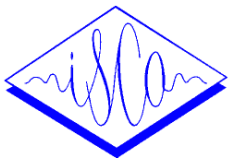
ISCA Archive
http://www.isca-speech.org/archive



3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

HARP: AN AUTONOMOUS SPEECH REHABILITATION SYSTEM FOR HEARING-IMPAIRED PEOPLE

Edmund Rooney[1], Fabrizio Carraro[1], Will Dempsey[1], Katie Robertson[1], Rebecca Vaughan[1], Mervyn Jack[1] and Jonathan Murray[2]

[1]Department of Electrical Engineering, University of Edinburgh,
80, South Bridge, Edinburgh EH1 1HN, Scotland;
[2]Department of Otolaryngology, Royal Infirmary of Edinburgh, Scotland

ABSTRACT

This paper describes the implementation and evaluation of the HARP system, a PC-based visual speech training aid for hearing-impaired people. The system offers teaching modules for a range of speech parameters, including intonation, rhythm, vowel quality and consonant production, and provides evaluative and diagnostic feedback to users. The development of the system is being assisted by consultation with hearing-impaired groups and speech and hearing therapists, and initial evaluations of a prototype aid have been completed.

1 INTRODUCTION

HARP is a two-year project funded by the European Community's TIDE programme (Technology Initiative for Disabled and Elderly people), with the aim of developing a speech rehabilitation system for hearing-impaired people. The system, based on an IBM-PC compatible microcomputer, is intended to provide visual feedback to assist speech training for the deaf. The system will be autonomous - that is, capable of being used without supervision by a therapist - and will be used to extend the services offered by therapists and teachers of the deaf in hospitals, clinics, schools and the home.

The HARP system development is exploiting technology built for teaching pronunciation to foreign language students. This technology, the SPELL system [2], provides visual feedback on a range of pronunciation features using an IBM-PC compatible, but is not specifically developed for use by those with limited auditory feedback.

The HARP project consortium consists of one academic partner and two industrial partners from Britain and France. The partners are:

- the Department of Electrical Engineering, the University of Edinburgh, Scotland
- Future Speech Systems (FSS) of Lanark, Scotland
- Agora Conseil of Grenoble, France.

The consortium is working closely with speech and hearing therapists, ENT specialists and teachers of the deaf from a number of institutions, to ensure that the design of the system meets the requirements of users, and that full evaluations are conducted throughout its development. These institutions include:

- the Department of Otolaryngology, Royal Infirmary of Edinburgh;
- Donaldson's School for the Deaf, Edinburgh;
- and the Institut des Jeunes Sourds, Chambéry, France.

The system is being developed to provide instruction for deaf speakers of both English and French in the first instance, though it is intended that the system will be modular enough to allow teaching materials in other languages to be added at a later stage.

2 DESIGN SPECIFICATIONS FOR THE HARP SYSTEM

2.1 Target groups

The HARP system is being developed with four main groups of hearing-impaired speakers in view, each of which shows rather different speech characteristics:

- *pre-lingually* deaf children and adults;
- *post-lingually* deaf children and adults;
- *elderly* deaf speakers;
- and *cochlear implant* patients.

Pre-lingually deaf speakers - that is, those with congenital hearing loss, or those who have suffered a hearing-impairment early in infancy, before the development of language - very rarely have normal speech, since they have never received the auditory feedback which would allow them to learn the correct articulatory patterns during language acquisition. Their speech is typically highly distorted as a result, with problems in pitch control, loudness, voicing, segmental timing, vowel quality and consonant articulation [5].

Post-lingually deafened speakers comprise those whose hearing loss has occurred after the onset of normal speech. Their speech impairment may be less serious than that shown by pre-lingually deaf speakers, but it can still result in intelligibility problems for listeners.

Deafness in the elderly is a common problem, and one which can not always be rectified by the provision of hearing aids. Elderly deaf speakers typically have problems with speech loudness and timing, again because they lack the feedback with which to monitor their speech production.

Cochlear implant patients form a very interesting group, since they have access to a form of auditory feedback, but need a considerable amount of therapy in order to interpret this feedback and relate it to their own speech production. The provision of a visual aid, capable of autonomous operation, could help enormously in this area.

Each group has rather different needs, and attempts are being made to take these differences into account in the provision of analysis modules and in the design of the user interface of the HARP system.

2.2 Autonomy of operation

The HARP system is intended to be autonomous, with users being able to train on the system without constant supervision by therapists. The ability of speakers to use the system without supervision frees the therapist to perform other types of work which cannot be done by a computer system, and allows a greater number of clients to be dealt with than would otherwise be the case. In addition, the speakers themselves benefit because they can have extra time to practise what has been taught inside the one-to-one therapy sessions: this increase in access to therapy is believed to be important in improving the "carryover" from such lessons into the speaker's everyday speech [1].

A truly autonomous system is required to be *evaluative*, providing not just raw visual feedback on the speaker's performance, but also some form of judgment of its quality, and diagnostic information which will help the speaker to correct their errors. The SPELL system on which the HARP development is based uses a range of similarity metrics to allow it to offer such evaluative and diagnostic feedback in all its existing analysis modules.

Autonomous systems must also be extremely accurate in their handling of errors, because the therapist cannot be there to identify and correct inadequate productions when the system itself lets them pass. This is a major problem for many systems, and is one of the main criticisms levelled at systems such as the IBM SpeechViewer. Particular attention will therefore be paid to strategies for dealing with errors in the development of HARP.

2.3 Acoustic versus physiological measures

One of the major decisions to be taken in the design of a speech training aid is on the nature of the input to the system. Acoustic systems, which form the majority, base all their measurements on the speech waveform and require only a microphone input, while physiologically-based systems (such as [3]) provide parameters which relate directly to the processes of speech production using a range of techniques such as electropalatography, electrolaryngography and airflow measurement. Physiologically-based systems have some advantages in the analysis of certain speech parameters, such as nasalisation, consonant production and laryngeal quality, but these advantages can be outweighed in practical use by the complexity, expense and relative immobility of the resulting hardware. The HARP system is therefore being developed using only acoustic inputs.

2.4 Courseware

Speech training systems vary widely in their provision of courseware to accompany the analysis and feedback software. Many provide only a series of independent teaching modules or games, often with little in the way of guidance as to how they can or should be used by the therapist. However, the success of computer-based training can depend critically on the conditions of use, and in particular on the integration of the system into a properly developed training curriculum [7]. The HARP system will therefore provide a range of courseware modules, starting with the acquisition of basic speech skills such as pitch and loudness control using isolated or sustained sounds, and progressing ultimately to the control of these parameters in connected speech.

3 HARP ANALYSIS SYSTEMS

The HARP system is making use of the technology developed within the SPELL project [2] for teaching foreign language pronunciation. The SPELL system performs analyses of intonation, rhythm, vowel quality and consonant production, and is capable of providing feedback on all four areas of speech, though not at present in real time. These capabilities are being enhanced and extended within the HARP project, to make the system suitable for the range of speech abilities found in the four hearing-impaired groups.

At the heart of the system is a Hidden Markov Model automatic segmenter [4], which has been tuned to the speech characteristics of the hearing-impaired. This is used to locate selected events within the user's speech input, and is also capable of detecting a number of pronunciation errors, using knowledge of the typical error patterns produced by speakers. The use of a segmenter to monitor the segmental content of the input allows the system to be fully evaluative and diagnostic.

3.1 Intonation

Intonation teaching in the HARP system depends on a fundamental frequency analysis performed by a time-domain pitch tracker, with considerable post-processing and interpolation to give smooth pitch contours. The fundamental frequency measures are normalised across speakers by the use of data on the speaker's long-term fundamental frequency behaviour (level and range) obtained during a short enrolment session. The pitch contour is aligned with a segmentation of the user's utterance, derived by the Hidden Markov Model segmenter. A decision metric then checks to see whether the user's contour shows the desired intonational features in the correct segmental locations, and provides feedback to show which parts of the contour are incorrect.

3.2 Rhythm

The rhythm analysis module also uses the HMM segmenter as the basis of its decision metric. Errors in the marking of syllables as strong or weak are detected by assessing the vowel quality and the duration of each syllable in the user's utterance. Feedback on the rhythmic status of each syllable is shown graphically, using a series of large or small blocks to represent strong and weak syllables.

3.3 Vowel analysis

The vowel analysis uses a method of cross-speaker normalisation [6] to obtain a two-dimensional representation of vowel quality from the first two formants and the fundamental frequency. These two acoustic dimensions (F2-F1 and F1-F0 differences, calculated on the Bark scale) correspond approximately to the articulatory dimensions of tongue frontness-backness and tongue height respectively, allowing feedback to the user to be presented with a simple pseudo-articulatory display which shows location of the speaker's vowel in relation to a set of vowel targets spread around the vowel articulatory area.

3.4 Consonant production

The analysis of consonants also makes use of the HMM segmenter, this time to perform categorical decisions on the quality of consonantal articulations. Models representing the desired target articulation and the typical error productions found in users' speech are made available to the segmenter for labelling the user's attempt. The decision of the segmenter as to which segment has been produced is used to provide categorical feedback, typically in the form of contrasting images showing the difference in word identity carried by the difference in sound.

4 PRELIMINARY EVALUATIONS

In order to involve future potential users of the HARP system in its development from the earliest possible stage, the initial specification of the system has been carried out in consultation with hearing-impaired speakers, speech therapists and teachers of the deaf in Britain and France. As part of this process, a series of preliminary evaluations of the existing system was held, in order to obtain reactions to the content, presentation and effectiveness of the available teaching modules.

These evaluations were held at several of the institutions collaborating with the project consortium, and a total of 11 hearing-impaired speakers (6 adults and 5 children) and 14 teachers and therapists took part. The evaluations involved demonstrations of the existing software, a "hands-on" trial by the hearing-impaired users, and completion of a comprehensive questionnaire. Users were also given ample opportunity for open discussion of the system.

A number of practical problems with the use of the software emerged during these demonstrations. These included:

- speakers' attempts exceeding the length of the speech buffer provided, owing to a much lower speaking rate than normal;
- the insertion of inappropriate pauses into the speaking materials, causing the automatic end-point detection program to cut off recordings prematurely;

- difficulties in gauging the correct level of loudness required to operate the system, causing overloading by speaking too loudly, or failing to trigger the signal acquisition software by not speaking loudly enough;
- problems with the placement of the headset microphone used with the system;
- fundamental frequency abnormalities which made estimating the level and range of fundamental frequency very difficult for some speakers; these abnormalities also caused problems for the vowel analysis, which is dependent on local fundamental frequency values for calculating the normalised vowel parameters.

Users made a number of suggestions for improvements in the system, including:

- the provision of an Introductory Module, which would instruct users in the operation of the system and provide an opportunity for some initial practice;
- the incorporation of real-time feedback into some of the teaching modules, in particular those dealing with fundamental frequency and vowel quality;
- the ability to monitor errors on more than one parameter simultaneously, to prevent the acceptance of utterances which were distorted in ways other than those being specifically trained;
- the provision of more explicit articulatory information to enable speakers to modify their productions.

Finally, therapists and teachers emphasized many times the need for *flexibility* in a training system such as this. Speakers vary widely in their needs, in their motivation, and in their ability to achieve the targets presented to them, and a system which cannot accommodate such diversity is likely to frustrate and discourage users from attempting to improve the quality of their speech.

5 CONCLUSION

This paper has described the implementation of the HARP speech training system, which is being developed for hearing-impaired speakers. The system, which is acoustically based and runs on an IBM-PC compatible, is intended to provide for a wide range of hearing disabilities, allowing for autonomous operation within a structured and comprehensive training curriculum. Initial specification of the system has been carried out in full consultation with hearing-impaired users and therapists, and preliminary evaluations have already suggested new ways of enhancing its provision.

Work on the system at present is concentrated on the development of real-time feedback modules, and improvements to the user interface, to make the system more suitable for use with the hearing-impaired.

Acknowledgment: This work is supported by the European Community's TIDE initiative, contract 1060.

6 REFERENCES

[1] L.E. Bernstein, M.H. Goldstein and J.J. Mahshie. "Speech training aids for hearing-impaired individuals: I. Overview and aims." *Journal of Rehabilitation Research and Development 25*, 53-62, 1988.

[2] S. Hiller, E. Rooney, J. Laver and M. Jack. "SPELL: an automated system for computer-aided pronunciation teaching." *Speech Communication., 13*, 463-473, 1993

[3] H. Javkin, N. Antonanzas-Barroso, A. Das, D. Zerkle, Y. Yamada, N. Murata, H. Levitt and K. Youdelman. "A motivation-sustaining articulatory/acoustic speech training system for profoundly deaf children." *Proc. IEEE ICASSP-93*, 145-148, 1993.

[4] F. McInnes, F. Carraro, S.M. Hiller and E.J. Rooney. "Evaluation and optimisation of a segmenter for a PC-based pronunciation teaching system." *Proc. Institute of Acoustics, 14*, 109-116, 1992.

[5] M.J. Osberger and N.S. McGarr. "Speech production characteristics of the hearing impaired." In *Speech and Language: Advances in Basic Research and Practice 8*, ed. N.J. Lass, 221-283, 1982.

[6] A.K. Syrdal and H.S. Gopal. "A perceptual model of vowel recognition based on the auditory representation of American English vowels." *J. Acoust. Soc. Amer. 68*, 1465-1475, 1986.

[7] C.S. Watson and D. Kewley-Port. "Advances in computer-based speech training: aids for the profoundly hearing-impaired." *Volta Review 91*, 29-45, 1989.