

# Η ΑΙΧΜΗ ΤΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΣΤΗΝ ΟΠΤΙΚΟΑΚΟΥΣΤΙΚΗ ΑΝΑΤΡΟΦΟΔΟΤΗΣΗ ΜΕΣΩ ΥΠΟΛΟΓΙΣΤΗ

*Αθανάσιος Χατζής – University of Sheffield, Department of Computer Science*

Email : [nassos@dcs.shef.ac.uk](mailto:nassos@dcs.shef.ac.uk) – HTTP : [www.dcs.shef.ac.uk/~nassos](http://www.dcs.shef.ac.uk/~nassos)

## Περίληψη

Στο άρθρο αυτό περιγράφουμε και συγκρίνουμε τις δύο βασικές τεχνικές οπτικοακουστικής ανατροφοδότησης που χρησιμοποιούνται σήμερα στο λογισμικό για λογοθεραπεία. Αυτές είναι : η αξιολόγηση και σύγκριση από τον υπολογιστή του προφερόμενου στόχου με πρότυπα και καταχωρημένα από άλλους ομιλητές παραδείγματα, και η οπτική αναπαράσταση σε πραγματικό χρόνο του ακουστικού σήματος πάνω σε χάρτη στις περιοχές του οποίου αναπαριστώνται φωνήματα. Κλείνουμε με μια σύντομη αναφορά στην πρόσφατη έρευνα η οποία διεξάγεται στον τομέα αυτό.

## 1. Εισαγωγή

Υπάρχουν πάρα πολλές κλινικές περιπτώσεις οι οποίες απαιτούν λογοθεραπεία, από τις πιο απλές όπως για παράδειγμα το ψεύδισμα έως τις πιο σύνθετες όπως είναι η κώφωση. Σε αυτές ο βασικός ρόλος του λογοθεραπευτή είναι να βοηθήσει τον ασθενή ώστε να βελτιώσει ή και να αποκαταστήσει την ομιλία του η οποία αποτελεί και το κύριο επικοινωνιακό όργανο του ασθενή με το περιβάλλον. Ένα από τα πιο δύσκολα και σημαντικά στοιχεία στο έργο αυτό είναι η ανατροφοδότηση του ασθενούς. Ακόμα και στις πιο σύγχρονες θεραπευτικές μεθόδους αυτή βασίζεται στις τρεις βασικές αισθήσεις όραση, ακοή, αφή. Ο λογοθεραπευτής ενισχύει, αξιολογεί, και κατευθύνει την ήδη υπάρχουσα ανατροφοδότηση του ασθενούς η οποία πραγματοποιείται από τον ίδιο άμεσα και αυτόματα κατά τη διάρκεια, και μετά το πέρας της ολοκλήρωσης κάθε προσπάθειας του να αρθρώσει σωστά ένα συγκεκριμένο τμήμα του λόγου. Εδώ ακριβώς βρίσκει πεδίο εφαρμογής ο υπολογιστής ως εργαλείο στα χέρια του ειδικού, αλλά και ως βοηθητικό μέσο ταυτόχρονης οπτικής και ακουστικής ανατροφοδότησης του ασθενούς.

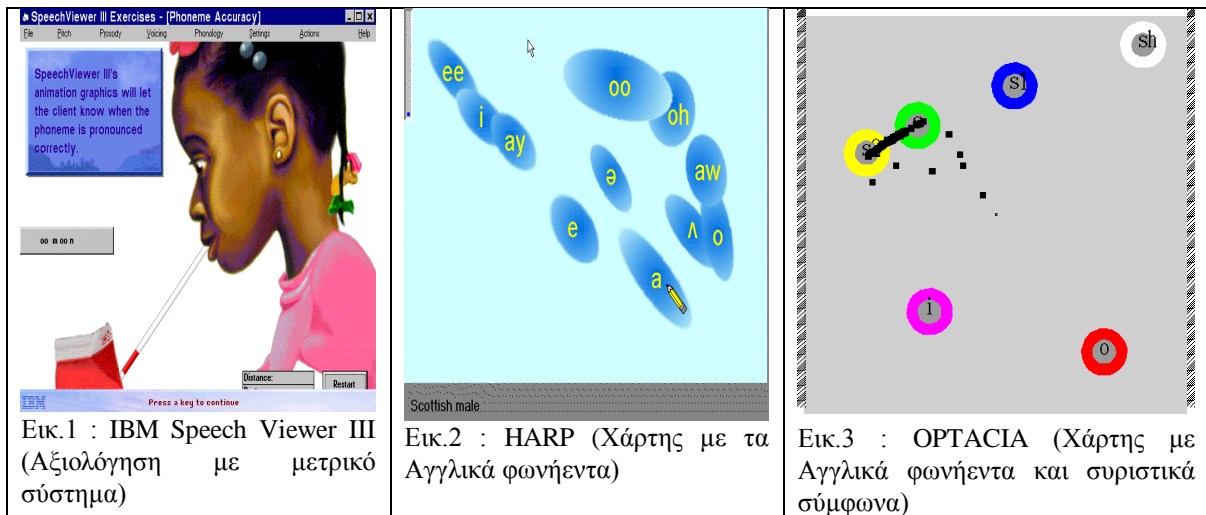
## 2. Ιστορικά στοιχεία

Θα μπορούσαμε να πούμε ότι η εφαρμογή των υπολογιστών στη λογοθεραπεία αρχίζει με το πρώτο ολοκληρωμένο σύστημα εκπαίδευσης στην ομιλία μέσω υπολογιστή από τους Nickerson και Stevens το 1976, [7], καθώς και με τον φασματογράφο ακουστικού σήματος, [6]. Σήμερα υπάρχει πληθώρα από εμπορικές εφαρμογές όπως το IBM Speech Viewer, [9], το ISTR, [5], το Video Voice, [1], το Accent Coach, τα οποία στοχεύουν στην εξάσκηση όλων των παραμέτρων στην παραγωγή ομιλίας (φώνηση, άρθρωση, τονισμός). Τα συστήματα αυτά αν και έχουν επιδοκιμαστεί από τους λογοθεραπευτές και τους ασθενείς τους σε πολλά από τα γυμνάσματα που παρέχουν δεν έχουν αποκτήσει την πλήρη αποδοχή για πολλούς λόγους σημαντικότερος των οποίων είναι η ακρίβεια και η ποιότητα της ανατροφοδότησης που δίνουν. Πιο συγκεκριμένα η ανατροφοδότηση πάνω στη θέση και στο τρόπο άρθρωσης συγκεκριμένου ήχου, συλλαβής, ή λέξης, παρουσιάζει ιδιαίτερο πρόβλημα σε αυτά τα συστήματα, [2].

## 2.1. Βασικές τεχνικές

Υπάρχουν δύο βασικές τεχνικές που χρησιμοποιούνται σήμερα στο λογισμικό για υπολογιστές σε συστήματα σαν και αυτά που αναφέραμε πιο πάνω :

α) η σύγκριση του παραγόμενου ακουστικού σήματος με ήδη αποθηκευμένα στατιστικά μοντέλα του στόχου προς μίμηση (template matching, hidden markov models). Σε αυτή τη περίπτωση έχουμε και μια αξιολόγηση με μετρικό σύστημα που βασίζεται στη σύγκριση αυτή, η οποία δίνεται αμέσως μετά την ολοκλήρωση της προσπάθειας του ασθενούς να παράγει τον στόχο (Εικ.1) .



β) η απεικόνιση του ακουστικού σήματος σε πραγματικό χρόνο, π.χ. κάθε 10 msec, σε χώρο δύο ή και τριών διαστάσεων. Ο ακουστικός αυτός χάρτης δημιουργείται από επιλεγμένους ήχους οι οποίοι συσχετίζονται με σημεία αναφοράς συγκεκριμένες θέσεις άρθρωσης, [4]. Στο χάρτη αυτό απεικονίζονται στόχοι που συμβολίζουν μεμονωμένους ήχους, π.χ. φωνήεντα (Εικ.2). Η ανάπτυξη και εξέλιξη της τεχνικής αυτής για την απεικόνιση συμφώνων μαζί με φωνήεντα (Εικ.3) αποτέλεσε αντικείμενο πολυετής έρευνας για τον γράφοντα και οδήγησε με αρκετή επιτυχία στην εφαρμογή της, [3], [4], [8], [2].

## 3. Διαφορές

Η σύγκριση και η επισήμανση των βασικών διαφορών μεταξύ των δύο αυτών τεχνικών θα μας δώσει και ορισμένα βασικά κριτήρια με τα οποία μπορούμε να αξιολογήσουμε την αποτελεσματικότητα της οπτικοακουστικής ανατροφοδότησης μέσω υπολογιστή.

### 3.1. Βιο-ανατροφοδότηση

Η τεχνική (β) ανήκει σε αυτή τη κατηγορία. Στην περίπτωση αυτή η απεικόνιση του ακουστικού σήματος γίνεται σε πραγματικό χρόνο και συσχετίζεται με την κίνηση και τα γεγονότα που λαμβάνουν χώρα στην οθόνη του υπολογιστή. Επίσης το ακουστικό σήμα παράγεται με συγκεκριμένες μεταβολές στις θέσεις των αρθρωτών (για παράδειγμα η συλλαβή /σα/ παράγεται εάν από τη θέση άρθρωσης του συριστικού /σ/ μετακινήσουμε προς τα κάτω την κάτω σιαγόνα έτσι ώστε να πάρουμε την θέση άρθρωσης του φωνήεντος /α/. Για το λόγο αυτό, σε πραγματικό χρόνο ο παρατηρητής της οθόνης συνδέει τα οπτικά γεγονότα, με το παραγόμενο ακουστικό σήμα και

παράλληλα με τις μεταβολές στις θέσεις άρθρωσης. Κάτι τέτοιο δεν είναι εφικτό με την τεχνική (α) η οποία δεν μπορεί να μας δώσει την λεπτομέρεια και την ακρίβεια που απαιτείται κατά τη διάρκεια της άρθρωσης,

### **3.2. Κατεύθυνση**

Είναι δυνατό να αντιστοιχίσουμε γειτονικές περιοχές του χάρτη με κοντινές θέσεις άρθρωσης προς αυτές που επιλέξαμε αναφορικά. Αυτό έχει ως αποτέλεσμα να μπορούμε να έχουμε κατεύθυνση προς τον στόχο κάτι που είναι πρακτικά αδύνατο να επιτύχουμε με την τεχνική (α).

### **3.3. Ποιοτική έναντι ποσοτικής ανατροφοδότησης**

Η τεχνική (α) επίσης σύμφωνα με τον ορισμό της βασίζεται σε ποσοτική αξιολόγηση του παραγόμενου τμήματος ομιλίας σε αντίθεση με την (β) η οποία δίνει μια ποιοτική αναπαράσταση του ακουστικού σήματος. Λογισμικό που χρησιμοποιεί την (α) απεικονίζει την κλίμακα μέτρησης γραφικά, σε μια διάσταση, ως μια απόσταση από έναν στόχο (Εικ.1). Στην περίπτωση αυτή ο ασθενής είναι δύσκολο ή αδύνατο να συνδέσει τις επιτυχημένες-αποτυχημένες προσπάθειές του να προσεγγίσει όσο πιο ακριβέστερα τον στόχο με τις διαφορετικές θέσεις, αξιολογήσεις του συστήματος, που λαμβάνει στην ευθεία μεταξύ του στόχου και της βάσης εκκίνησης (Εικ.1), ιδιαίτερα μάλιστα όταν η αξιολόγηση αυτή γίνεται μετά το πέρας της άρθρωσης. Αντίθετα οι αποστάσεις από την γραφική απεικόνιση ενός στόχου σε επίπεδο δύο διαστάσεων είναι σχετικές και απλά φανερώνουν τον συσχετισμό της παραγόμενης θέσης άρθρωσης με αυτή του στόχου (Εικ.2-3).

### **3.4. Επίπεδο εξάσκησης**

Η τεχνική (β) στο πρόσφατο στάδιο είναι κατάλληλη για εξάσκηση του ασθενούς σε μεμονωμένους ήχους και σε απλές συλλαβές, λέξεις. Εάν πρόκειται για πιο σύνθετες λέξεις καθώς και για προτάσεις η τεχνική (α) είναι πιο κατάλληλη.

### **3.5. Σταθερότητα στην άρθρωση και αντιπαραβολή**

Ίσως μια από τις πιο σημαντικές παραμέτρους που οδηγεί στην σταδιακή βελτίωση της άρθρωσης είναι και η συχνότητα με την οποία ο ασθενής προφέρει κάποιον στόχο. Στην τεχνική (α) πολλές διαφορετικές ακουστικά αρθρώσεις μπορούν να αξιολογηθούν το ίδιο, δηλαδή οπτικά η ίδια θέση στη απόσταση μεταξύ στόχου και βάσης. Έτσι όμως δεν γίνονται κατανοητές οι διαφορές που υπάρχουν. Στην τεχνική (β) αυτό είναι πιο δύσκολο να συμβεί λόγω της απεικόνισης που γίνεται στον χάρτη σε διαφορετικά σημεία.

## **4. Το πρόγραμμα ορθο-λογο-παιδεία (Ortho-Logo-Paedia, OLP)**

Το ερευνητικό πρόγραμμα OLP (<http://www.xanthi.ilsp.gr/olp/>), ξεκίνησε τον Ιανουάριο του 2001, στοχεύει στο να αξιολογήσει και να εφαρμόσει τις τεχνικές τις οποίες περιγράψαμε σε χαρακτηριστικές παθήσεις του λόγου. Το λογισμικό περιλαμβάνει όλα τα απαραίτητα εργαλεία για την συλλογή και επεξεργασία των δεδομένων καθώς και την δημιουργία μοντέλων, μαζί με πληθώρα από γραφικές διεπιφάνειες για τον χρήστη. Σημαντικό κομμάτι του προγράμματος είναι η από απόσταση συλλογή πληροφοριών και εκπαίδευση.

### **Ευχαριστίες**

Η έρευνα για την εξέλιξη του λογισμικού στο OLP καθώς και η κλινική εφαρμογή του επιδοτείται από το 5<sup>ο</sup> κοινοτικό πλαίσιο της Ευρωπαϊκής ένωσης (Quality of Life theme). Στο πρόγραμμα αυτό από την Ελλάδα συμμετέχουν, το κέντρο λογοθεραπείας (LOGOS) από την Θεσσαλονίκη, και το ινστιτούτο επεξεργασίας λόγου (ΙΕΛ) από την Αθήνα.

### **Βιβλιογραφία**

- [1] Fitzgerald M., Gruenwald A., Stoker R., (1989), “*Software review – Video Voice Speech Training System*”, Volta Review, vol. 89, pp. 171-173
- [2] Hatzis A., (1999), “*Optical Logo-Therapy (OLT), Computer-Based Audio-Visual Feedback Using Interactive Visual Displays for Speech Training*”, unpublished PhD thesis, Dept. of Computer Science, University of Sheffield.
- [3] Hatzis A., Green P.D., and Howard S. (1999), “*Visual Displays in Practical Auditory Phonetics Teaching*”, Phonetics Teaching & Learning Conference (PTLC’99), University College of London, U.K.
- [4] Hatzis, A. and 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.
- [5] Kewley-Port D., Watson C.S., and Cromer P.A., (1987). “*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.
- [6] Maki J. (1983), “*Applications of the speech spectrographic display in developing articulatory skill in hearing impaired adults*”, in Speech of the hearing impaired: research, training and personnel preparation, (Ed.) M.N. Osberger, University Park Press, Baltimore, MD.
- [7] Nickerson, R. S., Kalikow, D. N. & Stevens, K. N., (1976), “*Computer-aided speech training for the deaf*,” JSHD, 41, 120-132.
- [8] Öster AM, House D., Hatzis A., Green P.D. (2003), “*Testing a New Method for Training Fricatives using Visual Maps in the Ortho-Logo-Paedia Project (OLP)*”, fonetik, Vol.X
- [9] Pratt S., Heintzelman A.T., and Deming S. Ensrud (1993), “*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.
- [10] 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