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

E Rooney (1), F Carraro (1), W Dempsey (1), K Robertson (1), M Jack (1) & J Murray (2)

(1) Centre for Communication Interface Research, Department of Electrical Engineering, University of Edinburgh
(2) Department of Otolaryngology, Royal Infirmary of Edinburgh

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 project consortium comprises:

- the Centre for Communication Interface Research, Department of Electrical Engineering, 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, to ensure that the design of the system meets the requirements of users.

The HARP system development is exploiting technology built for teaching pronunciation to foreign language students. This technology, the SPELL system [5], provides visual feedback on features of intonation, rhythm, vowel quality and consonant production. These capabilities are being enhanced and extended within the HARP project, to make the system suitable for use by those with limited auditc.y feedback.

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:

Pre-lingually deaf speakers. Speakers 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 [11]. The HARP system will be aimed at both children and adult speakers in this category, with the greatest benefits expected for younger speakers [1,4].

Post-lingually deafened speakers. Speakers whose hearing loss has occurred after the onset of normal speech may have less serious impairment than that shown by pre-lingually deaf speakers, but it can still result in intelligibility problems for listeners. Again, the HARP system will be aimed at both children and adult speakers.

Elderly speakers. 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. Users of cochlear implants 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. Until recently, cochlear implant users were mostly adult speakers, but the numbers of children receiving implants are now increasing.

The different needs of each group will determine not only the provision of analysis modules but also the design of the user interface of the HARP system. Younger children and the elderly, for example, may lack the manual dexterity required to operate some computer systems; children also have less well developed language capabilities and a very restricted vocabulary, which limits the type of language used in both practice materials and instructions.

2.2 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 for input. This makes them relatively simple, cheap and non-invasive, and therefore a more practical option for many users: acoustic systems have in general been the main choice for commercially developed systems in the past [2]. The main drawback of using solely acoustic input is that a number of speech parameters such as nasalisation, consonant production and laryngeal quality are not easily obtained from the acoustic speech waveform: systems which desire to provide feedback on these parameters must either employ fairly complex and sophisticated analysis techniques, or supplement the acoustic information avilable elsewhere with one or more physiological techniques. Physiologically-based systems such as the Panasonic aid [6] overcome some of these problems by providing parameters which relate directly to the processes of speech production, using a range of techniques such as electropalatography, electrolaryngography and airflow measurement.

The disadvantages of physiologically-based systems are considerable, however. The equipment required to obtain the measurements is often delicate, expensive and difficult to adjust, rendering home use difficult if not impossible. In addition, some of the techniques are relatively invasive when compared with acoustic measurement, and there is resistance to their use from both clients and therapists. A final problem is that the measurements obtained may not always relate directly to the acoustic signal produced by the speaker, and assessment of the acceptability of the client's speech may therefore be difficult. For ease of use and maximum accessibility, therefore, the HARP system is being developed using only acoustic input.

2.3 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; this prospect is one of the main attractions of computer-based systems for speech and language pathology professionals, who in many cases are overloaded, and unable to spend the time required to perform routine practice drills with their clients.. 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 [3]. A third advantage is that autonomous computer-based systems can be used in the client's own home with their families: this is an important factor in stimulating the family's interest, and in increasing the speaker's motivation to improve their speech quality [8].

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. This implies the use of a norm or model of pronunciation, and the implementation of a metric capable of comparing the speaker's pronunciation against this model. The choice of this norm and the nature of the metric are perhaps the most difficult aspects of the design to get right, since the judgment of the machine may well differ from that of a human therapist [16]. 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 may not 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 [12] and the older frication indicators [7]. A system which allows speakers to continue to make errors, or which in the worst case actually reinforces bad productions by providing a "reward", can be extremely damaging to a speaker's development. The HARP system is based around a hidden Markov Model segmenter [9], which allows the system to monitor the client's speech for a range of segmental errors.

A related difficulty is the question of how to handle errors in speech features other than the specific one being practised at the time. Many systems pay attention to only one speech feature at a time, and are unable to monitor errors which develop in other aspects of speech production in the way that a human therapist would. Nickerson, Stevens and Rollins [10], who observed this phenomenon in evaluations of the Bolt, Beranek and Newman system, suggested that periodic evaluation sessions with the therapist should be held after every few training sessions to keep it in check. The HARP system will attempt to overcome this problem by giving the user access to several levels of feedback about their speech, though care will be taken to ensure that such information on the user's errors is presented in a way which does not discourage further attempts or undermine the speaker's confidence.

2.4 Courseware

A second requirement for autonomous operation is the existence of a body of motivating courseware which allows users to learn from the feedback they receive. The system should maintain their interest and avoid the frustration caused by the use of repetitive or limited materials, or insufficient feedback.

Existing speech training systems vary widely in their provision of courseware. 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 [15]. 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.

The lessons proposed for the HARP Courseware can be grouped into three major units, arranged in increasing order of sophistication:

Demonstration modules. The demonstration modules will provide the speaker's first experience of using the system, beginning with a tutorial how the system works and how it is operated. Issues such as the choice of microphone, microphone placement, navigating through the lessons using the keyboard, keypad or mouse, and the interpretation of some of the displays will also be dealt with. To familiarise them with the microphone, the speaker will receive simple feedback on the difference between speech and silence, for example, and on speech loudness, but there will be little explicit teaching at this stage, and judgments on the quality of the speaker's pronunciation will *not* be made. Rather, the aim will be to stimulate voice use and encourage the speaker to explore further.

Introductory modules. The introductory modules will allow the user to learn control of basic speech parameters such as pitch, loudness, duration, voicing and segmental quality. It is expected that the provision of *real-time* feedback will be most beneficial to users at this stage. Lessons will be used to build speakers' awareness of the problems in their speech, and to increase their motivation by providing achievable targets. Many of the lessons will use isolated or sustained speech sounds, and the system will tolerate a fairly high degree of error in those aspects of speech which are not explicitly being taught in the lesson.

Advanced modules. The advanced modules will allow the user to learn and practise a much wider range of speech skills, including a variety of vowel contrasts, basic intonational uses of fundamental frequency, rhythm and consonant production. The targets presented to the user will use real words and phrases rather than isolated or sustained sounds. At this stage, the system will be able to offer more constructive feedback on the speaker's pronunciation, and it will begin to provide feedback on errors in other aspects of speech production.

3. HARP ANALYSIS SYSTEMS

The HARP system is making use of the technology developed within the SPELL project [5] 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 [9], which is 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.

In response to requests from therapists and hearing-impaired users during initial evaluations of the existing system [13], a number of analysis modules with real-time feedback are being developed. The first of these to be completed is a real-time fundamental frequency module, which uses a time domain pitch tracker (the Schäfer-Vincent algorithm [14]) to produce estimates of fundamental frequency; amplitude measures, which are also produced by the same algorithm, will be used within various loudness teaching modules. Work on a real-time vowel quality module is scheduled to begin shortly.

4. MULTIMEDIA INTERFACE

The HARP system is being designed for use by a wide range of users, from children to elderly speakers, with a wide range of linguistic and physical abilities. The user interface is therefore being designed to ensure maximum accessibility and choice.

4.1 Multi-modal input

The HARP system uses a variety of means to allow users to communicate with it. At present, users can control the system using the keyboard, the mouse and a touch-sensitive screen. Other possible input modes being investigated include the use of a dedicated keypad, with colour-coded buttons offering a limited choice in each display window.

4.2 Multi-media output

One of the key design features of the HARP system is its use of developments in multi-media technology to provide a flexible, accessible and highly motivating teaching tool. To make the system fully autonomous, help and instructions are available from the system itself at all stages of a lesson. This help information is provided using multi-media video: at present, users can choose to receive this information in the form of subtitled speech (Figure 1), suitable for those with lip-reading ability or some residual hearing, or in the form of signing (Figure 2), restricted for the present to British Sign Language (BSL). These information video sequences are captured and encoded in MPEG form using a Vitec Videomaker board at 12 frames per second; this frame rate is adequate for signing, but rather slow for playback of the lip and face movements required for lip-reading, and it is therefore hoped to upgrade this facility to 24 frames per second in the near future.

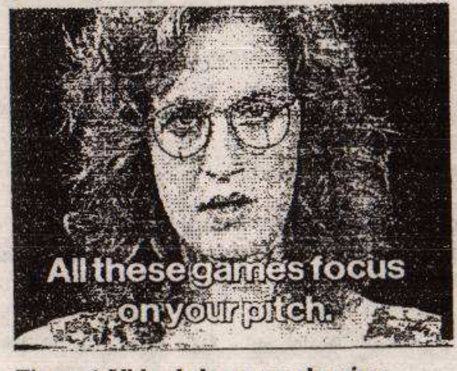

Figure 1 Video help screen showing subtitled speech

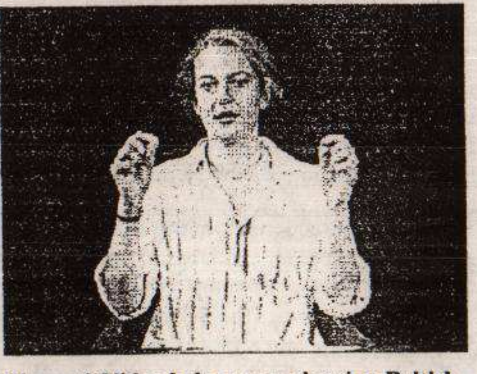

Figure 2 Video help screen showing British Sign Language

Multi-media video and computer animation are also used to provide feedback in a series of voice-controlled games. In the pitch module, for example, the user controls the playback of a set of short video sequences. In one set of games, the user's pitch controls the *speed* of playback of the sequence, high pitch corresponding to high playback speed. In another set, the user's pitch is correlated with *height*: in one sequence, for example, the user can "land" an airliner by lowering their pitch, or cause it to take off again by raising their pitch; while other sequences simulate the user rising or falling through space in synchrony with their rising or falling pitch. Other games use computer animation and graphics: users can fly an aeroplane over a series of obstacles, for example, or light up the bulbs on a Christmas tree to different levels.

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. Specification of the system is being carried out in full consultation with hearing-impaired users and therapists.

Work on the system at present is concentrated on the development of real-time feedback modules, and a flexible and accessible user interface using the latest developments in Multi-media technology. A series of user evaluations and trials is planned for second year of the project.

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

6 REFERENCES

[1] N Arends, D J Povel, E van Os, S Michielsen, J Claassen, and I Feiter, "An evaluation of the Visual Speech Apparatus", *Speech Communication*, vol. 10, pp. 405-414, 1991.

[2] J Bernstein, "Application of speech recognition technology in rehabilitation", in *Speech today and tomorrow: proceedings of a conference at Gallaudet University, September 1988*, ed. B.M. Virvan, pp. 181-187, Gallaudet University, Washington, 1989.

[3] 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.

[4] S Brooks, F Fallside, E Gulian, and P Hinds, "Teaching vowel articulation with the computer vowel trainer: Methodology and results", *British Journal of Audiology*, vol. 15, pp. 151-163, 1981.

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

[6] 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.

[7] R P Lippmann, "A review of research on speech training aids for the deaf", in *Speech and Language: advances in basic research and practice Vol 7*, ed. N.J. Lass, pp. 105-133, 1982.

[8] J J Mahshie, D Vari-Alquist, B Waddy-Smith, and L E Bernstein, "Speech training aids for hearing impaired individuals: III. Preliminary observations in the clinic and children's homes", *Journal of Rehabilitation Research and Development*, vol. 25, pp. 69-82, 1988.

[9] 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.

[10] R S Nickerson, K N Stevens, and A M Rollins, "The BBN computer-based system of speech training aids for the deaf: Current uses", Paper presented at the Research Conference on Speech Processing Aids for the Deaf, Washington D.C. May 1977.

[11] 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.

[12] S Pratt, A T Heintzelman, and S E Deming, "The efficacy of using the IBM Speech Viewer vowel accuracy module to treat young children with hearing impairment", *Journal of Speech and Hearing Research*, vol. 36, pp. 1063-1074, 1993.

[13] E Rooney, F Carraro, W Dempsey, K Robertson, R Vaughan, M Jack and J Murray, "HARP: an autonomous speech rehabilitation system for hearing-impaired people", *Proc. ICSLP 94, Yokohama*, 2019-2022, 1994.

[14] K Schäfer-Vincent. "Pitch period detection and chaining: method and evaluation". *Phonetica 40*, 177-202, 1983.

[15] 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.

[16] C S Watson, D J Reed, D Kewley-Port, and D Maki, "The Indiana speech training aid (ISTRA) I: Comparisons between human and computer-based evaluation of speech quality", *Journal of Speech and Hearing Research*, vol. 32, pp. 245-251, 1989.