

## FONÉTICA EXPERIMENTAL Y TECNOLOGÍAS DEL HABLA

### EXPERIMENTAL PHONETICS AND SPEECH TECHNOLOGIES

Juan María GARRIDO ALMIÑANA

*Grup de Lingüística Computacional (GLiCom)*  
*Universitat Pompeu Fabra*

#### RESUMEN:

Las relaciones entre Fonética y Tecnología del Habla han sido objeto de debate reiterado en la bibliografía de ambas disciplinas, debate en el que se ha puesto de manifiesto, por un lado, la innegable necesidad del conocimiento lingüístico en el desarrollo de estas tecnologías, y por otro, los problemas y limitaciones que ha supuesto en ocasiones para la Tecnología del Habla la falta de descripciones y modelos fonéticos adecuados, lo que ha llevado en muchos casos a la búsqueda de «camino alternativos», principalmente en la estadística y la probabilidad. Este artículo pretende ofrecer una reflexión personal sobre las relaciones entre ambas disciplinas, en el momento actual y en un futuro próximo, analizando, por un lado, en qué medida el conocimiento fonético puede contribuir al desarrollo de las tecnologías del habla en los próximos años, y por otro, de qué forma la Tecnología del Habla puede ayudar a mejorar los métodos y resultados del trabajo teórico en Fonética experimental.

PALABRAS CLAVE: Fonética, tecnologías del habla.

#### ABSTRACT:

The relations between Phonetics and Speech Technology have been the focus of a long discussion in the literature of both disciplines. This discussion has shown, on the one hand, the undeniable need of linguistic knowledge for the development of these technologies, and, on the other, the problems and limitations caused to Speech Technology by the lack of adequate phonetic descriptions and models, a fact that has caused in many cases the search for 'alternative paths', mainly in Statistics and Probability. This paper aims to offer a personal view about the relations between both disciplines, today and in the near future, by analyzing, on the one hand, to what extent phonetic knowledge can contribute to the development of Speech Technologies in the coming years, and, on the other, in which ways Speech Technologies can help to improve the methodologies and results of the theoretical work on Experimental Phonetics.

KEY WORDS: Phonetics, Speech Technology.

## 1. INTRODUCCIÓN

Las relaciones entre Fonética y Tecnologías del Habla han sido objeto de discusión desde el nacimiento de estas, hacia mediados del siglo XX, tanto desde el campo de la tecnología (Fant, 1983; Acero, 1995; Öhman, 2001; Greenberg, 2001; Ainsworth, 2005; Strik, 2005; entre muchos otros) como del de la Fonética (Llisterri *et al.*, 1999; Pols, 2001; Shriberg, 2005; Llisterri, 2007; Garrido, 2009; también entre otros). A lo largo de este debate, los ingenieros desarrolladores de Tecnología del Habla han señalado siempre la necesidad del conocimiento fonético para esta tarea (*We need a fifth generation of speech scientists rather than a fifth generation of computers*; Fant, 1983). Sin embargo, los intentos iniciales de desarrollar tecnologías del habla a partir de conocimiento fonético no obtuvieron resultados demasiado esperanzadores. En el campo de la conversión texto-habla, aproximaciones como la síntesis por reglas (Pattern Playback, Cooper, 1953; Holmes *et al.*, 1964), la síntesis articulatoria (Rosen, 1958; Cocker, 1968) o la síntesis por formantes (Dixon y Maxey, 1968; sistema INFOVOX, Carlson *et al.*, 1982), ofrecían resultados inteligibles, pero sin un nivel de naturalidad suficiente. Y en el caso del reconocimiento, las aproximaciones basadas en conocimiento lingüístico (Zue, 1983, 1985) tampoco dieron los resultados esperables. De ahí que desde el campo de la ingeniería se propusieran «camino alternativos», menos basados en conocimiento lingüístico, como la síntesis por concatenación de unidades (Olive, 1977; proyecto MBROLA, Dutoit, 1996), la síntesis a partir de corpus (Black y Campbell, 1995; Hunt y Black, 1996) o la síntesis mediante modelos ocultos de Markov (HMM), en el caso de la conversión texto-habla, o el uso de HMM y n-gramas para la creación de los modelos acústicos y textuales, respectivamente, en el caso del reconocimiento. El uso de estas técnicas, basadas en la estadística y la probabilidad, han permitido obtener unos resultados mucho mejores que las basadas en conocimiento lingüístico, con lo que estas han ido quedando poco a poco relegadas a un segundo plano, limitadas al ámbito de la investigación. Se ha discutido bastante sobre las causas de este rendimiento limitado de los sistemas basados en conocimiento lingüístico, y una de las más repetidas ha sido la falta de descripciones fonéticas lo suficientemente detalladas y formalizadas del habla en sus diferentes ámbitos en el momento de abordar su desarrollo. Puede decirse que, en general, el conocimiento teórico ha ido por detrás de las demandas de los ingenieros en este proceso. Así, por ejemplo, la investigación sobre invarianza fonética iniciada en los años 50 del siglo pasado (Liberman *et al.*, 1954) no tuvo continuidad más allá de los años 80 (Blumstein y Stevens, 1981), lejos de haber sido concluida; la definición de modelos descriptivos sobre la entonación iniciada en los años 80 del siglo pasado (Pierrehumbert, 1980; ‘t Hart *et al.*, 1990) es aún una tarea pendiente de ser completada; y aunque se han llevado a cabo ya algunos pasos (proyectos VILE I, VILE II y VILE P, por ejemplo; Battaner *et al.*, 2003, 2005), la descripción de la variación fonética, tanto a nivel segmental como suprasegmental, es aún una tarea mayoritariamente pendiente. Ello no quiere decir que en el desarrollo de esta tecnología se haya prescindido totalmente de la información

fonética, y de hecho en determinados componentes es un elemento esencial, pero es un hecho comúnmente aceptado que su contribución ha sido limitada.

Si la contribución de la Fonética al ámbito de la tecnología ha sido hasta ahora moderada, en el sentido opuesto, el desarrollo de la Fonética experimental ha estado desde siempre indisolublemente ligado a las mejoras tecnológicas. El trabajo en esa disciplina ha necesitado tradicionalmente de instrumentación y tecnología, primero mecánica, más tarde analógica (sistemas analógicos de grabación, espectrógrafos, electropalatógrafos), y finalmente digital (sistemas de grabación digital, como el DAT; programas de análisis acústico). La llegada de la tecnología digital ha supuesto una verdadera revolución en la forma de hacer Fonética, hasta el punto de que la Fonética experimental ha dejado de ser una disciplina de laboratorio para convertirse en «computacional», que puede ejercerse con tan solo un ordenador personal y los programas adecuados, tal como ya señaló Öhman:

I believe that the impressive progress in electronics, sound engineering and computer technology that we have witnessed in the last few decades provide linguists with several excellent ways of simplifying language study and making it more efficient (Öhman, 2001).

Sin embargo, el peso en el proceso experimental de esta nueva tecnología digital que se utiliza hoy en Fonética, y que puede denominarse también «del habla», no en el sentido en que el término se usa habitualmente, sino en otro más amplio de tecnología que permite el procesado del habla tanto para fines teóricos como aplicados, es aún menor que en otras disciplinas científicas: aún hoy, las mediciones se hacen manualmente, con lo que la cantidad de datos analizados es aún necesariamente reducida, además de que las mediciones quedan sometidas a la posibilidad de errores o imprecisiones, por que existe aún un amplio margen para la mejora.

El objetivo de esta ponencia no es tanto mirar al pasado de las relaciones entre Fonética y Tecnologías del Habla, ampliamente descritas por muchos otros especialistas anteriormente, sino plantear una visión personal, probablemente discutible, sobre su evolución en un futuro próximo, y la contribución de cada disciplina al desarrollo de la otra. Por un lado, cabe preguntarse si, a pesar de los fracasos anteriores, el conocimiento obtenido del trabajo experimental en Fonética puede aún contribuir al desarrollo de las tecnologías del habla en los próximos años, y si los expertos en Fonética experimental seguirán teniendo un papel en el desarrollo de esta tecnología. Y en el sentido contrario, esta por ver también hasta qué punto la Tecnología del Habla, entendida en un sentido más amplio de herramientas de procesado computacional del habla, puede ayudar a mejorar los métodos y resultados del trabajo teórico en Fonética experimental en un futuro próximo.

## 2. EL PAPEL DE LA FONÉTICA EXPERIMENTAL EN LA EVOLUCIÓN DE LAS TECNOLOGÍAS DEL HABLA

A día de hoy, las tecnologías del habla que se pueden denominar «tradicionales» (conversión texto-habla, reconocimiento del habla y gestión de diálogo) han alcanzado un grado de madurez suficiente como para permitir el desarrollo de aplicaciones que sean atractivas para el gran público, y por tanto rentables desde el punto de vista comercial. Probablemente el ejemplo más mediático sea el de los interfaces persona-máquina para dispositivos móviles, como Siri, de Apple, o Sherpa, de Google, pero también empiezan a generalizarse en el campo de la automoción (sistema SYNC, de Ford) o en páginas web, en forma de asistentes personales como el de Renfe. Existen también en el mercado otras aplicaciones no orientadas al gran público que incluyen tecnología del habla, como los productos de *speech analytics*, concebidos básicamente para empresas (Eureka, de CallMiner, por ejemplo), o los productos de ayuda a la discapacidad, como lectores de pantalla (por ejemplo, JAWS, de Freedom Scientific), aplicaciones de ayuda a la escritura (WordQ, de GoQ), o de aprendizaje de lectura y escritura y apoyo al estudio (Kurzweil 3000/Firefly, de Kurzweil Educational Systems; Ditres, de Rehasoft). Y la lista no es, ni mucho menos, exhaustiva. Puede afirmarse que la tecnología del habla ha dejado de ser una tecnología de laboratorio y está empezando a ser un negocio que genera empleo. Ahora bien, ¿empleo solo para ingenieros o también para expertos en Lingüística, y más concretamente en Fonética?

Las tecnologías del habla que se emplean en estas aplicaciones comerciales son una mezcla de conocimiento fonético y métodos estadísticos. Ejemplos de módulos o componentes que suelen emplear técnicas basadas en conocimiento fonético para su funcionamiento serían los módulos de procesado lingüístico del texto de entrada o los módulos de predicción de la Prosodia (duración segmental y F0), en el caso de la conversión texto-habla, o los diccionarios de pronunciación en los sistemas de reconocimiento. Determinadas tareas del desarrollo de estas tecnologías implica también una mezcla de conocimiento fonético y procesado estadístico, como el diseño y grabación y procesado de corpus para la síntesis o el reconocimiento, el diseño de las aplicaciones basadas en gestión de diálogo, o la evaluación de los sistemas antes de ser comercializados. Parece claro, pues, que en la actualidad las personas que trabajan en al menos determinadas partes del desarrollo de estas tecnologías deben manejar, de una u otra forma, un cierto conocimiento de la Fonética de la lengua o lenguas con la que se está trabajando. Los fonetistas tienen hoy día un papel en el **desarrollo** de tecnología del habla comercial, en aspectos como la incorporación de nuevos idiomas y desarrollo de voces sintéticas para sistemas comerciales, la adaptación de sistemas de reconocimiento para nuevos idiomas y situaciones, o el desarrollo de servicios basados en sistemas de diálogo, integrados, eso sí, el equipos interdisciplinarios en los que los ingenieros son hoy día mayoría.

Pero ¿y en el futuro? ¿Qué perspectivas se perfilan para la evolución de estas tecnologías en los próximos años? ¿Se intuyen nuevas necesidades de conocimiento fonético? Las líneas de investigación y desarrollo de nuevos productos en este campo muestran una tendencia a ir incorporando estas tecnologías al ámbito de nuestra vida

privada, con nuevas aplicaciones en el campo de la robótica (asistentes personales o juguetes, como AIBO, el pionero perro robótico de Sony, ya retirado del mercado), la domótica (control de electrodomésticos, calefacción, luz y otros elementos de una casa mediante la voz, con algunos productos ya en el mercado, como INVOX, de la empresa Vócali, o el Samsung Smart TV, que incorpora también entre sus funciones la navegación mediante voz) o la teleasistencia (con sistemas de ayuda en casa a personas mayores y enfermos que incorporen también asistentes virtuales multimodales semejantes a Denise, ya comercializado). Este salto a aplicaciones más pensadas para el ámbito privado implica nuevos retos tecnológicos, que tienen que ver, por un lado, con la mejora de la tecnología actual (sistemas de conversión texto-habla con un mayor grado de naturalidad y capaces de expresar emociones, sistemas de reconocimiento de habla mucho más robustos, pensados para usarse en entornos ruidosos o con habla informal y emotiva, y sistemas de diálogo mucho más flexibles y adaptados también a entornos informales), y que implicarán la mejora de determinados módulos de estos sistemas que manejan conocimiento lingüístico, como los modelos prosódicos en los conversores texto-habla, o los léxicos y modelos de lenguaje en los reconocedores. Pero también representarán la consolidación de nuevas tecnologías ahora emergentes, como la identificación de idioma (identificación en la señal de habla del idioma en se expresa un locutor), el reconocimiento automático de emociones (identificación de las emociones que transmite un locutor determinado a través del habla) o la diarización de hablantes (identificación en un diálogo de las intervenciones de los distintos participantes). Ante esta perspectiva, parece claro que determinadas descripciones fonéticas, como la del habla en diferentes situaciones (diálogos orientados a un objetivo, diálogos informales, del habla expresiva y emocional (en especial de sus aspectos prosódicos), o de la variación interlocutor e intralocutor (con especial atención a determinados colectivos, como personas ancianas o niños), podrán ser útiles en futuros desarrollos de Tecnología del Habla. Deberían ser, por tanto, objetivo preferente para los investigadores en Fonética experimental durante los próximos años.

Otra cuestión diferente es hasta qué punto los fonetistas del futuro podrán participar directamente en el desarrollo de estas nuevas tecnologías, o en la mejora de la actual. Dicho de otra manera, si en los próximos años cabe esperar un retorno a las técnicas basadas en conocimiento lingüístico (reglas, diccionarios, etc.) en detrimento de las utilizadas hoy en día (modelos estadísticos obtenidos a partir de grandes corpus anotados con la información necesaria). Se ha afirmado en varias ocasiones que la aproximación basada en técnicas estadísticas tiene un límite:

It would be a wishful dream to extract all this knowledge merely by computerized statistics, i.e., to collect a very large material of speech, give the computer some help for segmenting transcription, and look up and then just wait for the results to drop out. Many institutions are now developing such data banks for their research. This is a necessary involvement to make but satisfies a partial need only. We cannot store all possible patterns with table look-ups. To organize the data bank efficiently, we must rely on a continuing development of a model of speech production and generative rules on all levels up to the linguistic frame and down to an advanced vocal tract model which should include all what we know of aerodynamics and source filter interaction (Fant, 1983).

Sin embargo, en la era de los *big data*, del uso de grandes cantidades de datos para el estudio de cualquier aspecto, se hace difícil pensar que ese límite pueda alcanzarse en futuro próximo. Así, la aportación de los expertos en Fonética a la **investigación** en Tecnología del Habla, al desarrollo de nuevos prototipos aplicando técnicas basadas en conocimiento lingüístico, es más incierta. En todo caso, las aproximaciones basadas en conocimiento lingüístico siguen siendo una línea de investigación activa en determinados campos de investigación dentro de las tecnologías del habla, dadas las dificultades que presentan las aproximaciones por corpus para la solución de determinados problemas tecnológicos. Así ocurre, por ejemplo, con la síntesis de emociones, en el que la investigación utilizando técnicas paramétricas (basadas en conocimiento lingüístico) convive con la exploración de otras técnicas más probabilísticas, como la síntesis basada en HMM. El trabajo presentado en Laplaza y Garrido (2014) sería un ejemplo de aplicación de conocimiento lingüístico para la generación de habla sintética con emociones en una situación de lectura en voz alta de mensajes de chat, con resultados semejantes a los obtenidos con otras técnicas. En el campo del reconocimiento, el desarrollo de diccionarios de variantes de pronunciación (Fodge, 2014, por ejemplo) parece un campo en el que el uso de reglas y diccionarios puede ofrecer resultados incluso mejores que los obtenidos por aproximaciones basadas en corpus. Y finalmente, el diseño, recogida y anotación de corpus de habla, una tarea en la que los fonetistas han tenido hasta ahora un papel destacado, seguirá siendo una necesidad para el desarrollo de las futuras tecnologías del habla, con independencia de las técnicas usadas.

### 3. EL PAPEL DE LAS TECNOLOGÍAS EN LA EVOLUCIÓN DE LA FONÉTICA EXPERIMENTAL

Como ya se ha mencionado anteriormente, las nuevas «tecnologías del habla» para el procesado computacional del habla han supuesto un gran avance en los métodos utilizados en Fonética experimental. A día de hoy, no se concibe el trabajo en fonética experimental sin el uso de herramientas computacionales de procesado del habla. Sin embargo, el uso de este tipo de herramientas está aún en sus comienzos. En este sentido, la aparición de programas como Praat (Boersma, 2001) o R, que incluyen un lenguaje para la creación de scripts que facilita la automatización y personalización de determinados procesos, representan un paso muy importante, en la medida que facilitan el desarrollo de nuevas herramientas a investigadores sin conocimientos profundos de programación. Dos aspectos en los que probablemente el uso de estas tecnologías tendrá una especial incidencia en los próximos años será, por un lado, su uso para la anotación de grandes corpus, y por otro, la automatización de los procesos de extracción y análisis de datos fonéticos.

#### 3.1. *El desarrollo de grandes corpus anotados*

Los grandes corpus de habla anotados con información lingüística van a ser una necesidad tanto para la descripción teórica de los fenómenos del habla como para el

desarrollo de aplicaciones de Tecnología del Habla en un futuro próximo. Su desarrollo, sin embargo, requiere de un esfuerzo, especialmente en la fase de anotación, que solo será factible si se emplean herramientas computacionales para esta tarea. Hoy en día existen ya algunas herramientas automáticas que pueden emplearse en este proceso, como transcritores fonéticos (Segre, Pachès *et al.*, 2000; Saga; TexAfon, Garrido *et al.*, 2012), segmentadores (EasyAlign, Goldman, 2011; SPPAS, Bigi, 2012; SegProso, Garrido, 2013) o herramientas de anotación prosódica (MoMel, Hirst y Espesser, 1993; AuToBI, Rosemberg, 2010; MelAn, Garrido, 2010; Roseano y Fernández, 2013). El rendimiento de este tipo de herramientas aún no es lo suficientemente alto como para eliminar completamente la participación humana en esta tarea, pero sí permite obtener una primera versión de la anotación que limite esta intervención a una revisión, mucho más rápida que si todo el proceso se hiciera manualmente.

SPPAS es una herramienta desarrollada en el Laboratoire Parole et Langage de Aix-en-Provence (Francia) por Brigitte Bigi que permite obtener automáticamente diferentes niveles de anotación fonética para una grabación determinada: transcripción fonética, segmentación en sílabas, segmentación en grupos entonativos y anotación prosódica (MoMel). La figura 1 presenta un ejemplo de Textgrid obtenido con esta herramienta. Por su parte, MelAn es una herramienta para la estilización, anotación y modelización de curvas melódicas, desarrollada con Praat y R, que permite añadir una serie de tiers con la estilización y la anotación al TextGrid de entrada con la transcripción fonética y la segmentación en unidades prosódicas, tal como se ejemplifica en la figura 2.

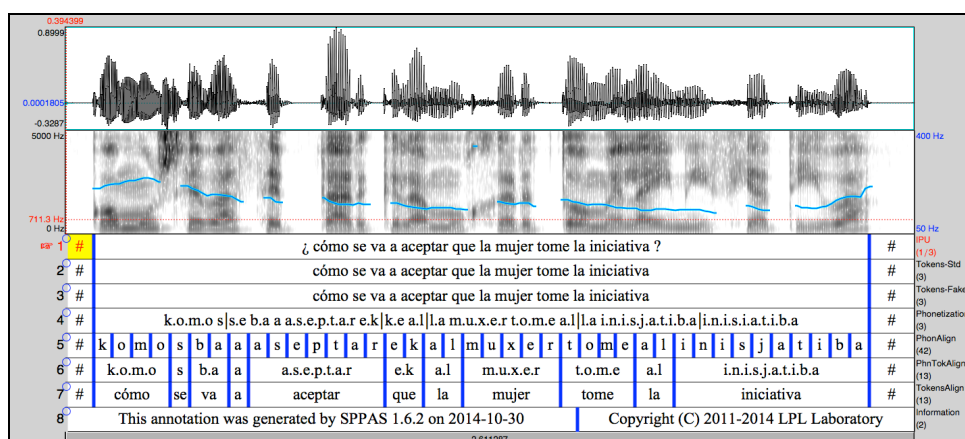


Figura 1. Fichero TextGrid del enunciado «¿Cómo se va a aceptar que la mujer tome la iniciativa?» pronunciado por una hablante femenina con la anotación prosódica automática generada por SPPAS

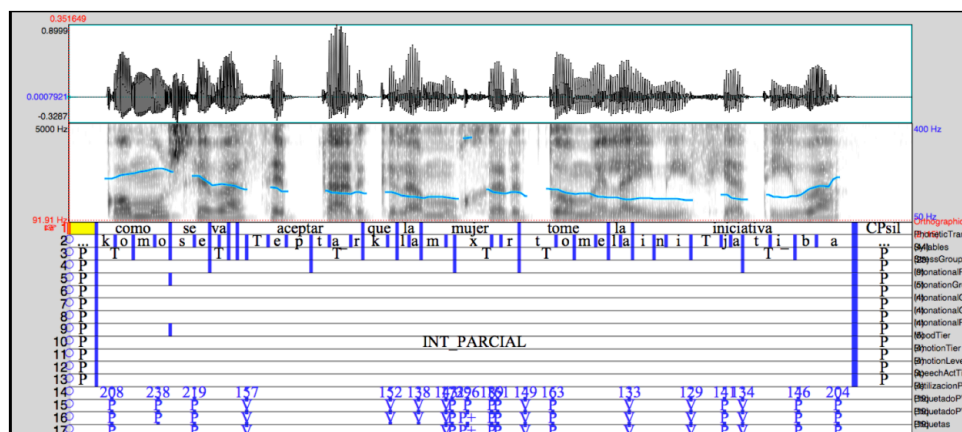


Figura 2. Fichero TextGrid del enunciado «¿Cómo se va a aceptar que la mujer tome la iniciativa?» pronunciado por una hablante femenina con la anotación prosódica automática generada por MelAn

Glissando (Garrido *et al.*, 2013) es un ejemplo de corpus de habla anotado en cuyo desarrollo se ha empleado este tipo de herramientas. Se trata de un corpus en español y catalán diseñado especialmente para estudios comparativos en el campo de la prosodia compuesto por más de 20 de horas de material grabado por idioma, una cantidad de material cuya anotación manual prácticamente inabordable por medios exclusivamente manuales. El corpus incluye la transcripción ortográfica y fonética alineada con la señal de habla, diferentes niveles de segmentación en unidades prosódicas (sílabas, grupos entonativos, grupos fónicos), y diferentes tipos de anotación prosódica. La mayor parte de esta anotación se obtuvo primero utilizando herramientas automáticas, y ha sido después revisada manualmente en su mayor parte.

### 3.2. Análisis automático de grandes corpus

De cara al estudio experimental del habla, disponer de este tipo de grandes corpus permitirá dar el alto de una Fonética experimental basada en el análisis de un conjunto reducido de datos, a menudo de un solo locutor, a experimentos con resultados obtenidos a partir de grandes corpus multilocutor y multiestilo, lo que amplía enormemente las posibilidades y objetivos de análisis que pueden plantearse: estudios sobre la invarianza y la variación fonética intra e interlocutor, comparación entre estilos de habla, entre lenguas... Pero para extraer la información necesaria de estos grandes corpus en un tiempo razonable será necesario también disponer de herramientas que permitan automatizar los procesos de obtención de datos y de análisis.



Patrón	Núm. de sílabas	Sílaba tónica	Modalidad	Posición	Num. de apariciones
PI0_VM0_VF0	1	1	AFIRM	FINAL_ENUNCIADO	186
VF0	1	1	AFIRM	FINAL_ENUNCIADO	174
PI0_VF0	1	1	AFIRM	FINAL_ENUNCIADO	137
VI0_VF0	1	1	AFIRM	FINAL_ENUNCIADO	98
VI0_PM0_VF0	1	1	AFIRM	FINAL_ENUNCIADO	67
0	1	1	AFIRM	FINAL_ENUNCIADO	26
VM0_VF0	1	1	AFIRM	FINAL_ENUNCIADO	17

Tabla 1. Ejemplo de listado de patrones obtenido con MelAn

Melan, la herramienta de anotación prosódica descrita en el apartado anterior, es también un ejemplo de sistema de obtención automática de datos fonéticos, en este caso sobre patrones entonativos. Además de anotar automáticamente el corpus, MelAn genera una serie de ficheros de texto con los inventarios completos de patrones de F0 detectados en el corpus de entrada. La información se presenta en forma de tablas que incluyen, además de la identificación del patrón, el número de apariciones del mismo en conjunto del corpus, como se ejemplifica en la tabla 1. Este método de obtención de datos se ha empleado en diferentes estudios teóricos sobre la descripción de los patrones entonativos del español y el catalán en diferentes situaciones de habla, como lectura neutra (Garrido 2012a,b,c), diálogos (Garrido y Rustullet, 2011) o emociones (Garrido, 2011).

#### 4. CONCLUSIONES: ALGUNAS PREDICCIONES DE FUTURO

De acuerdo con todo esto, parece probable que en el futuro se produzca una convergencia en los métodos de trabajo de Fonética y Tecnologías del Habla, basados en el uso de grandes corpus y de herramientas automáticas para la anotación y el procesado automático, como ya se hace de hecho en muchos procesos relacionados con el desarrollo de tecnologías del habla. De esta convergencia pueden beneficiarse ambas disciplinas, ya que permitirá obtener mejores descripciones teóricas de la invarianza y la variabilidad fonéticas, fruto del análisis de grandes cantidades de material fónico, que puedan emplearse para el desarrollo de tecnologías del habla basado en conocimiento lingüístico, y también mejores herramientas para el procesado del habla basadas en conocimiento lingüístico y fonético, que permitan mejorar la calidad de las tecnologías del habla actuales. Para ello será necesario, por un lado, disponer de corpus transcritos y anotados de diferente tipo (corpus de diálogos, de emociones, de habla de niños y ancianos...), y por otro, de herramientas de anotación y análisis automáticos cada vez más complejas, que sean el resultado de la implementación del conocimiento fonético y lingüístico en herramientas «inteligentes».

En el desarrollo de estos corpus y herramientas hay espacio para la participación de fonetistas, pero de fonetistas con un perfil más «computacional», con mayor dominio de las herramientas tecnológicas de procesado, no solo en lo que se refiere a su manejo sino incluso a su desarrollo (con Praat, por ejemplo), que le permita trabajar en la creación de nuevas herramientas o adaptarlas para sus propias necesidades de

investigación, tareas hasta ahora reservadas a ingenieros. Este nuevo perfil, que requerirá una nueva formación multidisciplinar aún por diseñar, capacitará a los nuevos fonetistas tanto para el trabajo en investigación básica como para el desarrollo de tecnologías del habla, bien en el mundo académico, bien en el de la empresa.

#### REFERENCIAS BIBLIOGRÁFICAS

- ACERO, A. (1995): «The role of phoneticians in speech technology», en Bloothoof, G., V. Hazan, D. Huber y J. Llisterrri, eds., *European studies in phonetics and speech communication*, Utrecht, OTS, 170-5.
- AINSWORTH, W. A. (2005): «Can phonetic knowledge be used to improve the performance of speech recognisers and synthesisers?», en Barry, W. J. y W. A. van Dommelen, eds., *The integration of phonetic knowledge in speech technology*, Dordrecht, Springer, 13-20, <[http://dx.doi.org/10.1007/1-4020-2637-4\\_2](http://dx.doi.org/10.1007/1-4020-2637-4_2)>.
- BATTANER, E., J. GIL, V. MARRERO, J. LLISTERRI, C. CARBÓ, M. J. MACHUCA, C. DE LA MOTA y A. RÍOS (2003): «VILE: Estudio acústico de la variación inter- e intralocutor en español», en *SEAF 2003. Actas del II Congreso de la Sociedad Española de Acústica Forense*, Barcelona, Sociedad Española de Acústica Forense, 59-70 [en línea]: <[http://liceu.uab.cat/~joaquim/phonetics/VILE/VILE\\_SEAF03.pdf](http://liceu.uab.cat/~joaquim/phonetics/VILE/VILE_SEAF03.pdf)>. [Consulta: 15/01/2015].
- BATTANER, E., C. CARBÓ, J. GIL, J. LLISTERRI, M. J. MACHUCA, N. MADRIGAL, V. MARRERO, C. DE LA MOTA, M. RIERA y A. RÍOS (2005): «VILE: Estudio acústico de la variación inter- e intralocutor en español», *Procesamiento del Lenguaje Natural*, vol. 35, 435-436 [en línea]: <[http://liceu.uab.cat/~joaquim/phonetics/VILE/VILE\\_SEPLN05.pdf](http://liceu.uab.cat/~joaquim/phonetics/VILE/VILE_SEPLN05.pdf)>. [Consulta: 15/01/2015].
- BIGI, B. (2012): «SPPAS: a tool for the phonetic segmentations of Speech», The eight international conference on Language Resources and Evaluation, Estambul, 1748-1755.
- BLACK, A. W. y N. CAMPBELL (1995): «Optimising Selection Of Units From Speech Databases For Concatenative Synthesis», *Proceedings of Eurospeech*, 581-582.
- BLUMSTEIN, S. E. y K. N. STEVENS (1981): «Phonetic features and acoustic invariance in speech», *Cognition*, vol. 10, 25-32, <[http://dx.doi.org/10.1016/0010-0277\(81\)90021-4](http://dx.doi.org/10.1016/0010-0277(81)90021-4)>.
- BOERSMA, P. (2001): «Praat, a system for doing phonetics by computer», *Glott International*, 5(9/10), 341-345.
- CARLSON, R., B. GRANSTRÖM y S. HUNNICUTT (1982): «A multi-language text-to-speech module», *Proceedings of ICASSP*, vol. 82, 1604-1607, <<http://dx.doi.org/10.1109/icassp.1982.1171433>>.
- COKER, C. H. (1968): «Speech synthesis with a parametric articulatory model», *Proc. Speech. Symp.*, Kyoto, paper A-4.

- COOPER, Franklin S. (1953): «Some instrumental aids to research on speech», en *Report of the fourth annual round table meeting on linguistics and language teaching*, Washington, D. C., Institute of Languages and Linguistics, Georgetown University, 46-53.
- DIXON, N. R. y H. D. MAXEY (1968): «Terminal Analog Synthesis of Continuous Speech Using the Diphone Method of Segment Assembly», *IEEE Trans. AudioElectroacoust.*, AU-16, 40-50, <<http://dx.doi.org/10.1109/tau.1968.1161948>>.
- DUTOIT, T., V. PAGEL, N. AGEL, N. PIERRET, F. BATAILLE y O. VAN DER VREKEN (1996): «The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes. Proc. ICSLP'96», *Philadelphia*, 3, 1393-1396, <<http://dx.doi.org/10.1109/icslp.1996.607874>>.
- FANT, G. (1983): «Phonetics and Speech Technology», *Proceedings of the Tenth International Congress of Phonetic Sciences*, 13-24.
- FODGE, K. (2014): *Introducing Spanish dialects in a linguistic processing module for improved ASR and novel speech synthesis capabilities*, trabajo de fin de máster, Departamento de Traducción y Ciencias del Lenguaje, Universidad Pompeu Fabra.
- GARRIDO, J. M. (2009): «Estudios humanísticos y tecnologías del habla. Presente y futuro», en de la Mota C. y G. Puigvert, eds., *La investigación en humanidades*, Madrid, Biblioteca Nueva, 171-82
- GARRIDO, J. M. (2010): «A Tool for Automatic F0 Stylisation, Annotation and Modelling of Large Corpora», *Speech Prosody*, Chicago.
- GARRIDO, J. M. (2011): «Análisis de las curvas melódicas del español en habla emotiva simulada», *Estudios de Fonética Experimental*, vol. XX, 205-255 [en línea]: <<http://stel.ub.edu/labfon/sites/default/files/XX-12-JMGarrido.pdf>>. [Consulta: 15/01/2015].
- GARRIDO, J. M. (2012a): «Análisis fonético de los patrones melódicos locales en español: patrones acentuales», *Revista Española de Lingüística*, vol. 42, 1, 79-107.
- GARRIDO, J. M. (2012b): «Análisis fonético de los patrones melódicos locales en español: patrones entonativos», *Revista Española de Lingüística*, vol. 42, 2, 95-125.
- GARRIDO, J. M. (2012c): «Glissando, un corpus anotat per a l'anàlisi de la prosòdia del català i del castellà. Descripció i primers resultats d'exploració», *Phonica*, vol. 8 [en línea]: <<http://www.publicacions.ub.edu/revistes/phonica8/documentos/880.pdf>>. [Consulta: 15/01/2015].
- GARRIDO, J. M. (2013): «ModProso: A Praat-Based tool for F0 Prediction and Modification», *Proceedings of TRASP 2013*, 38-41. [en línea]: <<http://aune.lpl.univ-aix.fr/%7Etrasp/Proceedings/19864-trasp2013.pdf>>. [Consulta: 15/01/2015].
- GARRIDO, J. M. y S. Rustullet (2011): «Patrones melódicos en el habla de diálogo en español: un primer análisis del corpus Glissando», *Oralia: Análisis del discurso oral*, vol. 14, 129-160.

- GARRIDO, J. M., Y. LAPLAZA, M. MARQUINA, C. SCHOENFELDER y S. RUSTULLET, (2012): «TexAFon: a multilingual text processing tool for text-to-speech applications», *Proceedings of IberSpeech 2012*, Madrid, Spain, noviembre, 21-23, 281-289.
- GARRIDO, J. M., D. ESCUDERO, L. AGUILAR, V. CARDEÑOSO, E. RODERO, C. DE-LA-MOTA, C. GONZÁLEZ, S. RUSTULLET, O. LARREA, Y. LAPLAZA, F. VIZCAÍNO, M. CABRERA y A. BONAFONTE (2013): «Glissando: a corpus for multidisciplinary prosodic studies in Spanish and Catalan», *Language Resources and Evaluation*, vol. 47, 4, 945-971, <<http://dx.doi.org/10.1007/s10579-012-9213-0>>.
- GOLDMAN, J.-Ph. (2011): «EasyAlign: an automatic phonetic alignment tool under Praat», *Proceedings of InterSpeech*.
- GREENBERG, S. (2001): «From here to utility - melding phonetic insight with speech technology», en *Eurospeech 2001 Scandinavia. Proceedings of the 7th European Conference on Speech Communication and Technology, 2nd Interspeech Event*, Aalborg, Denmark, 2485-8.
- HART, J. T., R. COLLIER y A. COHEN (1990): *A Perceptual Study of Intonation*, Cambridge, Cambridge University Press, <<http://dx.doi.org/10.1017/cbo9780511627743>>.
- HIRST, D. y R. ESPESSER (1993): «Automatic modelling of fundamental frequency using a quadratic spline function», *Travaux de l'Institut de Phonétique d'Aix*, vol. 15, 71-85.
- HOLMES, J. N., I. G. MATTINGLY y J. N. SHEARME (1964): «Speech synthesis by rule», *Language and Speech*, vol. 7, 127-143.
- HUNT, A. J. y A. W. BLACK (1996): «Unit selection in a concatenative speech synthesis system using a large speech database», *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-96)*, vol. 1, 373-376, <<http://dx.doi.org/10.1109/icassp.1996.541110>>.
- LAPLAZA, Y. y J. M. GARRIDO (2014): «Analysis and Synthesis of Emotional Speech in Spanish for the Chat Domain. A Parametric Approach», en Navarro-Mesa, J. L. *et al.*, eds.: *Advances in Speech and Language Technologies for Iberian Languages* (Proceedings of the Second International Conference, IberSPEECH 2014, Las Palmas de Gran Canaria, 19-21 de noviembre, 2014), *Lecture Notes in Computer Science*, vol. 8854 2014, Springer, 1-10, <[http://dx.doi.org/10.1007/978-3-319-13623-3\\_1](http://dx.doi.org/10.1007/978-3-319-13623-3_1)>.
- LIBERMAN, A., P. DELATTRE, F. COOPER y L. GERSTMAN (1954): «The role of consonant-vowel transitions in the perception of stop and nasal consonants», *Psychological Monographs*, vol. 68, 8, 1-13, <<http://dx.doi.org/10.1037/h0093673>>.
- LLISTERRI, J. (2007): «El papel de la fonética en las tecnologías del habla», en González González, M., E. Fernández Rei y B. González Rei, eds., *Actas do 3o congreso internacional de fonética experimental*, Santiago de Compostela, 24-26 de outubro de 2005, Santiago de Compostela, Xunta de Galicia, 23-37.
- LLISTERRI, J., L. AGUILAR, J. M. GARRIDO, M. J. MACHUCA, R. MARÍN, C. DE LA MOTA *et al.* (1999): «Fonética y tecnologías del habla», en Blecua, J. M., G. Clavería, C. Sánchez y J. Torruella, eds., *Filología e informática. Nuevas tecnologías en los*

- estudios lingüísticos*, Barcelona, Seminari de Filologia i Informàtica, Departament de Filologia Espanyola, Universitat Autònoma de Barcelona-Editorial Milenio, 449-79.
- ÖHMAN, S. E. G. (2001): «Why current speech technology is false phonetics», *Lund University, Department of Linguistics, Working Papers*, vol. 49, 180-183.
- OLIVE, J. P. (1977): «Rule synthesis of Speech from Diadic Units», *Proceedings of ICASSP-77*, 568-570, <<http://dx.doi.org/10.1109/icassp.1977.1170350>>.
- PACHÈS, P., C. DE LA MOTA, M. RIERA, M. P. PEREA, A. FEBRER, M. ESTRUCH, J. M. GARRIDO, M. J. MACHUCA, A. RÍOS, J. LLISTERRI, I. ESQUERRA, J. HERNANDO, J. PADRELL y C. NADEU (2000): «Segre: An automatic tool for grapheme-to-allophone transcription in Catalan», *Proc. of Workshop on Developing Language Resources for Minority Languages: Reusability and Strategic Priorities*, LREC, 52-61.
- PIERREHUMBERT, J. (1980): *The phonology and phonetics of English intonation*, tesis doctoral, MIT, Indiana University Linguistics Club.
- POLS, L. C. W. (2001): «Acquiring and implementing phonetic knowledge», en *Eurospeech 2001 Scandinavia. Proceedings of the 7th European Conference on Speech Communication and Technology, 2nd Interspeech Event*, Aalborg, 3-7 de septiembre 2001, K3-6.
- ROSEANO, P. y A. M. FERNÁNDEZ PLANAS (2013): «Transcripció fonètica i fonològica de l'entonació: una proposta d'etiquetatge automàtic», *Estudios de Fonética Experimental*, vol. XXII, 275-332 [en línea]: <<http://stel.ub.edu/labfon/sites/default/files/XXII-12-PRoseano.pdf>>. [Consulta: 15/01/2015].
- ROSEN, G. (1958): «Dynamic analog speech synthesizer», *Journal of the Acoustical Society of America*, vol. 30, núm. 3, 201-209, <<http://dx.doi.org/10.1121/1.1909541>>.
- ROSENBERG, A. (2010): «AuToBI - A tool for Automatic ToBI annotation», *Proceedings of Interspeech*, 146-149.
- SHRIBERG, E. (2005): «Spontaneous speech: How people really talk and why engineers should care», *Interspeech 2005-Eurospeech 2005*, Proceedings of the 9th European Conference on Speech Communication and Technology, Lisboa, 4-8 de september 2005, 1781-1784.
- STRIK, H. (2005): «Is phonetic knowledge of any use for speech technology?», en W. J. Barry y W. A. van Dommelen, eds., *The integration of phonetic knowledge in speech technology*, Dordrecht, Springer, 167-80, <[http://dx.doi.org/10.1007/1-4020-2637-4\\_10](http://dx.doi.org/10.1007/1-4020-2637-4_10)>.
- ZUE, V. W. (1983): «The use of phonetic rules in automatic speech recognition», *Speech Communication*, vol. 2, 181-186, <[http://dx.doi.org/10.1016/0167-6393\(83\)90023-7](http://dx.doi.org/10.1016/0167-6393(83)90023-7)>.
- ZUE, V. W. (1985): «The use of Speech Knowledge in Automatic Speech Recognition», *Proceedings of the IEEE*, vol. 73, núm. 11, <[1602-1615](http://dx.doi.org/10.1109/proc.1985.13342)>, <<http://dx.doi.org/10.1109/proc.1985.13342>>.