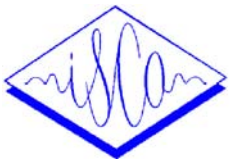

EUROSPEECH '97
5th European Conference on Speech Communication and Technology
Rhodes, Greece, September 22-25, 1997

Optical Logo-Therapy (OLT) : A Computer-Based Real Time Visual Feedback Application for Speech Training

A. Hatzis (1), P.D. Green (1), S.J. Howard (2)
(1) Speech and Hearing Group, Dept. of Computer Science, University of Sheffield,
Tel. +44-114-222 (1879) (1836), FAX. +44-114-278 0972, E-mail: a.hatzis@dcs.shef.ac.uk, p.green@dcs.shef.ac.uk
(2) Dept. of Human Communication Science, University of Sheffield,
Tel. +44-114-222 2448, FAX: +44-114-278 2403, E-mail: s.howard@sheffield.ac.uk

ABSTRACT

Traditional speech training methods can prove cumbersome because of the difficulty of providing the subject with good feedback, maintaining her/his motivation over long periods and stabilising the improvement in articulation. In this work we provide visual feedback based on displaying a trajectory in a 2-Dimensional 'phonetic space'. Data are presented from a small-scale efficacy study, which illustrate the use of OLT in speech therapy for misarticulated sibilant fricatives. Results for the contrasting articulations are compared and the potential of OLT as a therapeutic technique is discussed.

10.21437/Eurospeech.1997-302

1. INTRODUCTION

Today's Computer-Based-Speech-Training (CBST) systems provide feedback in various aspects of speech production such as voicing, intonation, and articulation with a number of different techniques. Although the feedback provided in various speech parameters of voicing such as timing, loudness and pitch is adequate, there are several limitations in the development and practice of articulation from such systems. OLT is a CBST system that focuses on the articulation of the subject and tries to combine kinaesthetic awareness with visual feedback.

Traditional methods such as electropalatography (EPG) [8] rely on the same principles but can only be used to investigate phones for which tongue-palate contact occurs, in many cases the tongue-palate proximity is not revealed. Furthermore EPG cannot be applied unless an artificial palate is built for each subject. On the other hand modern, acoustic measurement, CBST systems like Speech Viewer [9], VSA [7], ISTRA [10] and HARP [3] attempt to develop all the aspects of speech production and the monitoring of articulation is not so accurate. Moreover although the visual feedback provided in most of these systems is immediate they do not present the therapist with the exact location where the error occurs and the distance from the target is not clearly shown. Another important issue in many cases is that these systems cannot be individually tailored to cover different groups of speakers and or different pathological cases.

2. VISUALISATION OF SPEECH

It is a common technique of systems like the ones we referred to and other recent ones [4], [5], [6] to attempt to visualise the phonetic space or part of it, in two dimensions. The central idea is to project the acoustic signal so that the therapist and the subject can see the errors during the articulation of an utterance. This is the principle in the technique that is used to create an OLT phonetic map [1].

2.1 Creating an OLT Phonetic Map

There are three stages in the creation of an OLT phonetic map.

2.1.1 Learning Vector Quantisation (LVQ)

LVQ algorithms [2] are used to model the subset of phonetic space with a sufficient number of 9D reference vectors – 8 mel-scale frequency cepstral coefficients together with overall energy. These vectors represent all the speech data we have collected during the speech analysis.

2.1.2 Sammon mapping

A non-linear projection is then applied to reduce the 9D reference vectors to points in a 2D space. Sammon mapping [11] is based on fitting N points in the 2D space such that their inter-point distances approximate the corresponding inter-point distances in the 9D space.

2.1.3 Multi Layer Perceptron (MLP) mapping

Finally an MLP neural network is trained using the backpropagation algorithm to learn the non-linear relationship between 9D space patterns and 2D ones. In that way an incoming unknown input 9D vector is automatically transformed into a 2D one and it can be placed in any point on the map. The sequence of these consecutive points forms the speech trajectory of the acoustic signal.

2.2 The Interface and its Functionality

The OLT user interface currently deploys three windows:

- The map area where all the speech events occur.

- The control panel where one can modify several attributes of the speech trajectory, the kind of projection on the map, the recording and the playback of utterances and various others.
- The samples pool where pre-recorded, recorded utterances from the subjects are placed for testing and comparison on the map.

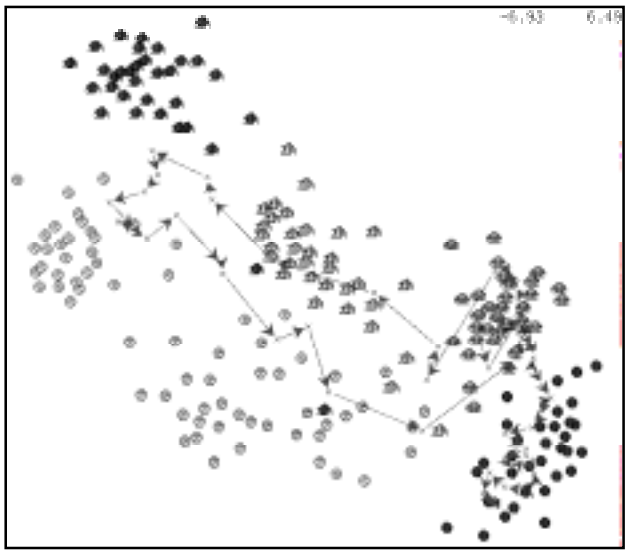

Figure 1 : Normal trajectory of /ee s u / with normal rate of speech.

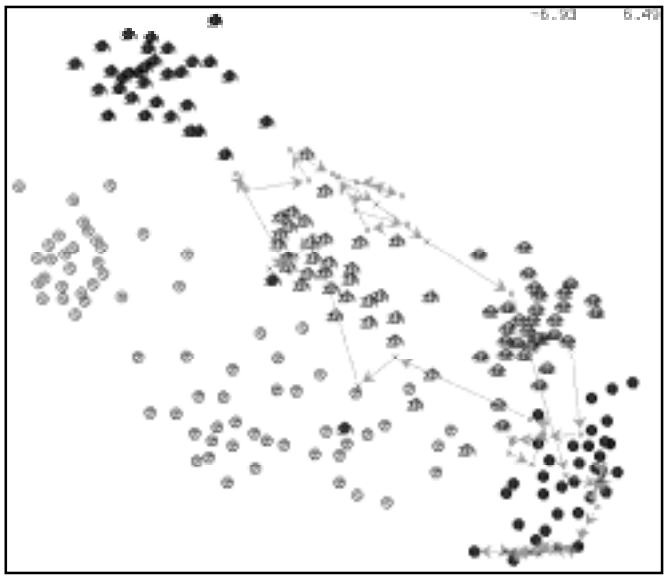

Figure 2: Abnormal trajectory of /ee s u/ with slow rate of speech.

The current version of OLT includes the real time recording routines for obtaining a real time visual feedback with the MLP mapping Fig [1] – [6].

3. EXPERIMENTAL RESULTS

We present results from two experiments, one with adults and one with children. Both of the experiments were aiming to evaluate the OLT phonetic maps by testing normal and abnormal trajectories and showing the differences of them. In both cases the system was applied to the same clinical problem, sibilant fricative lateralisation, but it was specialised to treat different subject groups, English male adults and English female children.

3.1 Adults Experiment

3.1.1 Collection of Speech Material

We constructed a map from six normal speakers, also male, adult and English Fig. [1]. To provide a fixed context for the fricative articulation, the utterances were spoken in a random sequence and were VCV utterances of the form /i X u/, where X is /s, sh, z, zh/. Each of the normal speakers recorded 11 repetitions for these sounds, 44 utterances in all. The data from four of the normal speakers were used to train the map and the rest were used to test the accuracy in classification.

3.1.2 Testing of the Map

Fig. [1], [2] show clearly the differences between the

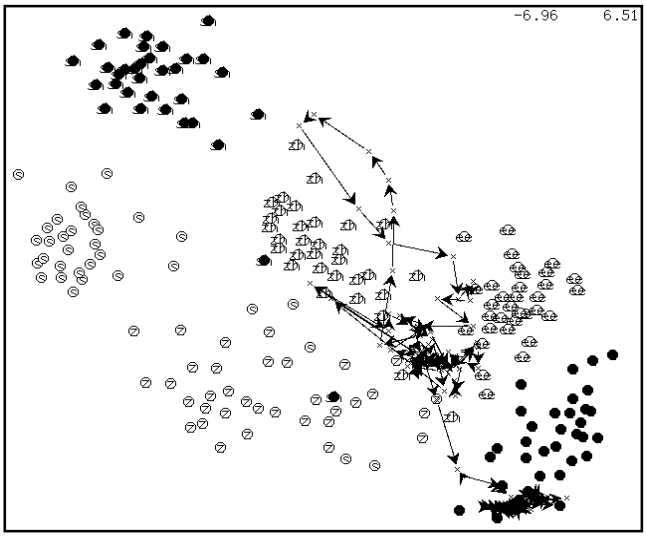

Figure 3: Abnormal trajectory of /ee s u / with normal rate of speech.

speech trajectory of a normal speaker and those of an abnormal subject. The impaired speaker is an adult English male. He was selected because a perceptual analysis of his speech suggested that all of the target sibilant fricatives /sh, s, z, zh/ were produced laterally,

rather than centrally. The VCV trajectories on the OLT are consistent with his 'lateralising problem". In particular he fails to reach the fricative target area of /s/ and moreover the quality of his speech deteriorates significantly when he tries to utter the same utterance in a slower rate Fig. [3].

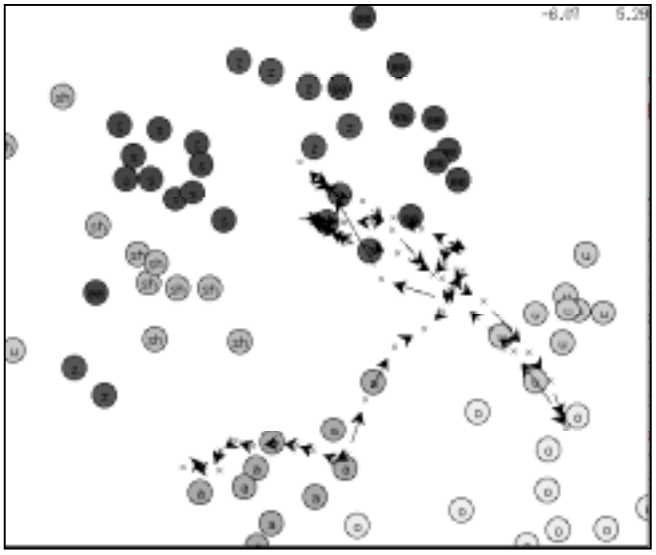

Figure 4: Normal speech trajectory - "a zoo"

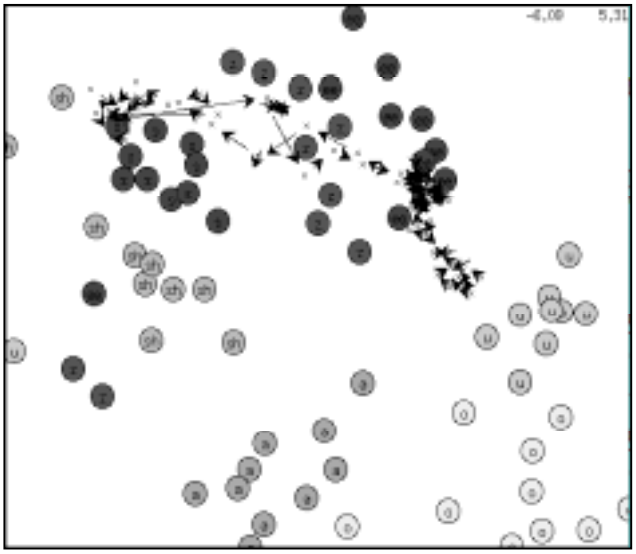

Figure 5: Abnormal speech trajectory - "a zoo"

3.2 Children Experiment

The collection of data from children is far more difficult than that of adults. Therefore for that purpose a special program was built to help with the recordings of children. In brief it was controlling a randomly generated sequence of simple pictures, which were displayed to the child together with a traffic light to control the start and end of the recording. Utterances were automatically assigned a unique identity tag and stored in the hard disk for future processing.

3.2.1 Collection of Speech Material

In order to find a suitable context for our sibilant fricatives and at the same time easily understandable by the children we chose the following utterances: "a sea, a zee, a sheep, a saw, a shore, a zorr, a zoo, a shoe, a suit". Thus we have a context on the form /A F V/ where F is /s, sh, z/ and V is /ee, o, u/. The plosive or stop sounds were removed. We recorded in total 18 normal children, 9 male and 9 female and each subject repeated each utterance 5 times. In addition, the subjects recorded isolated sounds of the above phones in order to contrast with the segmented speech. The initial vowel, consonant and final vowel segments of the utterance, were then labelled manually and mel-scale cepstral coefficients were taken every 10 msec. In this experiment we have taken the data from only the female speakers to create the map.

3.2.2 Testing of the Map

The OLT package was used with a seven years old girl with misarticulated sibilant fricatives, who consistently produced target alveolar and post-alveolar fricatives and affricates with lateral rather than central friction. Here we present a sample of speech from the articulated utterance "a zoo" Fig. [4]. The speech trajectory of the abnormal speech falls somewhere between the two areas of /s/ and /z/ during the transition and the frication Fig.

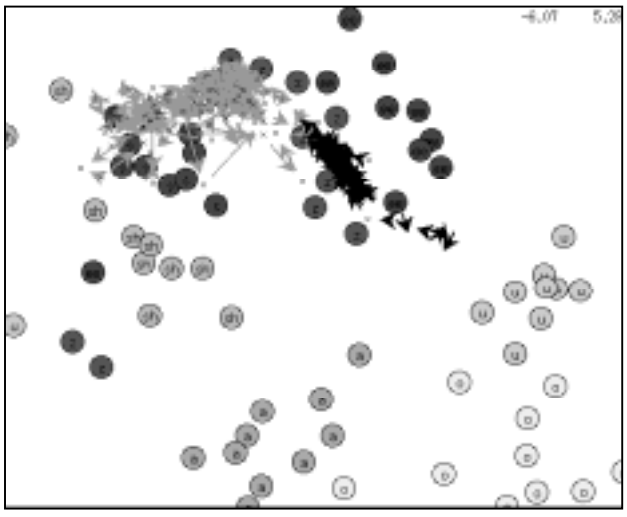

Figure 6: Normal vs Abnormal speech trajectory of the isolated sound of phoneme /z/.

[5]. Also it can be noticed that the signal is not starting from the area of /a/ because of the different quality of the phone pronounced. You are welcome to listen to the samples included with the CD version of the paper for a perceptual comparison. The abnormal frication of the /z/ sound can also be studied separately when comparing the normal and abnormal isolated speech trajectories. Again here we observe the failure of abnormal trajectory in reaching the target Fig. [6].

4. CONCLUSION

These are some observations during the testing of the OLT platform. In general both of the impaired subjects found the screen display easy to understand and they were able to relate the results of different articulatory configurations to the visual feedback they received from the screen. They enjoyed the visual feedback provided by the MLP mapping activity and they were clearly motivated by it to experiment according to the clinician's instructions.

A strength of the screen display used was that it contains a number of target phonemes, some of which (in particular the vowels) were produced normally by the abnormal speakers prior to therapy. This meant that there was always guaranteed some success during the therapy, which crucially helped to avoid the failure-avoidance behaviours encountered in the clinical context when a client has to focus on a speech behaviour which is clearly difficult for them. With OLT, activities can revolve around a number of sounds simultaneously, so that the child or the adult can be asked not only to produce sounds which are problematic for them, but also sounds which are easy for them. Thus it guarantees them some positive visual feedback, even in the initial stages of therapy, when consistent modification of the misarticulated sounds is still a source of difficulty.

4.1 Future Plans

OLT is still under an experimental stage. Our outmost goal is to provide a toolkit that will be customised on the studying of the articulation of a subject and providing suitable visual feedback.

We intend to address.

- Real time simple animated games based on the phonetic map.
- Other techniques including Neural Networks, dimensionality reduction, and classification for the creation of better phonetic maps.
- Personalised phonetic maps based on the speech disorder and potential improvement of the subject.

REFERENCES

[1] A.Hatzis, P.D. Green, S.Howard (1996), "Optical Logo-Therapy - (OLT). A computer based speech training system for the visualisation of articulation using connectionist techniques.", Proc. IOA. 18, 9, pp299-306.

[2] T. Kohonen, J. Hynninen, J. Kangas, J. Laaksonen, and Kari Torkkola. LVQ_PAK: The Learning Vector Quantisation Program Package. Technical report A30, Helsinki University, Faculty of Information Technology, Laboratory of Computer and Information Science, 1996

[3] E. Rooney, M. Jack, J. Lefevre and A. Sutherland. HARP-a speech training aid for the hearing impaired. In 2nd TIDE Congress, la Villette, Paris, April 1995.

[4] I. Nagayama, N. Akamatsu, T. Yoshino. Phonetic Visualisation for Speech Training System by Using Neural Network. In Proc. International Conference on spoken Language Processing (ICSLP94), pages 2027-2030, 1994

[5] V. Rodellar, V. Nieto, P.Gomez, D. Martinez and M. Perez. A Neural Network for Phonetically Decoding the Speech Trace. In Proc. International Conference on Spoken Language Processing (ICSLP94), pages 1575-1578, 1994.

[6] J. Reynolds and L. Tarassenko. Learning Pronunciation with the Visual Ear. Neural Computing and Applications, pages 169-175, 1993.

[7] N.Arends, D.J. Povel, S.Michielsen, J. Claassen, and I. Feiter. An evaluation of the visual speech apparatus (VSA). Speech Communication, 10:405-414, 1991.

[8] WJ. Hardcastle, FE. Gibbon, W.Jones. Visual display of tongue palate contact: Electropalatography in the assessment and remediation of speech disorders. Br J Disorders of Commun. 1991; 26: 41-74

[9] J. Ryalls. Comparison of two computerised speech training systems: Speech Viewer and ISTRA. Journal of Speech-Language Pathology and Audiology, 13(3):53-56, 1989

[10] D. Kewley-Port, C.S. Watson, and P.A. Cromer. The Indiana Speech Training Aid (ISTRA): A microcomputer-based aid using speaker-dependent speech recognition. In Synergy '87, The 1987 ASHF Computer Conference, Proceedings, pages 94-99, 1987.

[11] J.W. Sammon. A Nonlinear Mapping for Data Structure Analysis. IEEE Trans. on Computer, C-18, 5, pages 401-409, 1969.