Journal of Music, Technology & Education
Volume 9 Number 2
 Article. English language. doi: 10.1386/jmte.9.2.125_1

MANUEL PÉREZ-GIL, JESÚS TEJADA AND REMIGI MORANT
Universitat de València

A. PÉREZ-GONZÁLEZ DE MARTOS
Universitat Politècnica de València

Cantus: Construction and evaluation of a software solution for real-time vocal music training and musical intonation assessment

ABSTRACT

The development of the ability to sing or play in tune is one of the most critical tasks in music training. In music education, melodic patterns are usually learned by imitative processes (modelling). Once modelled, pitch sounds are then associated to a representation according to a syllabic system such as the Guidonian system – or an arbitrary single syllable – or western graphic notation system symbols. From a didactic standpoint, few advances have been made in this area besides the use of audio-supported guides and existing software, which use a microphone to analyse the input and estimate the pitch or fundamental frequency of the given tone. However, these programmes lack the necessary analytical algorithm to provide the student with precise feedback on their execution; and also they do not provide adequate noise-robust solutions to minimize the student assessment

KEYWORDS

musical intonation
music technology
software design
voice recognition
sight-singing
music education

error rate. The ongoing research discussed in this article focuses on Cantus, a new software solution expressly designed as an assessment and diagnosis tool for online training and assessment of vocal musical intonation at the initial stages of music education. Cantus software embodies the latest research on real-time analysis of audio stream, which permits the teacher to customize music training by means of recording patterns and embedding them into the programme. The study presented in this article includes the design, implementation and assessment of Cantus by music teachers. The pilot study for the software assessment includes a sample of 21 music teachers working at thirteen music schools in Valencia, Spain. These teachers worked with the software at their own pace for a week in order to evaluate it. Subsequently, a two-part questionnaire was filled in with (1) questions related to demographics, professional experience and the use of ITC; and (2) questions related to the software's technical and didactic aspects. The questionnaire also included three open questions related to Cantus, namely advantages, issues and suggestions. The results show an excellent reception by teachers, who consider this software as a highly adequate music training tool at the initial stages of music education.

1. INTRODUCTION

1.1 Teaching and learning musical intonation imitation, production and perception – Timbre

Sloboda (1985) proposes that music education should be structured in three stages: listening, imitation and extensive practice. Technology permits the combination of these three elements, providing models to listen to, imitate and practise. Technology can facilitate the learner's autonomy to find and recognize their own pace of practice with the aid of tools such as instant feedback on vocal performance, which allows for progressive improvement in the assessment of programmed exercises in music reading.

The imitation of models has been used widely in all areas of music teaching and is based on Bandura's cognitive model (2006). The imitation of melodic patterns has also been a subject of intensive research (Green 1990; Hermanson 1972; Small and McCachern 1983; Yarbrough et al. 1995). In addition, the use of recorded models for imitation has been a frequent subject of study (Galyen 2005; Hodges and Nolker 2011). However, most of these studies address musical instrument rather than vocal training. In support of the effectiveness of model imitation in intonation training, a recent study suggests that a choral group can gain more benefit from modelling the singing voice through sung examples than from an explanation of the performance.

A review of the studies on the relationship between perception and production in intonation suggests that there is no convergence between the two (Demorest 2011). However, another study contradicts this, showing the existence of the relationship (Watts et al. 2005). Along these lines, it has been found that performance is best when the vocal model has a similar range to that of the learners' vocal range, as in a male or female-child model (Demorest 2007; Yarbrough et al. 1995). Therefore, timbre can be seen as one of the most influential factors affecting performance – that is, music samples with similar timbres seem to facilitate accuracy in the perception and production processes of imitating models.

1.2 Teaching and learning music reading

Few theories focus specifically on the processes of music reading, and as a consequence, this field is not based on strong theoretical foundations (Lehmann and McArthur 2002). The lack of progress in this area is articulated in the following statement regarding the teaching and learning methods of western musical notation reading: '[…] we are still striving to find efficient ways to help students learn to read music' (Hodges and Nolker 2011: 63). In a study by Daniels (1986), no significant effects were found in the methods and materials used as accuracy predictors for music reading.

Melody, harmony and rhythm are combined in music. Despite this fact, each one is processed separately (Bengtsson and Ullen 2006; Parsons 2001; Schon and Besson 2002; Snyder 2000). Diverse studies suggest that intonation and rhythmic reading should be dissociated (Anvari et al. 2002; Bengtsson and Ullen 2006; Fasanaro et al. 1990). Gordon (2003) also proposes the separate study of rhythmic and melodic elements, although he does not present a comprehensive theory on reading music. Nevertheless, the first step in his theory is the process of imitation through modelling (Hodges and Nolker 2011).

There are investigations into the relationships between successful music sight-singing and other variables, such as saccadic eye movement (Goolsby 1994; Kopiez and Galley 2002), or mental images and the degree of brain specialization in rhythmic reading and intonation (Galera and Tejada 2012; Galera et al. 2013). Unfortunately, the conclusions of these studies have not provided much insight into the processes of music reading.

Another interesting aspect in this field is harmonic contextualization, which seems to facilitate college students' individual accuracy in sight-singing, although its effectiveness for young students or beginners is not clear (Hodges and Nolker 2011: 75). Lucas (1994) investigated the effect of harmonic context on accuracy in individual sight-singing exercises. The participants were high-school students assigned to three conditions: (1) just melody, (2) piano; and (3) vocal harmony. Results suggest that the use of harmonic context to support sight-singing is less effective in training and practice situations than in performance situations. This finding contrasts with an earlier study involving college students whose performances suggested that a harmonic context improves sight-singing accuracy, especially among less experienced students (Boyle and Lucas 1990). In any case, Gordon's theory (2003) on melodic contextualization deserves credit, since music is rarely monodic and requires the musician to listen to the harmonic context and be aware of the function of the notes in that given context. For example, singing a C in C major is not the same as singing a C in D major; in the first case the C is tonic and in the second case it is dominant seventh, merely a pause or tension to be resolved.

Sounds are not isolated entities, but rather cell-like melodic-harmonic tissue that should be studied in context rather than in isolation (Bamberger 1991). In this respect, music reading requires the perception of sound patterns rather than of individual sounds or sound intervals (Gromko 2004). The number of sounds that can be memorized by an individual when reading music has been studied and was determined to be about 7±2 sounds (Miller 1956). This finding has been confirmed in a study using singers as the research sample (Fine et al. 2006). With this in mind, a maximum of eight sounds were incorporated into the patterns to be used in perception and production activities in Cantus.

Meanwhile, Cassidy (1993) suggests that syllables or mnemonic devices can be used as effective aids in the development of reading skills when teaching western musical notation. In addition, Hodges and Nolker (2011) observe that although solmization is effective in general, the most effective method is yet to be determined.

Finally, it has been suggested that melodic patterns do not necessarily have to come directly from the songs children are learning, but should be of a similar level of difficulty (Gordon 2003). This author relies on harmonic functions in order to design patterns for training purposes, and thus the patterns are often presented as chord arpeggio melodies. In Cantus, isochronous sounds (patterns of between two and eight sounds without duration or accent) are used for performance imitation.

1.3 Pattern sequence

A number of principles have been suggested from research and theory concerning the characteristics of melodic patterns in intonation training. For example, it has been proposed that melodic patterns need not be taken from the songs the children are learning, but should reflect the same level of difficulty (Gordon 2003). Martin (1991) provides information about how to construct and select tonal patterns. Other authors have also examined the effect of several characteristics of tonal patterns. For example: most children, especially those between 3 and 5 years old, intonate better by echo-singing melodic contour rather than single specific intervals; the melodic contour, the number of different notes and the location of skips, were all found to have significant effects on the children's ability to echo-sing patterns accurately (Sinor 1985); and the combination of sounds in the pattern modifies its level of difficulty and therefore affects the accuracy in the ability to sing its contour or pitches (Reifinger 2009).

As previously mentioned, Cantus is intended to function as an aid in vocal intonation training by providing listen-and-repeat and notation singing activities. Cantus was designed to conform to the idea that any educative software devoted to music education should provide at least two options: a free itinerary (a syllabus determined by pupils or teachers) and a compulsory itinerary (a syllabus determined by the designers). On the one hand, Cantus could be considered itinerary free because it embodies a section that enables the teacher or the pupil to construct their own exercises. On the other hand, the software includes specific patterns in an explicit order. The majority of these patterns are based on Gordon's proposal (2003). In addition, the sequence of the patterns also takes into account the characteristics of Reifinger's study (2009). In these studies, the fundamental characteristics of the patterns include the use of not only tonal supports (*dms -do, mi, so*) but also the passing notes between them and repeated notes, especially in the first units of training. All in all, the determination of the sound patterns and their order – the teaching sequence – is very important for the design of any software intended for vocal intonation.

1.4 Educational technology applied to music reading and intonation

Up to the present time, technology applied to the training of the singing voice has been developed through research and has focused on the graphical display of waveforms emitted in real time, which is exemplified in projects like VOXed (Howard et al. 2003, 2007; Welch et al. 2004). In a review study on major support tools based on real-time feedback for singer training, analyses were carried out on systems such as Singad (Singing Assessment and Development System; Welch et al. 1989), Sing & See (Wilson et al. 2008) and WinSINGAD (Hoppe et al. 2006). The application of these systems has resulted in both quantitative and qualitative studies that suggest the effectiveness of real-time visual feedback on the improvement of singing skills. The results of these studies provided the impetus for including real-time visual feedback in Cantus.

Welch (1985a) had previously emphasized the advantages of real-time feedback in his theoretical model (Figure 1). The representation of Knowledge of Results (KR) during voice response offers the possibility of immediately modifying the response while observing its effect.

The SINGAD system was experimentally tested with 32 7-year-old children (Welch et al. 1989). Later, in the Albert and Sing & See systems, controllable vocal parameters were expanded to include singing training, pronunciation training, speech therapy and speech or timbre identification (spectrogram).

WinSINGAD (Howard et al. 2006, 2007) is a singing training software that allows for the study of parameters such as waveform and sound fundamental frequency (F0) input in relation to a timeline; short-term spectrum; narrowband spectrogram; spectral relation related to time; vocal tract area (VT, the distance between the glottis and the lips); and the average of min VT area versus time. Any of these parameters can be displayed on the screen in combination with any other. The effectiveness of this software was evaluated by a multidisciplinary group of singing teachers, singers, psychologists and scientists using an action research methodology.

It has been suggested that feedback in the form of a visual representation of vocal response helps singers to improve their pitch accuracy when training. Wilson et al. (2008) observed that although the immediate effects of concurrent visual feedback led to decreased performance because of the cognitive overload process, it eventually improved the singers' accuracy.

Until the present time, the use of computer programs for music reading has been limited, in many cases, to auditory aspects, interval recognition, musical

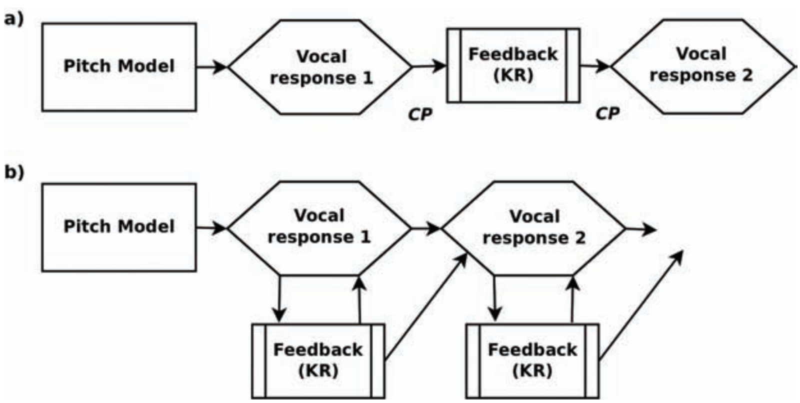

Figure 1: Welch's Model; (a) the modelling process by sequential imitation; and (b) the modelling process with feedback on real time can be observed (CP=Critical learning Period; KR=Knowledge of Results) (Howard et al. 2003).

dictations or western music notation note placement by naming notes on the staff. In addition, student performance assessment has been very limited, e.g., typing notes on a keyboard. Other computer software such as Singing Coach or mobile-phone software such as Sing Sharp or Learn to Sing require voice recording and include its visual representation, but do not include a systematic assessment process that can be useful to teachers for analysis or evaluation purposes.

2. DIDACTIC DESIGN

The Cantus didactic characteristics are as follows:

- Training by vocal imitation (Keetman 1974; Orff & Keetman [1950] 1954).
- Optional use of solfeggio syllables or neutral syllable (Cassidy 1993). The neutral syllable enables the learner to practise relative solfeggio. In Cantus, a neutral syllable ('nu') is used in some exercises, while solmization syllables are used in other exercises.
- Presentation of tonal patterns without rhythmic components (Bengtsson and Ullen 2006; Fasanaro et al.1990; Parsons 2001; Schon and Besson 2002).
- Contextualized exercises, which may be performed in several tonalities through a harmonic background with the tonic and fifth of each corresponding chord as a clear reference.
- Written exercises presented in seven clefs and configurable tempo from 40 to 100 bpm.
- Real-time visual feedback to aid in accurate intonation (see VOXed project: Hoppe et al. 2006; Howard et al. 2003, 2007; Howard et al. 2003; Welch et al. 2004; Wilson 2007).
- Several modes of melodic pattern representations in order to improve information processes (Colvin and Lyons 2010; Colvin and Mayer 2011).
- Assessment of exercises in each work unit. The evaluation provides additional information on the method of student imitation, that is, if they practice the pattern intonation by sight-singing or imitation.
- Teacher customization (creation and storing of exercises) to adapt the music content to each student's pace.
- A discrete number of sounds in each pattern (2–8 sounds) to avoid the overflow of working memory and facilitate memorization of the pattern (Fine et al. 2006; Miller 1956).
- The patterns are distributed in five work units to practise note reading and intonation on different clefs and tonalities. The sequence of work units and the organization of activities are based on music methods founded on tonal principles (Chevais 1941; Martenot 1957; Gordon 2003). In the first work unit, the learning sequence begins in the Major mode with an approach to the pentatonic scale (*so, la, so, mi, do*). The second work unit includes both passing and neighbour notes – upper and lower – thus introducing *re* and completing the pentatonic scale. In addition, the Minor modes are presented in a modal sequence as presented by Orff (Orff & Keetman [1950] 1954; Keetman 1974) and by Gordon (2003). The third work unit introduces the Dorian mode (with and without raised seventh), and in the fourth work unit, the Mixolydian mode is introduced. In the fifth work unit, tritones and chromatics are presented along with other altered sounds outside the tonality or modality.

The decision on which didactic characteristics to include in the design of Cantus was based on the analysis of three information sources: (1) data collected from several focus groups and a questionnaire among in-service music teachers in Valencia; (2) theoretical and empirical studies on music education and psychology, some of them previously explained and others mentioned before; and (3) the analysis of existing intonation assessment software.

3. IMPLEMENTATION

3.1 Architecture

In sight of the growing variety of devices and operating systems used for software, Cantus was based and developed on HTML5 technologies, providing cross-platform support. HTML5 is the fifth and latest version of the HTML standard. It offers new features that provide not only great media support but also enhanced support for the creation of web applications that allow for easier and more effective interaction between the user, his or her local data and the software on the server. Currently, the latest versions of all the major browsers (Chrome, Firefox, Opera, IE) support HTML5 APIs (Application Programming Interface), which can be employed with JavaScript. The following HTML5 APIs were particularly useful in the design and development of the software:

- The *canvas* element for an immediate 2D drawing mode.
- API Web Audio, a high-level JavaScript API for processing and synthesizing audio in web applications.
- An API File for handling file uploads and file manipulation.

Cantus is therefore a web application. A web application or web app is any type of software that runs in a web browser. Cantus was created in a browser-supported programming language (a combination of JavaScript, HTML and CSS) and relies on a web browser to render the application. It was developed as a total client-oriented application, meaning that the application is run entirely from the client's browser, which significantly simplifies the software's portability and installation.

3.2 Description of user interface and interaction with Cantus

In contrast to complex audio computer programs, Cantus was designed to be a user-friendly, intuitive program for non-expert users, allowing for simple, clear student interaction. Figure 2 shows the work units in Cantus. Figure 3 shows the main interface, with the exercise list on the left side and the music staff with the current exercise notation on the right side. The evaluation of results pops up in a new window after every exercise attempt. Additionally, detailed user evaluation statistics are gathered throughout the session and can be accessed at any time on a specific *evaluation results* page. Evaluation is made by contrasting the frequencies of the user's response versus the model frequencies. As seen in Figure 2, the software also provides a custom exercise creator where users and/or teachers can create their own set of exercises.

Human interaction begins when the user enters his or her name so the software can record all user activities and create the user's performance

Figure 2: Cantus graphic user interface. English translations are in blue.

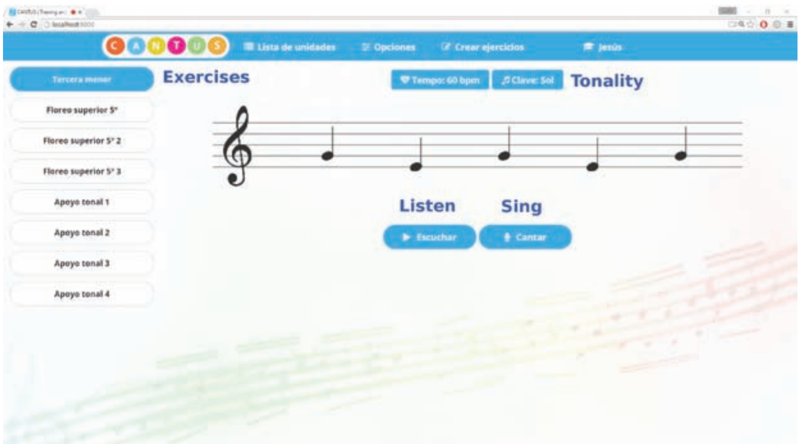

Figure 3: Cantus exercise user interface. English translations are in blue.

evaluation report. The user can then choose a work unit and the tonality in which he or she will perform the exercises (Figure 2). Next, a menu appears on the left of the screen, presenting the exercises in that unit (Figure 3). The user can freely choose an exercise from the menu on the right, and then click on the 'Listen' button. The user can see the notes light up on while listening to the model.

If the user chooses to click on the 'Sing' button, an audio-visual countdown is played in order to synchronize the user's response. Simultaneously, a sustained chord in the tonality of the exercise is heard in order to facilitate the intonation of the first note. The user then sings into the microphone. Given there is no rhythm track in the playback, the user is guided visually by observing notes which illuminate on the screen in order to synchronize his or her response. On completion of the exercise, the *evaluation results window* appears on the screen, showing the performance score along with a visual representation of the user's input (Figure 4). Evaluation is made by contrasting the frequencies of the user's response with the model's theoretical frequencies (see Section 3.3.2). At this point, the user can choose between four options

Figure 4: Cantus' evaluation results page. English translations are in blue.

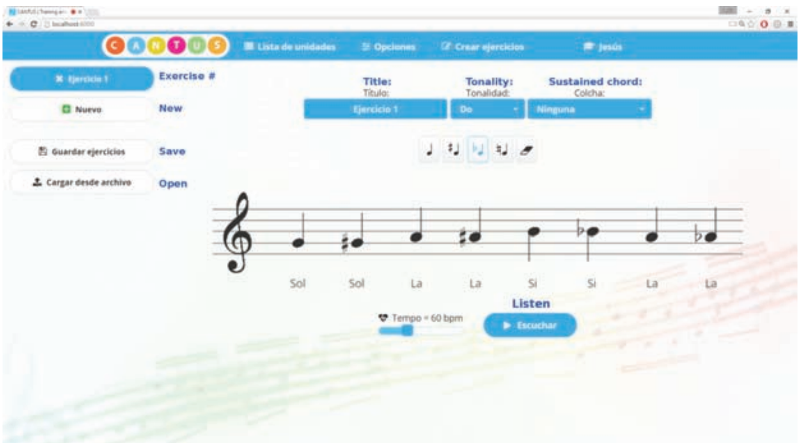

Figure 5: The exercise-creator window in Cantus. English translations are in blue.

by clicking on one of four buttons, which appear at the bottom right of the window: *Download* (to download the recorded audio of the performance), *Listen* (to listen to the user's recorded performance), *Repeat* (to repeat the exercise) and *Next Exercise* (to go on the next exercise in the work unit).

Figure 5 shows the exercise-creator window. The created exercises – a maximum of eight notes in length – can be saved individually or in a set and can be conveniently loaded from a small file.

3.3 Evaluation algorithms

Cantus automatically evaluates the pitch accuracy of the user's captured input for a given exercise. The evaluation is carried out in two steps. First, a Pitch Detection Algorithm (PDA) is used to estimate the pitch over the microphone-captured audio signal. Pitch mean values are then calculated and compared with the reference frequency values of the corresponding exercise notes. Details regarding both steps are given in the following sections.

3.3.1 PDA

The PDA is an algorithm designed to estimate the pitch or fundamental frequency of a quasi-periodic or virtually periodic signal, usually a digital recording of speech or a musical note or tone. In Cantus, a PDA is used to

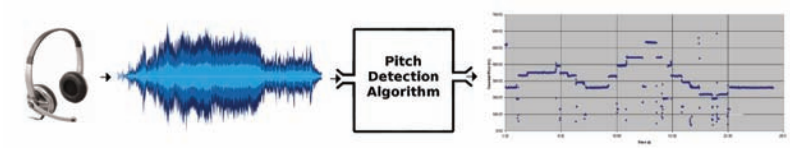

Figure 6: Pitch detection process overview.

calculate the pitch of what has been intoned (voice) or played (musical instrument) by the user in a given exercise. There are many different approaches and algorithms to perform this pitch estimation, which fall broadly into three different categories: time-domain, frequency-domain and spectral/temporal approaches. Cantus implements a time-domain PDA known as the *YIN algorithm*, which is based upon *autocorrelation*, a technique consisting of comparing signal segments with other segments, offset by a trial period, to find a match. YIN has proven to be robust and accurate for performing real-time pitch tracking (De Cheveigné and Kawahara 2002).

The pitch detection process (Figure 6) starts with audio capturing through the microphone. Once the audio signal buffer is ready, the YIN algorithm estimates the fundamental frequencies (F0) through time, which will later be used to evaluate the intonation accuracy.

3.3.2 Intonation Evaluation Algorithm

The Intonation Evaluation Algorithm receives input from both the exercise reference note frequencies and the estimated frequencies (F0) of the captured input signal, and then calculates an accuracy score as a figure between 0 and 10, measuring the closeness of the latter to the former. The estimated frequency for each exercise note is computed as the median of the note's 60 per cent central time-interval captured frequency.

The algorithm evaluates notes separately, each one counting as 10/*N*, where *N* is the number of notes of the particular exercise. If the difference between the reference and the estimated frequencies is lower than a third of a semitone, then the individual note score is maximum (10/*N*). Otherwise, the individual note score is reduced proportionally to the semitone difference ((10/*N*)·(1-*sdiff*)). This semitone difference is calculated in an octave-independent manner. This means the notes can be reproduced in any octave without affecting the resulting score, which is particularly useful in reinforcing the evaluation process from different audio sources (male voice, female voice, different musical instruments, etc.).

4. TEACHER EVALUATION

4.1 Sample

The pilot study sample, extracted for reasons of access and convenience, consisted of 21 music teachers (10 men and 11 women) between the ages of 25 and 55, working at thirteen different music schools and conservatories in the Valencian Community (mid-east of Spain) (Figure 7). The teachers come from diverse educational backgrounds and teach students between 11 and 15 years of age in the first four courses of elementary music studies. The sample was too small to make inferences on the data for external validity, but this was

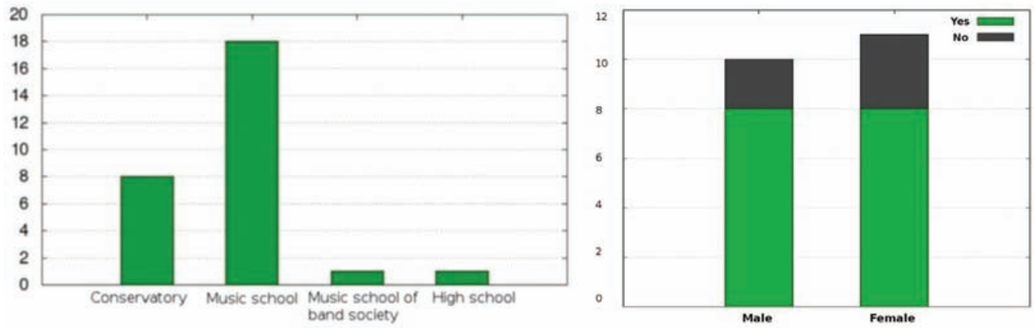

Figure 7: Left: type of institutions where the teachers teach (some teachers work simultaneously at two of them). Right: number of teachers using ITC in music classrooms.

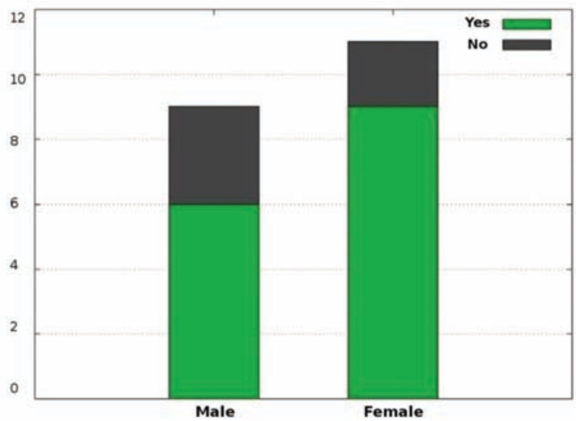

Figure 8: Teachers declaring to know other colleagues who use ITC in music classes.

considered unnecessary as the main objective of the experimental section was to make a preliminary probe of the software.

It was found that sixteen teachers (eight men and eight women) used technology when teaching music (Figure 7). Information regarding different aspects of ITC use was measured on a five-point scale ranging between 2 and 5 and showed the following data: ITC use for professional development showed a mean of 3.10, while the figure for ITC use in music teaching was 3.24, and the frequency of technology use in the teacher's personal lives was 3.76. It was also found that most of the computer programs used by the teachers were music score editors (e.g., Finale, Sibelius, Musescore, Noteflight), and, to a lesser extent, other kinds of software or applications such as Smartmusic, Audacity, Band in a Box and Internet blogs. In addition, fifteen teachers mentioned knowing other colleagues who use ITC in music teaching (Figure 8).

4.2 Instrument

A two-part questionnaire was designed in order to collect data and included (1) questions related to the sample's characteristics (demographic attributes, professional tendencies and ITC use) and (2) questions related to evaluation of the software, grouped into three – technical, didactic and global dimensions. The questionnaire also included three open questions related to Cantus, namely advantages, issues and suggestions. This instrument was validated by three independent judges that rated the items on two criteria (pertinence and item adequacy) using a five-point-scale (interjudge correlation r=0.82).

4.3 Variables

As mentioned above, a questionnaire was designed to collect data related to teachers' perceptions of Cantus. The questionnaire included several sets of questions:

1. Those related to professional aspects of the teacher (gender, age, course being taught and years of professional experience).
2. Those related to ITC use (frequency of use, programs used, experience in ITC use, current ITC use in music teaching and knowledge of other colleagues using ITC). These two groups of variables were used to characterize the sample.
3. Those related to the technical aspects of the software:
   - Efficacy
   - User friendliness
   - Versatility
   - Accessibility
   - Usability
   - Adequacy of the interface
   - Organization and navigation
   - Originality
   - Suitability of music content
   - Applicability to real music teaching and learning environments
4. Those related to didactic aspects included in the software:
   - Learning by modelling
   - Relevance of music content
   - Usefulness of content
   - Suitability of unit presentation order
   - Usefulness of neutral syllables for solmization
   - Suitability of pattern presentation order in each work unit
   - Usefulness of user input visualization
   - Usefulness of non-conventional representation
   - Usefulness of assessment grid
   - Usefulness of sight-reading
   - Usefulness of pattern creation
   - Sufficiency of program feedback
   - Adequacy of assessment
   - Sufficiency in the number of tonalities
   - Ability of Cantus to facilitate reflection on self-learning
   - Ability of Cantus to enhance motivation
5. Two additional questions were also formulated at the end of this set:
   - Would you recommend Cantus to your colleagues?
   - Do you believe that Cantus would have a positive impact on the development of students' intonation and competence in reading music?
6. Three open-ended questions were included at the end of the questionnaire in order to evaluate teachers' perceptions related to Cantus, namely advantages, issues and suggestions.

4.4 Procedure

Teachers were provided a general view of Cantus in a printed leaflet and an introductory video expressly recorded for this aim. They were later supplied with the program's electronic address (https://www.cantus.es), along with the questionnaire in PDF format in order to evaluate the software. They had one week to explore and test the software on the website and to fill in the questionnaire. After this period, the questionnaire was sent to the researchers by e-mail. Surprisingly, a significant number of teachers (21 of 25; 84%) sent the questionnaire approximately two weeks late.

5. DATA ANALYSIS

Data related to the teachers' professional and demographic characteristics are shown in the sample section. Figure 9 shows the mean and standard deviation for data collected in the set of variables regarding the technical aspects of Cantus.

All teachers rated each item highly – four points and above – with a discrete standard deviation. The most highly rated item was *applicability* in real learning–teaching environments, followed by *efficacy, originality* and *navigability*. The results do not show significant differences in the data with respect to gender. Overall, this data shows an excellent reception by teachers, who gave a high rating to the technical aspects of Cantus.

All the didactic aspects of Cantus were also very highly rated. *Usefulness of assessment grid* was the most highly rated item, followed by *Adequacy and usefulness of visualization input* and *Learning by modelling*. *Sufficiency in the number of tonalities* received the lowest rating and the highest number of comments (Figure 10).

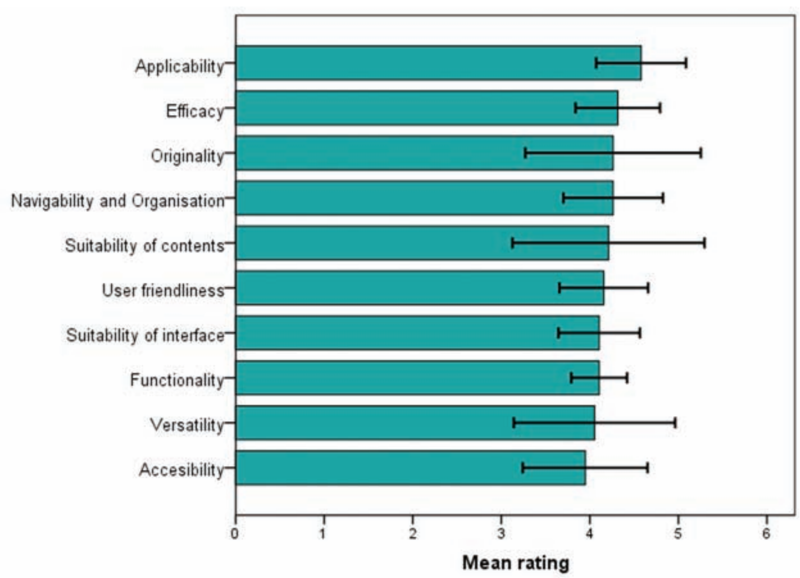

Figure 9: Technical evaluation of Cantus (+/- 1 SD in thick lines).

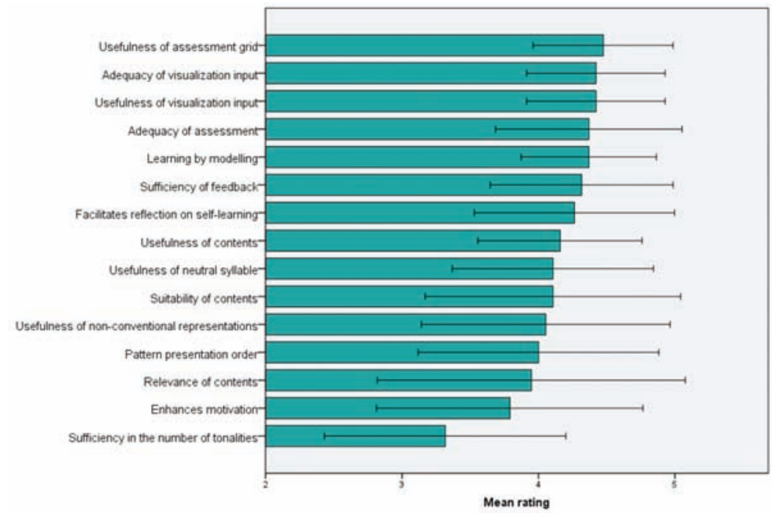

Figure 10: Didactic evaluation of Cantus (+/- 1 SD in thick lines).

Two additional questions were responded to by teachers at the end of this set:

1. Would you recommend Cantus to your colleagues?
2. Do you think that Cantus would have a positive impact on intonation development and competence in reading music?

It is important to mention here that all the teachers were ready to recommend Cantus to their colleagues. In addition, all of them considered that the program would have a positive impact on intonation development and ability in music reading. Finally, the teachers answered three open-ended questions about (1) the software's strengths; (2) the software's weaknesses; and (3) suggestions to enhance Cantus. The results, grouped in these categories, are as given below:

Strengths: The most frequently cited teacher response was related to training *autonomy* (that the software enables the pupil to work at home at their own pace, with no stress produced by the teacher or their classmates). Other responses were related to *self-assessment*: the software permits the pupils to hear and see music examples, to imitate them and auto-evaluate their own performance on the basis of their recording. The pupils get feedback from the exercises performed and may immediately view their assessment without the presence of the teacher.

Other strengths observed by the teachers were that

- Learning is by imitation.
- Ability to read music is not needed to use the software.
- The software is intuitive; it is motivating and easy to use.
- It provides visual feedback by using non-conventional representations.

Weaknesses: Most responses suggested that the number of tonalities (three, in the evaluated prototype) was insufficient to practise intonation. The teachers also felt that the tonality mode should be made clearer to the pupils by an ad hoc label. Other perceived weaknesses were as follows:

- The software does not include practise with minor relative modes of the patterns proposed in major tonalities.
- Younger pupils may not be motivated enough to practise intonation at home due to their age.
- The evaluation system is too demanding.
- The harmonic background sound (sustained chord) for exercise practice is not optimal.

Suggestions: Most teachers perceived that the number of tonalities was insufficient, and therefore that an additional number of tonalities be included. Further suggestions were as follows:

- Beginning intonation practice with fewer sounds (e.g., two sounds) and adding further tones as the students progress
- The addition of more scales
- The feature of not showing the name of the notes after the first unit (pupils should identify the notes on the staff, not by reading their names)
- The addition of further work units
- The addition of a window on the interface to select minor/major modes
- The addition of rhythm to the melodic patterns in advanced levels
- The addition of melodies at the end of each unit
- The modification of the assessment, changing it from obtaining points to achieving goals.

6. CONCLUSIONS

The main goal of this research was the development of a new software solution to facilitate tasks involved in the teaching and learning of the ability to sing in tune, or in other words, the processes of intonation. Therefore, Cantus was designed as an intonation-exercise generator tool that students could use both inside and outside the classroom. For this reason, several areas of theoretical and empirical research were considered and evaluated during its design, as mentioned in the introduction of this study. These characteristics are as follows: training by vocal imitation, pedagogical content distribution, the use of both solfeggio and neutral syllables, the dissociation of melody from the rhythmic component, the discretization of the number of sounds in each exercise, the use of musical modes, the use of a harmonic background, the inclusion of real-time feedback, the use of several representation modes, the achievement assessment in each work unit and the possibility of teacher-created exercises.

Taking into account the data obtained in the assessment, Cantus was highly evaluated by teachers for its technical elements and pedagogical dimension. It can therefore be suggested that this software may help music students develop automatic skills both in reading western tonal notation and in developing the motor skills required for correct voice intonation. If so, this pedagogical software could fill an important gap in the field of computerized music training as a training tool.

However, the content of this article is the result of an ongoing applied research. It is obvious that the next phase should be a quantitative and qualitative evaluation by students in a real context, along with an evaluation of the software's effect on music intonation ability and the role of emotions in training with the program. These tasks are currently in progress. Finally, Cantus was designed not only as a vocal intonation tool, but also as a musical instrument intonation tool. This is an area that is currently under development for string and brass instruments.

ACKNOWLEDGEMENTS

We would like to thank Bankia, Culturarts and Federació de Societats Musicals de la Comunitat Valenciana (The Federation of Musical Societies of Valencia) for funding this project, and no less importantly, we express our appreciation to all the teachers who participated in the pilot study.

REFERENCES

Anvari, S., Trainor, L., Woodside, J. and Levy, B. (2002), 'Relations among musical skills, phonological processing and early reading ability in preschool children', *Journal of Experimental Child Psychology*, 83: 2, pp. 111–30.

Bamberger, J. (1991), *The Mind Behind the Musical Ear: How Children Develop Musical Intelligence*, Cambridge, MA: Harvard University Press.

Bandura, A. (1986), *Social Foundations of Thought and Action: A Social Cognitive Theory*, Englewood Cliffs, NJ: Prentice-Hall Inc.

Bengtsson, S. and Ullen, F. (2006), 'Dissociation between melodic and rhythmic processing during piano performance from musical scores', *Neuroimage*, 30: 1, pp. 272–84.

Boyle, J. and Lucas, K. (1990), 'The effect of context on sight singing', *Bulletin of the Council for Research in Music Education*, 106: Fall, pp. 1–9.

Cassidy, J. (1993), 'Effects of various sight-singing strategies on non-music majors' pitch accuracy', *Journal of Research in Music Education*, 41: 4, pp. 293–302.

Chevais, M. (1941), *L'* éducation *musicale de l'enfance, vol. 3: Méthode active et directe/* 'Music education in the childhood, vol. 3: Direct and active method', Paris: A. Leduc.

Cheveigné, A. De and Kawahara, H. (2002), 'YIN, a fundamental frequency estimator for speech and music', *The Journal of the Acoustical Society of America*, 111: 4, pp. 1917–30, http://audition.ens.fr/adc/pdf/2002_JASA_YIN.pdf. Accessed 20 June 2015.

Colvin, R. and Lyons, Ch. (2010), *Graphics for Learning: Proven Guidelines for Planning, Designing, and Evaluating Visuals in Training Materials*, Hoboken, NJ, USA: Pfeiffer.

Colvin, R. and Mayer, R. (2011), *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*, Hoboken, NJ, USA: Pfeiffer.

Daniel, K. (1979), *Kodály Approach*, Delaware Water Gap: Mark Foster.

Daniels, R. (1986), 'Relationships among selected factors and the sight-reading ability of high school mixed choirs', *Journal of Research in Music Education*, 34: 4, pp. 279–89.

Demorest, S. (2007), 'Factors influencing the pitch matching of junior high boys', *Journal of Research in Music Education*, 55: Autumn, pp. 190–203.

—— (2011), 'Biological and environmental factors in music cognition and learning', in R. Colwell and P. Webster (eds), *MENC Handbook of Research on Music Learning*, New York: Music Educators National Conference, pp. 173–215.

Fasanaro, A., Spitaleri, D., Valiani, R. and Grossi, D. (1990), 'Dissociation in musical reading: A musician affected by alexia without agraphia', *Music Perception*, 7: 3, pp. 259–72.
Fine, P., Berry, A. and Rosner, B. (2006), 'The effect of pattern recognition and tonal predictability on sight-singing ability', *Psychology of Music*, 34: 4, pp. 431–47.
Galera, M. and Tejada, J. (2012), 'Lectura musical y procesos cognitivos implicados'/'Music reading: some of its cognitive processes', *Revista Electrónica de LEEME*, 29, pp. 56–82, http://musica.rediris.es/leeme/revista/galera&tejada12.pdf. Accessed 10 December 2015.
Galera, M., Tejada, J. and Trigo, E. (2013), 'Music notation software as a means to facilitate the study of singing musical scores', *Electronic Journal of Research in Educational Psychology*, 11: 1, pp. 215–38, http://www.investigacion-psicopedagogica.com/revista/articulos/29/english/Art_29_761.pdf. Accessed 10 December 2015.
Galyen, S. (2005), 'Sight-reading ability in wind and percussion students: A review of recent literature', *Update Applications of Research in Music Education*, 24: 1, pp. 57–70, http://upd.sagepub.com/content/24/1/57.full.pdf. Accessed 10 December 2015.
Goolsby, T. (1994), 'Eye movement in music reading: Effects of reading ability, notational complexity, and encounters', *Music Perception*, 12: 1, pp. 77–96.
Gordon, E. (2003), *A Music Learning Theory for Newborn and Young Children*, Chicago: GIA Publications.
Green, G. (1990), 'The effect of vocal modelling on pitch-matching accuracy of elementary school children', *Journal of Research in Music Education*, 38: 3, pp. 225–31.
Gromko, J. (2004), 'Predictors of music sight-reading ability in high school wind players', *Journal of Research in Music Education*, 52: 1, pp. 6–15.
Hermanson, L. (1972), 'An investigation of the effects of timber on simultaneous vocal pitch acuity of young children', doctoral dissertation, Columbia: Columbia University.
Hodges, A. and Nolker, B. (2011), 'The acquisition of music reading skills', in R. Colwell and P. Webster (eds), *MENC Handbook of Research on Music Learning*, Oxford: Oxford University Press, pp. 61–91.
Hoppe, D., Sadakata, M. and Desain, P. (2006), 'Development of real-time visual feedback assistance in singing training: A review', *Journal of Computer Assisted Learning*, 22: 4, pp. 308–16, https://www.researchgate.net/publication/220663218_Development_of_real-time_visual_feedback_assistance_in_singing_training_A_review. Accessed 23 June 2015.
Howard, D., Welch, G., Brereton, J. and Himonides, E. (2003), 'Towards a novel real-time visual display for singing training', in C. Manfredi (ed.), *Proceedings of the Models and Analysis of Vocal Emissions for Biomedical Applications 3rd International Workshop*, 10–12 December, Firenze: Firenze University Press, pp. 179–82.
Howard, D. M., Williams, J., Brereton, J., Welch, G. F., Himonides, E., Howard, A. and DeCosta, M. (2006), 'A mirror for sound: Introduction and practical exploration of the WinSingad software for teaching singing', in D. M. Howard, J. Brereton and H. Daffern (eds), *Abstracts of the 3rd International Conference on the Physiology and Acoustics of Singing, PAS3, 11–13 May, 33*, London: University of London. pp. 33.

Howard, D., Brereton, J., Welch, G., Himonides, E., DeCosta, M., Williams, J. and Howard, A. (2007), 'Are real-time displays of benefit in the singing studio? An exploratory study', *Journal of Voice*, 21: 1, pp. 20–34.

Keetman, G. (1974), *Elementaria: First Acquisition with Orff-Schulwerk*, Mainz: Schott.

Kopiez, R. and Galley, N. (2002), 'The musician's glance: A pilot study comparing eye movement parameters in musicians and non-musicians', in C. Stevens, D. Burnham, G. McPherson, E. Schubert and J. Renwick (eds), *Proceedings of the 7th International Conference on Music Perception and Cognition*, Adelaide, Australia: Causal Productions, pp. 683–86, musicweb. hmt- hannover.de/ kopiez/ICMPC7.pdf. Accessed 25 May 2015.

Lehmann, A. and McArthur, V. (2002), 'Sight-reading', in R. Parncutt and G. McPherson (eds), *The Science and Psychology of Music Performance*, Oxford: Oxford University Press, pp. 135–50.

Lucas, K. (1994), 'Contextual condition and sight singing achievement of middle school choral students', *Journal of Research in Music Education*, 42: 3, pp. 203–16.

Mann, A. (2008), 'The effect of vocal modelling on student achievement and attitude', doctoral dissertation, Eugene: University of Oregon, https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/8413/Mann_Betty_PhD_Spring2008.pdf?sequence=1. Accessed 10 September 2015.

Martenot, M. (1957), *Méthode Martenot*, Paris: Magnard.

Martin, B. A. (1991), 'The effects of hand signs, syllables, and letters on first graders' acquisition of tonal skills', *Journal of Research in Music Education,* 39: 2, pp. 161-170.

Miller, G. (1956), 'The magic number seven, plus or minus two: Some limits on our capacity for processing information', *Psychological Review*, 63: 3, pp. 81–97.

Orff, C. and Keetman, G. ([1950] 1954), *Das Schulwerk: Musik für Kinder/ Music for Children*, Mainz: Schott.

Parsons, L. (2001), 'Exploring the functional neuroanatomy of music performance, perception, and comprehension', *Annals of the New York Academy of Sciences*, 930, pp. 211–31.

Reifinger, J. (2009), 'An analysis of tonal patterns used for sight-singing instruction in second-grade general music class', *Journal of Research in Music Education*, 57: 3, pp. 203–16.

Schon, D. and Besson, M. (2002), 'Processing pitch and duration in music reading: A RT-ERP study', *Neuropsychologia*, 40, pp. 868–78. http://www.brainmusic.org/EducationalActivities/Schon_pitch2002.pdf. Accessed 10 October 2015.

Sinor, E. (1985), 'The singing of selected tonal patterns by preschool children', doctoral dissertation, Indiana: Indiana University.

Sloboda, J. (1985), *The Musical Mind: The Cognitive Psychology of Music*, Oxford: Oxford University Press.

Small, A. and McCachern, F. (1983), 'The effect of male and female vocal modeling on pitch- matching accuracy of first-grade children', *Journal of Research in Music Education*, 31: 3, pp. 227–33.

Snyder, B. (2000), *Music and Memory*, Cambridge, MA: MIT Press.

Watts, C., Moore, R. and McCaghren, K. (2005), 'The relationship between vocal pitch-matching skills and pitch discrimination skills in untrained accurate and inaccurate singers', *Journal of Voice*, 19: 4, pp. 534–43.

Welch, G. (1985a), 'Variability of practice and Knowledge of Results as factors in learning to sing in-tune', *Bulletin of the Council for Research in Music Education*, 85, pp. 238–47, https://www.researchgate.net/publication/284635485_Variability_of_practice_and_knowledge_of_results_as_factors_in_learning_to_sing_in_tune. Accessed 20 December 2014.

—— (1985b), 'A schema theory of how children learn to sing in tune', *Psychology of Music*, 13: 1, pp. 3–18.

Welch, G., Rush, C. and Howard, D. (1989), 'Real-time visual feedback in the development of vocal pitch accuracy in singing', *Psychology of Music*, 17: 2, pp. 146–57.

Welch, G., Howard, D., Himonides, E. and Brereton, J. (2003), 'Real-time feedback in the singing studio: An innovatory action-research project using new voice technology', *Music Education Research*, 7: 2, pp. 225–49.

Welch, G., Himonides, E., Howard, D. and Brereton, J. (2004), 'VOXed: Technology as a meaningful teaching aid in the singing studio', in R. Parncutt, A. Kessler and F. Zimmer (eds), *Proceedings of the Conference on Interdisciplinary Musicology (CIM04)*, (Graz/Austria, 15–18 April), Graz, Austria: Institut für Musikwissenschaft, Universität Graz, http://www.uni-graz.at/richard.parncutt/cim04/CIM04_paper_pdf/Welch_Himonides_Howard_Brereton_CIM04_proceedings.pdf. Accessed 1 November 2014.

Wilson, P., Lee, K., Callaghan, J. and Thorpe, W. (2008), 'Learning to sing in tune: Does real-time visual feedback help?', *Journal of Interdisciplinary Music Studies*, 2: 1 and 2, pp. 157–72, http://www.musicstudies.org/JIMS2008/articles/Wilson_JIMS_0821210.pdf. Accessed 10 August 2015.

Yarbrough, C., Morrison, S., Karrick, B. and Dunn, D. (1995), 'The effect of male falsetto on pitch matching accuracy of uncertain boy singers, grades K-8', *Update: Applications of Research in Music Education*, 14: 1, pp. 4–10.

SUGGESTED CITATION

Pérez-Gil, M., Tejada, J., Morant R. and Pérez, A. (2016), 'Cantus: Construction and evaluation of a software solution for real-time vocal music training and musical intonation assessment', *Journal of Music, Technology & Education,* 9: 2, pp. 125–144, doi: 10.1386/jmte.9.2.125_1

CONTRIBUTOR DETAILS

Manuel Pérez-Gil is an educator, researcher and musician with more than 30 years of experience in his field. He is currently a professor at the University of Valencia, Spain, and his research interest is the design of software for music education. He earned his doctorate from the University of Valladolid in 1993 with a thesis on the design and implementation of a set of computer programs for music instruction. He is a designer of software like Julivert, Tactus and Emolab.

Contact: Departament de Didàctica de l'Expressió Musical, Facultat de Magisteri, Avda, dels Tarongers, 4, 46022 València, Spain.
E-mail manuel.perez-gil@uv.es
Web address: http://www.uv.es/perezgil

Jesús Tejada is an educator, researcher and musician. He is professor of Music Technology at the University of Valencia, Spain. His research interests include the design of software for music education on a practical basis and the effects of multimodal presentation of musical information. He is designer of the software for rhythm training Tactus. He is currently working on the development of a software solution for intonation training for students of instrumental music.

Contact: Departament de Didàctica de l'Expressió Musical, Facultat de Magisteri, Avda, dels Tarongers, 4, 46022 València, Spain.
E-mail: jesus.tejada@uv.es
Web address: http://www.uv.es/jetegi

Remigi Morant is an educator, researcher and musician. He is Head of the Department of Education in Musical, Visual and Corporal Expression at the University of Valencia, Spain. He is Vice Chairman of the Federation of Musical Societies of Valencia. His research interest is music education in specialized music institutions.

Contact: Departament de Didàctica de l'Expressió Musical, Facultat de Magisteri, Avda, dels Tarongers, 4, 46022 València, Spain.
E-mail: remigi.morant@uv.es
Web address: http://remigimorant.blogspot.com

Alejandro Pérez is a computer science engineer in the field of artificial intelligence. He is a researcher in the Universitat Politècnica de València's Machine Learning and Language Processing Group, specializing in pattern recognition and statistical analysis. His interests include the application of Machine Learning technologies into different fields such as music education.

Contact: Machine Learning and Language Processing Group, Universitat Politècnica de València, Edifici 1F (DSIC), Camí de Vera, s/n. 46022 València, Spain.
E-mail: aperez@dsic.upv.es
Web address: http://www.alexdemartos.es