



SPEECH TRAINING FOR DEAF AND HEARING- IMPAIRED PEOPLE

ISAEUS consortium

Ramón García Gómez¹ (project coordinator), Ricardo López Barquilla¹, José Ignacio Puertas Tera¹, José Parera Bermudez¹, Marie-Christine Haton², Jean Paul Haton², Pierre Alinat⁴, Sofia Moreno³, Wolfgang Hess⁵, M^a Araceli Sanchez Raya⁶, Eduardo Alberto Martínez Gual⁶, Juan Luis Navas-Chaveli Daza⁶, Christophe Antoine², Marie-Madeleine Durel⁷, Genevieve Maugin⁸, Silke Hohmann⁵

1: UPM, 2: LORIA, 3: Teima Audiotex, 4: Thomson Marconi Sonar 5: U. Bonn, 6: Promi, 7: Centre Hospital Regional de Nancy, 8: Urapeda.

ramong@tdt.ssr.upm.es.

Abstract

In this paper we will describe the results of the ISAEUS project (TIDE DE 3004) achieved until now. The objective of this project is to develop a state of the art, cost effective, prototype for training deaf people in German, French and Spanish languages. We describe the user needs as understood for the users of ISAEUS: deaf persons and speech therapist. The paper also includes a brief description of the main blocks of the system: the acoustic-phonetic analyser in which will be based the exercises, the set of exercises that will be included, other features of the system and the real implementation. The paper will finalise with the description of the strategies for evaluating the effectiveness of this tool.

Introduction

In spite of the widely recognised necessity of having visualisation tools for representing speech that can help the deaf people to speak with good intelligibility and quality (by using visual feedback instead of the natural auditive feedback) the problem remains basically unsolved. The ISAEUS project addresses this problem. We are implementing a system for training deaf or partially deaf persons.

The first point that we have to realise is that the hearing and seeing have neither the same effective value nor the significance for the hearing impaired persons as for normal hearing persons. This fact has a profound impact: the sign language is the natural language of the deaf people. The sign language allows deaf people with major deafness to follow a normal evolution. We must recognise the strong interdependence between thinking and language. Consequently, the problem must be stated in a proper way: we have to address the problem of optimisation of the oral language for deaf people as a way of relating them with the oral speaking community.

We need to combine sign, written and spoken languages in a tool based on the standard PC that could be useful for deaf and partially deaf persons. The three languages will be used for providing information to the deaf. The system must also be accessible to a wide

range of ages: for children (the population most flexible within the deaf community) and for adults who have been oralised and who have the willingness of correcting their voices. The motivation of the user is an important requirement in order to maintain the attention enough time for obtaining advances in the intelligibility and quality of their voices. See the resume of the User's Needs.

Speech training of deaf persons is language dependent. In this sense one of the ISAEUS purposes is to implement exercises in three different languages: German, French and Spanish and compare the results.

In this paper we are going to describe the exercises, classified in six different levels, as envisaged by ISAEUS in the three languages. All the exercises will be based on the analysis of the speech signal by mean of the Speech Analyser, which will be described in the following section. The Speech Analyser provides a wide set of parameters both categorical and non-categorical. For each lesson we select from this set of parameters the more relevant ones to compare speech with the target schemes. In order to provide the visual feedback we need to represent the multidimensional time varying information about speech. The visual representation given by ISAEUS must be intelligible for the deaf people: after the sensibilisation phase the users will be able to modify their voices in order to achieve the target objectives. Different alternatives of mapping the multidimensional time varying information into a graphical representation will be given and how to provide the feedback advises for the deaf. The paper will conclude with the hardware and software description of ISAEUS system and the strategies for carry out the corresponding evaluation of the effectivity of ISAEUS as a training tool for the deaf people.

User's Needs

The user needs were defined based on three main sources of knowledge: the deaf people themselves, the speech therapists (represented in the ISAEUS project as three associations of deaf people), and commercial available systems (Speech viewer, VISHA and specially the SIRENE II).

The main points of the user's needs document are summarised as follows:

1. ISAEUS will include a set of exercises for training speech. It will also include tools for storing the results, personalising the exercises and configuring the microphone, the headphones and the vibrotactil device. It will extract graphical information about user's evolution.
2. ISAEUS will be an integrated system of tools.
3. It will also be able to make tonal audiometer measures.
4. ISAEUS will have protection measures to be sure that a user can not access to other user's data. Two ways of accessing the program will be implemented. The first way will be self-training way, where the user can access to data and to the exercises. This will be the way of the Speech Therapists to access to the system. The second way will be assisted by the Speech Therapists. In this mode the user will be able to access only to his/her account, and only the speech therapist will be able to give him/her access to the system.
5. ISAEUS system will integrate signed, written and oral languages in the explanations of the exercises.
6. The exercises will be structured into six levels with, at least, one exercise per level. The exercises will be based on the SIRENE II exercises that will be extended including exercises at higher levels of the speech and adding more relevant parameters at the low levels.
7. Two kinds of representations will be envisaged, one for speech therapists and adult trainees, and other for children. In the high levels as 5 and 6 the representations may be the same. These representations must be attractive to both groups.
8. Three types of feedback will be used: visual, one-channel vibrotactil and acoustic feedback.
9. The representation of loudness must be included in all exercises, placed in the same part of the screen in order to make simpler its control.
10. The use of scores to monitor the evolution in adults and the integration in simple visual games will help to motivate the users to continue the training.
11. ISAEUS will be controlled using mainly the mouse and its interface will be based in Windows95-style interfaces.
12. The exercises should include where possible demos of how to perform them.
13. In the exercises that use reference patterns, different models will be used for children, women and men.

14. Spectrograms and signals should be presented in a more complex way for speech therapist and specialists.

Levels of learning

The training sequence is divided in six levels. Each new level train new aspects with the previous ones trained in the previous levels. These are the levels:

1. Breath control and rhythm.
2. Training of vowels with basic prosody aspects (loudness, rhythm, and intonation).
3. Training of consonants in isolated way and in simple syllables.
4. Training of complex syllables and simple words
5. Training of complex words.
6. Training of sentences.

The system should exercise as early as possible, in a global way the interaction via the junction of all parameters, and not rely on individual parameters. The feedback should be made more complex including more learning parameters, the relevant ones for the learning process at the moment.

Structure of the exercises

An exercise can be described by a set of properties. The following list gives most of these properties that will be defined for each exercise, with examples:

- Title of the Exercise: Pronunciation of repetitive syllables
- Reference number 1.1.4
- Level of Learning: First (1st). Suprasegmental Aspects.
- Objective: Coordination of intonation, accent and rhythm.
- Prerequisite (conditions for accessing the exercise):
 - Be able to master the parameter involved separately
- Links with other existing or planned exercises
- Parameters and/or abilities to watch for:
 - Voicing, rhythm, intonation...
 - Rhythm-intonation coordination
 - Quality of the central vowel
- Feedback data that will be presented to the trainee (audio, video, tactile, on-line or off-line)
- Automaton of the exercise, as seen by the trainee, for example:
- Introductory explanation

- On-line simulated demonstration
- Performance by the trainee
- On-line visual display, audio or tactile feedback
- Off-line display of information
- Scoring from the valuation of the trainee's utterance
- Congratulations and recommendations
- Relevant selected parameters provided by the speech analyser.
- Methods for computing the feedback data from the selected parameters.
- Additional useful data
- Internal setting of the exercise
- Visual representation mode
- Methods for evaluating the trainee's utterance.
- Consequences when the exercise is finished (updatings, of the trainee's profile for example).
 - Conditions for stepping to another exercise.

Differences among the three languages

ISAEUS is to run in three languages, French, Spanish, and German. Its forerunner, SIRENE-2, was designed for French. Apart from the fact that all descriptions, instructions, and other text displays are designed in three languages, there are a number of language-specific items that will influence the speech analyser and the design of exercises. In the following there are a few examples.

- The vowel systems of French and German are similar with respect to vowel quality (except for the nasal vowels that do not exist in German). The Spanish vowel system is much simpler as it contains only the five vowels /i/, /e/, /a/, /o/, /u/. The quality of the mid vowels /e/ and /o/ differs from the corresponding vowels in French and German.
- Both Spanish and German have diphthongs, whereas French has not.
- In German vowel quantity is distinctive, i.e., there exists a long-short contrast. However, most short vowels (except for the /a/-/a:/ contrast) also differ in quality from their long counterparts. We will thus only be able to train the long German vowels with a sustained-vowel exercise. The short vowels must be trained simultaneously with speech rhythm and long-short contrasts.
- In contrast to French, in both Spanish and German word accent plays a major role. In both languages there are minimal pairs only distinguished by word accent. Exercises for German and Spanish must

thus take into account that across polysyllabic words the accent may fall on (almost) any syllable.

- Both Spanish and German have dorsal fricatives ([x] in Spanish; [x] and [ç] in German), whereas French has none of these.
- German phonotactics is considerably more complicated than that of French and Spanish with respect to syllable-initial and syllable-final consonant clusters. Up to three consonants is possible in both positions; in some cases up to five subsequent consonants are possible in syllable-final position. It is well known that the ability of articulating such consonant clusters degrades rather rapidly when no feedback is provided to the speaker via her/his auditory system. Special exercises on higher levels (with words and phrases) must thus be provided to train (or to preserve) the articulation ability for such sequences.

Front-end and speech analyser

The front-end performs three processes to the signal. The sampling frequency is 20 kHz. The three processes are the cochlear transform, the ambient noise estimation and the speech excitation determination. See [3].

The output of the front-end is transferred to the speech analyser that calculates a set of acoustic cues. Then it performs an extraction of phonetic parameters of the signal [3]:

Considerations about the visual feedback

To allow an efficient visual counter-reaction, it seems necessary to be as constant as possible in the representation of the parameters.

- Time always unfolds from left to right, if it is represented on a horizontal axis.
- The intensity is represented with a filling factor, that is the thickness of a line or the diameter of a circle.
- The fundamental frequency is represented as a point on a value scale on a vertical axis.
- A particular colour is attributed to an acoustic or phonetic characteristic, for example: breath in white, voice in red. For the same reasons, a model is always in dark grey. An activated model or an activated word to be pronounced will blink. The colour of success could be golden yellow...

Moreover, visual feedback has to be adapted according different factors:

- The type of trainee: adults or children, sign language or written language.
- The training situation: suprasegmental level or articulation level,
- The feedback mode: on line or off line,
- The necessity to give appreciation both about local aspects (quality of a sound) and global ones (intelligibility of the entire utterance), for example.

The development of visual displays must take into account all these factors.

Hardware and software architectures

ISAEUS is implemented as an end user application, to run on PC and Windows 9x Operating System family. The architecture of the system is designed to achieve the following goals:

- Real time implementation, the high quality ISAEUS approach is based upon a huge computational burden.
- Inexpensive solution. Including the fact that the deaf community does not update their computer frequently
- Easily update of software through a web server.

To deal with the two first points we have adopted two solutions:

- Early 97 PC 's, this one is based on an inexpensive DSP board to supplement the PC, designed for the project (TMS320C31@60MHz, 512Mw, ISA bus).
- For new generation computer (bought after 2000) a full host running version (CD + Sound board) will be available.

To allow an interactive design and easily update of the exercises ISAEUS software is a Win32, ActiveX control container, multimedia enabled, single document interface application.

Evaluation of the effectiveness of ISAEUS

An exercise must be the answer to an expectation from the specialists (specialised teachers, speech therapists) and must take place inside a global approach for speech re-education [1]. Thus, its design involves a committee composed of specialists, phoneticians and engineers. This committee takes into account the accuracy of the parameters provided by the speech processing modules, the real-time constraints

that may occur, ..., and proposes an instance of the frame described before.

The exercise is then developed by the engineers and is submitted to a validation step. Validation must address several aspects, especially the adequacy of the exercise to what was expected, its integration inside the existing exercises, and its acceptability by the users. Some basic principles must be checked:

- visual display and other feedback must help to perform the exercise when it needs real-time conditions;
- visual display and other feedback must help to understand the quality of the performance;
- performance analysis must use realistic criteria;
- recommendations at the end of the exercise must be linked, as much as possible, to articulatory advice, both motor and kinesthetic, etc.

The validation phase is performed according to a protocol established by the committee. It is based on questionnaires and observation grids, together with a subjective appreciation of the trainee's abilities. The major points concern height, intonation, timbre, intensity, breath, speech rate, and articulation.

Conclusions

As we explain the ISAEUS system is based on the SIRENE II, a well proven tool for training deaf persons, but includes more features of the speech analyser for each lesson and aims to develop lessons at the levels not included in the previous system. In other words our propose is to process more speech dimensions and also try to achieve higher levels (as defined in the paper). From the beginning of the project speech therapist and deaf persons maintain a strong relationship with the other ISAEUS partners. From this fruitful relationship we produces a definition of the system with some novelty characteristics that are not in other available systems. We envisage evaluating the system during the first half of the year 2000.

References

- [1] M.-C. Haton "Issues in Using Models for Self Evaluation and Correction of Speech" in *Computational Models of Speech pattern Processing*, K. Ponting, editor, Springer, 1999.
- [2] ISAEUS *deliverable* D2.1.1: User needs.
- [3] ISAEUS *deliverable* D3.1.1: Front-end Processing.