

# OPTICAL LOGO-THERAPY (OLT) : VISUAL DISPLAYS IN PRACTICAL AUDITORY PHONETICS TEACHING.

A.Hatzis (1), P.D. Green (1), S.Howard (2)

(1) Dept. of Computer Science, University of Sheffield, E-mail : [a.hatzis@dcs.shef.ac.uk](mailto:a.hatzis@dcs.shef.ac.uk), [p.green@dcs.shef.ac.uk](mailto:p.green@dcs.shef.ac.uk)

(2) Dept. of Human Communication Science, University of Sheffield, E-mail : [s.howard@sheffield.ac.uk](mailto:s.howard@sheffield.ac.uk)

WWW : [www.dcs.shef.ac.uk/~nassos/OLT](http://www.dcs.shef.ac.uk/~nassos/OLT)

## 1. Introduction

Despite the development of many audiovisual computer-based tools for teaching phonetics, there is still a lack of software which is capable of delivering our increasing knowledge and understanding of the field. Consider for example

how to teach the differences between an alveolar fricative /s/ and a palato-alveolar fricative  $\int$ / including difficult cases such as the distinction between these sounds articulated by a native English speaker and a native Greek speaker. While an experienced phonetician will be able to describe the production of these sounds accurately and demonstrate the distinction to a student, it is highly likely that in order to grasp the information the student will need to imitate the articulatory gestures and receive some sort of feedback. Learning the new skill can be reinforced on a trial-and-error basis once evaluative feedback is provided and there is a clear understanding of the target production.

To illustrate, Figure 1 shows the ‘speech cycle’ as a relationship between the phonetic sciences concerned with production (articulatory phonetics), transmission (acoustic phonetics) and perception (auditory/perceptual phonetics) [2]. Typically the teacher will produce a sequence of sounds to illustrate the difference between one articulatory gesture and another. The sounds are transmitted and perceived by the student who has to establish a cognitive relationship between what is heard and articulator movement.

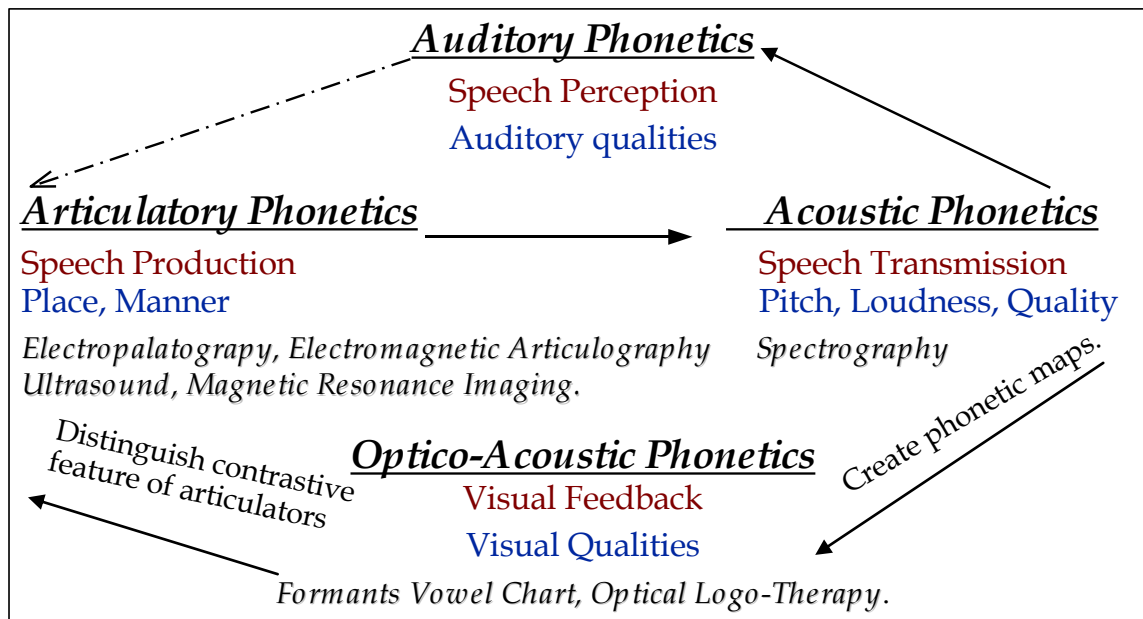


Figure 1 : *The speech cycle and the relationship with the Phonetics Science*

The way in which speech perception is related to articulator movement is a research area in which there is still much to be explored, particularly when the listener has a hearing impairment. The work reported here aims to provide additional help in navigating the speech cycle, by the provision of visual feedback. This new area might be called ‘optical-acoustic phonetics’. Our software system OLT provides this feedback, interactively and in real time.

## 2. Phonetic Maps

The lower part of Figure 1 shows the alternative route in the speech cycle provided by optical-acoustic phonetics. The first step is the creation of a ‘phonetic map’, a 2D display in which different sounds are mapped to different points. The idea of a phonetic map is not new and several attempts to produce such a display have been

reported. Perhaps the best-known is the formant vowel chart, which has been included in commercial speech training products such as the ‘Video Voice’ [1], and the ‘HARP’ [4]. A scheme based on Kohonen’s self-organising maps was reported by Reynolds and Tarassenko [3]. Earlier versions of OLT used Sammon mapping [5].

2D mapping based on formants requires reliable fundamental frequency and formant estimation, which is difficult to guarantee, and not suitable for all the consonants. Self organising maps and other techniques for visualisation of high-dimensional data on the other hand present problems for OLT because they do not give the teacher the freedom to design the map. For instance, the sound-class layout might be chosen to reflect a mid-sagittal view of the vocal organs. Furthermore, the map should not be cluttered and the sound clusters must not overlap or be too scattered. Finally, we need a way of rejecting sounds which do not sufficiently resemble the training data.

### 3. Phonetic Maps in OLT

The current version of OLT allows the teacher to create customised phonetic maps suited to the training problem and the needs of the student. The transformation from speech data to position on the phonetic map is performed by a multi-layer perceptron. The input to this MLP is 9-dimensional data vectors derived by cepstral analysis of manually-labeled utterances representative of the sounds which the map covers. Typically data from a few seconds of each target sound are used. Since the mapping will only be consistent for data similar to that used in training, it is necessary to automatically reject sounds which are not sufficiently similar to anything in the training set. The sensitivity of this rejection criterion can be controlled by the teacher.

For example, Figure 2 shows a map which includes targets for two vowels /i/ and /u/ and four fricatives /s/, /ʃ/, /z/ or /ʒ/. The training data was recorded from eight normal English male adult speakers producing the sounds in isolation. This display has been designed to teach Greek students how the English fricatives are articulated. It attempts to visualize the phonetic contrast between these sounds and provides real time audio-visual feedback for the student’s attempts. We have shown that OLT helps to make Greek speakers aware of the inter-lingual differences in fricative production, and allows them to reinforce their learning.

To make the feedback more captivating, particularly in another series of experiments with children, OLT provides real-time animation involving a cartoon aeroplane which flies between the targets. At any time when recording is live the system is in one of three states: the sound may be accepted for display, rejected or considered to be silence. These states correspond to a cartoon clown face which smiling, frowning or wearing a neutral expression. Evaluation of success can be displayed as a series of stars. This kind of display is shown in Figure 2. In these experiments, OLT was used in therapy for three children with fricative articulation problems, who had not responded to conventional treatment. Again, awareness and reinforcement resulted in consistent improvements in production.

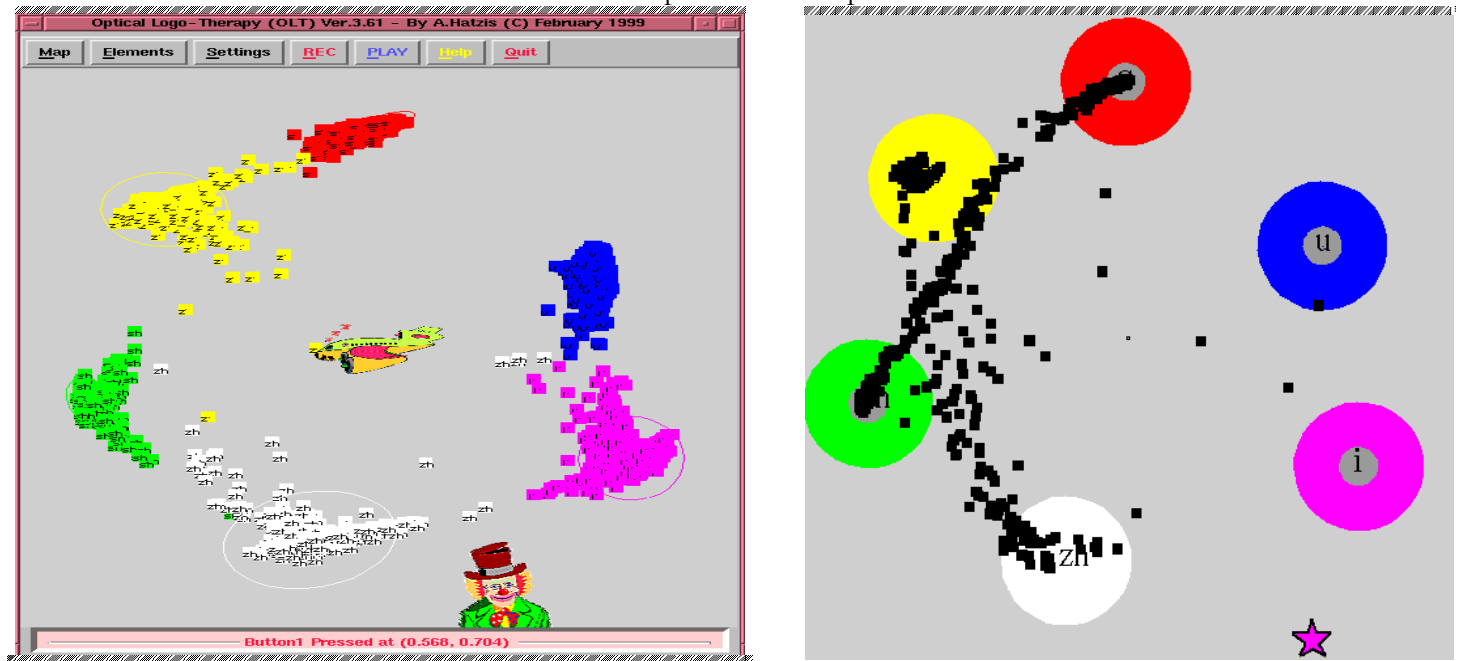


Figure 2 : *OLT phonetic map*

Figure 3 : *Tracing tongue retraction*

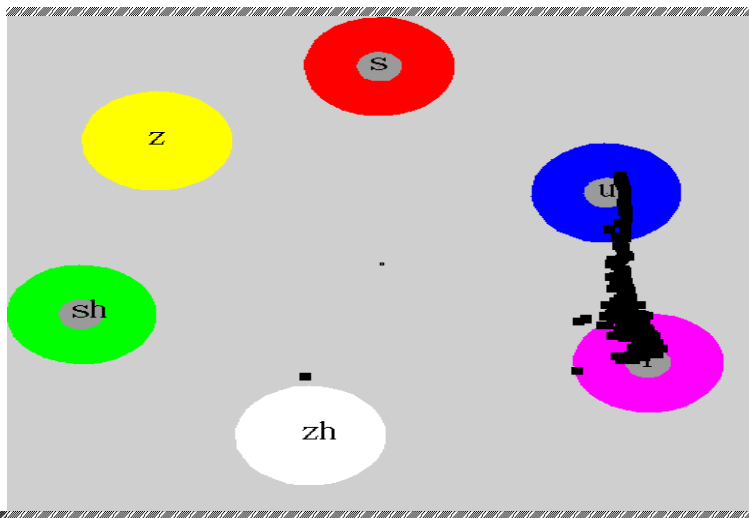


Figure 4 : *Tracing lip rounding*

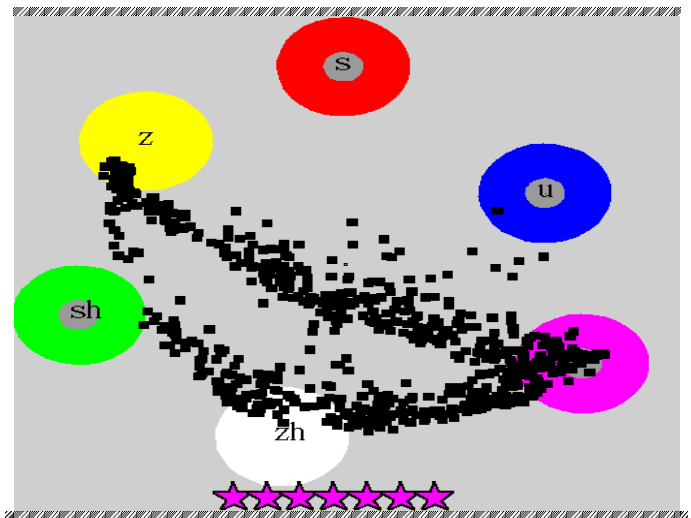


Figure 5 : *Tracing stricture*

#### 4. OLT displays and articulation contrasts

The quality of the speech training OLT facilitates depends on the system's ability to portray articulatory feature contrasts in a consistent way (Figures 3 to 5) and also make students visually aware of the difference between their past and present attempts to produce a certain sound (Figures 6 and 7). For instance

- **Tracing tongue retraction:** Figure 3 shows what happens when the speaker gradually retracts the tongue from a typical alveolar /s/ configuration to a palato-alveolar  $\text{ʃ}$  , and performs the same gesture but with voiced excitation. The sound is mapped into positions which move gradually from one center to the other.
- **Tracing lip rounding:** Figure 4 shows a similar progression between the /i/ and the /u/ vowel points achieved by varying lip rounding
- **Tracing Stricture:** In Figure 5 we experiment by narrowing the passage of the air flow and thus move from a vowel /i/ to a voiced fricative /z/ or  $\text{ʒ}$ .
- **Customising Phonetic Maps:** Figure 6 shows a map used in experiments with children and trained from normal speech. The child whose articulations are illustrated has a problem with production of /s/: instead of positioning the

tongue on the alveolar ridge he may palatalize or place the tongue by the lower incisors. Samples of these articulations are collected in training and then used to train a map with two new targets: S1 (palatal) and S2 (incisors) Figure 7. The child is provided with feedback which directly maps these variations.

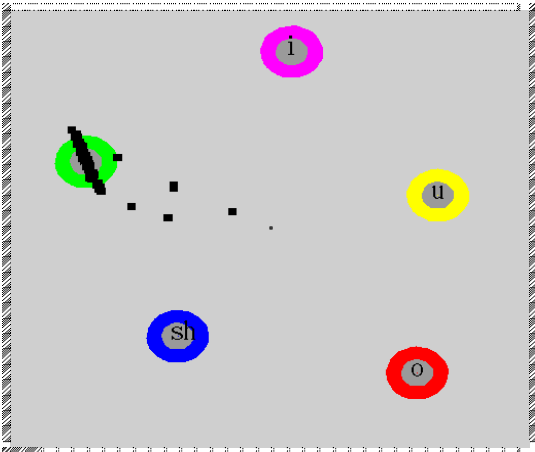


Figure 6 *Normal /s/ production on normal children's map with the blade of the tongue at alveolar ridge.*

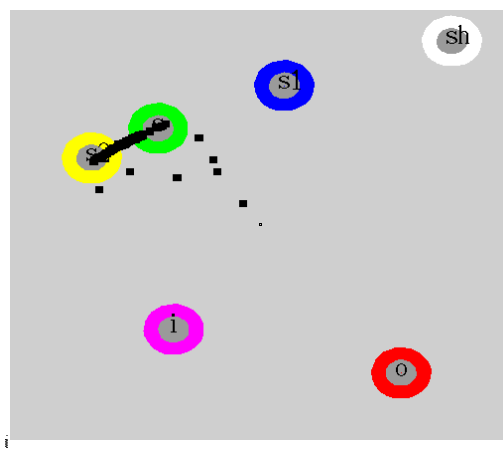


Figure 7 *Normal /s/ production on child's individualised map with the blade of the tongue at alveolar ridge.*

## 5. Future Plans

OLT has been applied in two difficult areas of phonetic teaching: speech therapy and accent modification, with promising results. In this paper we have outlined the phonetic aspects of OLT, concentrating on visualization of the relationship between articulation and acoustics.

In future work we aim to develop and extend the capability of OLT to portray this relationship, perhaps using a third dimension (since interdependent multi-dimensional changes can only be reduced to two dimensions by approximation). We also need to improve on the interface provided to build customised maps.

## Bibliography

- [1]. Fitzgerald M., Gruenwald A., Stoker R., (1989), "*Software review – Video Voice Speech Training System*", Volta Review, vol. 89, pp. 171-173
- [2]. Ball M.J., and Code C. (Eds), (1997), "*Electropalatography and its Clinical Applications*" Instrumental Clinical Phonetics, Whurr Publishers.
- [3]. Reynolds J., and Tarassenko L. (1993), "*Learning Pronunciation with the Visual Ear*", Neural Computing & Applications (1993) 1: pp 169 - 175
- [4]. Rooney E., Jack M., Lefevre J., and Sutherland A. (1995), "*HARP - A speech training aid for the hearing impaired*", 2nd TIDE Congress, La Villette, Paris, 26th-28th April 1995
- [5]. Hatzis A., Green P.D., and Howard S. (1997), "*OPTICAL LOGO-THERAPY (OLT) : A Computer-Based Real Time Visual Feedback Application for Speech Training.*", Proc. EUROSPEECH 1997, Vol 4, pp 1763-1766