (ODP). Oxford: OUP.
Weiner, Edmund & Upton, Clive (2000): "[hat], [hæt], and all that". *English Today* 61 (16/1): 44-46.
Wells, John C. (1982): *Accents of English*. Cambridge: CUP.
Wells, John C. (2008): *Longman Pronunciation Dictionary* (LPD). 3rd ed. Burnt Mill (Harlow): Pearson Education Ltd.
Windsor Lewis, Jack (2003): "IPA vowel symbols for British English in dictionaries". *Journal of the International Phonetic Association* 33/2: 143-152.
Windsor Lewis, Jack (2009): Review of *Longman Pronunciation Dictionary*, 3rd ed. *Journal of the International Phonetic Association* 39/2: 238-240.

CHAPTER 14. SOME RECENT CHANGES AND DEVELOPMENTS IN BRITISH ENGLISH

David Levey
University of Cadiz

1. Introduction

Spoken languages always change and will undoubtedly continue to do so; only dead languages remain stagnant. Chaucer's English was of course very different to Shakespeare's, which in turn was very different to the English we speak today. But languages do not only change over centuries, they may change relatively quickly, with developments noticeable within the lifetimes of their speakers. One just has to listen to old recordings from the 1940s and 1950s to hear how far British English has come in two or three generations.

In her 1957 televised Christmas broadcast, the Queen starts by wishing her subjects a "*Heppy* Christmas" with the TRAP vowel in the first syllable of 'Happy' being pronounced as an [ɛ] type vowel and lax [ɪ] used in the final syllable[1]. She then goes on to confide to her viewers that her own family "often gather round to watch television" pronouncing 'often' as [ɔ:]. The same revealing lexical items reoccur in her 2011 Christmas broadcast and are realised in a very similar, albeit slightly less pronounced, way.[2] These vowels sound somewhat outdated and *posh*, and it would be hard to find young people who choose to speak like that today. The TRAP vowel is now

[1] When describing vowels, I have used Wells keywords which are "intended to be unmistakeable no matter what accent one says them in" (Wells 1982: xviii).

[2] It has been claimed that the Queen's pronunciation has moved with the times and mellowed over the years. According to Harrington et al (2000: 927) "there has been a drift in the Queen's accent towards one that is characteristic of speakers who are younger and/or lower in the social hierarchy", although they go on to conclude that her vowels "still clearly set her aside from those of an SSB (standard southern-British) accent".

considerably lower and *happY* vowel tensing is considered the RP norm. The long CLOTH vowel has given way to a short back open vowel realisation of the type [ɒ] which has now become practically categorical.

2. Last year's model and this year's model

When Daniel Jones' *English Pronouncing Dictionary* was first published in 1917, it was largely received as a prescriptive reference book, despite the author's claims to the contrary:

> I am not one of those who believes in the desirability or the feasibility of setting up any one form of pronunciation as a standard for the English speaking world. […] those who think reforms or standards are necessary must be left with the invidious task of deciding what is to be approved and what is to be condemned (1917: ix).

Initial attempts to establish a reference standard had been made during the nineteenth century, but it was not really until the early twentieth century that a systematic approach was adopted. Daniel Jones is generally considered to be the founding father of modern English phonetics and largely responsible, albeit reluctantly, for establishing a standard model of English pronunciation. The pronunciation he described was that of the educated ruling classes, and the reference book he produced reflected their socially agreed norms. Their way of speaking was inevitably held to be the standard against which other accents were judged.

In his early work Jones used the term "Standard Pronunciation" or "Standard Southern Pronunciation" to refer to "the nearest approximation, according to the judgement of the writer, to the general usage of educated people in London and the neighbourhood" (1909: v). Then, after briefly toying with the name "Public School Pronunciation" in the first edition of his *English Pronunciation Dictionary* (1917), Jones finally opted for the term "Received Pronunciation" (RP) by the time the third edition came out in 1926.[3] His

[3] There is certain disagreement as to where and from whom the epithet "Received Pronunciation" originated. Following Parsons (1998), it is usually attributed to Alexander J. Ellis

definition of the pronunciation model was always a narrow one: "the families of Southern English persons whose men-folk have been educated at the great public boarding Schools" (Jones 1917: viii).

The society into which Jones was born has clearly changed and the model he was describing and recorded has also moved on. Standards, like social structures, inevitably alter with time and what is considered non-standard or unacceptable for one generation may become the "norm" of the next. In the past, to speak like the Queen (or King) of England was something to aspire to, and elocution lessons might have been sought to achieve this end so as to get on in society. Nowadays, however, a young person adopting that accent would run the risk of inviting ridicule in a new Britain where a less conspicuous "classless" standard is often sought which permits freer social movement.

Many of the British films and BBC broadcasts of the first half of the twentieth century may sound *frightfully* dated and often conjure up in our minds a bygone era with different values. Similarly, what was once the standard accent for stage and film thespians is rarely used by the modern actor, unless a period portrayal is required. Those speaking in that way today might be classified, somewhat disparagingly, as "luvvies".[4] Although now in disuse or confined to very small circles, these outmoded forms of speech are very much present in the British collective memory. Even though younger speakers today may not have had face-to-face contact with anyone speaking in that way, they are still aware of the underlying social implications and connotations. These days, this type of accent is principally reserved for

---

who, in the second half of the nineteenth century, wrote of "a received pronunciation all over the country, not widely differing in any locality, and admitting a certain degree of variety. It may be considered as the educated pronunciation of the metropolis, of the court and the bar". However, some writers such as Fisher (1993), trace it back even further, attributing it to John Walker who mentions the word "received" in his *Critical Pronouncing Dictionary and Expositor of the English Language* published in 1791. In the introduction, Walker establishes the criteria for his pronunciation model as those sounds: "which are the most generally received among the learned and polite" (1791: viii).

[4] Classically trained stage actors of the old school are colloquially known as "luvvies", apparently because of the stereotyped tendency to address people as "love" or "darling". "Luvvie" also refers to their way of speaking which often has posh, effusive, pompous and sensitive connotations.

comedic purposes, often to parody or ridicule pomposity, social superiority of former social class norms.[5]

3. What's in a name?

New attitudes, changes in social structures, education and mobility, as well as internal and external influences, not to mention the influence of the popular multimedia, have all contributed and shaped the way English is spoken today. As pronunciation changes, many question whether labels such as "Received Pronunciation", "BBC English", "the King's (or Queen's) English", "Oxford English" or "Public School English" are still meaningful or relevant today. It may once have been true that the above were practically synonymous, all implying "standard English", but standards and times have changed. The English spoken on the BBC today is not the same as it was 50 or 60 years ago, Oxford University is no longer only for Public School graduates, and Prince William's pronunciation is decidedly different from that of his father and his grandmother. Like other people of his generation, the Prince frequently uses glottal stops and fillers such as like, "kinda", "sort of", "yeah", and "and stuff" and expressions such as "chilling out".[6] The question is, if or when he eventually inherits the throne, could or should we say he speaks the King's English?...probably not.

In talks, papers and published works you will often hear that 3% of the British population speak RP. The source of this much cited figure can be traced back to Peter Trudgill (1974) who reached this conclusion based on his survey of 50 people over the age of 21 living in the city of Norwich. It is of course convenient to quantify a model, but such estimates or guestimates

[5] Numerous comedians have got comic mileage out of imitating the upper-class accent. These include Harry Enfield, whose Cholmondley Warner sketches parody the black and white documentaries and information films of the first half of the twentieth century, and Matt Lucas who plays Sir Bernard Chumley, a faded actor with aristocratic pretensions, in the series *Little Britain*. Armstrong and Miller's RAF airmen sketches derive humour from putting modern street language in the mouths of Second World War pilots with upper class accents.

[6] See, for example, Fearne Cotton's 2007 interview with Prince William and Prince Harry in the lead up to the *Concert for Diana*.

are misleading. Then and now these figures are meaningless unless clear definitions and judgement criteria are established. What do we mean by RP? Which RP are we referring to? Do we mean the RP of Daniel Jones? The RP of J.C.Wells? The RP of a new generation? All of them? We can no longer, if we ever could, talk about one homogenous form of RP.[7]

Many years have passed since the term was first coined and the British class society has changed, and continues to do so. In view of the extra baggage it carries, RP as an epithet for standard pronunciation is felt by many linguists to be inadequate and imprecise. Originally conceived in a rigid class society in which social mobility was difficult, it has arguably outgrown itself. However, despite recent efforts to relieve it of its duties, the term still endures, largely for want of a universally accepted alternative.[8]

John Wells, while always recognising certain limitations, has opted to work within the established terminology and simply update its contents. In his introduction to the *Longman Pronunciation Dictionary* (LPD) he explains that the British pronunciation model used is "a modernized version of the type known as Received Pronunciation or RP." He goes on to write:

---

[7] In *Accents of English*, John Wells writes that the name RP "is less than happy, relying as it does on an outdated mode meaning of *received* ("generally accepted"). But it is so well established that I have decided to retain it here" (1982: 117). This, however, led to a need to introduce modifications to the existing terminology such as: "C-RP" (constructed RP) to refer to the "official" pronunciation described in dictionaries, aimed as a standard for EFL and ESL learners, "N-RP" (native RP) to refer to the accent of those native speakers born and bred in an RP environment, "Adoptive RP" to refer to an RP pronunciation acquired after childhood (Wells, 1982: 283) and "Near RP" to refer to certain regional accents which, although falling outside the definition of RP, include very little in the way of regionalisms (Wells, 1982: 297-301). Furthermore, an often confusing array of modifying terms has been introduced by various authors to describe the social and generational shades of RP. In order to refer to an upper class or "posh" pronunciation, Wells (1982: 280), for example, chooses "U-RP" (as opposed to "mainstream RP"), whereas Windsor Lewis (1985: 253) refers to "conspicuous RP" and Cruttenden (1994: 80) adopts the term "refined RP". Gimson (1970) divided this category into generational subdivisions: "conservative RP" to refer to the pronunciation of the older generations and "advanced RP" for the speech of the younger speakers.

[8] Previous attempts to abandon RP and establish alternative labels have had limited repercussions. Trim (1961) referred to "English Standard Pronunciation", Wells and Colson (1971) used "Southern British Standard", Leitner (1982) talked of "Educated Southern English", while Windsor Lewis (1972 & 1985) argued for "General British".

> RP itself inevitably changes as the years pass. There is also a measure of diversity within it. Furthermore, the democratization undergone by English society during the second half of the twentieth century means that it is nowadays necessary to define RP in a rather broader way than was once customary (Wells 2008: xix).

LPD's main competitor, The *English Pronunciation Dictionary* (EPD), decided in 1997 to take what seemed to be a bold step and break with the tradition established in 1926 by Daniel Jones and discard the uncomfortable term RP. In the preface to the 15th edition, Roach and Hartman (1997) write: "the time has come to abandon the archaic name *Received Pronunciation*". Their decision, however, to replace it with "BBC English" inevitably invited criticism. To substitute one weighted term with another weighted term seemed to be a case of jumping out of the frying pan and into the fire. Needless to say, the BBC today has little to do with the BBC of yesteryear.

Matters were further complicated when in the mid 1980s an apparently new phenomenon, "Estuary English", emerged and was heralded by some as a potentially more representative pronunciation model. First coined in 1984 by David Rosewarne who described it as "modified regional speech...a mixture of non-regional and local south-eastern English pronunciation and intonation" (1984: 29), "Estuary English" (EE) has received considerably more popular attention than RP ever has. Apart from giving rise to easy to read humorous paperbacks such as Paul Coggle's *Do you Speak Estuary?* or Steve Crancher's *Dijja Wanna Say Sumfing*, various articles have appeared in British broadsheets as well as tabloids, sparking off debate about the state of the English language, and reviving the perennial concerns about declining standards in English pronunciation.[9] At the Tory Party Conference in 1995, Gillian Shepherd, the then Minister for Education, famously described "Estuary English" as a "bastardised version of Cockney",

---

[9] For "Estuary English", see, for example: Rosewarne (1984, 1994), Wells (1992, 1994a, 1994b, 1997, 1998), Maidment (1994), Coggle (1994), Parsons (1998), Haenni (1999), Altendorf (1999a, 1999b, 2003), Levey (2001), Przedlacka (2002), Mompean (2006). At the time of going to press most of these articles and other relevant information and references are available at http://phon.ucl.ac.uk/home/estuary.

calling it “slovenly” and placing the onus on teachers to maintain standards and eradicate it. [10]

As Wells rightly points out, “Estuary English” is nothing new:

> EE is a new name. But it is not a new phenomenon. It is a continuation of a trend that has been going on for five hundred years or more - the tendency for features of popular London speech to spread out geographically (to other parts of the country) and socially (to higher classes). The erosion of the English class system and the greater social mobility in Britain today means that this trend is more noticeable than was once the case (Wells 1997: 47).

From a sociolinguistic point of view, EE can be seen as a sort of middle ground between two extremes –London Cockney at one end and RP at the other.

> In the traditional British class society EE offers some sort of refuge […] Estuary English has, consciously or unconsciously been adopted by many as a means of avoiding being stereotyped, typecast or pigeonholed. The English public school boy, on one hand, in order not to be seen as *posh*, might rebel against the accent of his parents and consciously adopt an accent with *street cred* which will allow him more social mobility. To go all the way to the other extreme might sound artificial and forced. The Cockney, on the other hand, might *refine* his/her form of speech so as to avoid being categorised from the moment he/she opens his/her mouth. EE thus serves as a relatively safe no-man's-land (Levey, 2001: 263-4).

Changes in the media and the more relaxed attitudes of the BBC and other TV and radio stations have helped to make accent variation more acceptable. Whereas previously an RP accent may have been a pre-requisite for a job in broadcasting, regional accents have gained in acceptance and may be advantageous for certain types of programmes, particularly those aimed at younger audiences. In an effort to move with the times and to compete with the influx of new independent channels, there has been a noticeable swing towards a more popular form of broadcasting,

---

[10] Some argue that, in many cases, what is actually being produced is nothing more than an affected or pseudo-Cockney adopted by some well-educated middle class speakers looking for street credibility. Amongst those who have been accused of speaking what has become known as “Mockney” are Lily Allen and the Brit pop band Blur’s frontman Damon Albarn.

fronted by presenters with whom "the man in the street" can identify. The new breed of media personalities such as chat show host Jonathan Ross, chef Jamie Oliver, sports personality David Beckham or comedian Paul Merton, all of whom, to varying degrees, possess Estuary features, have given certain kudos to this pronunciation variety and have potentially become speech role models for the young.[11]

But "Estuary English" should not be viewed as a fixed point on a linear scale. Considerable overlap exists between EE and RP and EE and Cockney, and border lines are hazy, making it difficult to tell where one finishes and the next begins. Attempts have been made to define and demarcate those features which are considered Estuary rather than Cockney or RP (see, for example, Rosewarne: 1984 & 1994; Wells: 1992, 1994a, 1994b, 1998; Maidment: 1994). Although these serve as a guide, the distinctions are clearly open to debate and subject to change. To categorise accents and pronunciations in closed boxes in a world of intense demographic and social mobility is arguably unsafe, especially as what does not fit comfortably in a given box today may be admitted in the future.

Although features such as L-vocalisation, TH-fronting and T-glottaling may have originated in London, it would be wrong to suppose that they are demographically confined to this area. Many of the features attributed to EE have been noted outside the Thames Estuary area. Thus, rather than a self-contained local accent or dialect it would appear to form part of a more widespread sociolinguistic trend, reflecting, as John Maidment suggests, general changing attitudes towards language identity in Britain.

> An alternative explanation is that the perception of formality and informality has changed and that, in this post-modern age, it is quite acceptable to pick and mix accents. Perhaps, we ought to call this new trend Post-Modern English, rather than Estuary English. This is a suggestion I make with my tongue only slightly in my cheek (1994: 6).

---

[11] Recent research carried out at the University of Manchester and reported in various British media, including the BBC, *The Daily Mail* and *The Sun*, in April 2013, suggests that since his move to Los Angeles, David Beckham's speech "is getting posher". After listening to interviews of him before and after he went to the USA, Charles Boorman and Alix Roberts conclude that he is dropping his aitches far less and "nowadays speaks with more of a standard English accent".

4. Recent changes and developments

What follows is a compilation of some recent developments and phenomena which are prevalent in British English today yet which rarely make the EFL books. They are not new (phonetic features do not appear overnight out of thin air) – but have now become sufficiently widespread to be worthy of mention. Some might be considered RP, others EE, and others might be frowned at, but they are all undeniably present and therefore, at the very least, students should be made aware of their existence.

4.1. Monophthongisation

As the pace of modern life increases, speaking rates are speeding up (see, for example, Quené: 2008 & 2011), and as we speak faster, there is a tendency to reduce or shorten vowels and diphthongs. In most EFL books and pronunciation dictionaries [eə] is recommended or taught for SQUARE word sets (e.g. *fair hair*) and [ɪə] for NEAR (e.g. *near here*), although many, if not most, younger speakers today in southern Britain would probably produce monophthongs of the type [ɛ:] and [ɪ:] respectively or, if a diphthong is pronounced, the second element will tend to be very short.

In the case of CURE, although a diphthong [ʊə] is the model commonly used for learners of English, a long vowel of the type [ɔ:] is just as common if not more so. It is interesting to note that in the third edition of the *Longman Pronunciation Dictionary* (LPD3), the diphthong [ʊə] is recommended in the case of *cure, pure, tour*, yet [ɔ:] is suggested for *sure* and *poor*. One would expect that the shorter monophthong, for reasons of economy of speech, will eventually supersede the diphthong.[12]

---

[12] Data which appear in the different editions of the LPD suggest, however, that there may be some resistance. A comparison of the results of the 1988 and 1998 opinion polls which appeared in the 2nd edition of the dictionary suggested that the number of speakers who pronounce *sure* as [ʃɔ:] is gradually increasing (see Wells 2000: 752). While the diphthong [ʊə] is used by some older speakers, 60% of informants born after 1973 preferred [pɔ:]. This trend was repeated in an even more pronounced way for *poor* where 82% of speakers born after 1973 favoured the monophthong. It seemed therefore that [ʊə] in such cases was on its way to near-extinction. However, somewhat surprisingly, the figures and chart which appear in the 3rd edition show that in recent years the numbers of those who preferred [pɔ:] have levelled off and

4.2. Vowel Fronting

When we say that someone has a "plummy" accent or speaks "with a plum in his/her mouth" we are referring to someone with an upper class or *posh* accent. Why? Well try speaking with a small object in your mouth and you will find that as the front and central parts of your tongue and mouth are otherwise occupied, sounds will be produced further back in the mouth.[13] Very back realisations of /uː/ /ʊ/ /ɔː/ often sound quite old-fashioned, and in recent years these vowels are clearly moving forward (see de Jong et al. 2007). The GOOSE vowel tends to be produced as a centralised [ʉ] in a position considerably more advanced than Cardinal Vowel 8, where many books place it. This tendency, known as GOOSE fronting, was in fact recognised as a feature of modern RP in the 1980s (see Henton 1983), yet old habits die hard.

4.3. Yod Coalescence

Yod Coalescence is the name given to the process whereby [tj] [dj] clusters transform into [ʧ] [ʤ]. This is fairly commonplace in unstressed syllables across word boundaries and has been so for some time, especially when frequently used words such as *you* or *yours* are present (e.g. *what you need* [ˈwɒʧə ˈniːd], *bet your life* [ˈbeʧə ˈlaɪf], *would you like…?* [wʊʤə ˈlaɪk], *do you live?* [ʤə ˈlɪv]). It also occurs within words such as *adventure*, *nature, actually*, *education* and *gradual*, where it is increasingly unusual to hear the Yod pronounced today.

Yod Coalescence has spread in recent years to stressed positions (e.g. *tube* [ʧuːb], *Tuesday* [ˈʧuːzdi], *reduce* [rəˈʤuːs]), especially amongst younger speakers, making *dune, dew/due* and *dual* homophonous with *June, Jew*

have even dropped slightly to 74% (see Wells 2008: 627-8). Whether this slight fall is significant or not is difficult to say. Unfortunately, the word *sure* was not included in the 2007 poll which would have been useful to help compare and contrast results and see whether there is evidence to suggest that there is a possible shift back towards diphthong realisations. This, at first sight, seems unlikely, but stranger things have happened in spoken English!

[13] In George Bernard Shaw's play *Pygmalion*, Professor Higgins, during his elocution lessons with Eliza Doolittle, puts marbles in her mouth in an attempt to improve her pronunciation.

and jewel. This is a logical continuation of a well-established process but, perhaps because it is more conspicuous in stressed positions, some would classify this as non-standard. There is of course fundamentally no real reason why Yod Coalescence should be considered less acceptable in stressed positions than in unstressed positions. Negative judgements are often based on subjective views of "ugliness" which change with time.

Yod Coalescence may also occur when [s] and [z] are followed by [j] resulting in [ʃ] and [ʒ] (e.g. *tissue* ['tɪʃu:], *miss you* ['mɪʃə], *assume* [ə'ʃu:m] and *use your head* ['ju:ʒə 'hed]). However, some speakers may also include a yod after the coalescence resulting in ['tɪʃju:], ['mɪʃjə], [ə'ʃju:m] and ['ju:ʒjə 'hed]. In [stj] clusters it is not uncommon to hear an affricate pronounced after an initial fricative (*e.g. stupid* ['ʃtʃu:pɪd], *student* ['ʃtʃu:dənt] and *Estuary* ['eʃtʃəri].

Although Yod Dropping (e.g. *Tuesday* ['tu:zdi]) does occur in some southern British regional dialects such as in East Anglia (see, for example, Wells 1982: 338; Trudgill 1999: 133) and Cockney (Wells 1982: 331, Tollfree 1999: 174), it is still more commonly associated with General American English.

4.4. The American Influence

Given the appeal and influence of American popular culture and media in modern Britain, it is not surprising that General American has left its mark on British English. Differences between British and American English were once more defined but the dividing lines are beginning to blur.

While Americans are supposed to say ['i:ðə], British people are supposed to pronounce *either* as ['aɪðə]. According to Wells (2008: 265), this is still the case for RP, although these norms appear to be more relaxed amongst younger speakers. It is doubtful whether they can be considered regional markers any longer, and in Britain, it is simply a case of choosing between equally acceptable alternatives. I would suspect that most speakers today would find it difficult to say which of the two is British and which is American. Similarly, some consider the way the word *schedule* is pronounced to be an indicator of national origin. If the initial consonants are pronounced [ʃ], it is considered British English, while [sk] would be American. But the latter has become increasingly popular amongst young British speakers (see Wells 2008: 716-7), and I doubt whether, to most British ears today, one sounds particularly more American than another.

Word stress in words such as *advertising* and *controversy* has traditionally been a British/American shibboleth. An article which appeared in the Daily Telegraph on 5 February 2011 entitled "'ConTROversy' over changing pronunciations" began: "To language purists they may grate, but new ways of pronouncing words are spreading thanks to the influence of US culture". It went on to report the findings of a British library study which showed that the "American" stress patterns in these words are increasingly being adopted in Britain. In the case of the word *controversy*, three quarters of those who took part in the study placed the stress on the second syllable rather than the first. Although CONtroversy may still be the norm amongst RP speakers (see Wells 2008: 182), conTROversy is now so widespread that it has fast become the norm in Britain as a whole, and has no American associations whatsoever.

The complete elision of /t/ in *going to* (*gonna)* and *want to (wanna)* is often associated with American English. However, the omission of /t/ has also always been a feature of British English…after all, *The Beatles* did sing *I wanna hold your hand* in 1963 (although some might argue with the American market in mind). The dropping of /t/ (or in some cases replacing it with a tap) after nasals in words such as *Internet* or *International* remains quintessentially American, yet this process of elision in fast speech is finding its way into British English in words such as *twenty* or *plenty*.

The pronunciation of "t" as a tap or flap [t̬] (or [ɾ]) in intervocalic positions (e.g. *better, cotton*), which has always been one of the give-away features of American English, is also making clear inroads in British English. For many speakers, the tap offers a more comfortable alternative to the slower articulated alveolar plosive. They may or may not be aware they are producing it, and it is not unusual for different /t/ realisations to coexist in an individual's phonetic repertoire depending on context and environment.

It is interesting to note that Prince William, in the previously cited interview (see footnote 6), uses a mixture of realisations. He completely elides /t/ in the word *plenty* ("there's plenty of other people who've got worse problems than we have") while glottal stops are frequently used in word final positions, but not before vowels. In intervocalic environments he tends to use the 'American' tap. For example, agreeing with the interviewer's suggestion that there is a lot of "bad food eating" at university, he responds "yeah, there was a lot of that" ['jɛː ðə wəz ə 'lɒt̬əv 'ðæʔ].

4.5. T-Glottaling

T-Glottaling, the complete substitution of /t/ with the glottal stop [ʔ], has been a feature of southern British English for centuries although, until fairly recently, it was often brushed under the carpet as being an indication of poor education and breeding. Daniel Jones reflected (and perhaps still reflects) the view of many when, in advice to foreign learners, he wrote: "…the sound [ʔ] should in fact be avoided as much as possible. It is not a pleasant sound in itself, and is never necessary for the sense" (1909: 19). Once associated with low prestige lower class London and Glasgow accents, it was largely given a wide berth until the 1950s, when Christophersen (1952) recognised that the glottal stop had become embedded in the speech of RP speakers under the age of 40, and as such could no longer be simply ignored. In recent years, it would seem to have spread quickly across geographical and social boundaries, leading Trudgill (1999: 136) to claim that "the glottaling of intervocalic and word-final /t/ is one of the most dramatic, widespread and rapid changes to have occurred in British English".

The degree of acceptability of T-glottaling depends largely on how manifest or audible it is, and this principally depends on context and environment. Broadly speaking, the three possible environments are:

1. Before a consonant (e.g. *Great Britain* [greɪʔ 'brɪtən])

T-glottaling in this position is very common, and in words such as *Gatwick, football, seatbelt* and *suitcase*, I would say that a glottal release is standard practice in rapid speech. Similarly, when occurring across word boundaries (e.g. *get lost* ['geʔ 'lɒst]), it would go largely unnoticed and would therefore escape possible censure. Quite possibly, neither speaker nor listener will be aware that a glottal stop has been used.

2. Word-finally (e.g. *What* [wɒʔ])

When T-glottaling is produced at the end of the word before a pause, it becomes more conspicuous and therefore might be more frowned upon in certain circles. Nevertheless there has been a noticeable increase in the use of glottal stops in this environment, particularly with frequently used words like *right, but, what, got, that* and *it.*

3. Intervocalically (*e.g. What a pity* ['wɒʔ ə 'pɪʔi]).

The glottal stop becomes most noticeable between vowels across word boundaries (e.g. *Got a light?* ['gɒʔ ə 'laɪʔ]) or within a word (e.g. *better* ['beʔə]*, rotten* ['rɒʔən], *little* ['lɪʔəɫ]) and *city* ['sɪʔi]). It is in these environments that T-glottaling is most stigmatised, perhaps due to its associations with low prestige accents such as Cockney.

In the early 1980s, Wells considered pre-vocalic environments to generally fall outside the realms of what was considered mainstream, although he noted that glottalisation in word-final pre-vocalic positions is used by "some younger RP speakers" (1982: 261).[14] By the 1990s, Wells (1994a, 1994c, 1997) confirmed the presence in modern RP of word-final T-glottaling before vowels, but intervocalic T-glottaling remained "firmly excluded from RP" (1994c: 201). However, classifications of standard pronunciation models, whether the approach is descriptive or prescriptive clearly change. The emergence of the new watchword "Estuary English" has taken a certain degree of pressure off the somewhat overused and overworked term "Received Pronunciation", and it has become convenient to place traditionally non-standard features which are becoming increasingly common (e.g. T-glottaling) in this semi-acceptable category.[15]

---

[14] In 1982, Wells (1994a: 260-1) identifies the environments where T-Glottaling is acceptable in mainstream RP as:

- Word-finally before a true consonant (i.e. an obstruent) (e.g. *quite good*)
- Word-finally before a liquid or a semivowel (e.g. *quite likely*)
- Word-internally before a true consonant (e.g. *curtsey*)

[15] Wells (1992) suggests that the absence of pre-vocalic T-Glottlling, along with H-dropping, TH-fronting and MOUTH vowel monopthongisation distinguishes "Estuary English" from Cockney.

4.6. L-Vocalisation

When realising /l/ in pre-consonantal or pre-pausal positions, there is an increasingly common tendency to reduce or omit velar contact, leading to vocalic realisations in the close back region ranging from [ö] and [ɤ] to /ʊ/ or /u/.[16] Thus, *still ill* might be pronounced [stɪö ɪö]. This phenomenon, known as L-vocalisation, has been classified by many as a typical feature of "Estuary English" (see, for example, Rosewarne, 1984; Wells, 1992, 1997, 1998; Maidment, 1994; Altendorf, 1999b). However, in recent years it appears to be spreading both demographically and socially, although it remains stigmatised by some.

Negative attitudes towards what seems to be a natural phonetic progression, may again be due to its associations with the low prestige London Cockney accent and the class and education prejudices it has historically instilled. But many would argue that it is only a matter of time before L-vocalisation becomes standard practice, and indeed Jennifer Jenkins (2000: 139) argues that it already has: "The majority of RP speakers already pronounce pre-consonantal *dark* [ɫ] as [ʊ] in non-careful speech, although many would probably deny that they do so". She goes on to propose the vocalised /l/ as an acceptable alternative to the difficult dark /l/ in her *Lingua Franca Core* (LFC).[17]

From a phonetic point of view, it seems fair to say that L-vocalisation is a progression from velarisation where the contact of *dark* /l/ becomes gradually lighter until it disappears completely.[18] It could be argued that a

---

[16] Wells (1982: 259) points out that L-vocalisation has important implications for the English vowel system, offering possible phonemic status to new diphthongs such as [ɪʊ] or [ɪö] in *milk* or *pill* and [ɛʊ] or [ɛö] in *melt.*

[17] Jenkins argues that as dark [ɫ] does not exist in Welsh English, "it thus seems unreasonable to have 'higher' expectations of L2 speakers" (2000: 139).

[18] Just as the difference between clear and dark /l/ can not be viewed in simple binary terms, some would argue that there are degrees of vocalisation (Hardcastle and Barry, 1985; Wright, 1989: 358; Kerswill, 1995: 197). However, as Shockey (2003) rightly points out, from a strictly acoustic point of view, the difference between vocalised and velarised should, in fact, not be open to debate. If articulatory contact takes place, it should signify a consonant. On the contrary, if no contact takes place, a vocalic realisation is implied. For Shockey, the problem "lies in deciding how much tongue-palate contact can be allowed for a vowel and what it means for a consonant to be 'partially vocalised'" (2003: 35).

non pre-vocalic *clear* /l/ may evolve into a *dark* /l/, which in turn may eventually become vocalised.

4.7. TH-fronting

TH-fronting is the name given to the process of substituting dental fricatives /θ, ð/ with labiodental fricatives [f, v] (e.g. *think* [fɪŋk], *both* [bəʊf], *this* [vɪs], *mother* [ˈmʌvə]. Thus, minimal pairs such as *thin/fin three/free* and *than/van* may become homophonous. It is common in young British children, for whom "th" sounds may be difficult, and has traditionally been seen as a mild speech error which is usually corrected during adolescence. However, in a certain number of cases this may continue into adulthood, perceived either as a speech defect or seen as part of a regional variant.

Even if attitudes are changing, TH-fronting is generally considered a "non-standard" realisation, falling outside the confines of both RP and "Estuary English" (see, for example, Wells 1992). It is often associated with low prestigious lower class varieties and may be coupled with poor standards of education. In recent years, attitudes and the insistence on correct realisations of [θ] and [ð] have relaxed somewhat. In an increasingly accent-disperse and multi-cultural society, greater phonetic tolerance allows alternative realisations such as [f] and [v] to go unchecked and to gain in acceptance.

Although typical of London speech (see, for example, Wells, 1982: 328-30; Hughes and Trudgill, 1987: 44; Tollfree, 1999: 172), TH-fronting has radiated rapidly all over Britain. Of its development in Norwich, Trudgill writes:

> TH-fronting in modern England is a remarkable phenomenon. Formerly confined to the London area and to Bristol, it has in the 1980s and 1990s begun to spread enormously rapidly across England. It was totally absent from the 1968 Norwich corpus, but by the time of the 1983 survey it had made amazing inroads into the local dialect. None of the Norwich informants born before 1958 had this feature. Of informants born between 1959 and 1973, 70 per cent had some degree of loss of /θ/ and /ð/ through merger with /f/ and /v/, and 29 per cent had no instances of [θ] at all (1999: 137-8).

Dental fricatives [θ] and [ð] are relatively unusual in other world languages, and for many English language teachers it is not high on their list of priorities.[19] Jenkins (2000: 137) argues that the correct pronunciation of dental fricatives is not essential for mutual intelligibility for speakers of English as an International Language. If substituted, for example, by *f/v or s/z or t/d* effective communication and understanding would still be maintained. For this reasons she excludes them from her Lingua Franca Core, arguing that “a high level of difficulty coincides with a low level of salience for EIL intelligibility” (2000: 137).

4.8. H-dropping

H-dropping is the term used to refer to the elision of /h/ (e.g. *Did he have a happy honeymoon?* ['dɪdi 'æv ən 'æpi 'ʌnimu:n]. Although H-dropping is typically associated with the non-prestigious speech of the lower classes, Honey (1989: 44-5) cites several examples of use amongst the upper classes, including members of the Royal family such as the Queen Mother and Prince Charles.

As with T-glottaling, to be stigmatised, H-dropping must be noticeable. It is common to elide the “h” in frequently used weak forms of function words such as *he, him, his, have, has*. After consonants (e.g. *give him* ['gɪvɪm], *did he* ['dɪdi]), elision would usually go undetected. Even after vowels (e.g. *see him* ['si: ɪm], *she has* [ʃi əz]), although slightly more noticeable, H-dropping is so common that it is usually uncensored. H-dropping is most conspicuous at the beginning of words, particularly when the syllable is stressed. Thus, to drop your *aitches* in, for example, *he has a horrible headache* might be associated with Cockney English and invite possible reproach.

4.9. Intrusive R

Intrusive R is a sandhi or linking phenomenon. It is the reinterpretation of the linking R rule whereby [r] is inserted epenthetically to join two consecutive vowels, but in this case, there is strictly speaking no orthographic justification

[19] In the 1970s Gillian Brown (1974: 53) advised EFL teachers that “when time is short, it is probably not worthwhile spending time on teaching /θ/ and /ð/ if the students find them difficult”.

to do so (e.g. *saw a film* ['sɔ:r ə 'fɪlm]). This means that, while it is acceptable in non-rhotic accents to insert /r/ to create a liaison between the vowels in *lore is* ['lɔ:r ɪz], in the case of *law is* ['lɔ: ɪz], although the individual phonemes are identical, this liaison is deemed incorrect or non-standard. In the first case it would be classified as linking R, whereas in the second it would be classified as intrusive R. Although it may irritate the sharp-eared sticklers who may condemn it as a sign of bad English, to most people today, intrusive R is a fast speech phenomenon which is produced without thinking and goes largely unnoticed. It is interesting to note that Hannisdal's (2006) study found that, of the 30 BBC, Sky and ITV newsreaders she analysed, 27 used intrusive R to varying degrees.

4.10 A new R…or no R?

In the first half of the twentieth century, much importance was given to the correct pronunciation of /r/. Elocution exercises, consisting of sentences such as "around the rocks the ragged rascal ran", were designed to help the socially insecure tap their "R"s. Noel Coward (1899-1973), the English actor, playwright, singer and wit, is often associated with a type of upper-class speech which has all but disappeared. If you listen to recordings of him, one notices, among other things, his crisp tap [ɾ]. Today, this realisation is rare and has given way to an approximant [ɹ] in non-rhotic accents where there is little or no articulatory contact. This can prove problematic, not only for foreign learners, but for natives too, for it is one of the most difficult sounds for English children to acquire.

There are claims that in the south of England there is an allophonic realisation which some have baptised "Estuary R". According to Rosewarne (1984:29) "...in the typical Estuary realisation the tip of the tongue is lowered and the centre part raised to a position close to, but not touching, the soft palate" while for Coggle (1993), the realisation sounds something like "w". Partly because these descriptions are far from clear, Wells (1994c) and Maidment (1994) question both the possible place of articulation and its validity. Whether it is predominantly realised as a labiodental approximant [ʋ] or a velar approximant [ɰ] is debatable, but an approximant variant notably different to RP [ɹ] does appear to exist, particularly in the south of England. The question is whether the /r/ realisation of Southerners such as Jonathan Ross, Paul Merton and Roy Hodgson can be classified as a valid regional variant, or whether in some or all cases, it is simply a minor speech defect.[20] It is possible that through increased language tolerance, as is the case with TH-fronting, the typical infantile mispronunciation of approximant [ɹ] is left uncorrected by carers and educators. Some would argue that this is indicative of declining standards as less importance is given to correct articulation. However, whatever the reasons, as the number of role models increases, is it feasible that what was once seen as a speech defect may attain legitimate variant status in a near future? Who knows?

5. Conclusion

The considerations to be borne in mind when choosing a pronunciation model for EFL students have been debated at length (see, for example, Dziubalska-Kolaczyk & Przedlacka 2008; Mompean 2008). But, to a large extent this decision will be conditioned by the teacher's own pronunciation and the text books available. As commercial considerations come into play, these understandably tend to play it safe and don't stray too far from established norms (I know of no widely-available English teaching manual which claims to use "Estuary English" as a model).

But is it really necessary to opt for one model? Although it might be convenient and comfortable to simplify and standardise the English phonetic system for learning purposes, it is also important to make students aware that there are other realisations that they may come into contact with, whichever heading we decide to place them under.

---

[20] Jonathan Ross is "affectionately" known as "Wossy", because of the way he pronounces (or doesn't pronounce) his Rs. When Roy Hodgson was appointed England's football manager on the 2 May 2012 just before the European Championships, the Sun newspaper's much criticised headline read: "Woy gets England job. Bwing on the Euwos. We'll see you in Ukwaine against Fwance."

As a teacher, I am well aware that, in some cases, there is a difference between what books suggest I *should* say and what I *actually* say. I can reproduce diphthongs in words such as “parents” and “nearly” if required to, but, if left to my own devices, I would use monophthongs. In certain contexts I also confess to vocalising my *Ls* and replacing my *Ts* with glottal stops or even an “American” tap. So what English do I speak? I have a British accent rather than an American one. I think my English is pretty standard, but since most people would recognize that I am from the south of England I suppose I could not be classified as an RP speaker in the strictest sense, as, by its very definition, that model is not localised. I was born in London but I don’t speak Cockney nor do I sound like David Beckham.

Although attempts have been made to define and classify pronunciation models and their features, the truth is that people do not speak in models. Nor do they fit easily into pigeon holes. It is not unusual to find people who have some “RP” features, some “EE” features and others which might be said to belong to “American English”. We are exposed to a wider variety of accents than ever before and it is logical that these will influence our way of speaking. We should not forget that there is also a human tendency to adapt and accommodate, consciously or unconsciously, to the speech of others around us (if we are positively disposed to them).

But should we expose our students to some or all of the above mentioned features? Of course each must be considered individually (monophthongisation and TH-fronting, for example, are different kettles of fish), but, in principle, I see no real reason why not, as long as they are made aware of possible connotations and implications. Any innovation or departure from “established norms” has of course to be treated with a certain degree of caution, and the attitudes, possible prejudices and prestige/stigma judgements of the society we live in cannot be ignored. Whether we teach them or not will also clearly depend on many practical factors, such as the needs, level and maturity of our students as well as the course content and time available…and of course the teacher’s own pronunciation.

To show students *real* English in as many shapes and forms as possible is important. We are doing them no favours by overlooking new and not-so-new developments and departures which do not fit comfortably into established norms. It is a bit like the dilemma of whether or not to teach students slang and swear words. Teachers know they exist and that they are used all the time, yet it is often felt that the language classroom is not the place for them. In my experience, however, although students may be keen to know and recognize “deviant” forms, this does not mean they will actually want to use them. Furthermore, even if we choose not to teach “aberrant” language, students will invariably come into contact with it sooner than later anyway. In the past there has been a tendency to protect language learners from what were considered to be non-standard features and to shy away from exposing them to recent innovations. But, by doing this, we are perhaps doing students a disservice, since, when they come into contact with English outside the classroom, they may find a considerable and challenging difference between what they have been taught and what they actually hear.

REFERENCES

Altendorf, Ulrike (1999a): “Es(h)tuary English: Is English going Cockney?”. *Moderna Sprak 93(1)*: 1-11.

Altendorf, Ulrike (1999b): “Approaching the notion of ‘Estuary English’ /t/glottaling and /l/ vocalisation by the Thames Estuary”. In: CaritaParadis (ed.) 1999. *Recent trends in the pronunciation of English: social, regional and attitudinal aspects.* Stockholm: Almqvist & Wiksell, 15-31.

Altendorf, Ulrike (2003): *Estuary English. Levelling at the Interface of RP and South-Eastern British English*. Tübingen: Gunter Narr.

Brown, Gillian (1974): “Practical phonetics and phonology”. In: Allen, J.P.B. & Corder, S.P. (eds.). *The Edinburgh Course in Applied Linguistics Vol. 3: Techniques in Applied Linguistics.* Oxford: Oxford University Press, 24-58

Christophersen, P. (1952): “The glottal stop in English”. *In*: *English Studies 33*: 156-63.

Coggle, Paul (1993): *Do you Speak Estuary?* London: Bloomsbury.

Crancher, Steve (2003): *Dijja wanna say sumfing?*: A guide to Estuary English. Romford: Ian Henry Publications.

Cruttenden, Alan (1994): *Gimson’s Pronunciation of English*, 5th Edition, London: Arnold.

Dziubalska-Kolaczyk, K., & Przedlacka, J. (eds) (2008): *English pronunciation models: a changing scene.* Bern: Peter Lang.

Fisher, J.H. (1993): "The history of Received Pronunciation." In: U. Goebel, U & Lee, D. (eds.) *The Ring of Words in Medieval Literature*. Lewiston/ Queenston/Lampeter: Mellen, 41-61

Gimson, A.C. (1970) [1962]: *An Introduction to the Pronunciation of English* (2nd ed). London: Edward Arnold.

Haenni, Ruedi (1999): *The Case of Estuary English: supposed evidence and a perceptual approach*. Dissertation, University of Basel.

Hannisdal, Bente Rebecca (2006): *Variability and change in Received Pronunciation: A study of six phonological variables in the speech of television newsreaders*. Doctorate Dissertation. University of Bergen.

Hardcastle, W.J. & Barry, J.W. (1985): “Articulatory and perceptual factors in /l/ vocalisation in English”. *University of Reading Phonetics Laboratory Work in Progress no 5*: 31-44.

Harrington, Jonathan *et alii* (2000): “Does the Queen speak the Queen’s English?: Elizabeth II’s traditional pronunciation has been influenced by modern trends”. *Nature* 408: 927-928.

Henton, Caroline (1983): “Changes in the vowels of Received Pronunciation”. *Journal of Phonetics 11/4*: 353-371.

Honey, John (1989): *Does accent matter? The Pygmalion factor.* London: Faber and Faber.

Hughes, Arthur & Trudgill, Peter. (1987) [1979]: *English accents and dialects: An Introduction to social and regional varieties of British English.* London: Edward Arnold.

Jenkins, Jennifer (2000): *The Phonology of English as an International Language*. Oxford: Oxford University Press.

De Jong, Gea, *et alii* (2007): "The speaker discriminating power of sounds undergoing historical change: A formant-based study". *Proceedings of 16th ICPhS. Saarbrücken*, 1813–1816

Jones, Daniel (1984) [1909]: *The Pronunciation of English* Cambridge: Cambridge University Press.

Jones, Daniel (1917): *An English Pronouncing Dictionary* (1st edition). London: J.M. Dent

Jones, Daniel (1926): *An English Pronouncing Dictionary* (3rd edition). London: J.M. Dent.

Kerswill, Paul (1995): "Phonological convergence in dialect contact: evidence from citation forms". *Language Variation and Change 7*: 177-202.

Leitner, Gerhard (1982): "The consolidation of 'Educated Southern English' as a model in the early 20th Century". *International Review of Applied Linguitics 20*: 91-107.

Levey, David. (2001): "Is EFL ready for EE?" *In:* Levey, D. *et alii* (eds.) (2001). *English Language Teaching: Changing Perspectives in Context*. Cádiz: Servicio de Publicaciones Universidad de Cádiz, 261-72.

Maidment, John (1994): "Estuary English: Hybrid or Hype". Paper given at 4th New Zealand Conference on Language & Society, Lincoln University, Christchurch, New Zealand, August 1994. (http://phon.ucl.ac.uk/home/estuary/home.htm). [downloaded 25/03/2012]

Mompean, José Antonio (2006): "Estuary English: Revisiting the Debate on its status as a New Accent of English and Potential EFL Pronunciation *Model". Barcelona English Language and Literature Studies. Monograph on Empirical English Linguistics. 16*: 1- 13.

Mompean, José Antonio (2008): "Consumers' preferences and the Choice of English Pronunciation Models". *In:* Monroy, R. & Sánchez, A. (eds). *25 Years of Applied Linguistics in Spain: Milestones and Challeges*. University of Murcia, Spain, 959-964.

Parsons, Gudrun (1998): *From "RP" to "Estuary English".* M.A. dissertation, Hamburg University.(http://phon.ucl.ac.uk/home/estuary/home.htm). [downloaded 25/03/2012]

Przedlacka, Joanna. (2002): *Estuary English? A sociophonetic study of teenage speech in the Home Counties*. Bern: Peter Lang

Quené, Hugo. (2008): "Multilevel modelling of between-speaker and within-speaker variation in spontaneous speech tempo". *Journal of the Acoustical Society of America*, 123(2): 1104-1113.

Quené, Hugo. (2011): "Are we speaking faster?" *In*: Zonneveld, W *et alii* (eds.). *Sound and Sounds: Studies presented to M.E.H. (Bert) Schouten on the occasion of his 65th birthday*. Utrecht: UiL OTS, 161-171).

Roach, Peter & Hartman, James. (eds.) (1997): *English Pronouncing Dictionary* (15th Edition) Cambridge: Cambridge University Press.
Rosewarne, David (1984): “Estuary English”. *The Times Educational Supplement*, 19 October 1984: 29.
Rosewarne, David (1994): “Estuary English: tomorrow's RP”. *English Today* 37, 10/1: 3-8.
Shockey, Linda (2003): *Sound Patterns of Spoken English.* Oxford: Blackwell.
Tollfree, L.F. (1999): “South East London English: discrete versus continuous modelling of consonantal reduction”. In*:* Foulkes, P & Docherty, G. (eds.). *Urban Voices*: *Accent Studies in the British Isles.* London, Arnold, 163-184.
Trim, J.L.M. (1961): “English Standard Pronunciation”. *English Language Teaching* 16/1: 28-37
Trudgill, Peter. (1974): *The social differentiation of English in Norwich.* Cambridge: Cambridge University Press.
Trudgill, Peter. (1999): “Norwich: endogenous and exogenous linguistic change”. *In* Foulkes, P and Docherty, G. (eds.). *Urban Voices*: *Accent Studies in the British Isles.* London, Arnold, 124-140.
Walker, John (1791): *A Critical Pronouncing Dictionary and Expositor of the English Language.* London: Robinson and Cadell
Wells, John (1982): *Accents of English*. Cambridge: Cambridge University Press.
Wells, John (1992): “Estuary English?!?”. Handout for talk given at 1992 BAAP Colloquium, Cambridge.(http://phon.ucl.ac.uk/home /estuary/ home.htm). [downloaded 25/03/2012]
Wells, John (1994a): “Transcribing Estuary English” *Speech, Hearing and Language, 8, UCL Work in* Progress: 261-7.
Wells, John (1994b): “Recommendations for standardized phonetics of Estuary English”. Handout from a talk given in Heidelberg, November 1994. (http://phon.ucl.ac.uk/home/estuary/home.htm). [downloaded 25/03/ 2012]
Wells, John (1994c): “The Cockneyfication of RP?”. In: Melchers, G & Johannneson, L. (eds.). *Nonstandard varieties of language*. Stockholm: Almqvist & Wiksell, 198-205.
Wells, John (1997): “What is Estuary English?” *English Teaching Professional* 3: 46-7.
Wells, John (1998): “Pinning Down Estuary English”. Abstract from a lecture given in Lund, Sweden, April 1998. (http://phon.ucl.ac.uk/home/estuary/ home.htm). [downloaded 25/03/2012]
Wells, John (2000): *Longman's Pronunciation Dictionary* (2nd edition)*.* London: Longman.
Wells, John (2008): *Longman's Pronunciation Dictionary.* (3rd edition) London: Longman.
Wells, J.C. and Colson, G. (1971): *Practical Phonetics*. London: Pitman Publishing.

Windsor Lewis, Jack (1972): *A concise pronouncing dictionary of British and American English.* London: Oxford University Press

Windsor Lewis, Jack (1985): “British non-dialect accents”. *Zeitschrift für Anglistik und Amerikanistik* 33.3, 244-257

Wright, S. (1989): “The effects of style and speaking rate in /l/-vocalisation in local Cambridge English”. *York Papers in Linguistics 18*: 355-65.