Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

Presentation of a new EU project for speech therapy: OLP (Ortho-Logo-Paedia)

Öster, A-M.

journal: Proceedings of Fonetik, TMH-QPSR
volume: 44
number: 1
year: 2002
pages: 045-048

KTH Computer Science and Communication

http://www.speech.kth.se/qpsr

Presentation of a new EU project for speech therapy: OLP (Ortho-Logo-Paedia)

Anne-Marie Öster[1], David House[1], Athanassios Protopapas[2], and Athanassios Hatzis[3]
[1]Department of Speech, Music and Hearing, KTH, Stockholm, Sweden
[2]Institute for Language and Speech Processing, Athens, Greece
[3]Department of Computer Science, University of Sheffield

Abstract

This paper presents an overview of a newly started EU-funded research project and outlines the design of the speech therapy structure to be used within the project. The OLP (Ortho-Logo-Paedia) project aims at improving the quality of life of persons with articulatory impairments by applying a novel technological aid to speech therapy, by integrating this training with speech recognition technology and by making these facilities available over the Internet. The OLP project is an EU Quality of Life and Management of Living Resources project, coordinated by the Institute for Language and Speech Processing, Athens, Greece, with participation from France (Arches), Greece (Altec S.A. and Logos Centre for Speech-Voice Pathology), Spain (Universidad Politécnica de Madrid), Sweden (KTH), and the UK (Sheffield University and Barnsley District General Hospital).

Introduction and overview

This project proposes to apply a method to supplement (not replace) speech therapy for specific disorders at the articulation level based on an integrated computer-based system together with automatic speech recognition and distance learning. The key features of this proposal are: (a) therapy is based on real-time audio-visual feedback of client's speech productions, while therapy sessions are designed for each speech disorder separately and tailored specifically for each client; (b) speech production evaluation and interfacing to assistive technology is provided through automatic speech recognition based on statistical models of the data collected in the training/therapy phase; (c) web services provide remote collaboration and data collection for analysis and evaluation in diverse conditions.

These features correspond to the major system components:

1. OPTACIA developed from the Optical Logo-Therapy OLT, (Hatzis, 1999; Hatzis & Green, 2001) will visualise vocal tract configurations and trajectories through a tailored acoustic-to-articulation kinematic mapping in 2D or 3D. A therapist will be able to design a map, or select a predefined one, to suit an individual client. OPTACIA will thus provide the client with real-time visual feedback about her/his speech: articulator configuration will correspond to map position and articulator movement will correspond to map trajectory. It will be possible to reuse data collected during therapy sessions to retrain the mapping.

2. GRIFOS will be a speaker-dependent, small-vocabulary, automatic speech recognition system. Training data will be gathered through OPTACIA sessions, supplementing existing databases of material from similar client cases. In early sessions, GRIFOS will serve to set appropriate thresholds in OPTACIA to control acceptability of a client's speech productions and to analyse quantitatively these productions in syllable and word context during speech therapy. In later stages, it will primarily serve to evaluate client productions in continuous speech, with training material taken from OPTACIA sessions. For clients with severe articulation problems (e.g., dysarthria) unlikely to be fully resolved by therapy, GRIFOS will help stabilise production by providing feedback to increase production consistency rather than intelligibility per se. This will allow reliable interfacing to synthetic speech output and other assistive technology devices. Only speech synthesis (from already developed and available systems) will be used to test this functionality in the present project.

3. TELEMACHOS will apply distance-learning principles based on web database technology to provide the system's remote tutoring and monitoring ability. Telemedicine technology will enable wide application of the system

and remote data collection and evaluation. It will facilitate remote speech therapy sessions as well as diagnostic and remedial information sharing.

An overview of OLP functionality is illustrated in Figure 1 for the on-site case.

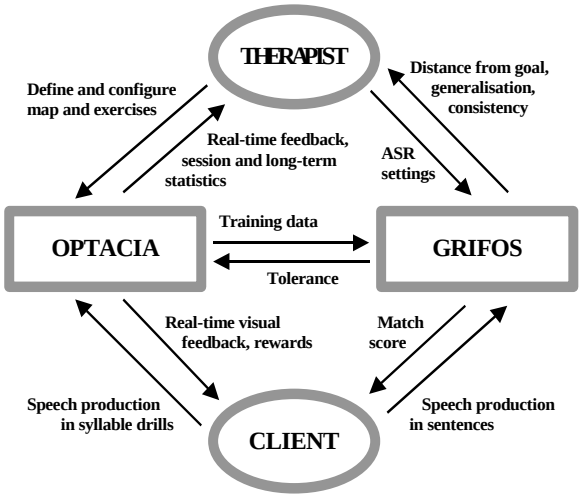

Figure 1. OLP system information flow in on-site application

Client populations

The term "articulatory impairment" encompasses a wide range of populations with diverse needs, their common point of reference being the imperative to improve their speech production skills through specialised therapy. The articulatory profiles of these subgroups and the therapy required in each case (both content and method) differ substantially but they all share the requirement for speech therapy.

It is important to define the groups of clients with speech disorders causing problems at the articulation level that are going to be judged suitable for treatment and rehabilitation from computer-based speech training systems. Diagnostic, acoustic and functional assessments are used to ensure a reliable description of the population. Partners in different countries primarily have access to patients with different types of pathologies. This will provide the project with a natural division of client groups into languages and pathologies thus enabling each clinical partner to concentrate on one particular pathology. Experiences from each language and pathology will then be collated to ensure that each specific approach can be modified according to language and pathology. Four basic types of pathologies will be addressed in the OLP project. These are dysarthria, craniofacial disorders, functional articulation disorders and hearing impairments.

Dysarthria is the most common acquired speech disorder affecting 170 per 100,000 population. It is a speech disorder resulting from impairment of the related central and/or peripheral nervous system or the oro-facial musculature. For those people with severe dysarthria, whose speech intelligibility is unlikely to be remediable by therapy, the OLP system will aim to stabilise speech production and make speech recognition and speech translation accessible.

The most frequent structural congenital craniofacia anomaly is cleft lip and palate, which occurs in approximately 1 per 700 births, presenting articulatory defects and velopharyngeal insufficiency.

The term functional articulation disorders refers to disorders that are not caused by identifiable abnormalities of an organ of the body. Special cases include consistent occurrence of misarticulation in particular phonological contexts only.

Speech therapy with prelingually deaf children aims at teaching them new speech patterns and at developing intelligible speech. An estimated 10% of the population has a hearing impairment, 2.5% of those being between 0–15 years of age.

Design of therapy

The expected effect of speech therapy is to establish automaticity and to transfer skills to untrained situations with the ultimate goal to expand the speaker's phonological system and to make the learner's best production be his or her most common production. Training exercises will be designed so that the training will be based on existing skills. Individual differences will determine the particular training material and order selected for training in a sensory-motor approach. Training exercises will be designed on different levels (e.g. segment, word, phrase, sentence) so that training on the segmental level will be expanded to the word level with the same phonemes. Training on the word level will then be expanded to the phrase and sentence level resulting in an expansion of the training of the same phonemes and words.

An example of this is training which starts with isolated speech sounds followed by repeated syllables, repeated alternated syllables with different stress patterns (to strengthen the client's mastery of breath control, intensity,

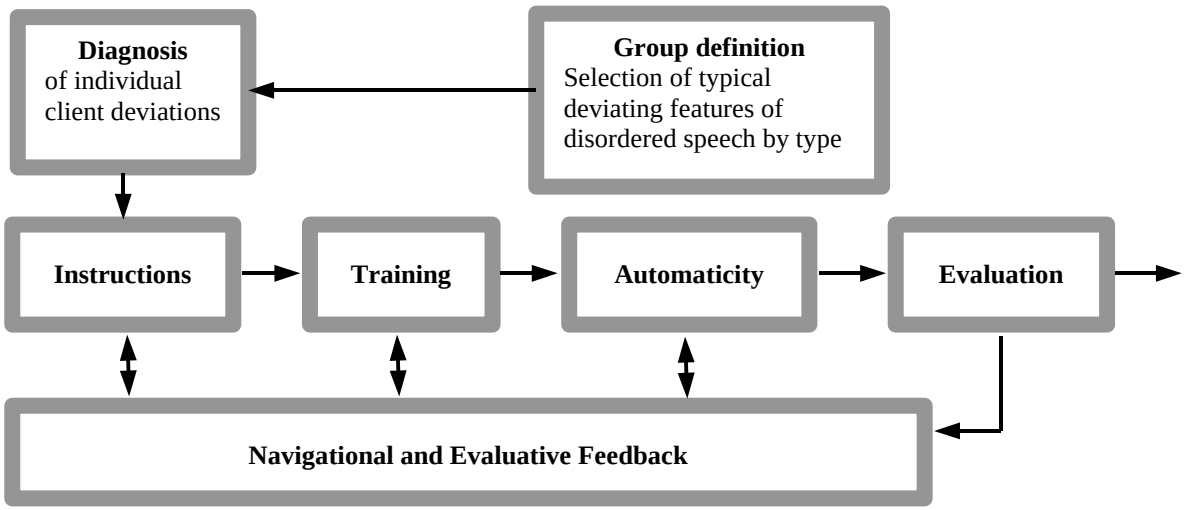

Figure 2. OLP therapy design components and information flow

pitch, duration, and voice quality), words containing the speech sounds of current interest and finally a short phrase containing the topical words.

Assessments should be made at every stage of this learning process through specially designed navigational and evaluative feedback. A training session should not exceed 30 minutes to be efficient. A complete training-period should contain about 10 sessions before the final evaluation. This evaluation should be done by the therapist by comparing a special recorded text material that the client reads before and after training. When it comes to small children who cannot yet read, this evaluative material must be based on an illustrative material. Figure 2 displays a diagram illustrating these principles.

Visual feedback

Previously many technical aids have been developed that used visual information as an alternative feedback for speech and hearing impaired children as well as in spoken L2-training (Gårding & Bannert, 1979; McAllister, 1986; Flege, 1989).

In early systems correct behaviour was indicated by a special lamp or by the deflection of a meter needle (Risberg, 1968). Some systems used oscilloscopes to reveal more detailed information of the acoustic patterns of speech (Fourcin & Abberton, 1971). Unfortunately, this type of technical aid was never evaluated within a pedagogical programme. Despite the fact that many of these aids have been reported to improve the speech of some children, the use of them was limited. This was probably due to the fact that the visual feedback provided by this type of technical speech training aid was difficult to understand, unnatural, unattractive, and had no motivational impact on the children.

A significant contribution was made by Nickerson & Stevens (1976) when they developed the first computer-based speech training system. Since then, the development of the micro-computer technique has offered more advanced computer-aided speech training programs that have enhanced the possibility for speech and hearing impaired persons to improve their pronunciation (Osberger et al., 1981; Yamada & Murata, 1991; Arends, 1993; Levitt, 1993; Javkin, 1994; Rooney et al., 1994; Öster, 1996, 1999). Most of these systems contain a microphone, an amplifier, and a speaker connected to a sound-card that allows the user to input, store and analyse speech and then display it and play it back. The software often contains several interactive programs that have been shown to be successful in assisting speech and hearing impaired children in achieving awareness and control over various speech attributes such as voicing, timing, pitch, and loudness as well as refining articulation and prosody.

Results and experiences from computer-based audiovisual speech training have shown that in order to be efficient the aid should provide a contrastive visual training, that is, the correct model of a reference speaker and the deviant production of the client should be shown simultaneously to be compared with each other. Moreover, the visual pattern must be attractive,

easily comprehensible and shown without delay. These findings will serve as a basis for the development of the OLP interface. These considerations will play a key role to the success and usability of the system.

Range of speech disorders and potential for therapy

The span of articulatory disorders represented by the client populations will allow for appropriate evaluation and validation of the materials and the OLP system. Speech therapy for speakers with dysarthria aims at stabilising production at syllable or word level to improve the consistency of production. For some speakers this will lead to greater intelligibility and may approach normal speech, while for others consistency of pronunciation will enable them to better use speech technology applications which are dependent on speech recognition and therefore require a consistent pronunciation. For prelingually and severely hearing impaired persons the aim of speech therapy is to teach them new speech patterns and increase the intelligibility of their speech. For children with cleft lip and palate and velopharyngeal incompetence the aim is to eliminate misarticulated speech patterns so that most of these speakers can achieve highly intelligible, normal speech patterns.

The variation between these population groups will be well suited for developing and testing various aspects of the OLP system including the different aims of the therapy and different dependencies on speech recognition, visual feedback and repetitive speech motor training. We anticipate that the within group variation will also serve the development and testing of the OLP system in the same manner. Finally it is the expectation of the OLP consortium that persons with other types of articulatory impairments will be able to benefit from the materials developed and that the system can easily be adapted to other languages.

Acknowledgements

This project is funded by the European Commission, Fifth Framework Programme, Quality of Life and Management of Living Resources, contract QLG5-CT-2001-01971.

References

Arends N (1993) *The visual speech apparatus, IvD/RES/9303/01*, Instituut voor Doven, Sint Michielgestel, The Netherlands.

Flege J E (1989). Using visual information to train foreign language vowel production, *Language Learning 38*, 365-407.

Fourcin A J & Abberton E R M (1971). First application of a new laryngograph, *Medical and Biological Review, 21*, 172-182

Gårding E & Bannert R (1979) Optimering av svensk uttalsundervisning. *Praktisk Lingvistik 1*, 1-9. Lund: Department of Linguistics, Lund University.

Hatzis A (1999). *Optical Logo-Therapy (OLT), Computer-based audio-visual feedback using interactive visual displays for speech training.* Dept. of Computer Science, University of Sheffield, UK.

Hatzis A & Green P D (2001). A two dimensional kinematic mapping between speech and acoustics and vocal tract configurations. *Workshop on Innovation in Speech Processing,(WISP'01),* Stratford-upon-Avon, UK.

Javkin H (1994). A new Speech Training System: Acoustic/Articulatory Data, Video Games and Synthesized Model Parameters. *2nd Int. Symposium on Speech and Hearing Sciences,*81-92. Osaka, Japan.

Levitt H (1993). The impact of technology on speech rehabilitation. *Proceedings of an ESCA Workshop on Speech and Language Technology for Disabled Persons.* Stockholm, Sweden.

McAllister R (1986). Tekniska hjälpmedel i Uttalsundervisningen: en delrapport. *PU-rapport 1986:1*, Stockholms universitet, Pedagogiska utvecklingsenheten.

Nickerson R S, Kalikow D N & Stevens K N (1976). Computer-aided speech training for the deaf, *JSHD, 41*, 120-132.

Osberger M J, Moeller M P & Kroese J M (1981). Computer-assisted speech training for the hearing impaired, *J. of the Academy of Rehabilitative Audiologists, 14*, 145-158.

Risberg A (1968). Visual aids for speech correction, *American Annals of the Deaf, 113,* 178-194.

Rooney E, Carraro F, Dempsey W, Robertson K, Vaughan R, & Jack M (1994). HARP: an autonomous speech rehabilitation system for hearing-impaired people. In *Proc. 1994 International Conference on Spoken Language Processing (ICSLP94),* 2019-2022. Yokohama, Japan.

Yamada Y & Murata N (1991). Computer Integrated Speech Training Aid, *Intern. Symposium on Speech and Hearing Sciences*, Osaka, Japan.

Öster A-M (1996). Clinical applications of computer-based speech training for children with hearing impairment. *Proceedings of ICSPL 96,* 157-160. Philadelphia, USA.

Öster A-M, Vicsi K, Roach P, Kacic Z, & Barczikay P (1999). A multimedia multilingual teaching and training system for speech and hearing impaired children – SPECO, *Proceedings of Fonetik 99,* 149-152. Gothenburg, Sweden.