European Cooperation in the field of Scientific and Technical Research - COST -

Secretariat

Brussels, 11 July 2006

COST 272/06

MEMORANDUM OF UNDERSTANDING

Subject : Memorandum of Understanding (MoU) for the implementation of a European Concerted Research Action designated as COST Action 2103 Advanced Voice Function Assessment

Delegations will find attached the Memorandum of Understanding for COST Action 2103 as approved by the COST Committee of Senior Officials (CSO) at its 165th meeting on 27/28 June 2006.

MEMORANDUM OF UNDERSTANDING
FOR THE IMPLEMENTATION OF A EUROPEAN CONCERTED RESEARCH ACTION DESIGNATED AS

COST Action 2103

'Advanced Voice Function Assessment'

The Signatories to this 'Memorandum of Understanding', declaring their common intention to participate in the concerted Action referred to above and described in the 'Technical Annex to the Memorandum', have reached the following understanding:

1. The Action will be carried out in accordance with the provisions of document COST 400/01 'Rules and Procedures for Implementing COST Actions', or in any new document amending or replacing it, the contents of which the Signatories are fully aware of.

2. The main objective of the Action is to combine previously unexploited techniques with new theoretical developments to improve the assessment of voice for as many European languages as possible, while acquiring in parallel data with a view to elaborating better voice production models.

3. The economic dimension of the activities carried out under the Action has been estimated, on the basis of information available during the planning of the Action, at 21 million EUR in 2006 prices.

4. The Memorandum of Understanding will take effect on being signed by at least five Signatories.

5. The Memorandum of Understanding will remain in force for a period of four years, calculated from the date of the first meeting of the Management Committee, unless the duration of the Action is modified according to the provisions of Chapter 6 of the document referred to in Point 1 above.

COST ACTION 2103

Advanced Voice Function Assessment

A. ABSTRACT

The present Action is a joint initiative of speech processing teams and the European Laryngological Research Group (ELRG). It is widely accepted that synergies between various complementary disciplines are a promising way to efficiently address the complexity of many current problems in research and development. Particularly, progress in the clinical assessment and enhancement of voice quality requires the cooperation of speech processing engineers and voice clinicians.

The main objective of the Action is to improve voice production models and analysis algorithms with a view to assessing voice disorders. New or previously unexploited techniques, such as high speed imaging of the vocal folds and recordings throughout the life span, will be combined with recent theoretical developments in order to improve modelling of normal and abnormal voice production including substitution voices. A deep understanding of the relationship between biomechanical changes at the level of the phonatory structures and the resulting alterations in the acoustical voice signal will lead to (1) better voice production models, (2) more accurate and clinically useful methods of investigation of voice quality, and (3) strategies for preventing occupational voice disorders in professional speakers.

Keywords: Voice quality, voice production models, biomechanics, speech and signal processing.

B. BACKGROUND

It is widely accepted that synergies between various complementary disciplines are a promising way to efficiently address the complexity of many current problems in research and development. Indeed, one of the goals of COST-funded research is to address multidisciplinary issues via the cooperation of experts focused on different topics.

Despite many years of effort devoted to developing algorithms for speech signal processing, and despite the elaboration of automatic speech recognition and synthesis systems, our knowledge of the nature of the speech signal is still limited. However, voice scientists and clinicians have developed analysis methods, which are inspired by the simple models and methods developed by speech signal processing engineers, for the assessment of voice disorders.

The limitations of existing models and methods are felt in both areas of expertise, i.e. speech signal processing applications and assessment of voice disorders. For example, the intervals are unknown within which signal model parameters must remain to represent signals with timbre that is perceived as natural. The efficient control of voice quality has important applications in modern text-to-speech synthesis systems (creating new synthetic voices, simulating emotions etc.). Voice clinicians, on the other hand, have expressed their disappointment with regard to the performance of existing methods for assessing voice quality. Major issues with current methods include robustness against noise, consistency of measurements, interpretation of estimated features from a speech production point of view, and correlation with perception.

COST Action 277 (Nonlinear Speech Processing) was an attempt to address limitations of existing speech models. Algorithms have been proposed and models of speech production have been suggested. The effort of the speech research community to include nonlinear signal processing techniques in speech analysis is an indicator of the complexity of the speech signal. However, many questions regarding the structure of speech signals remain unanswered: the estimation and interaction of individual descriptors of that structure, the way that these descriptors define the quality of voice etc. Any attempt to discover and describe the complex fine structure of the human voice relies on studies of the speech production mechanism and requires a better understanding of the physical processes involved.

In 1989, several European clinical research teams involved in research in laryngology and voice created the European Laryngological Research Group (ELRG http://www.elsoc.org/). The goal of ELRG is to coordinate and stimulate research on functional assessment of voice quality. In 2001, the research committee of ELRG proposed, and the European Laryngological Society (ELS) accepted, a protocol for the assessment of voice pathology and presentation of results from voice treatments. Seven European countries are currently using this protocol. Based on this protocol, databases of disordered voices have been compiled. Also, new instruments, such as high-speed laryngeal imaging systems, provide new ways for understanding the details of voicing mechanisms.

The above outline of activities in signal processing, laryngology and phoniatry suggests that the Action addresses issues that are equally important in both research areas and that cooperation between the two disciplines is not only desirable but necessary. Progress in speech signal processing may lead to improved acoustic features and provide new methods for quantitatively assessing voice disorders and voice quality in general. On the other hand, databases collected by clinicians will enable the interpretation of automatically extracted descriptors of the speech signal and, as well, lead to the development of models for the interaction of these descriptors.

Occupational health of professional voice users (such as teachers) is another important field in which a corroborative approach between each discipline is highly promising. The social relevance and economical impact of occupationally induced voice disorders is obvious: in the USA, the three million elementary and secondary school teachers represent the largest group of professionals who use their voice as a primary tool of trade. In a review of prevalence studies it is reported that at least 50% of teachers experience voice problems while the Educational Research Services and the National Center for Educational Statistics report that in the USA the annual national cost for teacher absenteeism related to voice problems amounts USD 567 million. Advanced technology for voice dosimetry and analysis within the working (teaching) situation needs to be developed (in a similar manner to the introduction of safety protocols for occupational noise-induced hearing loss). On the other hand, designing and implementing an intelligent audio-enhancement system for reducing vocal loading is a current project of one of the teams of the ELRG.

The only project related to voice pathology funded under the EU Framework programmes, was Ortho-Logo-Paedia (OLP) (FP5). The objective of OLP was to develop and test an integrated computer-based system to supplement conventional speech therapy. The focus was mainly on logopedic issues, providing people with speech disorders an individualised feedback and an interface to assistive technology for continuing the therapy outside the clinic. Currently, there is no EU-funded project that addresses issues of voice function assessment. At present, there is no EU instrument to coordinate common research activities of speech signal processing engineers and voice clinicians.

The research activities in the Action will have a basic and pre-competitive character. Until now, coordination of activities was performed via personal contacts and meetings in workshops and

scientific conferences. Given the need for the provision of feedback to policy makers (social services, legislators), for raising public concern about voice disorders and how they may affect quality of life, for normalising protocols for transnational use, and for pooling expertise from very different domains, it is crucial to network under an organising instrument such as COST, teams of engineers, researchers and clinicians working in the field of what has been termed 'vocology'.

C. OBJECTIVES AND BENEFITS

The main objective of the Action is to combine previously unexploited techniques with new theoretical developments to improve the assessment of voice for as many European languages as possible, while acquiring in parallel data with a view to elaborating better voice production models.

Progress in the clinical assessment and enhancement of voice quality requires the cooperation of speech processing engineers and laryngologists as well as phoniatricians. Specifically, this Action is a joint initiative of speech processing teams and the European Laryngological Research Group (ELRG).

Specific measurable objectives of the Action are the following:

1. Improve voice production models and analysis algorithms that impact on speech processing applications and assessment of voice disorders.

Combined analysis of voice data will enable the correlation of high-speed motion pictures of the vibration of the vocal folds with other physiological and/or acoustical signals. This opens up new possibilities for describing the biomechanics of normal and abnormal voicing. Such data will enable the development of better voice production models and may suggest better algorithms for speech synthesis, speech modification and speech signal modelling.

2. Develop accurate and clinically useful methods for investigation of voice quality in patients and professional speakers;

The improvement in acoustic analysis will provide clinicians with reliable cues of vocal timbre in continuous speech as a standard for use in clinical practice. It will also help voice clinicians and laryngologists assess the outcome of voice treatments (medication, functional voice therapy, phonosurgery). These developments have a direct impact on society, as they are a basic condition for improving quality of health care. For example, social activity of (totally or partially) laryngectomised patients is at present limited because substitution voices or artificial voice generators that are currently used enable the production of only poor-quality or non-human (mechanical, robotic) speech sounds. Work conducted in the framework of the Action is therefore geared towards developing improved substitution voices. Other examples include developing acoustic analysis/synthesis strategies to (a) improve the quality of life for patients with chronic voice problems and, (b) protect persons who use their voice as a primary tool of their occupation against developing voice disorders.

3. Build databases, particularly for testing and comparing these methods;

4. Assist European committees (particularly the Committee on Phoniatrics of the European Laryngological Society) in issuing guidelines and publishing protocols for several European languages.

Protocols will be designed and implemented for the assessment of the quality of substitution voices for laryngectomy patients and other forms of dysphonia with strong aperiodicity, for as many European languages as possible. This is of a major social relevance because these protocols enable comparisons and meta-analyses between, for example, different types of treatments for laryngeal malignancy.

D. SCIENTIFIC PROGRAMME

The Action addresses the problem of voice function assessment in an interdisciplinary way. Voice quality assessment refers to vocal function testing, which has today reached a critical stage in its development, because of the convergence of several semi-independent trends that may be observed in engineering, (psycho)physics, laryngology and occupational health, and speech therapy as well as in society at large.

Advances in speech processing have the potential to improve existing acoustic features and provide new features that quantitatively assess voice disorders and voice quality in general. The interpretation of these features from the perspective of voice production mechanisms enables extracting information about the pathophysiological and physical aspects of voice and speech.

To achieve the main objective of the Action, it is necessary to:

• Develop protocols for collecting categorical data on voice quality via self-evaluation of patients or via perceptual evaluations by clinicians for several European languages. The objective is to generalise and standardise clinical tools so that they may be used transnationally.

The protocol proposed by ELRG and approved by ELS in 2001 is a basic protocol for the assessment of voice pathology and of results of voice treatments. This was, however, applied and tested in a limited number of European countries. Collecting data (including self-assessment by patients) on a larger base is necessary. Moreover, improvements of the ELS protocol are desirable: for example, analysis of running speech instead of a sustained /a:/, and extensions to voice types other than type I voices (type I voices are nearly periodic).

• Build a multilingual database with normal and disordered voices representing a wide range of laryngeal pathologies. The goal is to provide developers and users of clinical software with reference data that forms the basis by which different methods may be compared. Databases have been central to the development of robust automatic speech and speaker recognition devices. The purpose of the voice disorder database is to provide a similar stimulus for clinical applications.

• Develop and evaluate acoustic cues of disordered vocal timbres that are reliable for sustained speech sounds and connected speech produced by severely hoarse speakers.

Acoustic analysis provides objective and non-invasive measures of vocal function. However, standard acoustical analysis is limited to stable fragments of sustained vowels and to perturbation cues to achieve a minimal clinically acceptable reliability. Only pseudo-periodic signals can be processed cogently. Most clinicians therefore request a robust extension of acoustic measurements to running speech (which is closer to habitual voice use).

• Adapt the recording of these clinical cues to portable devices for real-time in situ monitoring of professional speakers.

Occupational voice disorders have become an object of major interest in recent years, owing to the important social and economical implications of these disorders. Acoustical analyses are expected to constitute the core component of future portable systems that (1) preventively monitor the voice of normal speakers who belong to professional groups at risk of developing voice disorders (e.g. teachers, singers), and (2) warn against voice overuse and vocal stress. For occupational health purposes (and correlated risk evaluation), the use of voice dosimeters should be more widespread. Qualitative improvements of existing devices are firstly required. Such devices are expected to enable clinicians to monitor the voices of professional speakers (who have already developed voice problems) in their usual professional environment that may be very different acoustically from the laboratory or voice clinic.

• Develop statistical models for fusing several acoustic measurements into hybrid cues with a view to increasing the correlation between perceptual quality and numerical features of vocal timbre. The goal is to investigate whether automated recordings of voice quality features may support and refine (and perhaps, if high correlations are demonstrated, even replace) perceptual evaluations that are performed routinely in the clinic, but which are limited by inter-observer variability.

Currently, many speech decomposition techniques have been proposed in the context of speech analysis, speech coding and speech synthesis (e.g. prosody modification) while various forms of linear filtering and nonlinear transformations of speech were used for speech and speaker recognition. These forms of decomposition were mainly based on signal processing approaches or pure information theoretic arguments. These approaches to speech signal modification or compression are only loosely connected to the physical processes involved in voice production. The work in the Action concerned with the understanding of the complexity of speech signals is aimed at (1) developing speech decomposition methods that establish a connection between estimated components and the production system, and (2) reliably describing and tracking vocal function.

Also relevant to the Action is the analysis of speech events on a finely calibrated time scale. Such speech events, which are usually referred to as ‘fine structure’ in the speech waveform, arise from time-varying elements of the speech production mechanism and facilitate distinguishing between different speech sounds or voice types.

• Develop and evaluate mathematical models of voice production for normal and pathological cases (e.g. of chaotic vocal fold vibration) by means of biomechanical data (as for example, provided by high speed imaging).

With regard to voice production models and signal modelling the research of the Action is concerned with (1) the modelling of various voice phenomena, (2) the developing of methods to estimate in a robust way features characterising the voice production mechanism from the acoustic signal and, (3) studying the interaction between these features. Current proposed models will be reviewed and enhanced models will be designed.
Using biomechanical models one can study the physics of the production system with accuracy and realism. For example, high-speed imaging systems enable linking high-speed motion pictures of the vibration of the vocal folds to other physiological and/or acoustical signals. This opens up new possibilities for describing the biomechanics of normal and abnormal voicing mechanisms. Such data would enable the development of better voice production models and suggest better algorithms for speech synthesis, speech modification and speech signal modelling.

Finally, biomechanical data are also important to the electrophysiological exploration of the larynx. They will provide, in combination with mathematical models of voice generation, impetus towards

the development of hands-free devices that synthesise human-like voices for laryngectomised patients.

• Adapt existing psychoacoustic voice quality criteria used in audio coding algorithms to clinical use. The objective is to establish synergetic relationships between methods of subjective evaluation developed in different domains of application.

The current ELS protocol is not adequate for evaluating substitution voices or spasmodic dysphonia, for which the aspect of intelligibility and fluency requires additional investigation.

Acoustic analysis, possibly based on nonlinear approaches, and psychoacoustic-based algorithms applied in existing audio coding technology, are expected to contribute to the development of new quality criteria. It also should improve the understanding of the basic mechanisms of dysphonia and thus the underlying principles of phonosurgery.

The development of more accurate and robust objective/quantitative methods for assessing vocal function would also contribute to improving the assessment of the efficacy of competing forms of treatments.

E. ORGANISATION

A Management Committee (MC) consisting of Delegates from the Signatory countries will be responsible for coordinating all activity within the Action. To ensure the success of the Action, a smooth and effective collaboration between the two disciplines, speech processing and laryngology, is essential. The chairperson of the Action will have overall responsibility for the timely completion and delivery of reports to the COST Office. The MC will be supported by a secretariat or administrative office responsible for the financial issues. The MC will meet at least once every year.

The technical activities of the Action will be organised in the framework of five Working Groups (WGs). WGs are dedicated to achieving a set of objectives and follow the specific, strategically important, technical direction of the work plan. Each WG will have a designated coordinator who will monitor WG activities.

The chairperson, the vice-chairperson, and the coordinators of the WGs will form the Steering Committee (SC). The SC will meet regularly and not less that four times a year and it will act as the main driving force to ensure that the Action achieves its technical objectives. Meetings of the SC are open to all members of the MC of the Action. The last meeting each year of the SC will coincide with a meeting of the MC. During the joint meeting of the MC and SC, the SC will report on the technical progress and achievements within the Action. Based on this report, the MC will decide on new directions if so required. These directions will be the guidelines for the next meetings of the SC.

A key point for the success of the Action is the interaction between the two disciplines. Short-Term Scientific Missions (STSMs) and exchange of scientists and PhD students will be encouraged in the framework of the Action, providing them with the opportunity to use instrumentation currently available only in some laboratories (i.e. high-speed imaging). Applications for such missions and exchanges will be directed to the SC. During the MC meeting, the SC will make suggestions with regard to these activities to the MC for final approval.

Workshops or summer schools will be organised in parallel to the MC meetings, where the work in progress within each WG will be presented. Distinguished experts from Canada, Japan, Korea,

USA, etc. will be invited to participate in these events to strengthen the links and interactions of the Action with research centres outside Europe.

A web site will be set up to provide information about the goals of the Action to a wider audience and for communication within the Action. The secretariat, with the support from the SC, will maintain and update the web site.

Based on scientific challenges listed in Section E, the following Working Groups (WGs) will be implemented:

WG 1: Development of speech models based on physical processes involved in voice production. Study of the fine structure of the speech signal.

WG 2: Development of accurate and robust objective/quantitative methods for assessing vocal function.

WG 3: Development of improved protocols for the assessment of voice quality in several European languages. Recording and management of databases of voice disorders.

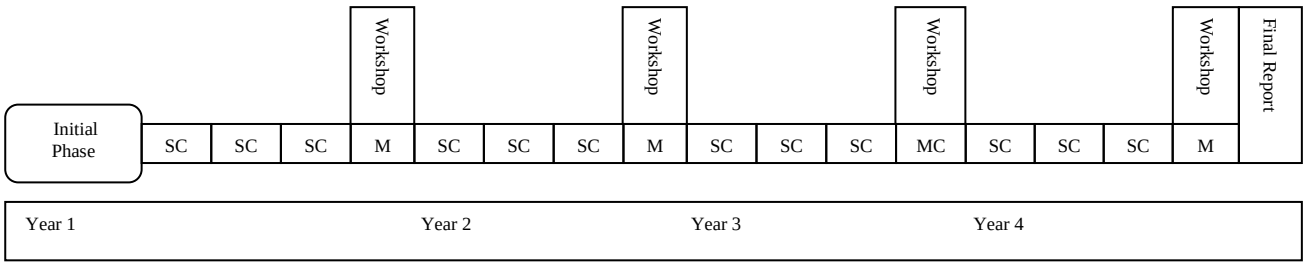
WG 4: Development of new instruments and devices for monitoring occupational voice disorders, and for generating human-like voices for laryngectomised patients based on improved acoustic analysis procedures and models.

WG 5: Coordination of actions in the Signatory countries for raising public concern about voice disorders and quality of life. Provision of feedback to policy makers (social services, legislators).

F. Timetable

The Action will coordinate a multidisciplinary group of university clinicians and speech scientists. The cooperation of several research teams with different backgrounds and priorities is a challenging task. It will be necessary to synchronise the efforts of the two communities: identify current problems in both domains, identify and collect existing resources in data and algorithms.

As indicated previously, concurrent with each meeting of the MC, a workshop will be organised where results obtained by each WG will be presented. The timetable of the Action is depicted in the figure below.



G. ECONOMIC DIMENSION

The following COST countries have actively participated in the preparation of the Action or otherwise indicated their interest: Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Lithuania, Netherlands, Slovenia, Sweden, United Kingdom.

On the basis of national estimates provided by the representatives of these countries, the economic dimension of the activities to be carried out under the Action has been estimated, in 2006 prices, at approximately 21 million EUR.

This estimate is valid on the assumption that all the countries mentioned above but no other countries will participate in the Action. Any departure from this will change the total cost accordingly.

H. DISSEMINATION PLAN

Research progress and results will be published in international peer-reviewed journals and conferences. COST Action events will be organised at selected conferences to report on the Action and to advertise COST in general. Every opportunity will be taken to promote the outcomes of the Action through special sessions at international conferences such as the International Conference on Audio Speech and Signal Processing (ICASSP), Interspeech, International Conference on Spoken Language Processing (ICSLP), and at the European Laryngological Society's workshops and conferences. Special issues in scientific journals will be produced (e.g. IEEE Speech and Audio Processing, Speech Communications, EURASIP *journals*, *Journal of Voice*, *European Archives of Oto-Rhino-Laryngology* etc.) in which invited articles and articles selected from the MC papers presented in the workshops organised by the Action will be published.

Advances in voice function assessment will be published in a book during the last year of the Action, where the new methods, extended protocols, and data will be presented while future directions in the domain will be discussed.

Meetings and annual summer schools or workshops will be organised for stimulating and promoting the research conducted in the Action.

The involvement of the European Laryngological Research Group (ELRG) in the Action provides a mechanism suited to the exploitation of the results, in a way that is most likely to produce practical impact. New concepts for the assessment of voice function and new instruments will be presented by members of ELRG directly to the appropriate audience (e.g. clinicians) in scientific meetings or through publications in specialised journals such as the *European Archives of Oto-Rhino-Laryngology* and *Head & Neck Surgery*, the *Journal of Speech and Hearing Disorders*, etc. In this way, results from the Action will gain maximal clinical acceptance.

In addition, COST Action outcomes will be transmitted and circulated via the European Laryngological Society (ELS). For example, approval of the extended protocols by ELS will ensure that these will be accepted by clinicians and applied in practice. The ELRG organises a specific symposium within every ELS congress; the next one is already scheduled for September 2006 in Nottingham (UK).

The Action will provide feedback, summary reports and publications to policy makers (e.g. legislators and social services) whenever this is required by these organisations. The Action will invest in disseminating information (scientific summaries, brochures etc.) to appropriate audiences such as associations of teachers, for raising public concern about voice disorders and quality of life. A dedicated WG has been established for this purpose.

A web site will provide information regarding the Action and publish events, news and ongoing research activities in the framework of the Action. It will also provide links to state-of-the-art reports and articles in scientific and technical journals published by organisations and researchers outside the Action, interim reports, guidelines, manuals, and final reports.