FCT project: PTDC/SAU-BEB/104995/2008

Title: Assistive Real-Time Technology in Singing

Start date: October 1st 2010
Duration: 3 years

1.0 Executive Summary (short version)

It is a fact that on the transition between the XX and XXI centuries, a period in technological history known as "information age", characterized by the ubiquity of the computer and inspired by the concept of "ambience intelligence"; the pedagogy of singing, the assessment of the quality of singing and the preventive vocal usage are poorly assisted by computers. This project proposal addresses this issue in an ambitious way by gathering together institutions, professionals and researchers from three complementary areas: singing pedagogy, engineering/signal processing, and medical/laryngology. The common purpose is to articulate knowledge and know-how from the different disciplines in order to design, implement and validate innovative technologies and methodologies that are useful to singing students, teachers or professionals, namely:

1) new technology-assisted pedagogic methodologies,
2) real-time visual feedback of relevant quality parameters of the singing voice, and
3) real-time monitoring and assessment of the singing voice with the purpose to prevent voice disorders.

In order to address these challenges, seven tasks have been planned that include the following specific goals:

- to promote a deep and structured knowledge concerning the voice production system, the correlation between subjective quality parameters of the singing voice (e.g., breathiness, clarity, vibrato, singer's formant) and objective acoustic features (e.g., jitter, shimmer, harmonics to noise ratio, harmonic irregularity and extension, closing/open coefficient of the glottal pulse), the correlation between objective acoustic features and voice disorders in singing,
- the design, realization and validation of biofeedback technologies in singing as well as technology-assisted teaching/learning methodologies,
- the design, realization and optimization of technologies allowing the real-time transcription of singing to musical score and including editing capabilities,
- the robust estimation of the glottal pulse in real-time from running singing and not only from sustained vowels as it is the rule with currently existing technology extracting information concerning the quality of the phonation or concerning the abnormal operation of the vocal folds,
- the design, realization and validation of technologies for the real-time assessment of the singing voice in order to monitor vocal stress, to detect risks of voice over-use and to prevent voice disorders.

2.0 Overview of the tasks of the project

This project proposal gathers expertise in the areas of singing pedagogy, engineering and laryngology, promotes pos-graduate research work benefitting from the synergy between different disciplines, and aims at providing singing students, teachers and professionals with solutions helping them to optimize singing learning and training, and to perform safely. This is the vision underlying seven tasks that target three mains realization areas:

1) real-time visual feedback of relevant quality parameters of the singing voice,
2) new technology-assisted pedagogic methodologies, and
3) real-time monitoring and assessment of the singing voice with the purpose to prevent voice disorders.

Three tasks are devoted to realization areas 1) and 2), three other tasks are devoted to the realization area 3), and another task is devoted to the management of the project. The seven tasks are as follows:

TASK1-correlation between subjective quality parameters of the singing voice and objective acoustic features,
TASK2-new technology-assisted methodologies in singing teaching/learning,
TASK3-singing to musical score transcription and music composition,
TASK4-correlation between objective acoustic features of the singing voice and voice disorders in singing,
TASK5-robust real-time glottal pulse estimation from running singing,
TASK6-real-time preventive assessment of the singing voice,
TASK7-management.

3.0 Partners and people

Faculdade de Engenharia da Universidade do Porto (FE/UP)
Rua Dr. Roberto Frias
4200-465 Porto

- Prof. Aníbal Ferreira, is the PI of the project and principal researcher from FEUP, and will coordinate the research activities involving FEUP in all tasks, and will also participate in TASK7,
- PhD student (BD1) to be recruited in the context of the project and will be especially involved in TASK1, TASK2, TASK3 and TASK6; these tasks require in particular excellent programming skills,
- PhD student (BD2) to be recruited in the context of the project, and will be especially involved in TASK2, TASK3, TASK4, TASK5 and TASK6,
- Nuno Fonseca, will act as a researcher and as a PhD student (not supported by the project budget) and will be especially involved in TASK1 and TASK3,
- Ricardo Sousa, will act as a researcher and as a PhD student and will be especially involved in TASK1, TASK2, TASK3, TASK4, TASK5 and TASK6,
- Susana Freitas, will act as a researcher and as a PhD student (not supported by the project budget) and will be especially involved in TASK1, TASK4 and TASK6.

Escola Superior de Música e das Artes do Espectáculo (ESMAE/IPP)
Rua da Alegria, 503
4000-046 Porto

- Prof. Daniela Coimbra, is the principal researcher from ESMAE and will coordinate the research activities involving ESMAE, specifically TASK1 and TASK2, and will also participate in TASK7,
- Prof. Rui Taveira, will act as a researcher and as a PhD student (not supported by the project budget) and will be especially involved in TASK1 and TASK2,

- MSc student (BM1) to be recruited in the context of the project, and will be especially involved in TASK1 and TASK2.

Faculdade de Medicina da Universidade do Porto (FM/UP)
Universidade do Porto - Alameda Prof. Hernâni Monteiro
4200-319 Porto

- Prof. Pais Clemente, is the principal researcher from FMUP and will coordinate the research activities involving FMUP, specifically TASK4, TASK5 and TASK6 and will also participate in TASK7
- PhD student (BD3), to be recruited in the context of the project, and will be especially involved in TASK4, TASK5 and TASK6.

Royal Institute of Technology (KTH)
Kungl Tekniska Högskolan
SE-100 4 STOCKHOLM

- Prof. Sten Ternström, will act as an advisor to the project research activities and will also be involved in the management (TASK7).

Universidade Católica Portuguesa (UCP)
Caminho da Palma de Cima
1649-023Lisboa

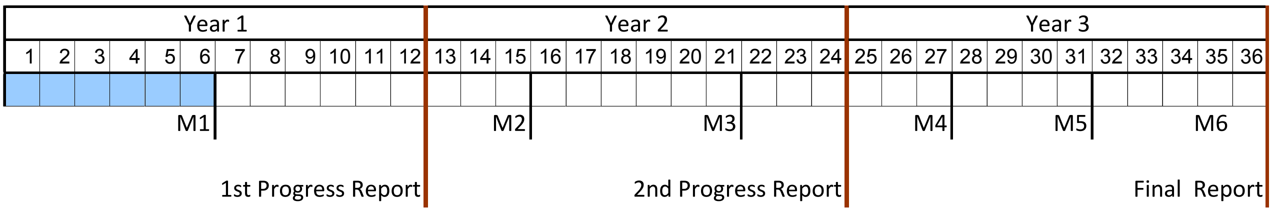
- Prof. Álvaro Barbosa (PhD), is the principal researcher from UCP and will coordinate the research activities involving UCP, specifically TASK3, and will also participate in TASK7,
- Prof. Sofia Serra (PhD student not supported by the project budget), teacher of singing at UCP, will be involved in TASK2,
- MSc (BM2) student to be recruited in the context of the project, and will be especially involved in TASK3.

4.0 TASK Specification

4.1 Task 1

Name: Correlation between subjective quality parameters of the singing voice and objective acoustic features
Coordination: ESMAE-IPP/FEUP
Duration: 6 months

Timeline



Task description
The objective of this task is to identify and characterize the most important stylistic/expressive perceptual parameters in singing, to investigate what objective

acoustic features correlate well with those parameters, and to develop efficient algorithms that are able to estimate them reliably and in real-time. This information is of paramount importance for TASK2 since the right features must be known and the right estimation algorithms must be implemented before a meaningful and useful visual representation is given to the associated perceptual parameters. Examples of perceptual parameters are pitch, brightness, warmth, clarity, vibrato, singer's formant, legato, portamento. Examples of possible acoustic features are fundamental frequency, power spectral density, spectral envelope, spectral balance, harmonic irregularity and extension, closing/open coefficient of the glottal pulse.

Most likely, new features will be found that serve better the objectives of this task. A strong possibility for this research is to include models of perception (i.e., psychoacoustic models) so as to selectively capture a representation of the acoustic information that is relevant to the auditory system, as it is acknowledged in the Memorandum of Understanding of an on-going Cost Action (2103) concerning "Advanced Voice Function Assessment" [cost2103]. Other inspiring contributions may arise from the area of auditory scene analysis [Bre90].

This research will be very interactive and experimental in nature and will involve singing students and teachers (ESMAE), digital signal processing engineers and PhD students (FEUP) who have a strong familiarity or research experience in the area. In particular, databases of singing voices will be structured in the context of this task, possibly using the voices of students and teachers at ESMAE.

Expected results
The main expected results of this task are: one report, one journal paper, and software models of estimation techniques of acoustic features.

Human resources

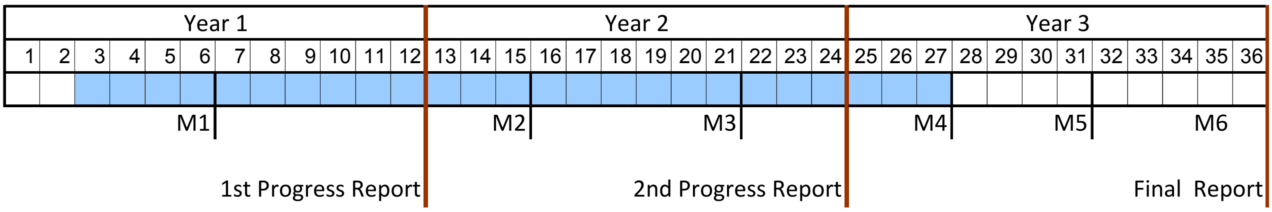
People	person-month
MSc. Stud. 1 (ESMAE)	4
PhD. Stud. 1 (FEUP)	5
Aníbal Ferreira	2
Daniela Coimbra	0.6
Nuno Fonseca	0.8
Ricardo Sousa	0.4
Rui Taveira	2
Susana Freitas	0.7
ALL	**15.5**

Relevant facts, remarks and progress so far
The PhD dissertation of Susana Freitas is strongly related to this task as it involves the correspondence between acoustic (objective) features of normal/pathological voices and perceptual (subjective) quality parameters. The opening of BM1 and BD1 scholarships must be immediate.

4.2 Task 2

Name: New technology-assisted methodologies in singing teaching/learning
Coordination: FEUP/ESMAE-IPP
Duration: 25 months

Timeline



Task description

The objective of this task is the design, realization and validation of interactive visual feedback technologies in singing as well as technology-assisted teaching, learning and training methodologies, particularly at beginner's level. Singer students and teachers from ESMAE (Prof. Rui Taveira), from UCP (Prof. Sofia Serra), and engineers will collaborate closely to enhance current singing learning, teaching and practicing methodologies with useful technologies providing objective visual feedback of singing expression, in addition to the natural auditory feedback. This combined and richer feedback of the singing voice will facilitate students to grasp better the subjective and objective dimensions of their singing exercises, which will make learning faster and more effective [Hop06]. In particular, the software environment should provide means to:
-detect objectively if the singing is out of tune,
-identify easily the vocal range profile of each singer and classify his/her singing voice register (e.g. bass, baritone, tenor, alto, soprano),
-visualize objectively the amplitude and frequency of the vibrato,
-visualize objectively the loudness of the singing voice,
-visualize objectively the continuity of the singing line (legato),
-visualize objectively the speed and quality of articulation,
-visualize objectively the metamorphosis of the vowels as a function of the fundamental frequency (i.e., as pitch increases, all vowels converge to the /a/),
-visualize objectively the dynamics and micro-dynamics of the singing voice (a property related to the style and musicality of the singing voice denoted for example by the variation and duration of the musical notes).

As a starting point, a software environment designed for singing visual feedback will be used that has been developed during the last four years under the coordination of the PI [SIS]. This software includes already competitive interactive functionalities involving the real-time visualization of the pitch of singing on a musical scale, detailed time and spectral representation, and pitch line to MIDI transcription (MIDI stands for Musical Instrument Digital Interface and consists in a protocol related to the symbolic representation of music). The results of TASK1 and also of TASK4 and TASK5 will be integrated in this environment, possibly after some re-design work in order to facilitate scalability.

In order to improve the interactive value of the visual feedback environment, whenever feasible from the point of view of signal processing, editing capabilities will also be added that will make it possible to transform the singing signal that has been recorded, in such as way as to correct or modify some desired expressive aspect, for example, to increase of reduce the vibrato extension. The capabilities will be available through an intuitive and easy-to-use touch screen menu of functionalities. Auto-scored singing exercises will also be supported by the visual feedback environment.

This task will motivate an intensive collaboration between engineers (FEUP) and singers (students and teachers at ESMAE and UCP) in order to validate the signal processing functionalities of the visual feedback environment, and in order to fine-tune and optimize the interaction and usability of its Graphical User Interface (GUI).

The GUI will be the object of careful design since the final users are mainly singing students, teachers and professionals. This means that the above goals can be successfully met only if the innovative solutions created as a result of the project operate in real-time, are non-invasive, operate with both running speech and singing, their utilization is intuitive, and if the GUIs are user-friendly. These constraints are quite severe and explain why the existing technological solutions for acoustic voice analysis for example, are regarded as disappointing by many voice clinicians [cost2103].

Expected results
The main results of this task are 4 reports (one report every six months), two journal papers, one international conference paper, and one prototype software environment.

Human resources

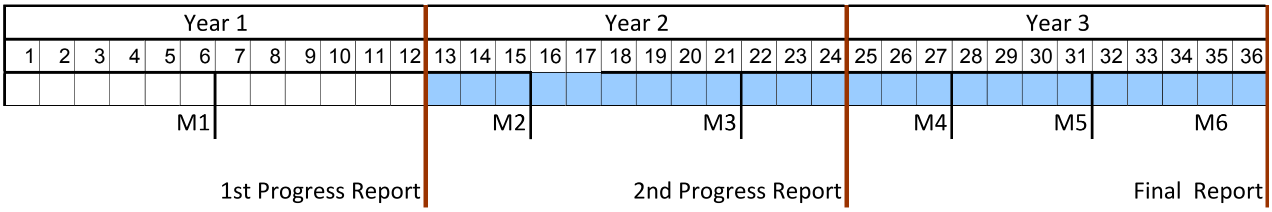
People	person-month
MSc. Stud. 1 (ESMAE)	2
PhD. Stud. 1 (FEUP)	11
PhD. Stud. 2 (FEUP)	5
Aníbal Ferreira	3
Daniela Coimbra	2.2
Ricardo Sousa	0.25
Rui Taveira	12.4
Sofia Serra	1.8
ALL	**37.65**

Relevant facts, remarks and progress so far
Prof. Rui Taveira has started already his PhD program at FEUP on Digital Media. A conversation with Sofia Serra will take place within the next two weeks to assess the possible articulation between this task and the topic of her PhD dissertation, and to plan the contribution of UCP to this task. It is important that a first version of the described interactive software is made available to ESMAE and UCP as soon as possible (which depends strongly on the involvement of BD1).

4.3 Task 3

Name: Singing to musical score transcription and music composition
Coordination: FEUP/UCP
Duration: 24 months

Timeline



Task description

To a significant extent, TASK3 runs in parallel with TASK2, and the objective is to specialize the software environment developed in the context of TASK2 in order to offer new interactive functionalities, namely (real-time) automatic transcription of singing to music score and didactic music composition using singing voice only as input. It should be emphasized that the input is an acoustic signal (the singing) that is captured by a microphone, then it is converted to a digital format (Pulse Code Modulated samples, i.e., linear PCM) which is used by the digital signal processing algorithms. Thus, in this task the input is non-semantic but the output is semantic because the music score identifies the music notes and their parameters.

In this case the MIDI (Musical Instrument Digital Interface) protocol will be used to represent the symbolic notation of music. Since in reality this consists of a set of musical parameters, MIDI is editable which means that after the transcription of singing to music score, the user is allowed to modify or correct the automatically recognized melody line and all its individual music notes. The final score can therefore be played back using any synthetic instrument that is allowed by the MIDI protocol. This is the basic didactic functionality that will be implemented in the context of this task. It will also be enhanced in order to allow the repetition of singing voice exercises that are assigned to different music instruments. Combining the results together leads to a practical music composition functionality taking singing voice as the only input. This task will bring together engineers (FEUP), musicians and interactive system designers (UCP), and will be developed in close collaboration with Casa da Música in Porto, whose representatives have already expressed their strong interest in such a scenario for workshops, in the perspective of the educational mission of Casa da Música. These workshops will be organized for the general public interested in learning the basics of singing, music notation and music composition.

Expected results

The main results of this task are 4 reports (one report every six months), one international conference paper, and one prototype software environment.

Human resources

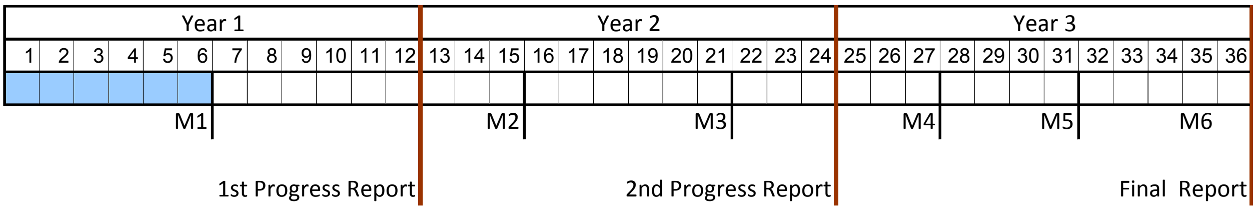
People	person-month
MSc. Stud. 1 (UCP)	6
PhD. Stud. 1 (FEUP)	10
PhD. Stud. 2 (FEUP)	12
Aníbal Ferreira	1.6
Álvaro Barbosa	1
Ricardo Sousa	0.25
Nuno Fonseca	1
ALL	**31.85**

Relevant facts, remarks and progress so far
An MSc student from FEUP has chosen the topic of this task for his dissertation. Work will start immediately with the study of the state-of-the-art and with code familiarization with the support of BD1.

4.4 Task 4

Name: Correlation between objective acoustic features of the singing voice and voice disorders in singing
Coordination: FMUP/FEUP
Duration: 6 months

Timeline



Task description
TASK4 will run in parallel with TASK1 and its objective it to identity what singing disorders are typical, what are the associated perceptual classification parameters, and to investigate what acoustic features correlate well with them. This information is extremely important as input to TASK6 whose objective is to use the acoustic signal in order to detect as early as possible, i.e., in a preventive perspective, risk factors (e.g., stress, tension) that could give rise to voice disorders.

Databases of both healthy singing and singing voices exhibiting voice pathologies like dysphonia or laryngeal lesions, are instrumental to this task. As it can be anticipated that these will be difficult to identify, contacts will be established with other research groups or organizations working on similar areas (e.g., members of Cost Action 2103 “Advanced Voice Function Assessment”, or members of the European Laryngological Research Group - http://www.elsoc.org/).

In order to understand the challenges involved and to devise possible new approaches to the problem, it is convenient to review a bit of the history of acoustic feature extraction. In our discussion, features are objective characteristics that are computed from an acoustic signal (spoken voice or singing voice) using digital signal processing techniques, after the signal has been captured by a microphone and converted to a digital format.

For more than 50 years at least, signal processing techniques have been extensively investigated and optimized in three main areas concerning spoken voice: coding/compression of speech, speech recognition and speech synthesis. Automatic speaker recognition has also received considerable attention in recent years but most frequently, the acoustic features used in this context are the same as those used in speech recognition [Sha99].

Comparatively, acoustic feature extraction for voice quality assessment has received little attention in recent years. Either the same acoustic features developed for speech coding or recognition have been used for voice quality evaluation, although with little success, or specific voice measures have been adopted with considerable more success [TIT94, Rei04]. Among these, three (jitter, shimmer and HNR) receive the largest consensus among the scientific community due to their consistent correlation with subjective parameters in sustained speech like roughness, breathiness, astheny and tension. Jitter refers to a short-term (cycle-to-cycle) perturbation in the periodicity of glottal pulses (i.e. the fundamental frequency of the voice) in the sustained phonation of a vowel, typically /a/. Shimmer refers to a short-term (cycle-to-cycle) perturbation in the amplitude of glottal pulses in sustained phonation of a vowel. The Harmonics-to-Noise ratio is a quality measure defined as the ratio between the energy of the harmonic components of a voiced vowel and the noise energy of a voiced vowel [TIT94]. Other acoustic features can be found in the literature but their relevance is not generally acknowledged by the scientific community and thus need to be confirmed in the context of this project and this task in particular.

Therefore, in this task acoustic features generally accepted by the scientific community as meaningful, will be first tested and correlations will be established using available databases. Then, new features such as harmonic irregularity/extension and closing/open coefficient of the glottal pulse [Leh07] (a research topic that will be addressed in TASK5) will be investigated.

The correlation of acoustic data with electroglottograph, laryngoscopic and stroboscopic information will also be important (as well as in TASK5) so as to conclude on functional/biomechanical profiles characterizing normal and abnormal voicing.

ORL doctors from FMUP and engineers from FEUP will collaborate in this task. PhD students who have already significant experience in acoustic-perceptual evaluation of voices will also be involved.

Expected results
The main expected results of this task are: one report, one journal paper, and software models of estimation techniques of acoustic features.

Human resources

People	person-month
PhD. Stud. 2 (FEUP)	2
PhD. Stud. 3 (FMUP)	5
Aníbal Ferreira	2
Pais Clemente	0.8
Ricardo Sousa	0.4
Susana Freitas	0.7
ALL	**10.9**

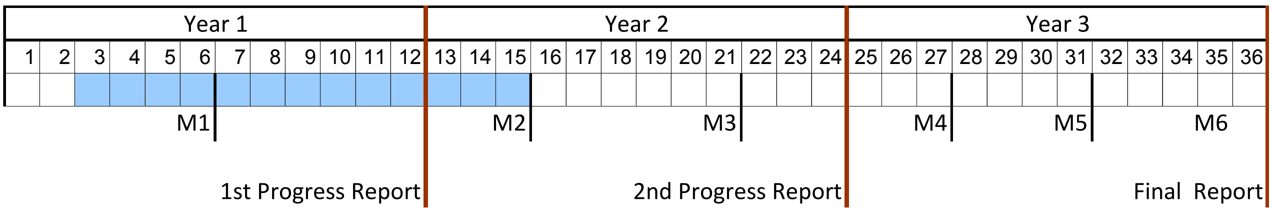
Relevant facts, remarks and progress so far

The opening of the BD3 scholarship (by FMUP) must be immediate.

4.5 Task 5

Name: Robust real-time glottal pulse estimation from running singing
Coordination: FEUP/FMUP
Duration: 13 months

Timeline



Task description

The objective of TASK5 is to develop a computational procedure that is able to estimate reliably and in a non-invasive way, the glottal pulse from running singing, in real-time. The glottal pulse is very important because it conveys quite relevant information regarding the physiological structure of the glottis and vibration pattern of the vocal folds [Ros07, Fou00, Wal07, Leh07]. In turn, these aspects determine the quality of the phonation, either in the perspective of artistic/aesthetic quality or in the perspective of healthy/non-healthy voice quality.

The objective of this task is quite ambitious as to our knowledge no solutions have yet been developed that use running singing [Sun03, Wal07], Also, the existing solutions are quite sensitive to the fundamental frequency of the voice, which indicates that the estimation in singing is likely to be more problematic than with speech. In order to estimate the glottal pulse from the acoustic signal (i.e., in a non-invasive way), an inverse filtering strategy is required [Wal07, Leh07]. Inverse filtering presumes the source-filter model (from Fant [Ros07]) of speech production and implies the reliable estimation of the vocal tract filter and lip radiation filter [Leh07]. This estimation presents practical challenges that are difficult to overcome with real sustained speech and even more difficult to address with pathological voice. Some good results are however achieved using some iterative procedure that starts with a parametric model of the glottal pulse (for example the Liljencrants-Fant model or the Rosenberg model

[Ros07]), then an estimate of the vocal tract and radiation filters is obtained which is then used to obtain an improved estimation of the glottal pulse. This approach is quite promising and it will be investigated with singing and will be adapted for real-time operation with non-stationary singing or speech. Very significant innovation results will be obtained in the context of this task that can be extended to other application scenarios where the economic value is considerable. For example, the results of the research carried out in the context of this task pave the way for the automatic remote assessment of the voice quality as when a patient calls to the hospital or clinic [Rei04]. Thus, this scenario justifies that an international patent application process be filled.

Although the objective is to develop a non-invasive procedure, invasive methods will be used to obtain data whose importance is central to complete the acoustic data in the definition of accurate models of the glottal pulse for different singing or spoken voice registers and health conditions. In particular, electroglottograph (EGG), laryngoscopic, and stroboscopic information will be captured in addition to the acoustic signal. This will be possible thanks to the participation of researchers from FMUP in this task (who are also ORL doctors), since only ORL doctors are allowed by the Portuguese law to perform these exams. Engineers (FEUP) will also be involved in this task.

The results of this task, in addition to the results of TASK4, are decisive for the success of TASK6.

Expected results
The expected outcomes of this task are 2 reports, software models and a patent application.

Human resources

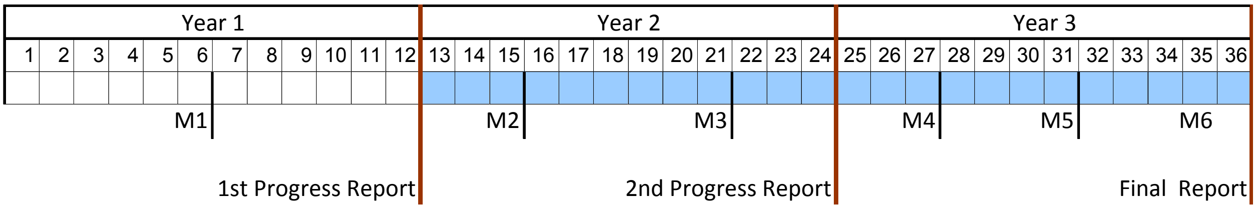
People	person-month
PhD. Stud. 2 (FEUP)	5
PhD. Stud. 3 (FMUP)	8
Aníbal Ferreira	3
Pais Clemente	1
Ricardo Sousa	0.25
ALL	**17.25**

Relevant facts, remarks and progress so far
Research work has already been developed and published regarding this task, concerning in particular the estimation of the relative delay of partials pertaining to the harmonic structure of a spoken or sung vowel. It is believed that this result may pave the way for a new frequency-domain approach to glottal pulse estimation in real-time and from running singing. A second-cycle MSc student (Sandra Dias, teacher of math) has selected the topic of this task as her dissertation topic.

4.6 Task 6

Name: Real-time preventive assessment of the singing voice
Coordination: FEUP/UCP
Duration: 24 months

Timeline



The objective of TASK6 is to develop a software environment (an extension of that developed in the context of TASK2) allowing singers to monitor their singing voice in real-time with the purpose to detect risk factors (i.e., stress factors or voice over-use factors) that could develop into voice disorders. The results of TASK 4 and TASK5 will be used to establish a safety margin between normal voice usage and incorrect or risky voice usage.

As in TASK2, it should be noted that a significant challenge that will be tackled in this task is not only to take full advantage of meaningful acoustic features, but to make it possible to extract them from running singing and not only from sustained vowels. This important advance paves the way for remote automatic assessment of the voice quality [Rei04] as it has been already highlighted in the description of TASK5.

This task will require extensive validation work and will involve researchers from FEUP (engineers) and FMUP (ORL doctors).

Expected results
The expected outcomes of this task are 4 reports, one journal paper, one international conference paper, and one prototype.

Human resources

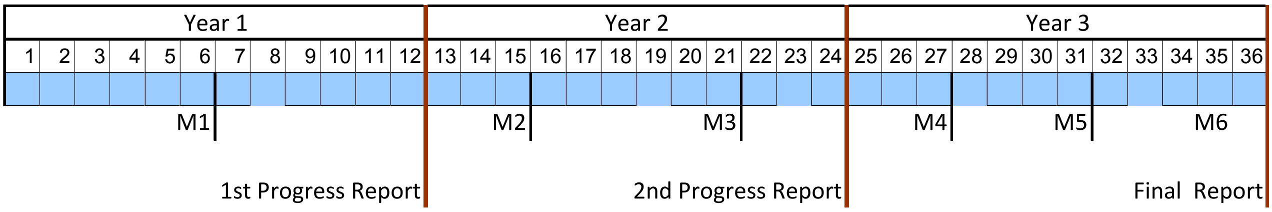
People	person-month
PhD. Stud. 1 (FEUP)	8
PhD. Stud. 2 (FEUP)	10
PhD. Stud. 3 (FMUP)	19
Aníbal Ferreira	2
Pais Clemente	1
Ricardo Sousa	0.25
Susana Freitas	0.4
ALL	**42.65**

Relevant facts, remarks and progress so far
Because this task is the natural continuation of task 4, the opening of the BD3 scholarship (by FMUP) is urgent and must be immediate.

4.7 Task 7

Name: Management
Coordination: FEUP/FMUP/UCP/KTH
Duration: 36 months

Timeline



Task description

This task is devoted to the management of the project. The principal researchers from FEUP (Aníbal Ferreira) , FMUP (Pais Clemente), ESMAE-IPP (Daniela Coimbra), and Prof. Sten Ternström from the Royal Institute of Technology (KTH), will be involved in the supervision of all activities of the project, will be responsible for detecting significant deviations in the execution relative to the plan, and will be responsible for making comments and giving advice so as to insure that collaboration is effective between the partner institutions, and that results are obtained within schedule. The principal researchers will meet twice a year and every year an internal workshop will be organized to presents results, to assess the progress of the activities of the project, and to devise corrective action if and when needed. Prof. Sten will attend all three internal workshops and, if necessary, will make an additional visit to the Partner Institutions in Porto during the first semester of 2012 in order to evaluate the progress of the activities and give qualified advice before the last workshop.

Professor Sten Ternström is a known scientist in the area of voice acoustics and is the Head of the Music Acoustics group within the Department of Speech, Music and Hearing at KTH, in Sweden. This Department is one of the most active research groups in the area of voice and singing research and home of eminent researchers, namely Prof. Johan Sundberg.

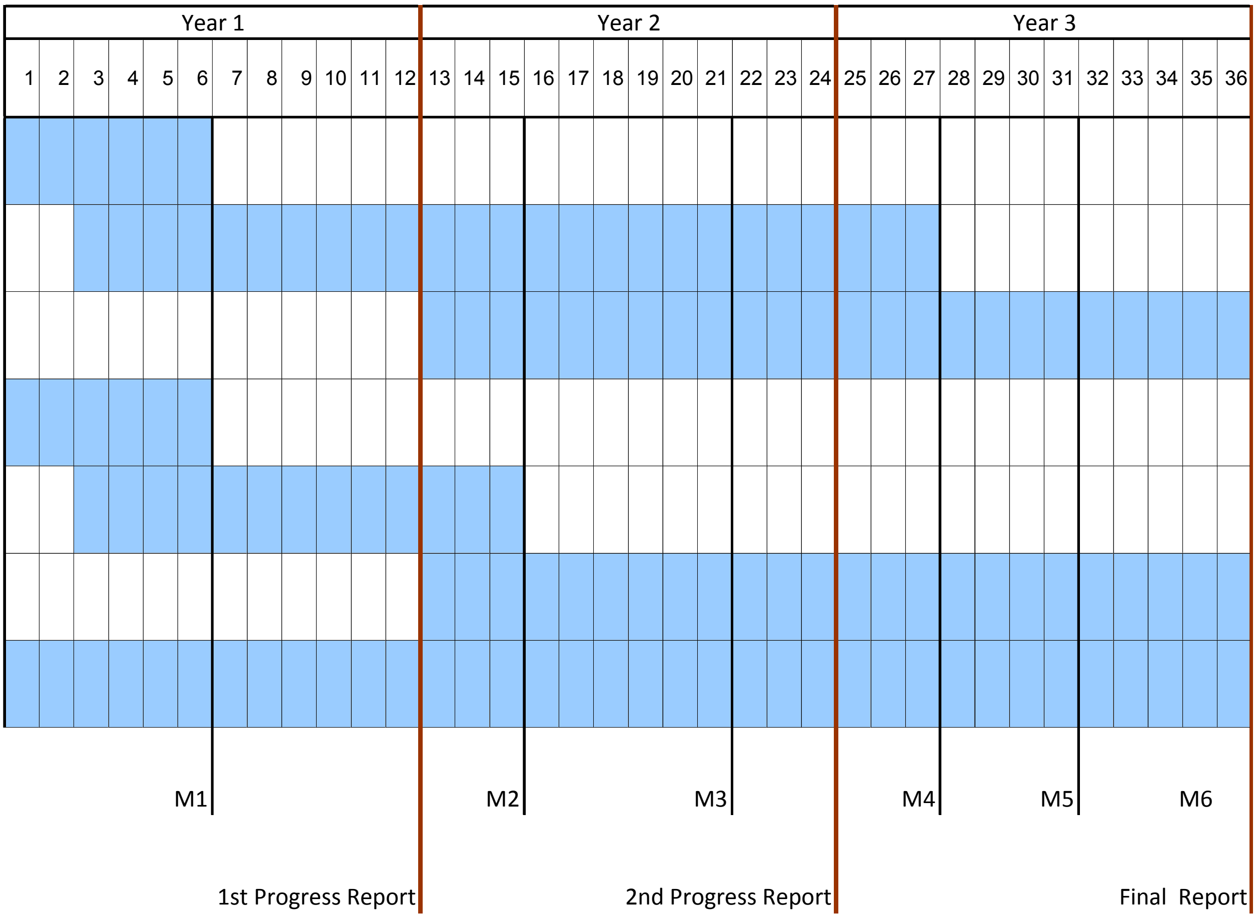
Expected results

The expected outcomes of this task are three progress reports.

Human resources

People	person-month
Álvaro Barbosa	0.8
Aníbal Ferreira	0.8
Daniela Coimbra	0.8
Pais Clemente	0.8
Sten Ternström	0.6
ALL	**3.8**

5.0 Overall Timeline



6.0 Milestone List

M1 - Milestone 1
Report describing important stylistic/expressive perceptual parameters in singing and its objective correlates. Software modules consisting of software implementation of algorithms allowing to extract objective acoustic features that correlate well with subjective quality parameters of the singing voice (TASK2), and with subjective quality parameters of the singing voice denoting voice disorders (TASK4).

M2 - Milestone 2
Report regarding the main realizations achieved in the context of Task 5.
Software module consisting in a software implementation of algorithms optimized for real-time operation and allowing to estimate the glottal pulse from running singing (TASK5).

M3 - Milestone 3
First version of the software environment complying with main functional objectives defined in TASK2 and ready to enter a phase of extensive validation.

M4 - Milestone 4
Report regarding the main realizations achieved in the context of Task 2.
Complete prototype of the software environment developed according to the objectives of TASK2 and after extensive validation and fine-tuning.

M5 - Milestone 5
First version of the software environment according TASK3 and ready enter a phase of testing and validation in collaboration with Casa da Música in Porto. First version of the software environment according to TASK6 and ready enter a phase and validation and fine-tuning.

M6 - Milestone 6
Complete prototypes of the software environments developed according to the objectives of TASK2 and TASK6, after extensive validation, optimization and fine-tuning.

7.0 References

[Cle08] 2008, C. Moura, L. Cunha, H. Vilarinho, M. Cunha, D. Freitas, M. Palha, S. Pueschel, M. Pais-Clemente (2008), Voice Parameters in Children With Down Syndrome, Journal of Voice, Volume 22, Issue 1, pp. 34-42.

[cost2103] 2006, Memorandum of Understanding (MoU) for the implementation of a European Concerted Research Action designated as COST Action 2103: Advanced Voice Function Assessment (2006).

[Ewa06] 2006, Niebudek-Bogusz, Ewa; Fiszer, Marta; Kotylo, Piotr and Sliwinska-Kowalska, Mariola (2006) Diagnostic value of voice acoustic analysis in assessment of occupational voice pathologies in teachers,Logopedics Phoniatrics Vocology,31:3,100—106.

[Fer01] 2001, Aníbal Ferreira (2001) Accurate Estimation in the ODFT Domain of the Frequency, Phase and Magnitude of Stationary Sinusoids, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 47-50.

[Fer02] 2002, Luís Gustavo Martins and Aníbal Ferreira (2002) PCM to MIDI Transposition, 112th Convention of the Audio Engineering Society, Paper 5524.

[Fer06] 2006, Ariel Rocha and Aníbal Ferreira (2006) Adaptive Audio Equalization of Rooms based on a Technique of Transparent Insertion of Acoustic Probe Signals, 120th Convention of the Audio Engineering Society, Paper 6738.

[Fer07] 2007, Aníbal Ferreira (2007) Static features in real-time recognition of isolated vowels at high pitch, J. Acoust. Soc. Am. Volume 122, Issue 4, pp. 2389-2404.

[Fer08] 2008, Anibal Ferreira, Filipe Abreu, Deepen Sinha, (2008), Stereo ACC Real-Time Audio Communication, 125th Convention of the Audio Engineering Society, Paper 7502.

[Fer08_2] 2008, Mara Carvalho and Aníbal Ferreira (2008) Real-Time Recognition of Isolated Vowels, Lecture Notes in Computer Science - Perception in Multimodal Dialogue Systems, Vol. 5078/2008, Springer156-167.

[Fer92] 1992, James Johnston and Aníbal Ferreira (1992) Sum-Difference Stereo Transform Coding, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. II-569 to II-572.

[Fer99] 1999, Aníbal Ferreira (1999) The Perceptual Audio Coding Concept: from Speech to High-Quality Audio Coding, Audio Engineering Society 17th International Conference on High-Quality Audio Coding, pp. 258-286.

[Fou00] 2000, Adrian Fourcin (2000) Precision Stroboscopy, Voice Quality and Electrolaryngography, Chapter 13 of 'Voice Quality Measurement', Kent R.D. and Ball M.J. (eds), Singular Publishing Group.

[Hop06] 2006
D. Hoppe, M. Sadakata & P. Desain (2006) Development of real-time visual feedback assistance in singing training: a review, Journal of Computer Assisted Learning 22, pp. 308–316.

[Lop08] 2008, José Lopes, Susana Freitas, Ricardo Sousa, Joaquim Matos, Filipe Abreu, Aníbal Ferreira (2008) A medida HNR: sua relevância na análise acústica da voz e sua estimação precisa, I Jornadas sobre Tecnologia e Saúde, Instituto Politécnico da Guarda.

[Rei04] 2004 Moran R., Reilly R.B., Lacy P. (2004) Voice Pathology Assessment based on a Dialogue System and Speech Analysis, Proc. of the AAAI Fall Symposium on Dialogue Systems for Health Communication.

[Sun03] 2003, Johan Sundberg (2003) Research on the singing voice in retrospect, Speech, Music and Hearing, KTH, Stockholm TMH-QPSR Volume 45: 11-22.

[Wak03] 2003, Gregory H. Wakefield (2003) Vocal Pedagogy and Pedagogical Voices, 2003 International Conference on Auditory Display.

[Wal07] 2007, Jacqueline Walker and Peter Murphy (2007) A Review of Glottal Waveform Analysis, Lecture Notes in Computer Science - Progress in Nonlinear Speech Processing, Volume 4391/2007, Springer.

[TIT94] 1994 INGO R. TITZE (1994) Workshop on Acoustic Voice Analysis - Summary Statement.

[Leh07] 2007, L. Lehto, M. Airas, E. Björkner, J. Sundberg, P. Alku (2007) Comparison of Two Inverse Filtering Methods in Parameterization of the Glottal Closing Phase Characteristics in Different Phonation Types, Journal of Voice, Volume 21, Issue 2, pp. 138-150.

[Bre90] 1990, Bregman, Albert S. (1990) Auditory Scene Analysis: The Perceptual Organization of sound. The MIT Press.
[Sha99] 1999, O`shaughnessy (1999) Speech Communications : Human and Machine, John Wiley.
[SIS] 2008, SingingStudio - SEEGNAL Research, Lda.
[VOS] 2008, VoiceStudio - SEEGNAL Research, Lda.