

Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης
Τμήμα Μηχανικών Μουσικής Τεχνολογίας & Ακουστικής

**Ανάπτυξη και αξιολόγηση βάσης δεδομένων
ομιλίας καταγεγραμμένης μέσω οπτικών
τεχνικών λέιζερ**

Αρά Μοβσεσιάν

Επιβλέποντες καθηγητές :

Δρ. Ευθύμιος Μπακαρέζος
Δρ. Νεκτάριος Παπαδογιάννης
Δρ. Παναγιώτης Ζέρβας

Ρέθυμνο, 2014

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τους τρεις επιβλέποντες καθηγητές μου για τις πολύτιμες συμβουλές και καθοδηγήσεις που μου δώσανε για την εκπόνηση αυτής της διπλωματικής εργασίας, καθώς επίσης και όλους όσους συμμετείχαν στα στατιστικά πειράματα για τον χρόνο που μου αφιέρωσανε αλλά και για την υπομονή τους. Μεγάλες ευχαριστίες θα ήθελα όμως να αποδώσω και στον Νικόλαο Τσαγκαράκη για την σημαντική βοήθεια του σε διάφορα θέματα που σχετίζονταν με την εργασία μου και στην οικογένεια μου για την αδιάκοπη υποστήριξη και ενίσχυση που μου παρείχαν τα τελευταία 23 χρόνια σε κάθε ευκαιρία ανεξαρτήτου του γεγονότος.

Περιεχόμενα

Ευχαριστίες.....	3
Περιεχόμενα.....	4
Περίληψη	6
Abstract.....	6
Εισαγωγή	7
Στοιχεία θεωρίας.....	9
2.1 Η Ανθρώπινη Ομιλία	9
2.1.1 Το φωνητικό όργανο	9
2.1.2 Συχνотικό περιεχόμενο ανθρώπινης ομιλίας	13
2.1.3 Οι ιδιοσυχνότητες του φωνητικού οργάνου (Formants)	14
2.2 Τα Συναισθήματα στην Ομιλία	15
2.2.1 Τα δύο κανάλια της ανθρώπινης επικοινωνίας	17
2.2.2 Η σημασία της προσωδίας στη συναισθηματική ομιλία	20
2.2.3 Η σημασία της ποιότητας της φωνής στη συναισθηματική ομιλία.....	20
2.3 Ανίχνευση ήχων εξ' αποστάσεως με οπτοηλεκτρονικές τεχνικές.....	22
2.3.1 Η λειτουργία των lasers.....	22
2.3.2 Ανίχνευση ήχων με οπτοηλεκτρονική διάταξη εκτροπής δέσμης	26
2.4 Βελτίωση ηχητικών καταγραφών	28
2.4.1 Η αντήχηση και η αντιμετώπιση της.....	28
2.4.2 Το προφίλ της αντήχησης και η αντηχητική ουρά.....	29
2.4.3 Αποθορυβοποίηση	31
2.4.4 Το προφίλ της αποθορυβοποίησης.....	32
2.4.5 Η καμπύλη μείωσης και το κατώφλι αποθορυβοποίησης	33
Πειραματικές τεχνικές και διαδικασία	35

3.1 Οπτοηλεκτρονική διάταξη και διαδικασία μετρήσεων.....	35
3.2 Επεξεργασία καταγεγραμμένου ηχητικού υλικού	38
3.2.1 Διαχωρισμός ηχητικού υλικού ως προς την καταληπτότητα της ομιλίας	39
3.2.2 Εφαρμογή αλγόριθμων αφαίρεσης αντήχησης και αποθορυβοποίησης	41
3.3 Πειραματική διαδικασία στατιστικών ακροάσεων	43
3.3.1 Συνθήκες κατά την ακρόαση	43
Επεξεργασία και ανάλυση δεδομένων.....	46
4.1 Παρατηρήσεις του στατιστικού πειράματος	46
4.1.1 Αξιολόγηση του εννοιολογικού περιεχομένου των ηχογραφήσεων	46
4.1.2 Αξιολόγηση συναισθηματικής πληροφορίας ηχογραφήσεων	47
4.2 Ανάλυση και σχολιασμός των στατιστικών μετρήσεων	49
4.2.1 Ανάλυση των μετρήσεων	50
4.2.2 Σχολιασμός των αποτελεσμάτων	55
Συμπεράσματα.....	57
Βιβλιογραφία	61

Περίληψη

Στην παρούσα διπλωματική εργασία υλοποιείται μια οπτοηλεκτρονική διάταξη ανίχνευσης ήχων εξ' αποστάσεως, που βασίζεται στην τεχνική εκτροπής δέσμης laser από δονούμενη επιφάνεια. Με την χρήση αυτής της διάταξης καταγράφεται ηχητικό υλικό από μια πλήρως χαρακτηρισμένη βάση δεδομένων συναισθηματικής ομιλίας. Εν συνεχεία το καταγεγραμμένο ηχητικό υλικό εισέρχεται σε αλγόριθμους βελτίωσης. Έπειτα από την ολοκλήρωση της επεξεργασίας διεξάγεται μια σειρά από ειδικά σχεδιασμένα στατιστικά πειράματα με την συμμετοχή 10 ακροατών όπου τους ζητείται να αναγνωρίσουν το λεξιλογικό και το συναισθηματικό περιεχόμενο της ομιλίας στην επεξεργασμένη και την μη επεξεργασμένη εκδοχή των αρχείων. Η ανάλυση των απαντήσεων γίνεται με σκοπό την εξαγωγή συμπερασμάτων για την επίδοση των αλγορίθμων, της οπτοηλεκτρονικής διάταξης, και των ακροατών σε συνθήκες υψηλής και χαμηλής καταληπτότητας, μέσω σύγκρισης γραφημάτων ράβδων.

Abstract

In the present study an optoelectronic arrangement for the remote detection of sound is developed, based on the laser beam deflection technique. Using this, sound material from a fully characterized speech data base is detected. The recorded material is then processed using appropriate algorithms. After the processing, the recorded material is used in a series of specially designed statistical experiments in which a total of 10 listeners are asked to recognise the lexical and emotional content of speech in both the processed and unprocessed versions. The answers are analysed in order to extract useful conclusions used to evaluate the performance of the algorithm, the optoelectronic arrangement, and the listeners, under conditions of high and low intelligibility, while the results are presented in appropriate bar graphs.

Κεφάλαιο 1^ο

Εισαγωγή

Η απομακρυσμένη καταγραφή ήχου υπήρξε αντικείμενο προβληματισμού ανά την επιστημονική κοινότητα για πολλές δεκαετίες λόγω των περιορισμών που έθετε το μηχανικό μέρος των μικροφώνων στην λειτουργία και την ευαισθησία τους. Η ανίχνευση ήχων σε μακρινές αποστάσεις ήταν σχεδόν αδύνατη με την υπάρχουσα κατασκευή των μικροφώνων και έτσι έπρεπε να παρουσιαστεί μια νέα καινοτόμα ιδέα η οποία θα προσέφερε την δυνατότητα βέλτιστης ευαισθησίας και θα ήταν ικανή να καταγράψει ηχητικό υλικό σε μεγάλη απόσταση. Αυτές τις ιδιότητες τις παρουσιάζουν οι οπτοηλεκτρονικές διατάξεις οι οποίες χρησιμοποιούν μια δέσμη laser. Θεωρητικά αυτές οι διατάξεις είναι ικανές να πετύχουν αρκετά καλή πιστότητα ηχογράφησης σε μεγάλες αποστάσεις λόγω της έντονης κατευθυντικότητας της δέσμης, της μονοχρωματικότητας της αλλά και της έντασής της.

Στην παρούσα εργασία υλοποιείται μια τέτοιου είδους διάταξη, η οποία βασίζεται στην τεχνική εκτροπής δέσμης laser (laser beam deflection) από δονούμενη επιφάνεια, συνέπεια πρόσπτωσης σε αυτήν ηχητικού υλικού που προέρχεται από μια πλήρως χαρακτηρισμένη βάση δεδομένων συναισθηματικής ομιλίας. Ο απώτερος σκοπός είναι να διαπιστωθεί κατά πόσο στο καταγεγραμμένο ηχητικό υλικό από μια τέτοια απλή διάταξη είναι δυνατός ο εντοπισμός συναισθημάτων, αφού βέβαια εφαρμοστεί σχετικά απλή διαδικασία επεξεργασίας βελτίωσης σημάτων.

Συγκεκριμένα τα κεφάλαια αναπτύσσονται ως εξής: στο κεφάλαιο 2 κτίζεται ένα θεωρητικό υπόβαθρο που αφορά όλες εκείνες τις έννοιες που είναι συνυφασμένες με την εργασία αυτή. Πιο συγκεκριμένα παρουσιάζεται η λειτουργία και οι ιδιαιτερότητες της ανθρώπινης ομιλίας, οι τεχνικές εξ' αποστάσεως ανίχνευσής της και οι τεχνικές αποθρομβοποίησης / αφαίρεσης αντήχησης του καταγεγραμμένου ηχητικού υλικού. Στο κεφάλαιο 3 αναλύεται η αρχή λειτουργίας της οπτοηλεκτρονικής διάταξης που υλοποιήθηκε καθώς επίσης και η διαδικασία που ακολουθήθηκε κατά την καταγραφή του ηχητικού υλικού.

Μεταγενέστερα γίνεται ένας διαχωρισμός των περισσότερο καταληπτών αρχείων από αυτά με την μικρότερη καταληπτότητα, και στην συνέχεια εισέρχεται το σύνολο του ηχητικού υλικού στους αλγόριθμους αφαίρεσης αντήχησης και αποθρομβοποίησης. Κατά την

ολοκλήρωση του κεφαλαίου περιγράφεται λεπτομερώς η πειραματική διαδικασία των στατιστικών ακροάσεων. Στο κεφάλαιο 4 παρατίθενται όλες οι παρατηρήσεις και οι επισημάνσεις που εξήχθησαν από τις ακροάσεις, καθώς επίσης και όλα τα στατιστικά δεδομένα υπό την μορφή γραφημάτων ράβδων. Στο κεφάλαιο 5 συνοψίζονται όλες οι διαπιστώσεις και τα συμπεράσματα που δημιουργήθηκαν κατά την διάρκεια και ιδιαίτερα κατά την ολοκλήρωση των πειραματικών διαδικασιών και προκύπτει ένα τελικό πόρισμα για την συνολική επίδοση όλων των επεξεργασιών και διεργασιών που αφορούσαν τους αλγόριθμους, το ηχητικό υλικό και την οπτοηλεκτρονική διάταξη. Κλείνοντας αυτή την ενότητα, πραγματεύονται όλες οι πιθανές ενέργειες που θα έδιναν μια προοπτική βελτίωσης στην λειτουργία της οπτοηλεκτρονικής διάταξης και θα βοηθούσαν μακροπρόθεσμα στην απόπειρα μιας νέας απομακρυσμένης ηχογράφησης.

Κεφάλαιο 2^ο

Στοιχεία θεωρίας

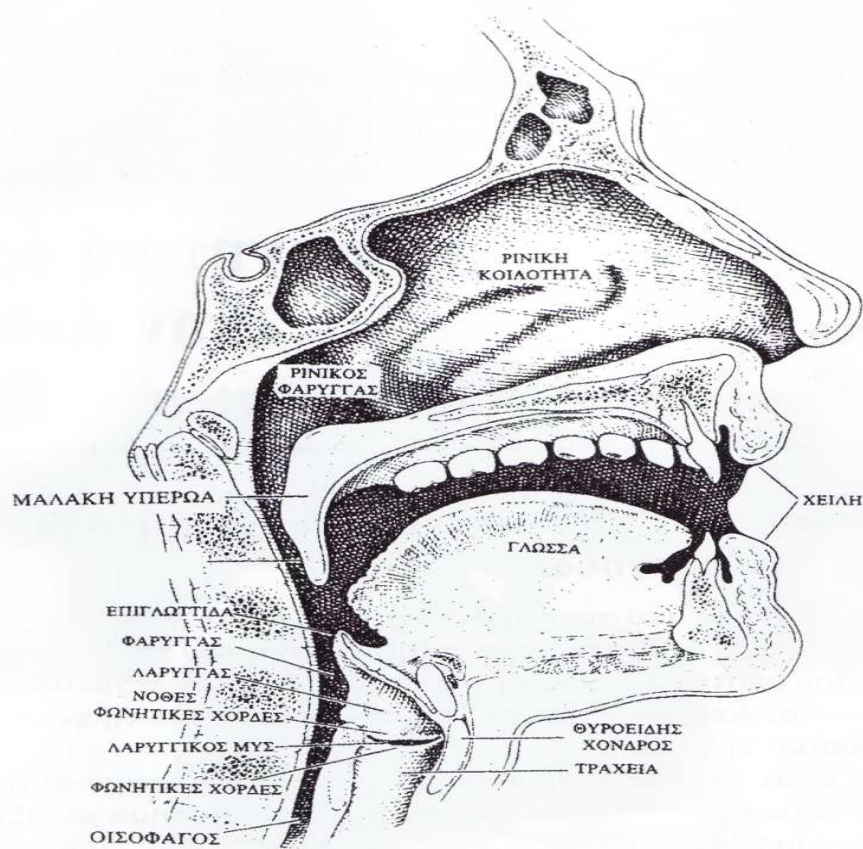
Στο κεφάλαιο αυτό παρουσιάζονται, σε θεωρητικό επίπεδο, όλες εκείνες οι έννοιες που αφορούν και άπτονται της εργασίας αυτής. Συγκεκριμένα, θα παρουσιαστεί η λειτουργία και οι ιδιαιτερότητες της ανθρώπινης ομιλίας, οι τεχνικές εξ αποστάσεως ανίχνευσης της και οι τεχνικές αποθορυβοποίησης / αφαίρεσης αντήχησης από καταγεγραμμένο ηχητικό υλικό.

2.1 Η Ανθρώπινη Ομιλία

2.1.1 Το φωνητικό όργανο

Το όργανο της φωνής περιλαμβάνει τους πνεύμονες ,τον λάρυγγα ,τον φάρυγγα ,τη μύτη και το στόμα (Εικ. 1).

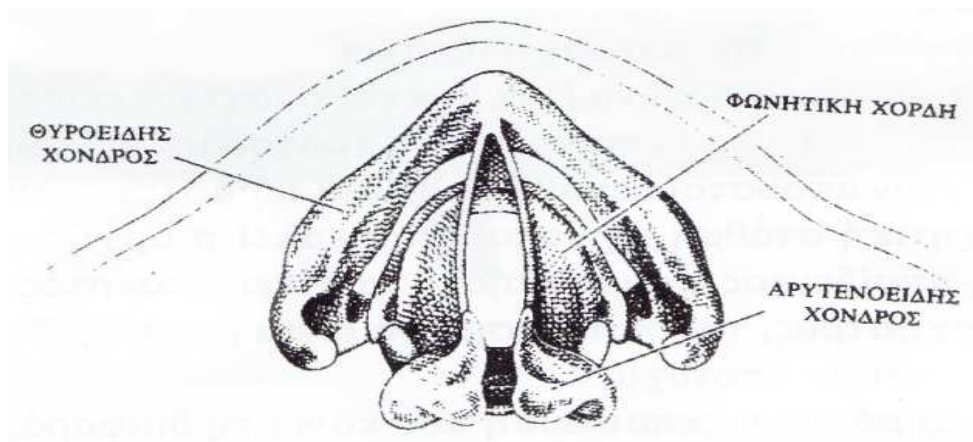
Η κύρια φωνητική λειτουργία των πνευμόνων είναι να δημιουργούν αυξημένη πίεση αέρα εξαιτίας της οποίας προκαλείται μια ροή αέρα. Ο αέρας περνά διαμέσου της γλωττίδας, που είναι ένα άνοιγμα στην βάση του λάρυγγα, ανάμεσα στις 2 φωνητικές χορδές. Το μπροστινό άκρο κάθε μιας φωνητικής χορδής συνάπτεται με το θυρεοειδή χόνδρο ή το μήλο του Αδάμ. Το πίσω άκρο καθεμιάς φωνητικής χορδής συνάπτεται στον έναν από τους 2 μικρούς αρυτενοειδείς χόνδρους, οι οποίοι είναι ευκίνητοι και ή ανοίγουν τις φωνητικές χορδές κατά την αναπνοή ή τις σμίγουν τεντώνοντας τες (Εικ. 2).



Εικόνα 1. Το φωνητικό όργανο αποτελούν οι πνεύμονες, ο λάρυγγας, ο φάρυγγας, το στόμα και η μύτη. Ο λάρυγγας, ο φάρυγγας και η στοματική κοιλότητα (στους έρρινους ήχους και η ρινική κοιλότητα) αποτελούν τον φωνητικό σωλήνα.

Κατά την ομιλία οι φωνητικές χορδές ενεργούν με διάφορους τρόπους. Από μια εντελώς κλειστή θέση με την οποία διακόπτουν τη ροή του αέρα μπορούν να ανοίξουν ξαφνικά (περίπτωση ήπιας μορφής βήχα). Μια ενδιάμεση θέση παρατηρείται στον ήχο "χ" όπου το ρεύμα αέρα διακόπτεται ελαφρά καθώς περνά μεταξύ των φωνητικών χορδών.

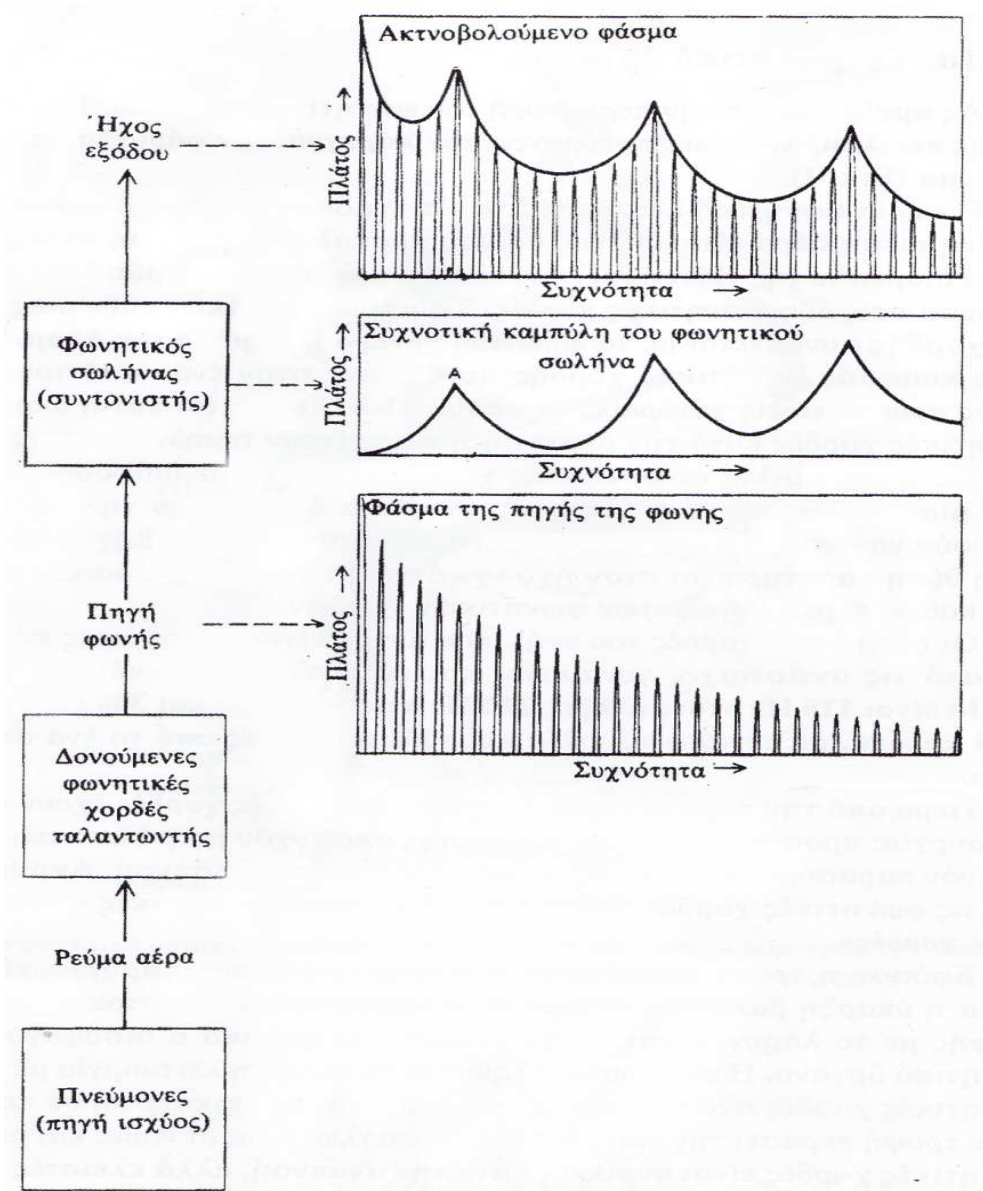
Οι φωνητικές χορδές του ενήλικου άντρα είναι μακρύτερες και βαρύτερες από τις αντίστοιχες γυναικείες και παιδικές. Οι συχνότητες κατά την ομιλία είναι 110 Hz στους άνδρες 220 Hz στις γυναίκες και 300 Hz στα παιδιά κατά κανόνα, με μεγάλη ποικιλία και ιδιαιτερότητες από το ένα άτομο στο άλλο.



Εικόνα 2. Οι φωνητικές χορδές παραμένουν ανοικτές κατά την αναπνοή και κλειστές κατά την παραγωγή της φωνής με το πλησίασμα των αρυτενοειδών χόνδρων.

Οι φωνητικές χορδές βρίσκονται στο κάτω μέρος του λάρυγγα, που έχει μορφή σωλήνα και ο οποίος είναι προσαρμοσμένος μέσα στον φάρυγγα, την πλατύτερη κοιλότητα που οδηγεί από το στόμα στον οισοφάγο. Το ανώτερο τμήμα του φάρυγγα, η μαλακή υπερώα (ουρανίσκος), είναι μεταφορικά η πύλη της ρινικής κοιλότητας. Όταν ο ουρανίσκος είναι υπερυψωμένος, το πέρασμα προς την μύτη κλείνει και ο αέρας εξέρχεται από το στόμα. Ο λάρυγγας μαζί με το φάρυγγα και το στόμα σχηματίζουν το φωνητικό όργανο, ένα συντονισμένο χώρο, κάτι σαν τον ηχητικό σωλήνα ενός πνευστού οργάνου ή το αντηχείο του βιολιού.

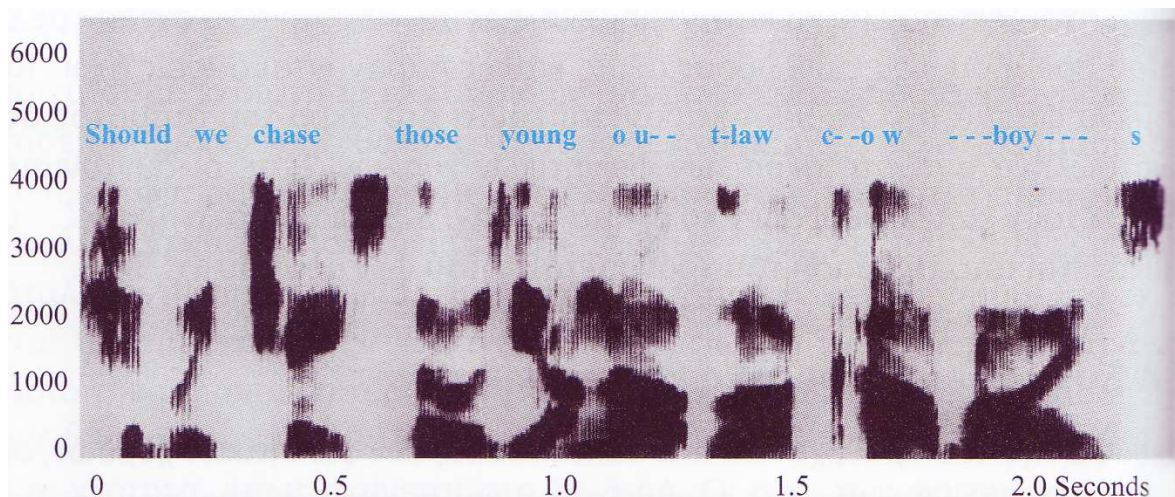
Η μορφή του φωνητικού οργάνου καθορίζεται από τις θέσεις των κινητών μερών: χείλη, σιαγόνα, γλώσσα και λάρυγγα. Κινήσεις των χειλιών, των σιαγόνων και του λάρυγγα διαστέλλουν ή συστέλλουν το φωνητικό όργανο σε συγκεκριμένα σημεία. Το μήκος του αυξάνει όταν τα χείλη ωθούνται προς τα έξω ενώ μειώνεται κατά την πραγματοποίηση των αντίστροφων κινήσεων.



Εικόνα 3. Λειτουργικά το όργανο της φωνής αποτελείται από την πηγή ισχύος, τον ταλαντωτή και τον συντονιστή. Το αέριο ρεύμα απο τους πνεύμονες διακόπτεται περιοδικά από τις δονούμενες φωνητικές χορδές. Ο παραγόμενος ήχος έχει ένα φάσμα πλούσιο σε αρμονικούς το πλάτος των οποίων φθίνει ομοιόμορφα καθώς αυξάνει η συχνότητα. Η αέρια στήλη στον φωνητικό σωλήνα έχει χαρακτηριστικούς τρόπους δόνησης ή συντονισμούς, που ονομάζονται Formants (Α,Β,Γ) και οι οποίες διαμορφώνουν τους αρμονικούς του φάσματος, δημιουργώντας έτσι το φάσμα του εκπεμπόμενου ήχου από τα χείλη.

2.1.2 Συχνοτικό περιεχόμενο ανθρώπινης ομιλίας

Οι ήχοι φωνής που προέρχονται από ταλαντώσεις των φωνητικών χορδών, οι άφωνοι ήχοι που προέρχονται από τυρβώδη ροή και κροτικοί ήχοι που ξεκινούν κοντά στα χείλη, σχηματίζουν ενοποιημένοι τους ήχους ομιλίας. Καθώς μιλούμε οι χαρακτηριστικές ιδιοσυχνότητες μετατοπίζονται στη συχνότητα καθώς τα χείλη, το σαγόι, η γλώσσα, και ο ουρανίσκος αλλάζουν θέση για να σχηματιστούν οι επιθυμητές λέξεις. Το αποτέλεσμα είναι η πολυπλοκότητα της ανθρώπινης φωνής, η οποία είναι εμφανής στο φασματογράφημα της εικόνας 4. Οι πληροφορίες που μεταδίδονται μέσω της ομιλίας συνίστανται από ένα μοτίβο συχνοτήτων και έντασης που μετατοπίζονται γρήγορα με τον χρόνο. Αξίζει να σημειωθεί στην εικόνα 4 ότι υπάρχει μικρή ποσότητα ενέργειας ομιλίας πάνω από τα 4 kHz. Παρότι δεν φαίνεται στο φασματογράφημα, υπάρχει αντίστοιχα σχετικά μικρή ποσότητα ενέργειας ομιλίας και κάτω από τα 100 Hz. Γίνεται πλέον κατανοητό για πιο λόγο τα φίλτρα παρουσίας έχουν μέγιστο στην περιοχή 2 με 3 kHz. Διότι εκεί συντονίζονται οι ανθρώπινοι ηχητικοί σωλήνες.



Εικόνα 4. Φασματογράφημα ήχου μιας πρότασης ομιλούμενης από αντρική φωνή.

2.1.3 Οι ιδιοσυχνότητες του φωνητικού οργάνου (Formants)

Το φωνητικό όργανο είναι ένας συντονιστής και η εκπομπή του ήχου από ένα ακουστικό συντονιστή εξαρτάται κατά πολύ από την συχνότητα. Ήχοι με τις ιδιοσυχνότητες κάθε συντονιστή εξασθενούν πολύ λιγότερο από άλλους ήχους και γι' αυτό ακτινοβολούνται με μεγαλύτερο σχετικό πλάτος ή με μεγαλύτερη σχετική ακουστότητα από άλλους ήχους. Όσο μεγαλύτερη είναι η συχνοτική διαφορά ενός ήχου από την ιδιοσυχνότητα του συντονιστή, τόσο πιο εξασθενημένα ακτινοβολείται ο συγκεκριμένος ήχος.

Το φωνητικό όργανο έχει 4 με 5 χαρακτηριστικές ιδιοσυχνότητες, που ονομάζονται formants. Οι πολλοί αρμονικοί του λαρυγγικού ήχου, που διοχετεύθηκαν στο φωνητικό όργανο, το διασχίζουν με ευκολία η οποία εξαρτάται από την συχνότητα τους. Όσο πλησιέστερα βρίσκεται ένας αρμονικός προς κάποια formant συχνότητα, τόσο πιο ενισχυμένο είναι το πλάτος του στο άνοιγμα των χειλιών.

Η παρουσία των formants διαταράσσει την ομοιόμορφη κλίση της περιβάλλουσας του φάσματος του λαρυγγικού ήχου, επιβάλλοντας κορυφές στην θέση των formant συχνοτήτων. Αυτή η διαταραχή στην περιβάλλουσα του φάσματος του λαρυγγικού ήχου προκαλεί διακεκριμένους ήχους κατά την ομιλία. Συγκεκριμένες formant συχνότητες εκδηλώνονται στο ακτινοβολούμενο φάσμα σαν κορυφές τις περιβάλλουσας και αυτές οι κορυφές είναι χαρακτηριστικές συγκεκριμένων ήχων.

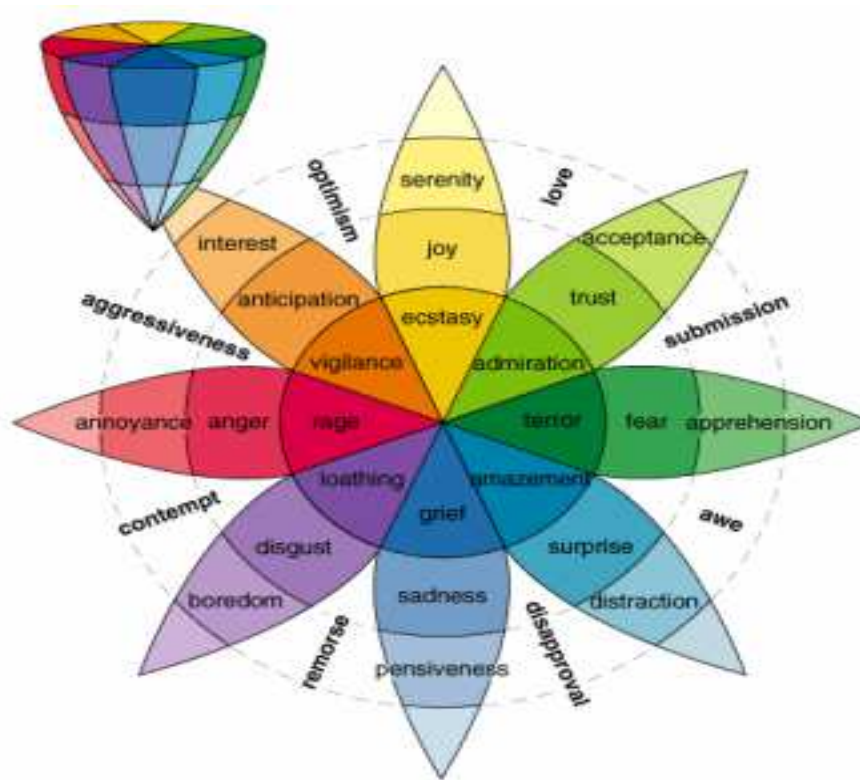
Οι συχνότητες formant προσδιορίζονται από την μορφή του ηχητικού οργάνου. Εάν το φωνητικό όργανο ήταν ένας τέλειος κύλινδρος κλειστός στην γλωττίδα και ανοικτός στα χείλη με μήκος 17,5 cm (αυτό είναι το μήκος του φωνητικού οργάνου στους ενήλικους άνδρες), τότε οι 4 πρώτες formants θα είναι κοντά στα 500,1500,2500 και 3500 Hz. Δοθέντος ενός μακρύτερου ή κοντότερου φωνητικού οργάνου, αυτές οι βασικές συχνότητες είναι κάπως χαμηλότερες ή υψηλότερες. Οποιαδήποτε μεταβολή στη διατομή του φωνητικού σωλήνα γίνεται αιτία μετατόπισης των επιμέρους formant συχνοτήτων. Ο φωνητικός σωλήνας στενεύει και διευρύνεται με πολλούς και πολύπλοκους τρόπους και, στενεύοντας σ' ένα μέρος, επηρεάζει τη συχνότητα όλων των formants ποικιλοτρόπως. Υπάρχουν, όμως, τρία κύρια εργαλεία για τη μεταβολή της μορφής του φωνητικού σωλήνα κατά τέτοιο τρόπο, ώστε η συχνότητα μιας συγκεκριμένης formant μετατοπίζεται προς μια συγκεκριμένη κατεύθυνση.

Αυτά τα εργαλεία είναι η σιαγόνα, ο κορμός της γλώσσας και η άκρη της γλώσσας. Το άνοιγμα της σιαγόνας, που στενεύει τον φωνητικό σωλήνα προς την γλωττίδα και τον διευρύνει προς τα χείλη, επηρεάζει αποφασιστικά τη συχνότητα της πρώτης formant αυξάνοντας την, όσο το άνοιγμα γίνεται μεγαλύτερο. Η δεύτερη formant συχνότητα είναι ιδιαίτερα ευαίσθητη στο σχήμα του κορμού της γλώσσας. Η τρίτη formant συχνότητα επηρεάζεται από την θέση της άκρης της γλώσσας. Καθώς κινούνται τα διάφορα όργανα προφοράς με διάφορους τρόπους, παρουσιάζονται μεταβολές στις συχνότητες των 2 χαμηλότερων formants μέσα σε ένα σημαντικό ευρύ διάστημα, το οποίο στους ενήλικες άνδρες κυμαίνεται κατά προσέγγιση από τα 250-700 Hz για την πρώτη formant και από τα 700-2500 Hz για την δεύτερη formant.

Κάθε συγκεκριμένη στάση άρθρωσης αντιστοιχεί σε ένα σύνολο συχνοτήτων formants, οι οποίες με τη σειρά τους συνδέονται με έναν συγκεκριμένο ήχο φωνήεντος. Πιο συγκεκριμένα, οι συχνότητες formant ενισχύουν κάποιες αρμονικές συχνότητες του λαρυγγικού ήχου καθιστώντας τις κορυφές, που χαρακτηρίζουν την περιβάλλουσα στο φάσμα καθενός φωνήεντος.

2.2 Τα Συναισθήματα στην Ομιλία

Κατατάσσουμε τα συναισθήματα σε δύο κύριες κατηγορίες: τα πρωτεύοντα (βασικά) και τα δευτερεύοντα (παραγόμενα) συναισθήματα. Τα *πρωτεύοντα ή βασικά συναισθήματα* είναι αυτά που κατά κύριο λόγο βιώνονται από όλα τα κοινωνικά θηλαστικά και σχετίζονται με συγκεκριμένους τρόπους έκφρασης (εκφράσεις προσώπου, τάσεις συμπεριφοράς, φυσιολογικά πρότυπα). Οι επιστήμονες δεν έχουν ακόμα καταλήξει στο ποια υπό-ομάδα συναισθημάτων αποτελούν τα βασικά συναισθήματα και στο πόσα είναι (κατηγοριοποιήσεις αναφέρουν από δύο έως και εννέα συναισθήματα). Υπάρχει όμως συμφωνία στο ότι τα συναισθήματα του φόβου και του θυμού είναι βασικά, αντίθετα δεν είναι σίγουρο ότι είναι βασικά τα δύο πιο συνήθη συναισθήματα της χαράς και της λύπης. Τα βασικά συναισθήματα είναι σημαντικά γιατί αντιπροσωπεύουν μοναδικές κινητήριες τάσεις και συμπεριφορές και είναι επίσης εύκολα αναγνωρίσιμα από φυσιολογικής πλευράς.



Εικόνα 5. Τρισδιάστατο μοντέλο του Robert Plutchik. Περιγραφή της σχέσης μεταξύ των συναισθημάτων με μια αναλογία χρωμάτων. Η κάθετη διάσταση του κώνου είναι η ένταση του συναισθήματος και ο κύκλος αντιπροσωπεύει το βαθμό ομοιότητας μεταξύ των συναισθημάτων. Στο κεντρικό κύκλο βρίσκονται τα βασικά συναισθήματα όπως ορίζονται στη συγκεκριμένη προσέγγιση και παρουσιάζονται αντιδιαμετρικά ως ζευγάρια αντιθέτων. Τα συναισθήματα που βρίσκονται στο λευκό κομμάτι είναι μίγμα δύο βασικών συναισθημάτων (δύαδες βασικών).

Υπάρχει μια μεγάλη βιβλιογραφία στα χαρακτηριστικά σημάδια που φανερώνουν το συναίσθημα και μέσα στην ψυχολογική επιστήμη και πέρα από αυτή. Το φωνητικό στοιχείο είναι μια από τις θεμελιώδεις εκφράσεις των συναισθημάτων, ίσης αξίας με την έκφραση του προσώπου. Όλα τα θηλαστικά έχουν συναισθήματα και μπορούν να τα μεταβιβάσουν από τα φωνητικά χαρακτηριστικά. Οι άνθρωποι μπορούν να εκφράσουν τα συναισθήματά τους φωνάζοντας, γελώντας, κλαίγοντας, χρησιμοποιώντας επίσης και τα πιο λεπτά χαρακτηριστικά της ομιλίας. Συνεπώς, στην έκφραση και την κατανόηση των συναισθημάτων, διαφορετικοί τύποι πηγών πρέπει να εξεταστούν. Η συναισθηματική εκδήλωση έχει επίσης ένα ευρύ φάσμα

σωματικών συστατικών, συμπεριλαμβανομένου του καρδιακού ρυθμού, της ειδικής αντίστασης του δέρματος, της θερμοκρασίας, της διαμέτρου της κόρης των ματιών και της δραστηριότητας των μυών. Τα παραπάνω έχουν χρησιμοποιηθεί ευρέως για να προσδιορίσουν τις σχετικές με συναισθήματα καταστάσεις όπως για παράδειγμα στην ανίχνευση ψέματος. Εντούτοις, βασικές μελέτες αναπτύσσουν την ιδέα ότι τα σημάδια του συναισθήματος είναι σαφέστερα όταν προσδιορίζονται μέσω της έκφρασης του προσώπου και της φωνής.

2.2.1 Τα δύο κανάλια της ανθρώπινης επικοινωνίας

Η ανθρώπινη ομιλία παρουσιάζει δύο ευρείς τύπους πληροφορίας. Έχουν διαχωριστεί δύο κανάλια επικοινωνίας : Το πρώτο κανάλι μεταδίδει σαφή γλωσσικά μηνύματα, τα οποία μπορεί να αφορούν κάποιο θέμα αλλά και κανένα, και το δεύτερο κανάλι μεταδίδει υποκρυπτόμενα μηνύματα με πληροφορία για τους ίδιους τους ομιλητές (προσωδία) . Η προσωδία είναι μια πολύπλοκη σύνθεση φυσιολογικών (άρθρωση, μηχανισμός παραγωγής ομιλίας) και φωνητικών διεργασιών τα οποία υιοθετούνται από τον ομιλητή προκειμένου να εκφράσει την διάθεση του, να εκφέρει τις υποθέσεις και την προσοχή του. Αποτελεί ουσιαστικά ένα παράλληλο κανάλι πληροφορίας της καθημερινής λεκτικής επικοινωνίας του. Το σημασιολογικό περιεχόμενο ενός γραπτού ή προφορικού μηνύματος καλείται ως δήλωση (denotation), ενώ το συναισθηματικό και η πληροφορία μέσω της έμφασης που προσδίδονται από τον ομιλητή, ή αντιλαμβάνονται από τον ακροατή συγκροτούν το συμπέρασμα (connotation) του μηνύματος. Από την οπτική γωνία του ακροατή, η προσωδία πραγματώνεται μέσα από την συστηματική αντίληψη και ανάκτηση των προθέσεων του ομιλητή βασιζόμενος στα παρακάτω ερεθίσματα:

- *Παύσεις*: για να ορισθούν οι ενδοπεριοδικές φράσεις και για την αποφυγή ελλείψεως αέρα
- *Τονικότητα*: η αντίληψη του ρυθμού ταλάντωσης των φωνητικών χορδών (θεμελιώδης συχνότητα F0) σε συνάρτηση με το χρόνο από τον ακροατή.

- *Κατανομή/σχετική διάρκεια:* διάρκειες φωνημάτων, συγχρονισμός και ρυθμός
- *Ένταση:* σχετικό πλάτος ήχου.

Η τονικότητα είναι εκείνο το συστατικό της προσωδία που μεταφέρει το μεγαλύτερο νοηματικό φορτίο. Καθώς μιλάμε, συστηματικά μεταβάλουμε την θεμελιώδη συχνότητα προκειμένου να εκφράσουμε τα συναισθήματα μας για αυτό το οποίο αναφερόμαστε, ή για να κατευθύνουμε την προσοχή του ακροατή σε συγκεκριμένα σημεία του μηνύματος που εκφέρει. Αν προσπαθήσουμε να εκφέρουμε ένα γραπτό μήνυμα με σταθερή τονικότητα και σταθερές παύσεις (ή χωρίς παύσεις) μεταξύ των λέξεων, το αποτέλεσμα είναι αφύσικο.

Τόσο η γλωσσολογία όσο και η τεχνολογία έχουν επενδύσει μεγάλες προσπάθειες να κατανοήσουν το πρώτο, σαφές κανάλι, αλλά το δεύτερο (προσωδία) δεν είναι και τόσο κατανοητό. Η κατανόηση του συναισθήματος κάποιου άλλου ανθρώπου είναι συνδεδεμένη με το δεύτερο, υποκρύπτον, κανάλι επικοινωνίας. Η προσωδία λέει στους ανθρώπους το πως πρέπει να λάβουν τη πληροφορία που μεταδίδεται από το σαφές κανάλι. Φυσικά σε αυτό το σημείο υπάρχει και η μεγαλύτερη κοινωνική επιρροή στο πως λαμβάνεται το μήνυμα, η υποκρυπτόμενη πληροφορία είναι σχετική για τον καθένα και πολύ περισσότερο για διαφορετικές κουλτούρες. Η σχέση των δύο καναλιών γίνεται πιο φανερή όταν έχουμε να κάνουμε με πρωτεύοντα συναισθήματα που εκφράζονται πλήρως από τον ομιλητή και δεν είναι τόσο φανερή στις μιμησιαστικές εκφράσεις του συναισθήματος. Η γνώση για τον τρόπο που συμβάλει στο κανάλι η προσωδία συνεχώς αυξάνεται και φαίνεται λογικό να εξεταστεί ως μέρος ενός ευρύτερου πεδίου που έχει να κάνει με την κατανόηση της κατάστασης του ομιλητή. Για παράδειγμα η ίδια σειρά λέξεων μπορεί να εκφράσει και να μεταδώσει τελείως διαφορετικά μηνύματα αν ληφθεί υπόψη και η υποκρύπτουσα πληροφορία. Η ίδια φράση μπορεί να ληφθεί ως αστείο, ως ερώτηση, ή ως μια επιθετική πρόκληση (π.χ. «Υποθέτω πως πιστεύεις ότι η ομάδα θα κερδίσει τον αγώνα»). Η γνώση της κατάστασης του ομιλητή σε αυτή τη περίπτωση είναι απαραίτητη για να συνεχιστεί η συζήτηση. Η έρευνα αποδεικνύει επαρκώς ότι ορισμένες συναισθηματικές πληροφορίες εσωκλείονται μέσα στην προφορική επικοινωνία που στηρίζεται μόνο στις ακουστικές πτυχές του λεκτικού σήματος. Μεγάλος αριθμός μελετών έχουν αφιερωθεί στο να καθοριστεί ποιες ακουστικές ιδιότητες, που περιλαμβάνονται στην προφορική έκφραση, είναι οι πιο συμφέροντες υποψήφιοι για να ανιχνεύσουν κατηγορηματικά τη συναισθηματική

κατάσταση του ομιλητή. Μια σημαντική έρευνα είναι αυτή των Murray και Arnott , των οποίων τα αποτελέσματα, παράγουν τις καλύτερες ακουστικές ιδιότητες για την ανίχνευση των βασικών συναισθημάτων, παρουσιάζονται στον πίνακα 1.

	Θυμός	Ευτυχία	Λύπη	Φόβος	Απέχθεια
Ρυθμός ομιλίας	Λίγο γρήγορος	Γρήγορος ή αργός	Λίγο αργός	Πολύ γρήγορος	Πάρα πολύ γρήγορος
Μέση τιμή pitch	Πάρα πολύ υψηλή	Πολύ υψηλή	Λίγο χαμηλή	Πάρα πολύ υψηλή	Πάρα πολύ χαμηλή
Διακύμανση pitch	Πολύ μεγάλη	Πολύ μεγάλη	Λίγη στενή	Πολύ μεγάλη	Λίγο μεγάλη
Ένταση	Υψηλή	Υψηλή	Χαμηλή	Κανονική	Χαμηλή
Ποιότητα φωνής	Φανερές	Φανερές	Ηχηρή	Μη-κανονική φωνή	Μουρμούρα, τονικότητα απο το στήθος
Αλλαγές στο pitch	Απότομες στην έμφαση	Ομαλές, ανοδική διακύμανση	Καθοδική διακύμανση	Κανονικές	Ευρείς καθοδική διακύμανση στα όρια
Αρθρωση	Νευρική	Κανονική	Ελλιπής	Ακριβής	Κανονική

Πίνακας 1 Τα βασικά συναισθήματα και η σχέση τους με τις παραμέτρους ομιλίας (Murray and Arnott, 1996)

Ακουστικές παράμετροι, όπως η F0, η ένταση και η διάρκεια, που έχουν φαινομενικά σαφείς σχέσεις με τα πιο κύρια αντιληπτά χαρακτηριστικά της ομιλίας έχουν λάβει την περισσότερη προσοχή. Αυτές οι παράμετροι τείνουν επίσης να είναι οι ευκολότερες να αναλυθούν. Αντίθετα, είναι λιγότερο μελετημένα τα χαρακτηριστικά του φάσματος του λεκτικού σήματος, όπως αποκαλύπτονται από τη δομή των formants, και τη γενική δομή του μέσου φάσματος. Τέτοιες πτυχές του λεκτικού σήματος, περισσότερο σχετικές με τις ποιοτικές ιδιότητες φωνής, δεν έχουν και τόσο προφανή αντιληπτά συστατικά και είναι δυσκολότερο και χρονοβόρο να αναλυθούν. Εντούτοις, κάποια στοιχεία δείχνουν ότι η προσθήκη των ποιοτικών

γνωρισμάτων της φωνής θα μπορούσε να προσφέρει καλύτερη υπολογιστική δύναμη κατά τη διάκριση μεταξύ των συναισθημάτων συγκρινόμενη με τη διάκριση μόνο με προσωδιακά χαρακτηριστικά.

2.2.2 Η σημασία της προσωδίας στη συναισθηματική ομιλία

Η προσωδία είναι ο κύριος μεταφορέας της κατάστασης του ομιλητή και ως εκ τούτου είναι σημαντική στην ανάκτηση των πληροφοριών που είναι θεμελιώδεις για την επικοινωνία. Ο όρος «προσωδία» αναφέρεται σε όλες τις τάξεις διαφοροποιήσεων που συμβαίνουν στο ύψος της φωνής και στην ενέργεια και που περιέχουν λεκτική σημασία. Κάποιες μελέτες έχουν προσπαθήσει να καταδείξουν πόσο κάποιες ελαφρώς διαφοροποιημένες έννοιες εισάγονται στη λεκτική επικοινωνία από τις προσωδιακές παραλλαγές και πώς οι διαφορές στη μορφή της προσωδίας υποδεικνύουν το βαθμό αβεβαιότητας ή βεβαιότητας όσον αφορά την επίγνωση του ομιλητή, τα συγκεκριμένα συναισθήματα, την ένταση των συναισθημάτων, και τις επιπτώσεις άλλων ταυτόχρονα εμφανιζόμενων συναισθημάτων.

2.2.3 Η σημασία της ποιότητας της φωνής στη συναισθηματική ομιλία

Οι περισσότερες μελέτες, που εστιάζουν στην έκφραση των συναισθημάτων, έχουν επικεντρωθεί στα γνωρίσματα της προσωδίας. Αν και αυτές οι έρευνες παράγουν επιτυχή συναισθηματική έκφραση στη συνθετική ομιλία με το χειρισμό της θεμελιώδους συχνότητας και της διάρκειας, εμφανίζεται μια έλλειψη εκφραστικότητας λόγω της απουσίας ποιοτικών εκτιμήσεων της φωνής. Εκτός από την επιρροή της συναισθηματικής κατάστασης στις προσωδιακές παραμέτρους (F0, ποσοστό ομιλίας), έχει αποδειχθεί ότι η συναισθηματική κατάσταση επηρεάζει και τη μορφή του γλωττιδικού παλμού (π.χ. ανοικτή - κλειστή αναλογία των φωνητικών χορδών), την αναλογία αρμονικότητας προς θόρυβο (HNR), τη φασματική ενεργειακή διανομή και συσχετίζεται με κάποιον τρόπο με ανωμαλίες στη φωνή.

Παρατηρήσεις από έρευνες που έγιναν με σκοπό να ανιχνεύσουν τις ιδιότητες των φυσιολογικών διαδικασιών κατά τη διάρκεια της παραγωγής των διαφορετικών συναισθημάτων υποδεικνύουν ότι το γλωττιδικό άνοιγμα απεικονίζει αρκετά καλά τις διαφορές μεταξύ του θυμού και των ουδέτερων συναισθηματικά καταστάσεων. Επίσης το συναίσθημα της ευτυχίας περιλαμβάνει περισσότερο το στοιχείο της αναπνοής (breathy) σε σύγκριση με αρνητικά ή/και ουδέτερα συναισθήματα. Επομένως, υποθέτεται ότι η ευτυχία παράγεται με μια υψηλότερη υπό-γλωττιδική πίεση από τις θυμωμένες ή ουδέτερες εκδηλώσεις.

Από την άλλη μεριά, το μέγεθος του πρώτου formant εξαρτάται από το βαθμό του γλωττιδικού ανοίγματος κατά τη διάρκεια ενός ανοικτού κύκλου. Οι αναπνοές και η τραχύτητα χρησιμοποιούνται επίσης ως παράμετροι που περιλαμβάνουν το συναίσθημα και υπολογίζονται από την αναλογία αρμονικότητας προς θόρυβο (HNR). Προτάσεις που εκφράζονται με θυμό παρουσιάζουν σημαντικά υψηλότερο HNR από τις ουδέτερες εκφράσεις. Σε μια πειραματική μελέτη που διεξήχθη από τους M. Kienast και W.F. Sendlmeier (2000) αποκαλύφθηκε ότι η ανάλυση των formant των φωνηέντων παράγει χρήσιμα εργαλεία για τη διάκριση συναισθήματος. Οι προτάσεις που εκφράζουν το φόβο, τη θλίψη ή την πλήξη χαρακτηρίζονται από μια μετατόπιση formant προς μια κεντρική θέση σε όλα τα διαφορετικά φωνήεντα.

Τα αποτελέσματα που επιτυγχάνονται με την ανάλυση των formants των προτάσεων με θυμό παρουσιάζουν διαφορές από την ανάλυση για τη θλίψη, το φόβο και την πλήξη. Στις θυμωμένες προτάσεις παρατηρήθηκε μια μετατόπιση των φωνηέντων σε ένα πιο ακραίο μέρος του διαγράμματος και για αρσενικούς και για θηλυκούς ομιλητές. Στις προτάσεις εκφρασμένες με ευτυχία η ανάλυση των formant συχνοτήτων αποκάλυψε μια αύξηση και στο πρώτο και στο δεύτερο formant. Αυτό θα μπορούσε να εξηγηθεί με μια γενική μείωση του μήκους του φωνητικού καναλιού. Ο περιορισμός του μήκους θα μπορούσε να προκληθεί με το άπλωμα των χειλιών όπως π.χ. κατά τη διάρκεια του χαμόγελου στην ευτυχία. Ένας συμπληρωματικός παράγοντας για αυτήν την επίδραση μπορεί επίσης να είναι μια αύξηση του λάρυγγα, η οποία συνοδεύεται συχνά από μια υψηλή θεμελιώδη συχνότητα. Μια θέση αύξησης του λάρυγγα στην ευτυχή ομιλία είναι πιθανή, επειδή η ευτυχής ομιλία αναφέρεται να έχει γενικά μια μεγάλη αύξηση στην θεμελιώδη συχνότητα.

Συμπερασματικά, τόσο τα γνωρίσματα προσωδιακού όσο και ποιοτικού χαρακτήρα της φωνής επηρεάζονται από τη συναισθηματική κατάσταση του ομιλητή. Ο MufRAY αποκαλύπτει

ότι τα γνωρίσματα της καμπύλης της $F0$ είναι οι σημαντικότερες παράμετροι στη διαφοροποίηση μεταξύ των βασικών συναισθημάτων, ενώ η ποιότητα της φωνής είναι σημαντική στη διαφοροποίηση μεταξύ των δευτερευόντων συναισθημάτων. Επίσης σημείωσε ότι τα ακουστικά σημάδια για την αναγνώριση των βασικών συναισθημάτων είναι διαπολιτισμικά και κοινά παγκοσμίως, αντίθετα όμως εκείνα των δευτερευόντων συναισθημάτων είναι εξαρτώμενα και διαφοροποιούνται για κάθε πολιτισμό και λαό.

2.3 Ανίχνευση ήχων εξ' αποστάσεως με οπτοηλεκτρονικές τεχνικές

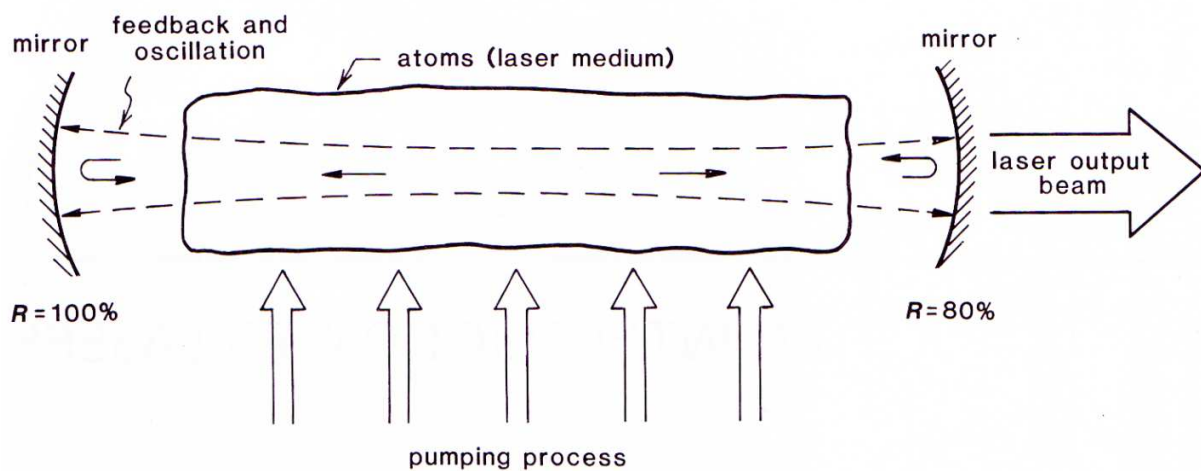
Για την καταγραφή του ήχου έχουν αναπτυχθεί και χρησιμοποιηθεί διάφορες μέθοδοι αναλόγως της ιδιαιτερότητας και του ζητούμενου σε κάθε περίπτωση. Ιδιαίτερο ενδιαφέρον παρουσιάζει η εξ' αποστάσεως καταγραφή του ήχου, όπως, για παράδειγμα, σε περιπτώσεις που η απευθείας καταγραφή του ήχου είναι αδύνατη ή μη-επιθυμητή. Σε τέτοιες περιπτώσεις έχουν βρει εφαρμογή οπτοηλεκτρονικές μέθοδοι ανίχνευσης βασισμένες σε χρήση δέσμης laser. Μια τέτοια τεχνική εφαρμόστηκε στη παρούσα πτυχιακή και εδώ παρουσιάζονται οι βασικές αρχές στις οποίες αυτή βασίζεται. Συγκεκριμένα, παρουσιάζεται η έννοια και λειτουργία ενός laser, και οι δυνατότητες και τρόπος χρήσης τους σε ανίχνευση ήχων με το συγκεκριμένο παράδειγμα της διάταξης εκτροπής δέσμης laser από δονούμενο αντικείμενο, που ήταν και η μέθοδος που ακολουθήθηκε.

2.3.1 Η λειτουργία των lasers

Τα lasers είναι συσκευές που παράγουν και ενισχύουν σύμφωνη ακτινοβολία στην υπέρυθρη, ορατή, ή υπεριώδη περιοχή του ηλεκτρομαγνητικού φάσματος. Τα lasers λειτουργούν χρησιμοποιώντας μια γενική αρχή που αρχικά επινοήθηκε στις συχνότητες μικροκυμάτων, όπου και είχε ονομαστεί *μικροκυματική ενίσχυση* με εξαναγκασμένη εκπομπή ακτινοβολίας ή δράση *maser*. Όταν επεκταθεί σε οπτικές συχνότητες αυτό φυσικά γίνεται ενίσχυση φωτός με εξαναγκασμένη εκπομπή ακτινοβολίας ή δράση *laser* (*light amplification by stimulated emission*

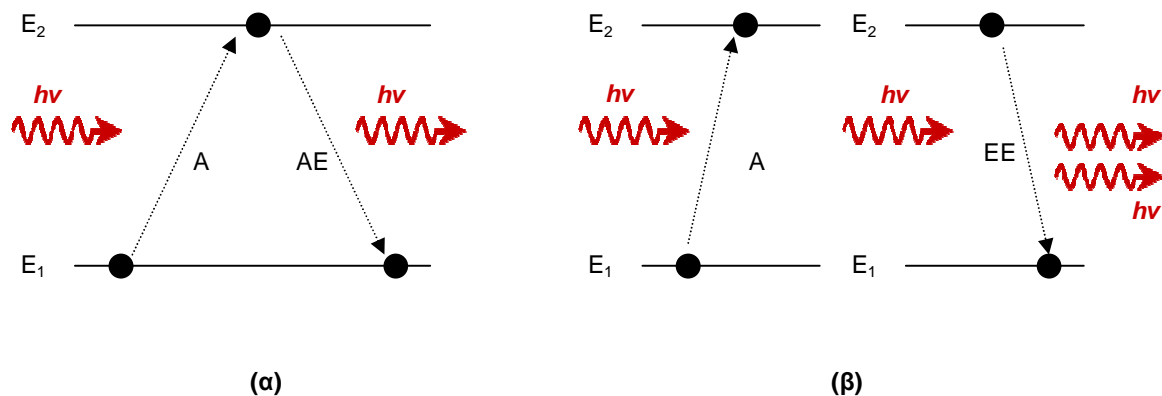
of radiation). Η μελέτη των laser και οι επιστημονικές τους εφαρμογές συχνά αναφέρονται ως το πεδίο της *κβαντικής ηλεκτρονικής*.

Τα lasers έχουν μεγάλη ποικιλομορφία χρησιμοποιώντας πολλά διαφορετικά υλικά, πολλά διαφορετικά ατομικά συστήματα, και πολλές διαφορετικές τεχνικές άντλησης (pumping) και διέγερσης. Η δέσμη της ακτινοβολίας που εκπέμπουν τα laser έχει αξιοσημείωτες ιδιότητες κατευθυντικότητας, φασματική καθαρότητα και ένταση. Αυτές οι ιδιότητες έχουν ήδη οδηγήσει σε μια τεράστια ποικιλία εφαρμογών, και απ' ότι φαίνεται θα συνεχίσουν να αναπτύσσονται νέες στο μέλλον. Τα ουσιώδη στοιχεία μιας συσκευής laser, όπως φαίνεται στην εικόνα 6 είναι : α) το ενεργό υλικό ή μέσο (medium) του laser που μπορεί να είναι στερεό, υγρό ή αέριο, β) μια διαδικασία διέγερσης (άντλησης) σε υψηλότερα ενεργειακά επίπεδα, που μπορεί να γίνει με διάφορους τρόπους, για παράδειγμα οπτικά (μέσω ενός άλλου laser ή ειδικών λαμπών), ηλεκτρικά κτλ., και γ) κατάλληλα στοιχεία (καθρέπτες) οπτικής ανάδρασης που επιτρέπουν μια δέσμη να «αναπηδήσει» εμπρός και πίσω επαναλαμβανόμενα διασχίζοντας το μέσο του laser (σε μια κατάσταση ταλάντωσης) σχηματίζοντας με τα παραπάνω ένα «οπτικό αντηχείο».



Εικόνα 6. Τα ουσιώδη στοιχεία μιας συσκευής laser.

Για την έννοια της εξαναγκασμένης εκπομπής, και ιδίως σε αντιδιαστολή με την αυθόρμητη εκπομπή, ας θεωρήσουμε την εικόνα 7.



Εικόνα 7. (α) Διαδικασία αυθόρμητης εκπομπής (AE), και (β) εξαναγκασμένης εκπομπής (EE). A – απορρόφηση, $h\nu$ – φωτόνιο, με μαύρο κύκλο απεικονίζεται ένα ηλεκτρόνιο.

Στην αριστερή μεριά (εικόνα 7.α) φαίνεται ο μηχανισμός της αυθόρμητης εκπομπής: ένα φωτόνιο ενέργειας $h\nu$, όπου h η σταθερά του Planck και ν η συχνότητά του, προσπίπτει σε ένα μέσο όπου ένα, για παράδειγμα, ηλεκτρόνιο βρίσκεται στη θεμελιώδη στάθμη του, ενέργειας E_1 . Στη θεώρηση που απεικονίζεται, το φωτόνιο αυτό έχει ενέργεια $h\nu = \Delta E = E_2 - E_1$, έτσι ώστε απορροφούμενο (A) από το ηλεκτρόνιο μπορεί να το διεγείρει σε υψηλότερη στάθμη ενέργειας E_2 . Εκεί αν μείνει ανεπηρέαστο το ηλεκτρόνιο θα παραμείνει για κάποιο χρόνο, και έπειτα θα μεταβεί πάλι στην στάθμη E_1 , *αυθόρμητα*, εκπέμποντας ένα φωτόνιο (AE), πάλι ενέργειας $h\nu = \Delta E$.

Στην δεξιά μεριά (εικόνα 7.β) φαίνεται ο μηχανισμός της εξαναγκασμένης εκπομπής που συμβαίνει σε ένα laser: πάλι ένα φωτόνιο ενέργειας $h\nu$, απορροφάται (A) από ένα ηλεκτρόνιο το οποίο και μεταβαίνει σε στάθμη ενέργειας E_2 . Αν κατά το χρόνο που το ηλεκτρόνιο βρίσκεται στην στάθμη ενέργειας E_2 προλάβει και προσπέσει ένα δεύτερο φωτόνιο, τότε αυτό θα «εξαναγκάσει» (EE) το ήδη διεγερμένο ηλεκτρόνιο να μεταβεί στην στάθμη ενέργειας E_1 εκπέμποντας ταυτόχρονα ένα δεύτερο φωτόνιο, πανομοιότυπο με το αρχικό. Με αυτό τον τρόπο «ενισχύεται» η ακτινοβολία (αρχικά: 1 φωτόνιο, τελικά: δύο όμοια φωτόνια).

Αν τώρα το σύστημα αυτό βρίσκεται μέσα σε καθρέπτες που είναι τοποθετημένοι σε κάθε άκρο του, τότε τα φωτόνια αναπηδούν μπρος και πίσω μεταξύ αυτών καθρεπτών, και κάθε φορά που περνούν από το μέσο παράγονται περισσότερα, έχουμε δηλαδή περαιτέρω ενίσχυση κάθε φορά. Εάν η καθαρή ενίσχυση laser ανάμεσα στους καθρέπτες, λαμβάνοντας υπόψη κάθε σκέδαση ή άλλες απώλειες, υπερβαίνει την καθαρή απώλεια ανάκλασης στους ίδιους τους

καθρέπτες, τότε σύμφωνες οπτικές ταλαντώσεις θα δημιουργηθούν στο σύστημα, όπως και σε κάθε άλλο ηλεκτρονικό ταλαντωτή. Θεωρώντας, τώρα, ότι ο ένας από τους δύο καθρέπτες είναι 100% ανακλαστικός ενώ ο άλλος μερικώς ανακλαστικός (π.χ. 80% όπως φαίνεται στην εικόνα 6) τότε μετά από διαδοχικές ανακλάσεις κάποια φωτόνια θα διαπερνούν τον μερικώς ανακλαστικό καθρέπτη, σχηματίζοντας έτσι τη δέσμη laser.

Όπως έχει ήδη αναφερθεί, τα παραπάνω συνιστούν ένα «οπτικό αντηχείο» ή αλλιώς μια κοιλότητα συντονισμού, η οποία παίζει σημαντικό ρόλο στην λειτουργία του laser. Είναι προφανές ότι και λόγω της γεωμετρίας του όλου συστήματος, μόνο εκείνα τα φωτόνια που διαδίδονται πολύ κοντά κατά μήκος του άξονα της κοιλότητας, θα περνάνε έξω από τις πλευρές της κοιλότητας, και θα συνιστούν τη δέσμη laser. Γι αυτό το λόγο οι συσκευές laser εμφανίζουν δέσμες με υψηλό βαθμό κατευθυντικότητας. Επίσης, και λόγω της φύσεως παραγωγής του φωτός, οι πηγές laser εμφανίζουν υψηλό βαθμό συμφωνίας, σχεδόν όλα δηλαδή τα φωτόνια είναι σύμφωνα μεταξύ τους.

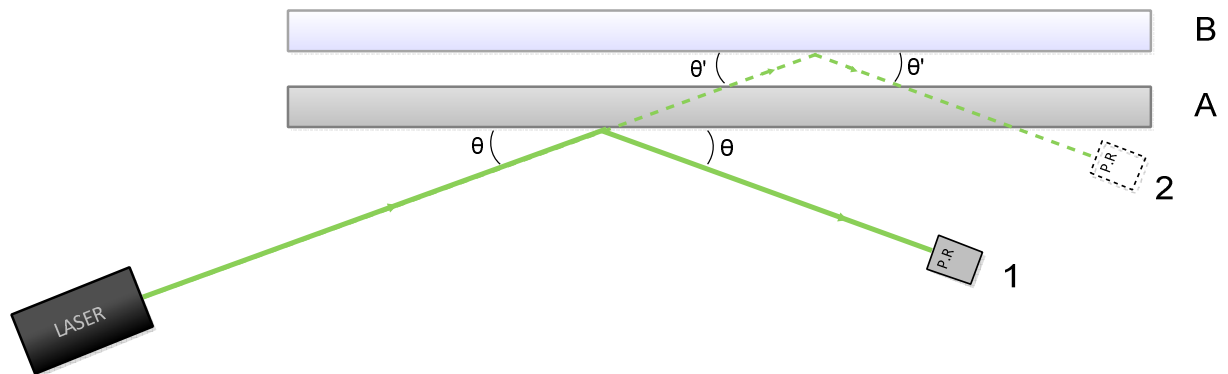
Τέλος, εμφανίζουν υψηλή μονοχρωματικότητα, δηλαδή εκπομπή σε πολύ στενή φασματική γραμμή. Εξάλλου, η διαταραχή που διαδίδεται μέσα στην κοιλότητα παίρνει την μορφή στάσιμων κυμάτων που καθορίζεται από την απόσταση (L) των καθρεφτών. Η κοιλότητα αντηχεί όταν υπάρχει ακέραιος αριθμός (m) μισού μήκος κύματος, που καλύπτει την περιοχή ανάμεσα στους καθρέπτες. Θα πρέπει να υπάρχει ένας δεσμός σε κάθε καθρέπτη, και αυτό μπορεί να γίνει μόνο όταν το L είναι ακέραιο πολλαπλάσιο του $\lambda/2$ (όπου $\lambda = \lambda_0/n$). Επομένως υπάρχει ένας άπειρος αριθμός πιθανόν διαμηκών τρόπων ταλάντωσης της κοιλότητας, ο καθένας με μια ξεχωριστή συχνότητα ν_m . Οι αντηχητικοί τρόποι της κοιλότητας είναι σημαντικά στενότεροι σε συχνότητα από το εύρος ζώνης της κανονικής αυθόρμητης ατομικής μετάδοσης. Αυτοί οι τρόποι, ακόμα και αν η συσκευή είναι κατασκευασμένη ούτως ώστε να υπάρχουν 1 ή 2, θα είναι αυτοί που θα έχουν μια διάρκεια μέσα στην κοιλότητα, και ως εκ τούτου η αναδύομενη δέσμη περιορίζεται σε μια περιοχή κοντά σε αυτές τις συχνότητες. Με άλλα λόγια, η ακτινοβολούσα μετάβαση διαθέτει ένα σχετικά ευρύ φάσμα συχνοτήτων από της οποίες η κοιλότητα θα διαλέξει και θα ενισχύσει μόνο ορισμένες στενές περιοχές και εάν είναι επιθυμητό, ακόμα και μία μόνο περιοχή. Αυτή είναι και η προέλευση και της έντονης μονοχρωματικότητας του laser.

Εκτός της κατηγοριοποίησης των laser ανάλογα με το ενεργό υλικό τους, ένας άλλος σημαντικός διαχωρισμός γίνεται ανάλογα με το αν το φως laser εκπέμπεται σε παλμούς ή όχι. Όταν η εκπομπή του φωτός γίνεται με συνεχή τρόπο, τότε μιλάμε για laser συνεχούς μήκους κύματος (CW). Όταν όμως, μέσω κατάλληλων τεχνικών, υπάρχει εκπομπή σειράς παλμών φωτός, τότε μιλάμε για παλμικά laser. Στην περίπτωση αυτή, η επαναληψιμότητα των παλμών (repetition rate), δηλαδή κάθε πότε εκπέμπεται ένας παλμός, μπορεί να κυμαίνεται από λίγα Hz έως κάποιες δεκάδες MHz. Αυτό όμως που παίζει σημαντικότερο ρόλο, και εν τέλει είναι αυτό που διαφοροποιεί ένα laser σε παλμικό, είναι η διάρκεια του κάθε παλμού laser. Αυτή μπορεί να είναι από λίγα nanosecond (ns, 1 nanosecond = 10^{-9} second) έως λίγα femtosecond (fs, 1 femtosecond = 10^{-15} second). Η μικρή διάρκεια παλμών των παλμικών laser τα καθιστά ικανά να πετυχαίνουν πολύ υψηλή ισχύ και πολύ υψηλές εντάσεις (όταν εστιάζεται η δέσμη).

2.3.2 Ανίχνευση ήχων με οπτοηλεκτρονική διάταξη εκτροπής δέσμης

Οι τεχνικές ανίχνευσης κίνησης μέσω οπτικών διατάξεων laser προσφέρουν μεγαλύτερη χωρική και χρονική διακριτική ικανότητα σε σχέση με τις αντίστοιχες μηχανικές τεχνικές, και είναι τα τελευταία χρόνια σε συνεχή εξέλιξη και μεγάλη χρήση λόγω ακριβώς αυτών των πλεονεκτημάτων τους, που πηγάζουν από τη φύση του φωτός laser. Αυτές οι οπτοηλεκτρονικές τεχνικές ανίχνευσης αλλά και οι οπτικές διατάξεις καθ' εαυτές ποικίλουν ανάλογα με τις ανάγκες και τις ιδιαιτερότητες της ζητούμενης μέτρησης. Στην παρούσα εργασία θα εστιαστεί η προσοχή σε μια από τις πιο απλές αλλά θεμελιώδεις οπτοηλεκτρονικές τεχνικές ανίχνευσης κίνησης μέσω laser, η οποία βασίζεται στην εκτροπή (deflection) μιας δέσμης laser από δονούμενη επιφάνεια διεγερμένη από προσπίπτοντα ηχητικά κύματα. Μια τέτοια διάταξη υλοποιήθηκε στα πλαίσια της παρούσας εργασίας και παρουσιάζεται αναλυτικά στο αντίστοιχο κεφάλαιο.

Η αρχή λειτουργίας μιας τέτοιας διάταξης απεικονίζεται στην εικόνα 8.



Εικόνα 8. Αρχή λειτουργίας διάταξης ανίχνευσης ανακλώμενης δέσμης λέιζερ από δονούμενη επιφάνεια. (A) επιφάνεια σε ηρεμία, (B) επιφάνεια μετατοπισμένη λόγω διέγερσης από ηχητικά κύματα.

Όταν μια δέσμη laser προσπίπτει πάνω σε μια ικανά ανακλαστική επιφάνεια που βρίσκεται σε κατάσταση ηρεμίας (A) τότε η ανακλώμενη δέσμη ακολουθεί πορεία σε γωνία ίση με τη γωνία πρόσπτωσης (1), και εν τέλει ανιχνεύεται από ένα φωτοευαίσθητο στοιχείο, για παράδειγμα από μια φωτοαντίσταση. Εν συνεχεία, εάν ηχητικά κύματα τα οποία παράγονται από μια ηχητική πηγή προσπέσουν επάνω στην επιφάνεια η κατάσταση της αλλάζει από ακίνητη σε δονούμενη (B). Ως συνέπεια της δόνησης αυτής, η ανακλώμενη δέσμη laser θα ακολουθήσει διαφορετική πορεία, ανάλογα με την χρονική στιγμή κατά την οποία θα ανακλαστεί από την δονούμενη επιφάνεια, πάλι όμως υπό την ίδια γωνία ανάκλασης. Αποτέλεσμα αυτού είναι πάνω στο φωτοευαίσθητο στοιχείο να προσπέσει ένα διαφορετικό τμήμα της δέσμης, άρα και να μεταβληθεί το σήμα εξόδου του στοιχείου αυτού. Οι αλλαγές είναι ευθέως ανάλογες με τις αλλαγές λόγω δόνησης, και αφού η δόνηση είναι συνέπεια των προσπιπτόντων ηχητικών κυμάτων στην επιφάνεια, οι αλλαγές θα είναι ανάλογες της φύσεως των ηχητικών κυμάτων. Συμπερασματικά, η εκτροπή της ανακλώμενης δέσμης εμπεριέχει πληροφορία για τον ήχο.

2.4 Βελτίωση ηχητικών καταγραφών

Στην καταγραφή του ήχου συνήθως απαιτείται επεξεργασία του καταγεγραμμένου ηχητικού περιεχομένου, που κατά κύριο λόγο έχει να κάνει με την λεγόμενη αποθορυβοποίηση, δηλαδή την αφαίρεση του θορύβου με ειδικές τεχνικές. Επιπρόσθετα, σε περιπτώσεις καταγραφής ήχου σε κλειστούς χώρους, σημαντικό ρόλο παίζει η αφαίρεση της ανεπιθύμητης αντήχησης, πάλι με ειδικές τεχνικές. Παρακάτω παρουσιάζονται τα βασικά σημεία αυτών των διαδικασιών.

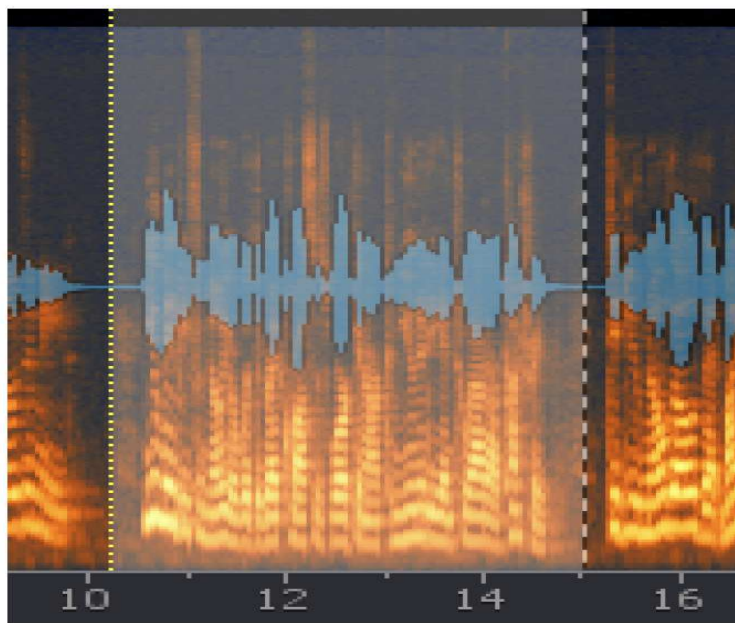
2.4.1 Η αντήχηση και η αντιμετώπιση της

Η αντήχηση είναι ο περιβάλλοντα ήχος της ηχητικής ενέργειας καθώς χάνει ενέργεια στον χώρο. Ο ήχος σε ένα δωμάτιο έχει 2 συνιστώσες: Τον απευθείας ήχο και την αντηχητική ουρά. Η αντηχητική ουρά εξασθενεί σε ενέργεια κατά την πάροδο του χρόνου ανάλογα με διάφορους παράγοντες, όπως το μέγεθος του χώρου, τα υλικά μέσα σε αυτόν, και την γεωμετρία της κατασκευής του. Η αντηχητική ουρά έχει επίσης μερικές σημαντικές χρονικές συνιστώσες: Της πρώιμες ανακλάσεις. Οι πρώιμες ανακλάσεις είναι οι ταχύς απόηχοι ενός απευθείας ήχου από μια κοντινή επιφάνεια. Είναι συχνά διαφορετικοί από την υπόλοιπη αντηχητική ουρά επειδή έχουν πολύ ενέργεια αλλά έχουν γρήγορο τέλος. Οι πρώιμες ανακλάσεις τυπικά αποτελούν τα πρώτα 5 με 10 ms μιας αντηχητικής ουράς. Εάν ένας ακροατής (ή ένα μικρόφωνο) είναι κοντά στην πηγή του άμεσου ήχου, ο άμεσος ήχος γίνεται περισσότερο αντιληπτός από την αντήχηση. Εάν ένας ακροατής είναι πιο μακριά από τον άμεσο ήχο, περισσότερη αντήχηση γίνεται αντιληπτή σε σχέση με τον απευθείας ήχο. Ο ακροατής εάν απομακρύνεται διαρκώς από τον απευθείας ήχο σε έναν χώρο θα διασχίσει τελικά ένα όριο όπου το σήμα με αντήχηση γίνεται ισόποσα αντιληπτό και εμφανές με το άμεσο σήμα.

Ένας αλγόριθμος αφαίρεσης της αντήχησης ,επεξεργάζεται τον ήχο σύμφωνα με την αντηχητική /άμεση αναλογία (wet/dry ratio) που ανιχνεύεται στο σήμα. Χρειάζεται να μάθει πληροφορίες για τον υπο εξέταση ήχο έτσι ώστε να μπορεί να προτείνει ορισμένες ρυθμίσεις

συχνότητας και εξασθένησης του χρόνου, ή μπορεί ο χρήστης να κάνει μια προσωπική εκτίμηση.

2.4.2 Το προφίλ της αντήχησης και η αντηχητική ουρά



Εικόνα 9. Για να βρεθεί η βέλτιστη ρύθμιση για το υπό εξέταση σήμα σύντομα, θα πρέπει να βρεθεί ένα μέρος του σήματος διάρκειας 5 δευτερολέπτων, που ξεκινά με θόρυβο και έχει εξίσου απευθείας και αντηχητικό σήμα.

Ένας αλγόριθμος αφαίρεσης αντήχησης μπορεί να προτείνει κάποιες ρυθμίσεις βασισμένες στο σήμα που έχει εισάγει ο χρήστης. Θα πρέπει να βρεθεί ένα χρονικό πλαίσιο της τάξεως των 5 με 10 sec του σήματος όπου ξεκινά με θόρυβο και έχει άμεσο ήχο αλλά και αντηχητική ουρά.

Το απευθείας σήμα, η αντηχητική ουρά και ο θόρυβος είναι όλα σημαντικά γιατί βοηθάνε τον αλγόριθμο να καταλάβει τον εισαγόμενο ήχο και να ρυθμίσει τους ελεγκτές του κατάλληλα. Θα πρέπει να καταλάβει την αναλογία του "χρήσιμου" σήματος σε σχέση με το σήμα με αντήχηση, το πόσο διαρκεί η αντηχητική ουρά, και που βρίσκεται το δάπεδο θορύβου του σήματος (για να αποφθεχθεί η υπερβολική επεξεργασία).

Η λειτουργία της εκμάθησης αναλύει το σήμα και καθορίζει την αναλογία καθαρού και αντηχητικού σήματος ανά συχνότητα του συνολικού σήματος, καθώς και το συνολικό ποσοστό της εξασθένησης της αντήχησης. Όταν η λειτουργία εκμάθησης ολοκληρωθεί, το προφίλ της αντήχησης και το μήκος της ουράς θα οριστούν αυτόματα στις ιδανικές τιμές τους. Η λειτουργία της εκμάθησης μπορεί να εφαρμοστεί σε οποιοδήποτε αντηχητικό ήχο. Εάν παρουσιαστεί πρόβλημα στην εύρεση καλών αποτελεσμάτων μετά την χρήση της λειτουργίας της εκμάθησης μπορεί να δοκιμαστεί :

- Η εκμάθηση πάνω σε παροδικά (transients) ευρυζωνικά σήματα, όπως τύμπανα, χειροκρότημα ή βήχα
- Η εκμάθηση πάνω σε οποιοδήποτε ηχητικό σήμα είναι αντηχητικό
- Η εκμάθηση πάνω σε μεγαλύτερο χρονικό διάστημα : Οι περισσότερες αντηχήσεις μπορούν να αναλυθούν μέσα σε μερικά δευτερόλεπτα, αλλά μερικές αντηχήσεις μπορεί να απαιτούν ανάλυση μέχρι και 10 sec .

Το προφίλ της αντήχησης ελέγχει το ποσό της δράσης του αλγόριθμου αφαίρεσης που εφαρμόζεται ανά ζώνη συχνοτήτων. Αυτές οι ρυθμίσεις θέτονται αυτόματα απο την ρύθμιση της εκμάθησης. Εάν μια ομάδα απο αντηχητικούς τόνους είναι περισσότερο προεξέχουσα σε ένα σήμα, θα πρέπει να αυξηθεί η ο έλεγχος για την συγκεκριμένη ζώνη. Γενικότερα θα πρέπει οι συγκεκριμένες ρυθμίσεις να έχουν τεθεί έτσι ώστε να ταιριάζουν με την αντήχηση που είναι εξ αρχής παρούσα στο σήμα. Για παράδειγμα εάν η αντήχηση χρειάζεται περισσότερο χρόνο για να παρακμάσει (ή είναι πιο εμφανής) σε μια συγκεκριμένη ζώνη, θα πρέπει η ρύθμιση να τεθεί σε μεγαλύτερα επίπεδα. Επίσης, αυτές οι ρυθμίσεις μπορούν να χρησιμοποιηθούν για να αντιμετωπισθούν περισσότερο οι συντονισμένες ομάδες ή αυτές με το χαρακτηριστικό "κουδούνισμα". Το μήκος της αντηχητικής ουράς ελέγχει την εξασθένηση της αφαίρεσης της αντήχησης. Αυτή η ρύθμιση είναι μια προσέγγιση του RT60, το ποσοστό του χρόνου που χρειάζεται για ένα αντηχητικό σήμα να μειωθεί σε πλάτος κατά 60 dB. Αυτό ορίζεται αυτομάτως απο την ρύθμιση της εκμάθησης. Εάν οι αντηχητικές ουρές επιστρέφουν μετά την επεξεργασία ή οι πρώιμες ανακλάσεις είναι πολύ εμφανής, θα πρέπει να αυξηθεί αυτή η ρύθμιση ενώ εάν οι αντηχητικές ουρές, τα δάπεδα θορύβου και γενικά ο ήχος ακούγονται υπέρ-

επεξεργασμένα και "θαμπά", θα πρέπει να μειωθεί η ρύθμιση. Αυτή η ρύθμιση μπορεί να τεθεί στην ελάχιστη τιμή για να επεξεργαστούν οι πρώιμες ανακλάσεις.

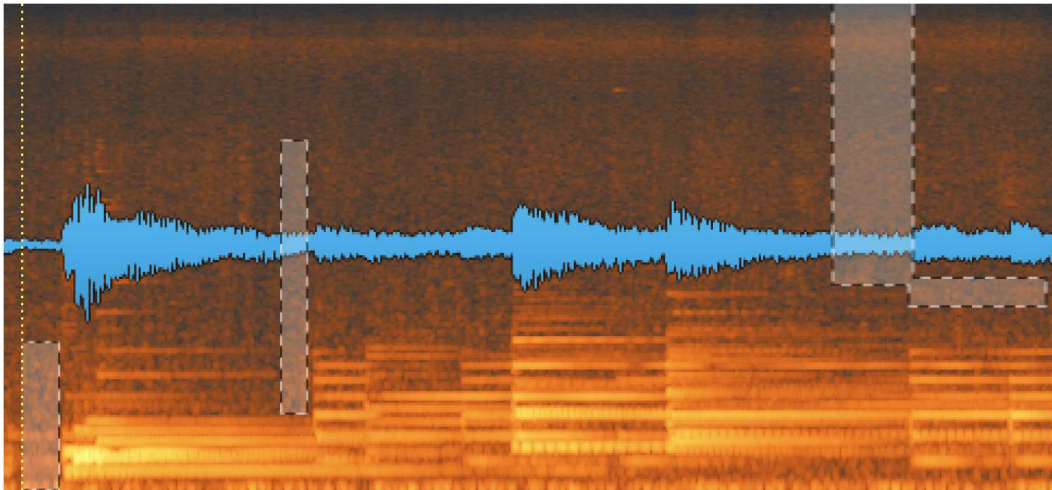
2.4.3 Αποθορυβοποίηση

Ο θόρυβος είναι ένα ανεπιθύμητο ηλεκτρικό σήμα το οποίο συνυπάρχει με το ακουστικό σήμα. Είναι ένας ήχος ανεξακρίβωτου ύψους, φτιαγμένος από τυχαίους συνδυασμούς όλων των συχνοτήτων του ακουστικού και του μη ακουστικού φάσματος. Ο θόρυβος έχει μη περιοδική κυματομορφή γιατί δεν παρέχει ενδείξεις για το ύψος του και είναι είτε ενδογενής δηλαδή παράγεται μέσα στα ηλεκτρονικά κυκλώματα και συσκευές, είτε εξωγενής οπότε προέρχεται από εξωτερικές πηγές. Ο θόρυβος επίσης διαχωρίζεται σε 2 κύριες κατηγορίες. Αυτή των τυχαίων θορύβων και αυτή των θορύβων στατικού χαρακτήρα όπως ο θόρυβος τύπου hiss και hum όπου συναντιούνται στην ακουστική ηλεκτρονική.

Αρκετοί αλγόριθμοι αποθορυβοποίησης δίνουν την δυνατότητα ελέγχου του στατικού θορύβου σε μία ηχογράφιση. Ο στατικός θόρυβος δεν αλλάζει σημαντικά σε στάθμη ή φασματική μορφή καθ' όλη την διάρκεια της ηχογράφησης. Ο στατικός θόρυβος μπορεί να δημιουργείται από "σφύριγμα" ταινίας (tape hiss), μικροφωνικό βόμβο (hum), θόρυβο από κινητήρα κάμερας, ανεμιστήρα και από συστήματα κλιματισμού. Επίσης, δίνει τον αναγκαίο έλεγχο για την αντιμετώπιση προβληματικού θορύβου με ξεχωριστούς ελεγκτές για τονικό και ευρυζωνικό θόρυβο, διαχείριση των τεχνουργημάτων αποθορυβοποίησης (denoising artifacts), και μια διεπαφή(interface) επεξεργασίας για τον έλεγχο της μείωσης του θορύβου σε όλο το φάσμα συχνοτήτων. Ο υπό εξέταση αλγόριθμος χρησιμοποιεί έναν φασματικό αποθορυβοποιητή. Χτίζει ένα φασματικό "αποτύπωμα" του θορύβου, και αφαιρεί αυτόν τον θόρυβο όταν η συχνότητα ενός σήματος πέφτει κάτω από το όριο που έχει καθοριστεί από το φασματικό αποτύπωμα.

2.4.4 Το προφίλ της αποθουροβοποίησης

Η αποθουροβοποίηση χρειάζεται να "μάθει" τον τύπο του θουρύβου που επρόκειτο να αφαιρεθεί από την ηχογράφηση για να δώσει τα βέλτιστα αποτελέσματα. Για να εκπαιδευτεί η μονάδα αποθουροβοποίησης, θα πρέπει να προσδιοριστεί ένα τμήμα της ηχογράφησης που περιέχει μόνο θούρυβο, χωρίς κανένα χρήσιμο ηχητικό σήμα. Συχνά, αυτά τα σημεία είναι στην αρχή και στο τέλος ενός αρχείου, αλλά μπορεί επίσης να είναι κατά την διάρκεια παύσεων της ομιλίας. Θα πρέπει να σημειωθεί πως είναι δυνατόν να δημιουργηθεί ένα προφίλ θουρύβου από πολλαπλές απομονωμένες περιοχές. Αυτό είναι χρήσιμο εάν υπάρχει ένα αρχείο όπου είναι αδύνατο να επιλεγεί αρκετός ήχος όπου μόνο ο θούρυβος είναι παρόν για να χτίσει το προφίλ. Για παράδειγμα, εάν πρέπει να αποκατασταθεί ένα αρχείο όπου κάποιος μιλάει κατά την διάρκεια ενός θουρύβου, υπάρχει η δυνατότητα να επιλεγεί θούρυβος σε συχνότητες όπου υπάρχει απουσία ομιλίας. Μετά την δημιουργία ενός καλού προφίλ, θα πρέπει να προσαρμοστεί το ρυθμιστικό της μείωσης εωσότου ο θούρυβος στο υπο-εξέτασιν σήμα να είναι στην επιθυμητή στάθμη. Θα πρέπει να εφαρμοστεί μια όσο το δυνατόν μικρή μείωση και αργά να ανέβει το επίπεδο μείωσης μέχρι το "δάπεδο" θουρύβου να είναι σε ένα αποδεκτό επίπεδο.

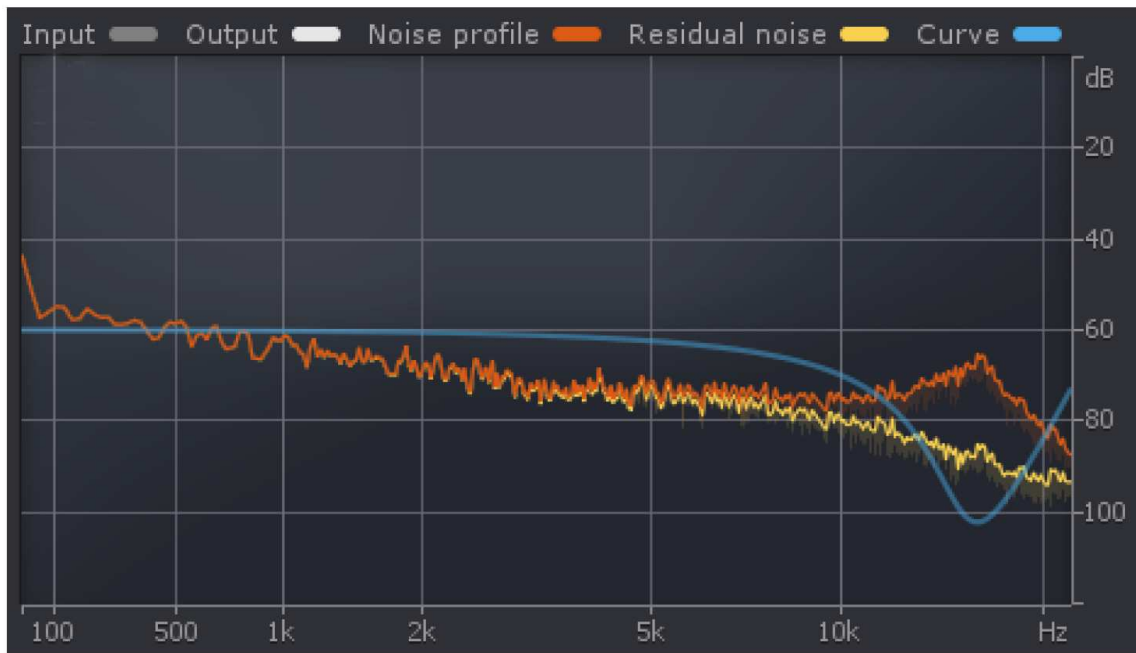


Εικόνα 10. Διαλογή πολλαπλών τμημάτων θουρύβου για την κατασκευή ενός καλύτερου προφίλ θουρύβου.

2.4.5 Η καμπύλη μείωσης και το κατώφλι αποθορυβοποίησης

Μόλις επιτευχθεί ένα καλό ποσό μείωσης, θα πρέπει να προσαρμοστεί και η ρύθμιση για το κατώφλι αποθορυβοποίησης. Κατά τη ρύθμιση του κατωφλίου, θα πρέπει ο χρήστης να εντοπίσει τυχόν τεχνουργήματα (artifacts) που δημιουργούνται λόγω των αλγορίθμων επεξεργασίας και αποθορυβοποίησης. Χαμηλές τιμές κατωφλίου θα αντιμετωπίσουν το περισσότερο σήμα ως χρήσιμο σήμα, και είναι χρήσιμες αν είναι αναγκαίο να διατηρηθεί ηχητικό υλικό που υποχωρεί σταδιακά σε "ήσυχο θόρυβο" (όπως οι ορχηστρικές ουρές). Υψηλές τιμές κατωφλίου είναι χρήσιμες για την επίτευξη περισσότερης μείωσης σε ήχους παροδικής φύσης (όπως κρουστά ή φωνή).

Εάν τα αποτελέσματα χρειάζονται βελτίωση, μπορούν να γίνουν προσαρμογές στο ποσό της αποθορυβοποίησης κατά μήκος του συχνοτικού φάσματος. Αυτό βοηθά στην βελτίωση της σαφήνειας σε υψηλές συχνότητες, την αντιμετώπιση βόμβων χαμηλής συχνότητας και γενικός προσαρμόζει την επεξεργασία του ήχου στις ανάγκες του χρήστη. Ο χρήστης μπορεί να σχεδιάσει μια χρήσιμη καμπύλη μείωσης δημιουργώντας ένα-δύο σημεία πάνω στο φάσμα και στην συνέχεια να προσαρμόσει τις θέσεις τους. Χαμηλότερα σημεία χαμηλώνουν το "δάπεδο" θορύβου ενώ υψηλότερα σημεία το ανεβάζουν. Η παράμετρος του κατωφλίου ελέγχει τον διαχωρισμό πλάτους των στάθμεων του θορύβου και του χρήσιμου σήματος. Υψηλότερες ρυθμίσεις κατωφλίου μειώνουν περισσότερο θόρυβο, αλλά καταστέλλουν και συνιστώσες που βρίσκονται σε σήματα χαμηλών στάθμεων. Ένα χαμηλότερο κατώφλι διατηρεί τις λεπτομέρειες που βρίσκονται στα χαμηλά επίπεδα στάθμεων, αλλά υπάρχει περίπτωση να οδηγήσει σε διαμόρφωση του θορύβου από το σήμα. Η ανύψωση του κατωφλίου μπορεί να γίνει ξεχωριστά για τονικά και τυχαία κομμάτια θορύβου. Μια καλή προεπιλογή είναι τα 0 dB.



Εικόνα 11. Η παραπάνω καμπύλη μείωσης απεικονίζει στην τιθάσευση ενός επίμονου θορύβου στις υψηλές συχνότητες χωρίς να επηρεάζεται το υπόλοιπο αρχείο ήχου.

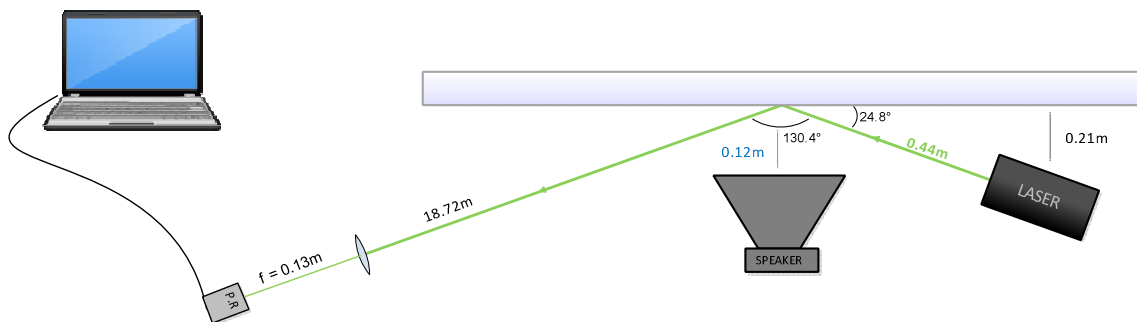
Κεφάλαιο 3^ο

Πειραματικές τεχνικές και διαδικασία

Στην παρόν κεφάλαιο αναλύεται η υλοποιηθείσα οπτοηλεκτρονική πειραματική τεχνική ανίχνευσης ήχων εξ αποστάσεως, η διαδικασία που ακολουθήθηκε στην καταγραφή του εκπεμπόμενου ηχητικού υλικού, καθώς επίσης και τα ιδιαίτερα γνωρίσματα αυτού. Εν συνεχεία πραγματοποιείται ένας διαχωρισμός των περισσότερο καταληπτών ηχητικών υλικών, εφαρμόζονται οι αλγόριθμοι αποθορυβοποίησης και αφαίρεσης αντήχησης, και περιγράφεται η πειραματική διαδικασία των στατιστικών ακροάσεων.

3.1 Οπτοηλεκτρονική διάταξη και διαδικασία μετρήσεων

Για τις πειραματικές μετρήσεις ανίχνευσης φωνής εξ' αποστάσεως υλοποιήθηκε οπτοηλεκτρονική διάταξη (εικόνα 12) βασισμένη στην εκτροπή δέσμης λέιζερ από δονούμενη επιφάνεια.



Εικόνα 12. Η οπτοηλεκτρονική πειραματική διάταξη που υλοποιήθηκε. PR – φωτοαντίσταση, f – συγκλίνων φακός

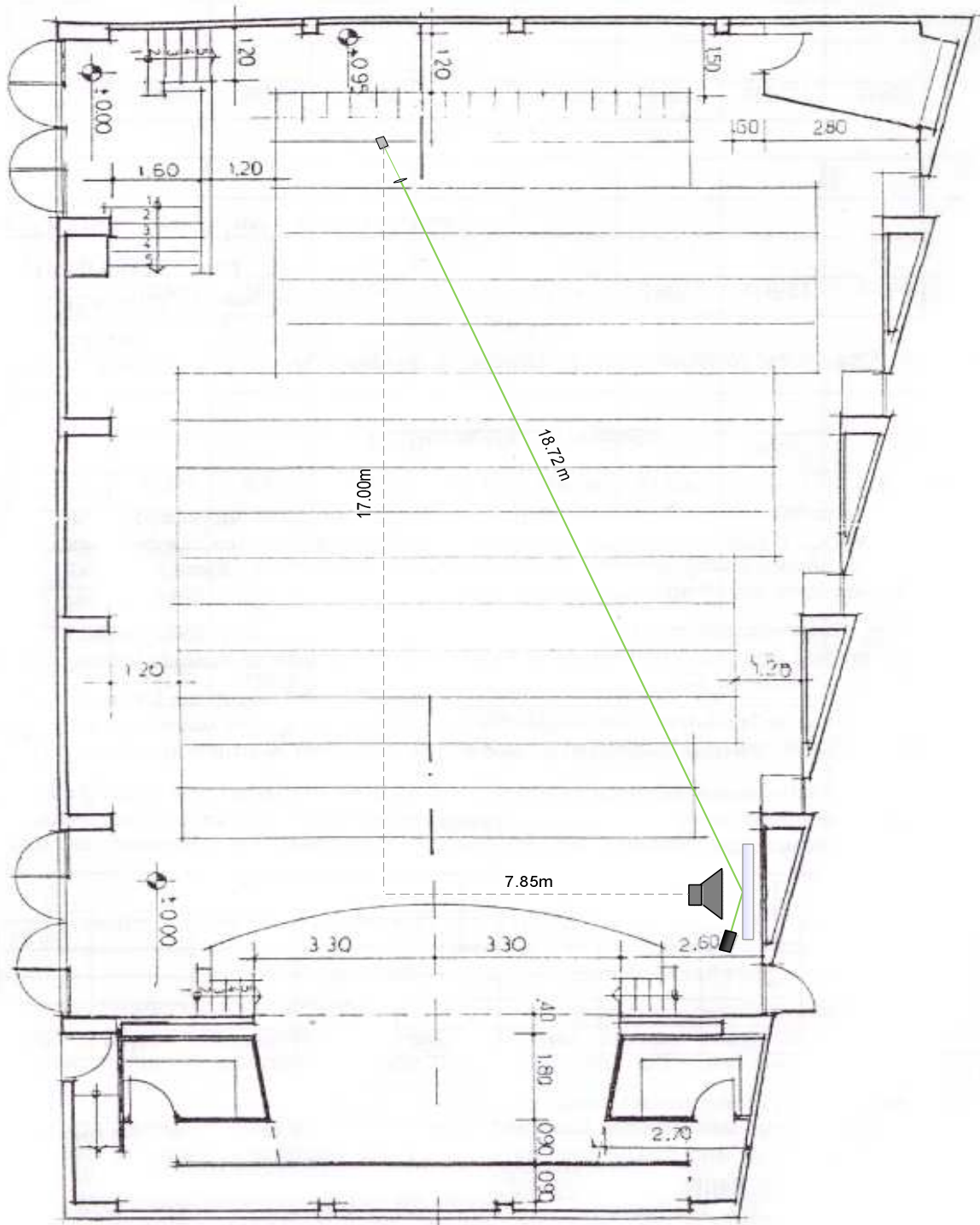
Χρησιμοποιήθηκε laser στερεού μέσου (Nd:YAG) διοδικής άντλησης, συνεχούς μήκους κύματος (CW), με ονομαστική μέση ισχύ 500 mW και μήκος κύματος $\lambda = 532\text{ nm}$. Για την

ανίχνευση της δέσμης laser χρησιμοποιήθηκε εμπορικά διαθέσιμη φωτοαντίσταση Θειούχου Καδμίου (CdS) συνδεδεμένη μέσω κατάλληλου βύσματος με την είσοδο «Line In» φορητού Η/Υ. Επί προσθέτως χρησιμοποιήθηκε συγκλίνων φακός εστιακής απόστασης $f = 15 \text{ cm}$.

Οι μετρήσεις πραγματοποιήθηκαν στο αμφιθέατρο του Τμήματος Μηχανικών Μουσικής Τεχνολογίας και Ακουστικής, και στην εικόνα 13 δίνεται αρχιτεκτονική κάτοψη του αμφιθεάτρου που περιέχει την χωροθέτηση της διάταξης και διαστασιολογικές πληροφορίες.

Καθώς στο αμφιθέατρο δεν υπήρχε ανακλαστική επιφάνεια (π.χ. υάλινο παράθυρο) χρησιμοποιήθηκε για αυτό το σκοπό διπλό διάφανο plexiglass, , διαστάσεων $0.80\text{m} \times 1.10\text{m} \times 0.003\text{m}$, πυκνότητας 0.118kg/m^3 . Το plexiglass στηρίχθηκε κατάλληλα σε υπάρχουσα επιφάνεια, και με τρόπο ώστε να είναι ελεύθερο να δονείται απρόσκοπτα λόγω της πρόσπτωσης σε αυτό ηχητικών κυμάτων με ελεγχόμενο τρόπο.

Δοκιμάστηκαν διάφορες γεωμετρίες (σχετικές θέσεις ηχητικής πηγής / ανακλαστικής επιφάνειας / προσπίπτουσας δέσμης laser) με την βέλτιστη γεωμετρία να παρουσιάζεται στην εικόνα 12. Τέλος, θα πρέπει να σημειωθεί πως οι μετρήσεις πραγματοποιήθηκαν σε συνθήκες σχετικής συσκότισης, ώστε να ελαχιστοποιηθεί η επίδραση του περιβάλλοντος φωτισμού στις μετρήσεις.



Εικόνα 13. Αρχιτεκτονική κάτοψη του αμφιθεάτρου με την χωροθέτηση της πειραματικής διάταξης.

Χρησιμοποιήθηκε ηχείο 2 δρόμων (μοντέλο 8030A εταιρείας Genelec) για την εκπομπή αρχείων ήχου προς την ανακλαστική επιφάνεια, δια μέσου ενός φορητού μέσου αναπαραγωγής μουσικής (iPod 5^{ης} γενιάς, εταιρείας Apple), το οποίο ήταν ρυθμισμένο να λειτουργεί σε «επίπεδη» κατάσταση (flat). Συνέπεια της πρόσπτωσης ήχου στην ανακλαστική επιφάνεια ήταν η κίνηση αυτής ακολουθώντας τις διακυμάνσεις (συχνοτικό περιεχόμενο) του ήχου. Αυτό μεταφράζεται σε αλλαγές του φωτός λέιζερ που προσπίπτει πάνω στη φωτοαντίσταση οι οποίες και καταγράφονται από το φορητό Η/Υ σε μορφή αρχείων .wav με συχνότητα δειγματοληψίας 44100 Hz και 24 bit (2^{24} στάθμες κβαντισμού)

Τα αρχεία ήχου ήταν αποσπάσματα ομιλίας μορφής .wav και προέρχονταν από πλήρως χαρακτηρισμένη βάση δεδομένων ομιλίας (GrES). Αυτή η βάση δεδομένων εμπεριείχε 5 φακέλους όπου ο κάθε φάκελος εξέφραζε και ένα ξεχωριστό συναίσθημα. Η κάθε συναισθηματική κατάσταση (θυμός, φόβος, χαρά, λύπη, ουδέτερο) είχε συγκεντρωμένα 40 .wav αρχεία ήχου με ουδέτερες προτάσεις ως προς το κείμενο για να μην επηρεάσει την κρίση του ακροατή το περιεχόμενο του κειμένου αλλά ο τρόπος παραγωγής της πρότασης.

Αυτές οι 40 προτάσεις όπου είχαν ηχογραφηθεί και για τις 5 συναισθηματικές καταστάσεις από το ίδιο άτομο (γυναίκα ομιλία), συνενώθηκαν σε ένα ενιαίο .wav αρχείο για κάθε συναισθηματική κατάσταση μέσω κατάλληλου λογισμικού και εν συνεχεία τροφοδοτήθηκαν στο ηχείο έχοντας ως αποτέλεσμα να εκπέμπεται η μια πρόταση μετά την άλλη κρατώντας τους χρόνους απουσίας της ομιλίας στην αρχή και στο τέλος κάθε πρότασης όπως ήταν τα αρχεία πριν την ενιαία συνένωση.

3.2 Επεξεργασία καταγεγραμμένου ηχητικού υλικού

Τα αρχεία που προέκυψαν με την εξ' αποστάσεως καταγραφή μέσω της τεχνικής deflection, χρησιμοποιήθηκαν για περαιτέρω επεξεργασία και ανάλυση με σκοπό την εξαγωγή χρήσιμων δεδομένων. Κάθε ξεχωριστής συναισθηματικής κατάστασης αρχείο .wav, εισήχθη σε εμπορικά διαθέσιμο λογισμικό σε κατάλληλα σχεδιασμένο κώδικα κοπής-διαχώρισης, με σκοπό τον ξεχώρισμα των αρχικών συνιστωσών που αποτελούσαν το συνενωμένο αρχείο. Την εξαγωγή δηλαδή 40 .wav αρχείων για κάθε συναισθηματική κατάσταση τα οποία θα ήταν πανομοιότυπα

με τα αρχικά (αυθεντικά) ως προς τους χρόνους απουσίας της ομιλίας στην αρχή και το τέλος κάθε πρότασης, όπως ήταν δηλαδή πριν την ενιαία συνένωση. Θα πρέπει να σημειωθεί πως αμέσως μετά τον διαχωρισμό των αρχείων, εφαρμόστηκε μια επεξεργασία κανονικοποίησης στα προκύπτοντα σήματα που επέτρεψε την μεγιστοποίηση των τιμών πλάτους χωρίς την παρουσία ψαλιδισμού. Η παραπάνω διαδικασία της κοπής-διαχώρισης, είχε ως στόχο την δημιουργία μιας αντίστοιχης βάσης δεδομένων ομιλίας όπως αυτή της GrES που χρησιμοποιήθηκε στο πείραμα. Συνεπώς, συγκρίνοντας τις δύο βάσεις δεδομένων συναίσθημα προς συναίσθημα και αρχείο προς αρχείο θα υπάρχει πλήρης αντιστοίχιση και ταύτιση της χρονικής διάρκειας, του συναισθήματος και του λεξιλογικού περιεχομένου.

3.2.1 Διαχωρισμός ηχητικού υλικού ως προς την καταληπτότητα της ομιλίας

Έπειτα από την ολοκλήρωση της παραπάνω διαδικασίας και της δημιουργίας μιας επεξεργασμένης βάσης ομιλίας, πραγματοποιήθηκε μια αξιολόγηση των αρχείων που προέκυψαν ως προς την καταληπτότητα και την μη-καταληπτότητα τους, από τον συγγραφέα. Θα πρέπει σε αυτό το σημείο να σημειωθεί πως ο αρχικός διαχωρισμός σε καταληπτό και μη καταληπτό ηχητικό υλικό έγινε από τον γράφοντα με βάση την υποκειμενική του αντίληψη και αξιολόγηση. Εν τέλει, πραγματοποιήθηκε ο διαχωρισμός των αρχείων σε καταληπτά και μη-καταληπτά εξετάζοντας κάθε συναίσθημα και μοιράζοντας τα αρχεία που αποτελούσαν το κάθε συναίσθημα, στην κατηγορία που άνηκε το καθένα.

Μεγάλο ρόλο στην καταληπτότητα έπαιξε το συχνοτικό περιεχόμενο των αρχείων. Παρατηρήθηκε πως σε συναισθήματα με χαμηλό συχνοτικό περιεχόμενο και μέτρια ένταση όπως αυτό του ουδέτερου, του φόβου και μερικός της λύπης, ήταν δυσκολότερη η καταληπτότητα στην ακρόαση λόγω της έλλειψης συχνοτικής καθαρότητας που τα διέκριναν σε εκείνη την περιοχή και σε συναισθήματα με υψηλή ένταση και έντονη παρουσία υψηλού συχνοτικού περιεχομένου παρατηρήθηκε μια σχετική καταληπτότητα. Εδώ θα πρέπει να επισημανθεί πως την καταληπτότητα ενός αρχείου δεν την έκρινε μονάχα η παράμετρος του συναισθήματος.

Υπήρξαν πολλές περιπτώσεις αρχείων που άνηκαν για παράδειγμα, στην ουδέτερη συναισθηματική κατάσταση όπου το λεξιλογικό περιεχόμενο τους περιείχε έντονες συλλαβές με ταχύτατη ατάκα έχοντας ως αποτέλεσμα να κάνουν διακριτή ολόκληρη την λέξη και πολλές φορές ολόκληρη την πρόταση συνήθως χρησιμοποιώντας την λογική των συμφραζόμενων παρότι υπήρχε μια γενικότερη υποβάθμιση της διακριτότητας λόγω της συχνοτικής ασάφειας που αναφέρθηκε παραπάνω, όπως επίσης υπήρξαν περιπτώσεις όπου αρχεία που άνηκαν στις πιο διαυγείς κατηγορίες σαν αυτές της χαράς και του θυμού έχαναν την καταληπτότητα τους λόγω κάποιων λέξεων που εκ φύσεως περιέχουν πολλές ηχηρές συλλαβές ενώ η γενική ηχητική εικόνα τους δεν προμήνυε κάποια υποβάθμιση διακριτότητας.

Τα αρχεία που προέκυψαν από την ηχογράφιση μέσω της υλοποιηθείσας διάταξης, καταγράφηκαν σε ένα μεγάλο χώρο και αυτό είχε ως αποτέλεσμα την έντονη αντήχηση η οποία εμφανίστηκε στα τελικά αρχεία. Πραγματοποιώντας μια σύντομη ακρόαση μπορεί κάποιος να διαπιστώσει την έντονη παρουσία πρώιμων ανακλάσεων στις συλλαβές της ομιλίας και μιας μικρής αντηχητικής ουράς στις παύσεις και στις ολοκληρώσεις των προτάσεων. Επιπροσθέτως, η φωτοαντίσταση που χρησιμοποιήθηκε ήταν αρκετά χαμηλού κόστους με αποτέλεσμα να προσθέτει ένα μικρό "φύσημα" θορύβου όπου θα μπορούσε κάποιος να το εκλάβει ως θόρυβο βάθους, που όμως δεν επηρέαζε σημαντικά την καταληπτότητα. Επίσης, το ηχείο αν και εξέπεμπε σε μια επίπεδη κατάσταση όσο αναφορά το συχνοτικό φάσμα, λόγω της μεγάλης έντασης εκπομπής των ηχητικών κυμάτων αλλά και τις αρκετά κοντινής απόστασης του με την επιφάνεια δόνησης, ενίσχυε περαιτέρω τις ήδη ηχηρές συλλαβές των φράσεων έχοντας ως διαδοχικό αποτέλεσμα την ενίσχυση της γενικής αντήχησης και της μοιραίας πρόσθεσης της στα τελικά αρχεία. Όλες οι παραπάνω διαπιστώσεις έπαιξαν σημαντικό ρόλο στην διαμόρφωση του τελικού ηχητικού αποτελέσματος των ηχογραφημένων αρχείων κάτι που επηρέασε τελικώς την καταληπτότητα τους και συμπερασματικά την κριτική ικανότητα του γράφοντα για τον διαχωρισμό τους στις δύο κατηγορίες.

Αξιοσημείωτο είναι το γεγονός το οποίο υποδεικνύει πως κατά τον διαχωρισμό των αρχείων όλων των συναισθημάτων, ενώ σε όλες τις συναισθηματικές καταστάσεις υπήρχαν καταληπτά και μη-καταληπτά αρχεία, σε αυτή του θυμού παρουσιάστηκε καταληπτότητα σε όλα τα αρχεία ανεξαρτήτως και έτσι εντάχθηκε ολόκληρη στην κατηγορία καταληπτής ομιλίας. Αυτό οφείλεται στην μεγάλη εντασιακή στάθμη εκπομπής και στην ενισχυμένη ζώνη υψηλού

συχνοτικού περιεχομένου που εμπεριέχει το συναίσθημα του θυμού στην παραγωγή της φωνής χωρίς την παρουσία ηχηρών συχνοτήτων στην χαμηλή και μεσαία περιοχή που θα υποβάθμιζαν όπως προαναφέρθηκε το αποτέλεσμα, καθώς επίσης και στις απότομες αλλαγές της διακύμανσης του pitch που βοηθούν το ανθρώπινο σύστημα ακοής στην αποσαφήνιση των προτάσεων.

3.2.2 Εφαρμογή αλγόριθμων αφαίρεσης αντήχησης και αποθορυβοποίησης

Μετά την ολοκλήρωση της συγκέντρωσης των αρχείων στις δύο βασικές κατηγορίες, οργανώθηκαν οι εργασίες για την αφαίρεση της αντήχησης και της αποθορυβοποίησης τους. Για την διαδικασία της αφαίρεσης της αντήχησης όπως και της αποθορυβοποίησης χρησιμοποιήθηκε εμπορικά διαθέσιμο λογισμικό επεξεργασίας ήχου. Οι ρυθμίσεις στην αφαίρεση της αντήχησης τέθηκαν ξεχωριστά για κάθε συναισθηματική κατάσταση λόγω της ιδιαιτερότητας του κάθε συναισθήματος στην συχνότητα, την ένταση, γενικώς στα προσωδιακά χαρακτηριστικά της ομιλίας του αλλά και της συμπεριφοράς του σε έναν μεγάλο αντηχητικό χώρο. Για παράδειγμα, ένα προφίλ αντήχησης που έχει βρεθεί να λειτουργεί για το συναίσθημα του θυμού δεν είναι σίγουρο ότι θα λειτουργεί και για το ουδέτερο συναίσθημα ή το συναίσθημα της χαράς.

Γι αυτό τον λόγο δημιουργήθηκαν 5 προφίλ αντήχησης για τις 5 ομάδες συναισθημάτων. Για κάθε συναίσθημα επιλέχθηκε ένα αρχείο ομιλίας διαμέσου του οποίου θα βρισκόταν το προφίλ της αντήχησης που θα αντιπροσώπευε όλα τα αρχεία του συναισθήματος. Αυτό το αρχείο "αντιπρόσωπος" θα έπρεπε να τηρεί της προϋποθέσεις για την δημιουργία ενός λειτουργικού προφίλ αντήχησης όπως αναφέρθηκε στην θεωρητική ενότητα της αφαίρεσης της αντήχησης καθώς επίσης και θα έπρεπε να έχει τα πιο χαρακτηριστικά στοιχεία του συναισθήματος που αντιπροσωπεύει ως προς την προσωδία αλλά και ως προς την ποιότητα της φωνής. Μετά την εύρεση του προφίλ αντήχησης για κάθε συναισθηματική κατάσταση ξεχωριστά με την χρήση της παραπάνω μεθόδου, αποθηκεύθηκαν οι ρυθμίσεις με κατάλληλα χαρακτηρισμένα ονόματα και έλαβε χώρα η μαζική επεξεργασία των αρχείων του κάθε συναισθήματος και για τις δύο βασικές κατηγορίες. Μετά την ολοκλήρωση της αφαίρεσης της αντήχησης, τα αρχεία, εισήχθησαν στον αλγόριθμο αποθορυβοποίησης του λογισμικού.

Στόχος αυτή την φορά ήταν η αφαίρεση του θορύβου που ήταν στο "βάθος" των καταγεγραμμένων αρχείων και είχε ευρυζωνικό χαρακτήρα. Με την εφαρμογή της αφαίρεσης της αντήχησης καταπολεμήθηκε αρκετός θόρυβος ο οποίος είχε προστεθεί ως μέρος των πρώιμων ανακλάσεων και ως εκ τούτου η εφαρμογή της αποθορυβοποίησης ως ξεχωριστός αλγόριθμος επεξεργασίας είχε έναν πιο αισθητικό χαρακτήρα, λόγω του ότι θα καταπολεμούσε τον εναπομένον υπολειμματικό θόρυβο που δεν κατάφερε να εντοπιστεί και να αφαιρεθεί από τον αλγόριθμο της αφαίρεσης της αντήχησης αλλά και το χαμηλό δάπεδο θορύβου το οποίο είχε προέλθει από τα ηλεκτρονικά κυκλώματα των μέσων καταγραφής (φωτοαντίσταση, καλωδίωση, κάρτα ήχου υπολογιστή) και ήταν στατικού ευρυζωνικού χαρακτήρα.

Ακολουθήθηκε μια παρόμοια τεχνική στην δημιουργία του προφίλ θορύβου όπως και στον αλγόριθμο αφαίρεσης της αντήχησης, πιο συγκεκριμένα, επιλέχθηκε ένα αρχείο ήχου από το οποίο θα σχηματιζόταν μια εκτίμηση του προφίλ του υπάρχοντος θορύβου με την μόνη διαφορά ότι η παραπάνω διεργασία έγινε από ένα αρχείο και μόνο για όλα τα συναισθήματα. Αυτό συνέβη επειδή όλα τα αρχεία ανεξάρτητος είχαν τον ίδιο στατικό θόρυβο βάθους και παρόμοιο υπολειμματικό θόρυβο, συνεπώς η εύρεση ενός αρχείου που θα είχε τους παραπάνω θορύβους ευδιάκριτα και σε προσβάσιμα σημεία εντοπισμού κατά μήκος του χρόνου (στην αρχή και το τέλος του αρχείου) θα επαρκούσε και θα ήταν αντιπροσωπευτικό για όλα τα υπόλοιπα αρχεία ομιλίας όλων των συναισθημάτων. Έπειτα από τον εντοπισμό του ενδεδειγμένου αρχείου, πραγματοποιήθηκε μια χρονική σάρωση για την επιλογή και την αποθήκευση ενός βέλτιστου προφίλ θορύβου για την επερχόμενη μαζική χρήση του σε όλες τις κατηγορίες συναισθημάτων αλλά και καταληπτότητας. Μετά την ολοκλήρωση και της παραπάνω διαδικασίας αποθορυβοποίησης, πραγματοποιήθηκε μια συγκέντρωση όλων των αρχείων στις εξής ομάδες:

- "Καταληπτή Ομιλία"
- "Μη-Καταληπτή Ομιλία"
- "Καταληπτή Ομιλία (επεξεργασμένη)"
- "Μη-Καταληπτή Ομιλία (επεξεργασμένη)"

Οι πρώτες δύο ομάδες αποτελούνται από τα καταγεγραμμένα αρχεία κατευθείαν από την διάταξη laser, τα οποία έχουν διαχωριστεί ανάλογα με το πόσο καταληπτά είναι ενώ οι επόμενες δύο ομάδες είναι τα πανομοιότυπα αρχεία των πρώτων δύο ομάδων τα οποία έχουν εμπλακεί στις ανάλογες επεξεργασίες αφαίρεσης αντήχησης και αποθορυβοποίησης κρατώντας τον αρχικό διαχωρισμό της καταληπτότητας και της μη καταληπτότητας.

Αυτή η δημιουργία τεσσάρων βάσεων δεδομένων συναισθηματικής ομιλίας είχε ως απώτερο σκοπό την διεξαγωγή κάποιων ακουστικών πειραμάτων με την συμμετοχή 10 ακροατών όπου στόχος ήταν η εξαγωγή χρήσιμων στατιστικών πληροφοριών όσον αφορά την ποσοστιαία καταληπτότητα του λεξιλογικού περιεχομένου των φράσεων της ομιλίας, την ποσοστιαία καταληπτότητα του εκφραζόμενου συναισθήματος, καθώς επίσης και την σύγκριση μεταξύ επεξεργασμένων και μη επεξεργασμένων αρχείων και στις τέσσερις κατηγορίες.

3.3 Πειραματική διαδικασία στατιστικών ακροάσεων

Όπως προαναφέρθηκε οι ακροάσεις διεξήχθησαν με την συμμετοχή 10 ακροατών, από τους οποίους ζητήθηκε να αναγνωριστούν οι φράσεις που άκουγαν, ως προς το λεξιλογικό περιεχόμενο και στην συνέχεια να κάνουν μια εκτίμηση του συναισθήματος του οποίου αντιλαμβάνονταν. Συγκεκριμένα, στον κάθε ακροατή αντιστοιχούσε 1 τεστ για κάθε μία από τις 4 βάσεις δεδομένων συναισθηματικής ομιλίας, το οποίο περιλάμβανε την αναγνώριση του λεξιλογικού και του συναισθηματικού περιεχομένου κάθε φράσης.

3.3.1 Συνθήκες κατά την ακρόαση

Οι πειραματικές ακροάσεις διεξήχθησαν σε πολλούς διαφορετικούς χώρους όμως σε κανέναν δεν παρουσιάστηκε πρόβλημα τυχαίου εξωτερικού θορύβου ή κάποιου είδους θορύβου βάθους. Σε όλες τις ακροάσεις ανεξαιρέτως χρησιμοποιήθηκε ένα ζευγάρι ακουστικών (headphones) ανοιχτού τύπου το οποίο προτιμήθηκε έναντι της αναπαραγωγής των φράσεων μέσω μεγαφώνου/ηχείου λόγω της σημασίας που κατείχε στην συγκέντρωση της προσοχής των

ακροατών, στην αποφυγή των ανακλάσεων στο δωμάτιο ακρόασης (στην περίπτωση χρήσης μεγαφώνου) και συνεπώς, στην ανάδειξη της εγκυρότητας των αποτελεσμάτων. Η ένταση εκπομπής των αρχείων ρυθμιζόταν σε κάθε ακρόαση επί τόπου ανάλογα το επίπεδο στάθμης όπου ένιωθε άνετα ο κάθε ακροατής. Για την διευκόλυνση των πειραμάτων παρασχέθηκαν στους ακροατές κόλλες χαρτιού για να υπάρξει η δυνατότητα σημειώσεων κατά την διάρκεια των ακροάσεων. Τα περισσότερα από τα αρχεία ήταν σύντομες φράσεις προσεγγιστικά 5 λέξεων όμως υπήρχαν και αρχεία τα οποία ήταν μεγάλες προτάσεις οι οποίες προϋπόθεταν την χρήση σημειώσεων εάν δεν κατάφερναν οι ακροατές να τις απομνημονεύσουν λόγο του όγκου του λεξιλογικού περιεχομένου τους. Έτσι κατά την διάρκεια της ακρόασης, επιτράπηκε στους ακροατές να σημειώσουν επί τόπου αυτό που δεν προλάβαν να ή δυσκολευόντουσαν να μάθουν χωρίς όμως να χρησιμοποιήσουν την σημείωση σαν βοηθητική παροχή σε μια μελλοντική ακρόαση. Εδώ θα πρέπει να σημειωθεί πως οι ακροατές άκουγαν το λεξιλογικό περιεχόμενο μιας πρότασης, πάνω απο μια φορά αφού κάθε πρόταση υπήρχε σε όλες τις συναισθηματικές καταστάσεις και έτσι εάν κράταγαν κάποια σημείωση για μια πρόταση σε μια από τις 5 συναισθηματικές καταστάσεις την πρώτη φορά, τους απαγορευόταν να χρησιμοποιήσουν αυτή την σημείωση στην μελλοντική ακρόαση της ίδιας πρότασης στα επόμενα 4 συναισθήματα.

Το κάθε αρχείο εκπεμπόταν μία με 2 φορές αναλόγως το αίτημα του ακροατή αλλά υπήρχαν και περιπτώσεις όπου παρεχόταν και μια τρίτη ακρόαση όποτε ο γράφων του κειμένου έκρινε πως είναι σημαντική και αναγκαία. Η τρίτη ακρόαση εφαρμοζόταν σε μονολεκτικές φράσεις του τύπου "Ναι", "Όχι", "Ευχαριστώ" όπου υπήρχε παρουσία έντονης ατάκας και ταχύτητας στην άρθρωση, οπότε ο ακροατής δεν προλάβαινε να αντιληφθεί την παραγόμενη λέξη ειδικότερα αν το αρχείο άνηκε στις λιγότερα καταληπτές κατηγορίες συναισθημάτων όπως είναι το ουδέτερο συναίσθημα και το συναίσθημα του φόβου και της λύπης. Τα αρχεία που χρησιμοποιήθηκαν σε κάθε ένα από τα τεστ, εκπεμπόντουσαν σε τυχαία σειρά για να μειώνεται η επαναληψιμότητα των χαρακτηριστικών της χροιάς της φωνής του κάθε συναισθήματος ομιλίας αλλά και για να αποφεύγεται ο κορεσμός λόγο της επαναληπτικής ακρόασης του ίδιου συναισθήματος, κάτι που θα επηρέαζε σημαντικά την κρίση των ακροατών στην εξαγωγή των αποτελεσμάτων τους.

Στους ακροατές παρεχόντουσαν τα 5 διαθέσιμα συναισθήματα για να υπάρξει μια πιο σαφής κατεύθυνση των επιλογών τους όσο αναφορά την συναισθηματική κατάσταση του

εκπεμπόμενου λόγου και υπήρχε ένα είδος βαθμονόμησης της επίδοσης τους όσο αναφορά το λεξιλογικό περιεχόμενο των φράσεων. Πιο συγκεκριμένα, πραγματοποιούταν η εκπομπή των φράσεων και μετά από μία έως 3 ακροάσεις, ο ακροατής αναπαρήγαγε το λεξιλογικό περιεχόμενο που κατάφερε να αντιληφθεί καθώς επίσης και το συναίσθημα της παραγωγής της ομιλίας που αισθάνθηκε. Εν συνεχεία, ο συγγραφέων βαθμολογούσε την επίδοση του λεξιλογικού περιεχομένου της φράσης σύμφωνα με μια προγενέστερη βαθμονόμηση που είχε πραγματοποιηθεί,

Perfect	Άριστα	80% - 100%
Well Enough	Λίαν Καλώς	50% - 70%,
Approximately	Κατά Προσέγγιση/Καλώς	20% - 40%
None	Ουδόλως	0% - 10%

Πίνακας 2. Ποσοστιαία αντιστοίχιση βαθμονόμησης

και τέλος, σημείωνε το συναίσθημα που αισθανόταν ο ακροατής. Η βαθμολόγηση της επίδοσης του λεξιλογικού περιεχομένου των ακροατών γινόταν με έναν δυσανάλογο τρόπο δηλαδή η κάθε πρόταση υπόκειντο σε διαφορετική και μοναδική μεταχείριση χωρίς να εισέρχεται σε διαδικασίες εξίσωσης με τις υπόλοιπες. Για παράδειγμα, εάν ένας ακροατής κατάφερε να εντοπίσει τις 13 από τις 15 λέξεις μιας πρότασης θα βαθμολογούταν με "Άριστα", ενώ εάν κατάφερε να εντοπίσει τις 2 από τις 4 λέξεις μιας πρότασης δεν θα βαθμολογούταν με "Άριστα", λόγο της ανομοιότητας του πλήθους των λέξεων, του βαθμού δυσκολίας και συνεπώς του διαφορετικού ποσοστού επιτυχίας.

Τελικώς, όλες οι παραπάνω επισημάνσεις συντέλεσαν στην δημιουργία ενός εύπορου περιβάλλοντος ακροάσεων που στόχο είχε την πλήρης και βέλτιστη αποτύπωση των αποτελεσμάτων για την μελλοντική χρήση τους σε στατιστικά μοντέλα πινάκων και γραφικών παραστάσεων με σκοπό την σύγκριση και την αξιολόγηση των εξεταζόμενων δεδομένων των τεσσάρων βάσεων δεδομένων ομιλίας.

Κεφάλαιο 4^ο

Επεξεργασία και ανάλυση δεδομένων

Στην παρούσα ενότητα παρουσιάζονται, αναλύονται και αξιολογούνται τα αποτελέσματα της πειραματικής διαδικασίας. Στα πλαίσια της ανάλυσης αυτής παρατίθενται ανάλογα γραφήματα και πίνακες όπου παρουσιάζεται η απόδοση της διάταξης.

4.1 Παρατηρήσεις του στατιστικού πειράματος

Όπως προαναφέρθηκε και στην παραπάνω ενότητα, κατά την διάρκεια της αξιολόγησης, κάθε ακροατής άκουγε την κάθε φράση μια φορά για κάθε συναισθηματική κατάσταση (το σύνολο 5 φορές). Η διαδικασία αυτή έλαβε χώρα και για τα 4 σύνολα φράσεων.

4.1.1 Αξιολόγηση του εννοιολογικού περιεχομένου των ηχογραφήσεων

Η παραπάνω σχεδίαση της αξιολόγησης του συστήματος εμπεριέχει ένα «σφάλμα». Λόγω του γεγονότος ότι το κείμενο των φράσεων είναι το ίδιο για κάθε συναισθηματική κατάσταση είχε σαν αποτέλεσμα την σε ένα βαθμό εξοικείωση των ακροατών με το περιεχόμενο των ηχογραφήσεων. Έπειτα από κάθε ακρόαση η αντικειμενικότητα της κρίσης τους θα μειωνόταν λόγω της εξοικείωσης με το περιεχόμενο και την χροιά της εκφωνήτριας. Προκειμένου να εξισορροπηθεί η προαναφερθείσα παρενέργεια, ζητήθηκε από τους ακροατές να προσπαθήσουν να είναι όσο το δυνατόν πιο αντικειμενικοί και να κρίνουν άμεσα την ηχογράφιση με σκοπό την μείωση της επενέργειας της πρότερης τους γνώσης (πχ γνώση του λεξιλογικού περιεχομένου, τα προσωδιακά χαρακτηριστικά της φωνής και την χροιά). Το παραπάνω εγχείρημα αν και αρκετά σύνθετο βγήκε εις πέρας από ένα αρκετά μεγάλο ποσοστό των ακροατών όπου έκριναν με κάθε πιθανή και διαθέσιμη αντικειμενικότητα το άκουσμα τους, παραβλέποντας όσο ήταν δυνατόν προηγούμενη πληροφορία για τα χαρακτηριστικά του αρχείου.

Ένα μεγάλο ποσοστό των ακροατών όταν δεν κατόρθωνε να εξάγει ξεκάθαρη πληροφορία για το λεξιλογικό περιεχόμενο των φράσεων συμπλήρωνε τα δυσνόητα σημεία της φράσης με λέξεις οι οποίες υπάγονταν στην λογική των συμφραζόμενων. Ειδικότερα, επέλεξαν να αναπληρώσουν τα κενά (κατανόησης) που παρουσιαζόντουσαν στην εκφορά μιας φράσης με λέξεις οι οποίες έδιναν την δυνατότητα να προκύψει μια λογική δήλωση με συγκεκριμένο νόημα. Ένα άλλο πρόβλημα που αντιμετωπίστηκε ήταν η παρουσίαση περιπτώσεων ακροατών όπου δεν γνώριζαν την εννοιολογική σημασία λέξεων, γεγονός που δυσχέραινε το έργο τους στην εξαγωγή ορθών απαντήσεων. Αξιοσημείωτο είναι ότι ενώ άκουγαν τις ίδιες προτάσεις αρκετές φορές, εμφανίζονταν περιπτώσεις ακροατών οι οποίοι προσέθεταν ή αφαιρούσαν άρθρα αλλά και ολόκληρες λέξεις μετά την πρώτη ακρόαση της πρότασης. Εν ολίγης, αναδομούσαν το λεξιλογικό περιεχόμενο της πρότασης εκ νέου σύμφωνα με αυτό που υπέθεταν ότι αντιλαμβάνονταν. Αυτό οφείλεται πιθανότατα στο γεγονός της μερικής απρόβλεπτης αλλοίωσης που δημιουργήθηκε από τους αλγόριθμους επεξεργασίας σε έναν μικρό αριθμό αρχείων όπου μερικές συλλαβές είτε υποχωρούν αιφνίδια από μια προκαθορισμένη εντασιακή στάθμη σε πολύ χαμηλά επίπεδα (Ducking) είτε παρουσιάζεται ένα συχνοτικό φίλτρο το οποίο σαρώνει στιγμιαία την χαμηλομεσαία περιοχή της κρίσιμης συλλαβής που έχει ως διαδοχικό αποτέλεσμα την διαστρέβλωση της αρχικής φράσης ως προς το λεξιλογικό περιεχόμενο.

4.1.2 Αξιολόγηση συναισθηματικής πληροφορίας ηχογραφήσεων

Η αξιολόγηση της συναισθηματικής πληροφορίας των ηχογραφήσεων που προέκυψαν από την διάταξη έγινε με την ίδια διαδικασία με αυτή που περιγράφηκε προηγουμένως. Κατά την διεκπεραίωση των πειραμάτων αξιολόγησης ένα πρόβλημα το οποίο παρατηρήθηκε ήταν αυτό της σύγχυσης του συναισθήματος του φόβου με αυτό της λύπης και της ουδέτερης κατάστασης. Αυτό συνέβη κατά κύριο λόγο επειδή το βιμπράτο που εμπεριέχετε στο συναίσθημα του φόβου στην ομιλία και το καθιστά αναγνωρίσιμο και ευδιάκριτο από τα υπόλοιπα χαρακτηριστικά όλων των άλλων συναισθημάτων, δεν ήταν έντονου χαρακτήρα στην GrES βάση δεδομένων εξαρχής και έτσι ηχογραφήθηκε αρκετά φτωχά έχοντας ως αποτέλεσμα την υποβάθμιση του. Η ασάφεια που δημιουργήθηκε στους ακροατές λόγο της απουσίας της

ευκρίνειας του βιμπράτο, τους οδήγησε σε μία κατάσταση διλήμματος για το πιο συναίσθημα είναι αυτό που αντιλαμβάνονται και πρέπει να επιλέξουν. Σε αντιδιαστολή με την προαναφερθείσα κατάσταση, τα συναισθήματα του θυμού, της χαράς και η ουδέτερη κατάσταση γινόντουσαν αμέσως αντιληπτά ακόμα και σε καταστάσεις έντονης ασάφειας ως προς τη καταληπτότητα αν και εδώ θα πρέπει να επισημανθεί πως παρουσιάστηκαν κάποιες περιορισμένες καταστάσεις στις οποίες και οι 10 ακροατές συνέχισαν τα συναισθήματα του θυμού και της χαράς. Αυτό οφείλεται στον γρήγορο ρυθμό ομιλίας, την υψηλή μέση τιμή pitch, την μεγάλη διακύμανση του pitch, την υψηλή ένταση και την απουσία της ηχηρότητας στην ποιότητα της φωνής που έχουν και τα δύο αυτά συναισθήματα.

Ένα αξιοσημείωτο γεγονός είναι το ότι ένα μέρος των ακροατών που συμμετείχαν στα ακουστικά πειράματα, όταν δεν κατόρθωναν να εξάγουν κανένα απολύτως συμπεράσματα ως προς την συναισθηματική κατάσταση του ομιλητή, κατοχύρωναν την φράση ως ουδέτερης συναισθηματικής κατάστασης. Πιθανότατα εξέταζαν όλα τα γνωστά και διακριτά χαρακτηριστικά του κάθε συναισθήματος και εφόσον δεν αντιλαμβάνοντουσαν κάποιο ίχνος της παρουσίας τους στην εξεταζόμενη φράση εξέθεταν την φράση ως ουδέτερη με την μέθοδο της εις ατόπου απαγωγής. Τέλος, θα πρέπει να αναφερθεί πως η ψυχολογική και η πνευματική κατάσταση των ακροατών επηρέαζε σημαντικά τις επιδόσεις τους. Μια άσχημη ψυχολογική κατάσταση είχε ως αποτέλεσμα την μείωση της προσήλωσης και του ενδιαφέροντος και την εκδήλωση γρήγορου κορεσμού, ενώ μια καλή ψυχολογική και πνευματική κατάσταση είχε ως επακόλουθο την οξυμένη αντίληψη και προσοχή στα αναπαραγόμενα αρχεία και μια γενικότερη παραγωγικότητα και αποδοτικότητα στην εξαγωγή των αποτελεσμάτων.

Στην παρακάτω ενότητα γίνεται μια παρουσίαση των στατιστικών δεδομένων που προέκυψαν από της ακροάσεις 10 ατόμων σε 4 ομάδες αρχείων υπό την μορφή γραφημάτων ράβδων. Στην συνέχεια αξιολογούνται και σχολιάζονται περαιτέρω τα αποτελέσματα της κάθε ομάδας και πραγματοποιείτε σύγκριση μεταξύ των. Όλες οι εξεταζόμενες ομάδες καθώς και τα υποσύνολα τους είναι τα εξής:

➤ "Καταληπτή Ομιλία"

α) Ως προς το λεξιλογικό περιεχόμενο

β) Ως προς το συναισθηματικό περιεχόμενο

➤ "Μη-Καταληπτή Ομιλία"

α) Ως προς το λεξιλογικό περιεχόμενο

β) Ως προς το συναισθηματικό περιεχόμενο

➤ "Καταληπτή Ομιλία (επεξεργασμένη)"

α) Ως προς το λεξιλογικό περιεχόμενο

β) Ως προς το συναισθηματικό περιεχόμενο

➤ "Μη-Καταληπτή Ομιλία (επεξεργασμένη)"

α) Ως προς το λεξιλογικό περιεχόμενο

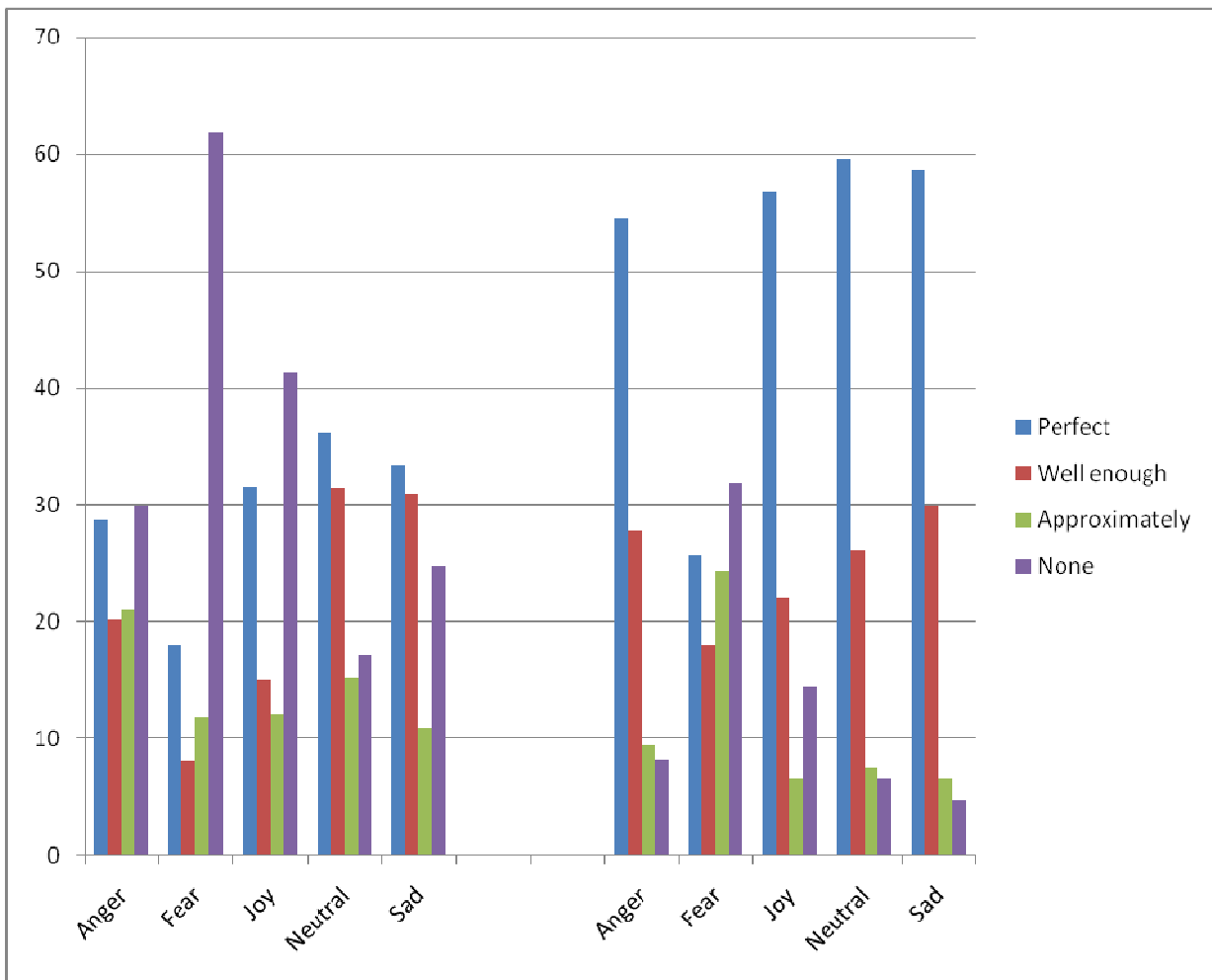
β) Ως προς το συναισθηματικό περιεχόμενο

4.2 Ανάλυση και σχολιασμός των στατιστικών μετρήσεων

Εισαγωγικά, πριν ξεκινήσει η ανάλυση των δεδομένων, θα πρέπει να αναφερθούν ορισμένα στοιχεία που αφορούν την δομή των γραφημάτων. Στα γραφήματα που αφορούν το λεξιλογικό περιεχόμενο των φράσεων, παρουσιάζεται το ποσοστό επιτυχίας % (άξονας Y) στην εύρεση του λεξιλογικού περιεχομένου, ανά συναίσθημα (άξονας X), και σε κάθε συναίσθημα υπάρχει μια υποκειμενική βαθμονόμηση αναλόγως το ποσοστό ευστοχίας στο σύνολο των λέξεων μιας πρότασης (χρώματα ράβδων). Η βαθμονόμηση της ευστοχίας έχει αναλυθεί εκτενέστερα σε προηγούμενη ενότητα. Στα γραφήματα που αφορούν το συναισθηματικό περιεχόμενο των φράσεων, ο άξονας X εκφράζει τα 5 στοχευόμενα συναισθήματα (target emotions), ο άξονας Y εκφράζει το ποσοστό επιτυχίας % , και τα χρώματα των ράβδων εκφράζουν τα εκτιμώμενα συναισθήματα (estimated emotions) από τους ακροατές. Κατά την

διάρκεια των αξιολογήσεων οι λέξεις "Άριστα", "Λίαν Καλώς", "Κατά Προσέγγιση/Καλώς", και "Ουδόλως" που αποτελούν την βαθμονόμηση της ευστοχίας των λεξιλογικών περιεχομένων των προτάσεων, αντικαταστάθηκαν με τις φράσεις "Πλήρης Ευστοχία", "Μερική Ευστοχία", "Έλλιπής Ευστοχία", " Πλήρης Αστοχία" αντίστοιχα.

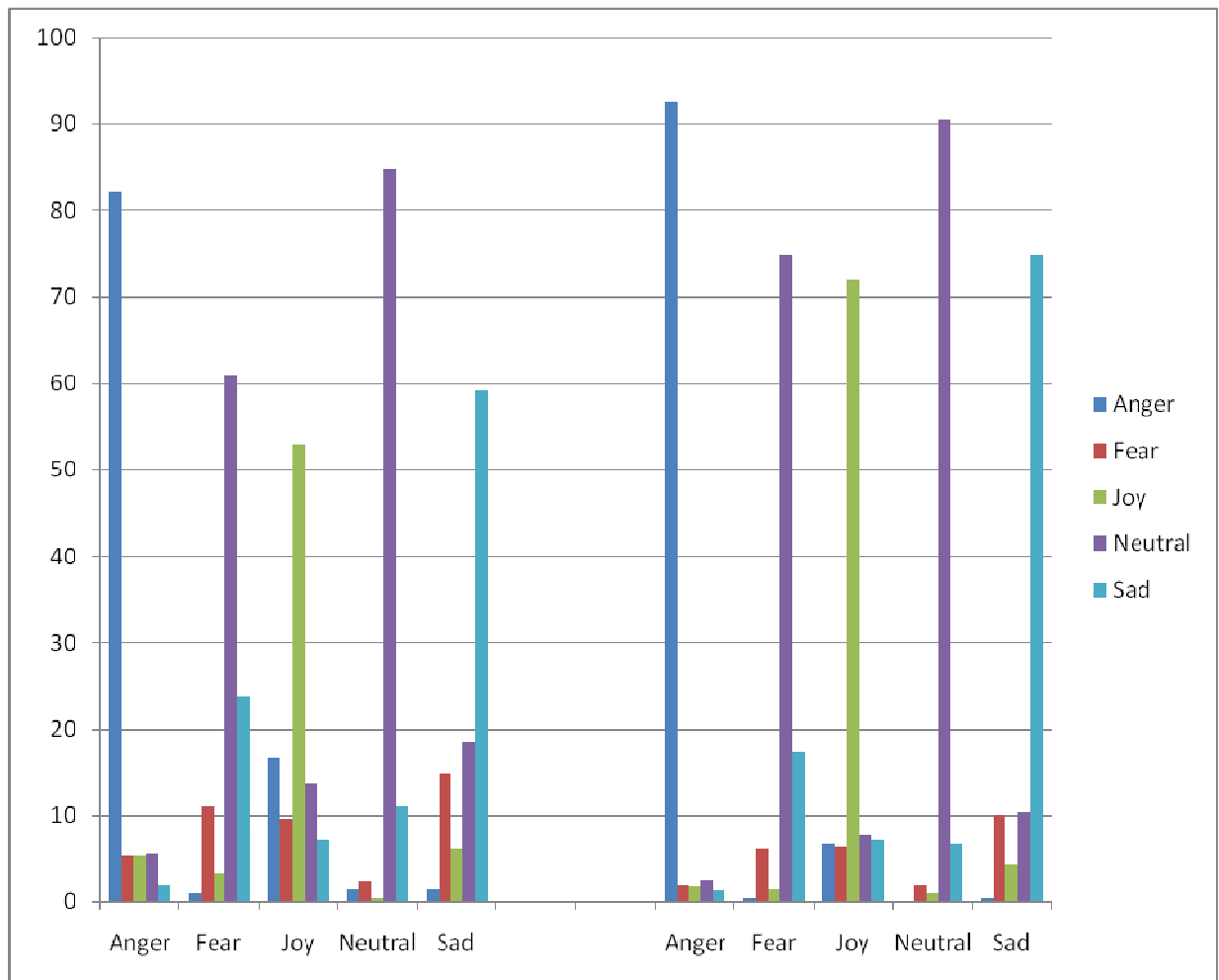
4.2.1 Ανάλυση των μετρήσεων



Εικόνα 14. Καταληπτή Ομιλία ως προς το λεξιλογικό περιεχόμενο (αριστερά χωρίς επεξεργασία, δεξιά με επεξεργασία)

Παρατηρώντας το γράφημα της καταληπτής ομιλίας ως προς το λεξιλογικό περιεχόμενο διαπιστώνει κανείς πως στα συναισθήματα του θυμού, του φόβου και της χαράς υπάρχει ένα μεγάλο ποσοστό της τάξεως του 40% πλήρους αστοχίας ενώ το ποσοστό της πλήρους ευστοχίας κυμαίνεται περίπου στο 25%. Το υπόλοιπο 35% αποτελείται από μια μέτρια και ελλιπείς ευστοχία. Από την άλλη πλευρά, στα συναισθήματα του ουδέτερου και της λύπης υπάρχει ένα ποσοστό πλήρους ευστοχίας της τάξεως του 35% ενώ παρουσιάζεται πτώση κατά το ήμισυ στο ποσοστό της πλήρους αστοχίας σε σύγκριση με τα προηγούμενα 3 συναισθήματα, όπου φτάνει το 20%. Αξιοσημείωτη είναι επίσης και η άνοδος του ποσοστού της μέτριας και της ελλιπής ευστοχίας όπου φτάνει έως και το 45% της συνολικής διαβάθμισης της ευστοχίας.

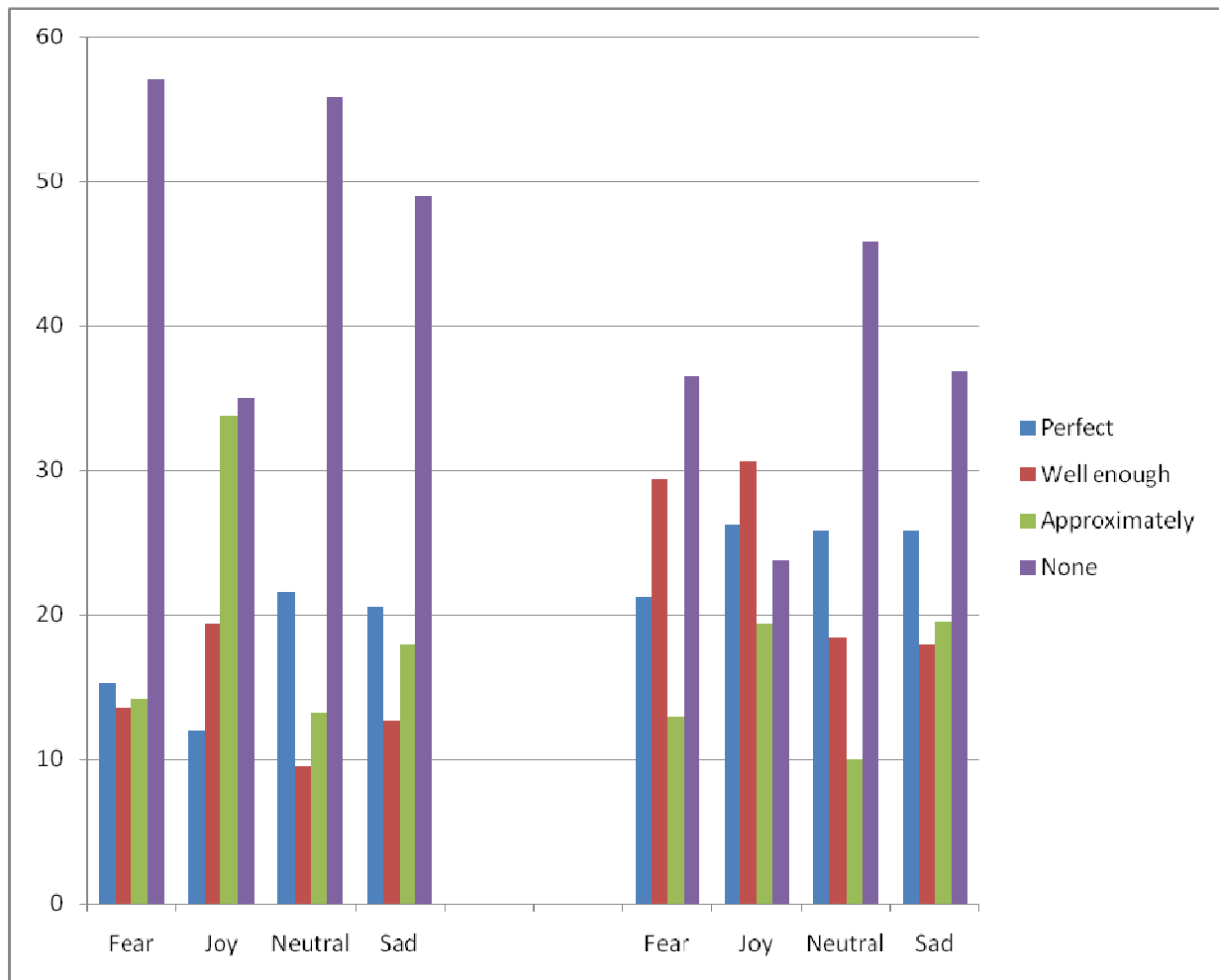
Παρατηρώντας το γράφημα της επεξεργασμένης καταληπτής ομιλίας ως προς το λεξιλογικό περιεχόμενο μπορεί εύκολα κάποιος να διαπιστώσει την υπέρμετρη αύξηση των απόλυτα εύστοχων απαντήσεων σε όλα τα συναισθήματα σε σύγκριση με την ομόλογη της που δεν έχει οδηγηθεί σε διαδικασίες βελτίωσης. Υπάρχει μια αύξηση της τάξεως περίπου του 20 % στις απόλυτα σωστές απαντήσεις ενώ εμφανίζεται μια μείωση στις απόλυτα άστοχες απαντήσεις περίπου κατά 22%. Άνοδος σημειώθηκε και στις απαντήσεις μέτριας ευστοχίας η οποία ήταν αντιστρόφως ανάλογη της μείωσης των απαντήσεων ελλιπής ευστοχίας.



Εικόνα 15. Καταληπτή Ομιλία ως προς το συναισθηματικό περιεχόμενο (αριστερά χωρίς επεξεργασία, δεξιά με επεξεργασία)

Εξετάζοντας το γράφημα της καταληπτής ομιλίας ως προς το συναισθηματικό περιεχόμενο διαπιστώνεται πως σε όλα τα συναισθήματα εκτός αυτού του φόβου εμφανίζονται μεγάλα ποσοστά ευστοχίας, ταύτισης δηλαδή του στοχευόμενου συναισθήματος με το εκτιμώμενο συναίσθημα των ακροατών. Απο αυτή την ευστοχία εξαιρείτε το συναίσθημα του φόβου όπου το σύνολο των ακροατών το εκτίμησε σε ένα μεγάλο ποσοστό ως ουδέτερο και σε ένα μικρότερο ως λύπη, για τους λόγους της μη διακριτότητας των κύριων χαρακτηριστικών του όπου έχουν αναφερθεί αναλυτικότερα στην ενότητα των παρατηρήσεων των ακουστικών

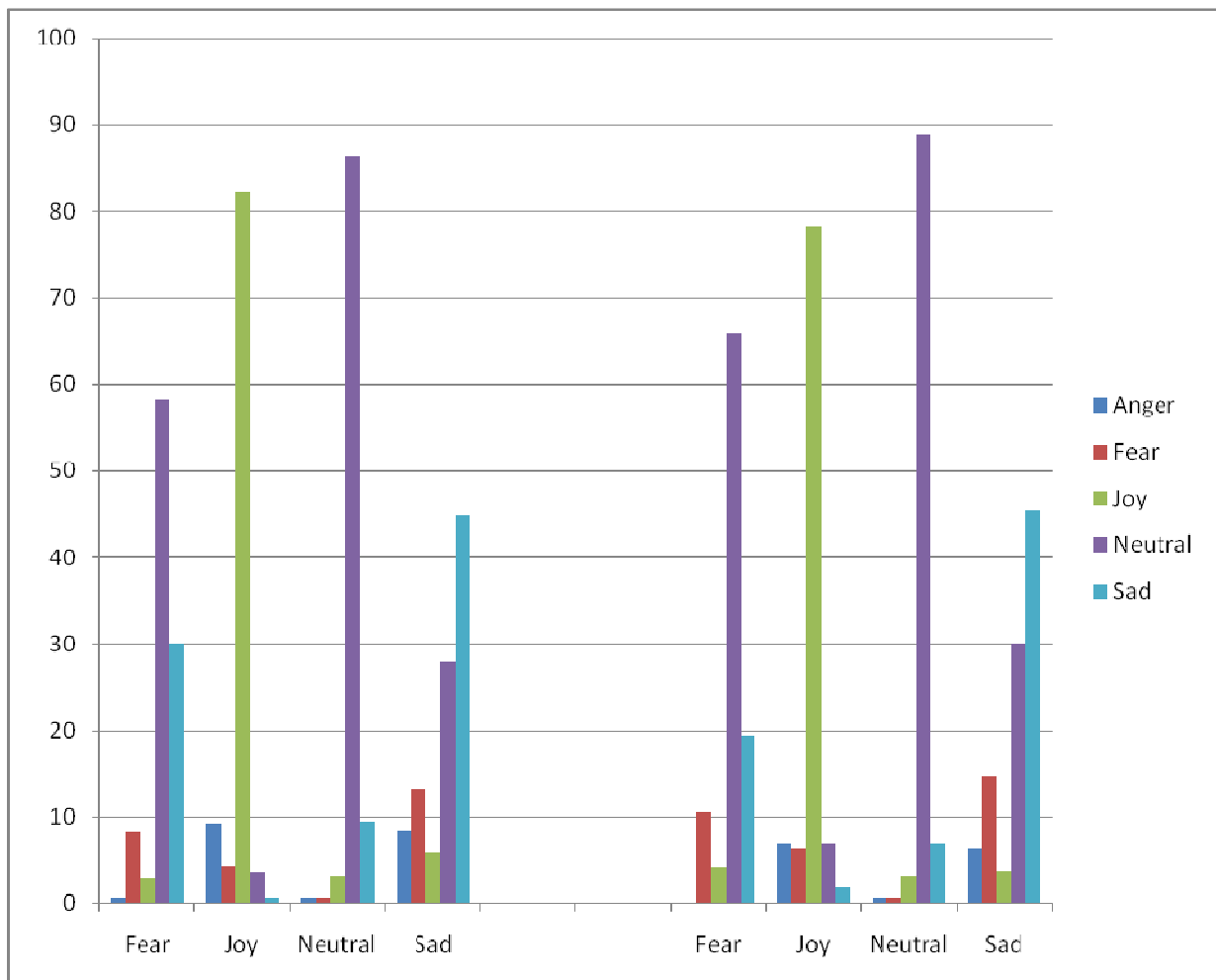
πειραμάτων. Στο γράφημα της επεξεργασμένης καταληπτής ομιλίας ως προς το συναισθηματικό περιεχόμενο υπάρχει μια σχεδόν πανομοιότυπη εκτίμηση του συναισθήματος με αυτήν της μη επεξεργασμένης ομιλίας της, ωστόσο τα αποτελέσματα παρουσιάζονται σε μεγαλύτερο ποσοστό επιτυχίας ιδιαίτερα στα συναισθήματα της χαράς και της λύπης .



Εικόνα 16. Μη-Καταληπτή Ομιλία ως προς το λεξιλογικό περιεχόμενο (αριστερά χωρίς επεξεργασία, δεξιά με επεξεργασία)

Στο γράφημα της μη- καταληπτής ομιλίας ως προς το λεξιλογικό περιεχόμενο υπάρχει μια ολοφάνερη αύξηση της πλήρους αστοχίας στην εξαγωγή των λεξιλογικών περιεχομένων των

προτάσεων στα 4 εξεταζόμενα συναισθήματα σε ένα ποσοστό που κυμαίνεται περίπου στο 50% της συνολικής διαβάθμισης ευστοχίας. Επίσης εμφανής είναι και η πτώση του ποσοστού των πλήρως εύστοχων απαντήσεων που βρίσκονται κάτω από το 20% των συνολικών απαντήσεων. Εξετάζοντας το γράφημα της επεξεργασμένης μη- καταληπτής ομιλίας ως προς το λεξιλογικό περιεχόμενο παρατηρείται πως σε όλα τα συναισθήματα η πλήρως άστοχη επιλογή έχει υποχωρήσει αρκετά πιο χαμηλά από το 50% των απαντήσεων ενώ η πλήρως εύστοχη επιλογή σε όλα τα συναισθήματα είναι άνω του 20% των συνολικών απαντήσεων.



Εικόνα 17. Μη-Καταληπτή Ομιλία ως προς το συναισθηματικό περιεχόμενο (αριστερά χωρίς επεξεργασία, δεξιά με επεξεργασία)

Στην μη- καταληπτή ομιλία ως προς το συναισθηματικό περιεχόμενο βλέπουμε πως οι εκτιμήσεις των συναισθημάτων δεν έχουν καμία ουσιαστική διαφορά με της εκτιμήσεις στα καταληπτά αρχεία, αφού οι ακροατές εξακολουθούν να είναι σε θέση να εντοπίζουν το σωστό συναίσθημα με εξαίρεση πάντα το συναίσθημα του φόβου, ακόμα και κάτω απο συνθήκες έντονης ασάφειας ως προς το λεξιλογικό περιεχόμενο της ομιλίας.

Τέλος, βλέποντας το γράφημα της επεξεργασμένης μη- καταληπτής ομιλίας ως προς το συναισθηματικό περιεχόμενο, εξάγεται το συμπέρασμα της βελτίωσης του ποσοστού ευστοχίας σε όλα τα συναισθήματα.

4.2.2 Σχολιασμός των αποτελεσμάτων

Έπειτα απο την ολοκλήρωση της στατιστικής ανάλυσης των ακουστικών μετρήσεων, παρατηρήθηκε μια γενικότερη βελτίωση σε όλες τις κατηγορίες και τις υπο-ομάδες των βάσεων δεδομένων ομιλίας. Από τα γραφήματα των επιδόσεων στο λεξιλογικό και συναισθηματικό περιεχόμενο των φράσεων παρουσιάζεται με εμφανής ποσοστιαία βελτίωση σε όλες περιπτώσεις και διαβαθμίσεις κάτι που οφείλεται στην βέλτιστη λειτουργία των αλγορίθμων αφαίρεσης της αντήχησης και αποθορυβοποίησης. Ενδεικτικά, στο γράφημα της καταληπτής ομιλίας όπου εξετάζεται η ευστοχία στην αναπαραγωγή του λεξιλογικού περιεχομένου, στην επεξεργασμένη εκδοχή παρατηρείται πως η πλήρης εύστοχες απαντήσεις έχουν μεγιστοποιηθεί και οι πλήρως άστοχες απαντήσεις έχουν μειωθεί σημαντικά σε σύγκριση με την μη-επεξεργασμένη εκδοχή.

Εξετάζοντας τα γραφήματα που αφορούν την εύρεση του συναισθηματικού περιεχομένου των καταληπτών φράσεων διαπιστώνεται πως υπάρχουν μεγάλα ποσοστά επιτυχίας σχεδόν σε όλα τα συναισθήματα. Οι ακροατές έκαναν τις σωστές εκτιμήσεις για τα στοχευόμενα συναισθήματα και είχαν επιπλέον αύξηση της επιτυχίας τους όταν εφαρμόστηκαν οι κατάλληλες επεξεργασίες των αλγορίθμων. Η αρχική επιλογή τους επιβεβαιώθηκε ακόμη παραπάνω όταν ξεκαθαρίστηκε το πιο συναίσθημα εκπέμπετε, περισσότερο λόγο της εφαρμογής της αφαίρεσης της αντήχησης και την επίτευξη περεταίρω ευκρίνειας.

Περνώντας στα γραφήματα που μελετούν την ευστοχία του λεξιλογικού περιεχομένου στην μη- καταληπτή ομιλία, φαίνεται πως στην μη επεξεργασμένη ακρόαση παρουσιάζονται

υψηλές ποσοστιαίες στάθμες στην πλήρη αστοχία. Εδώ φαίνεται ξεκάθαρα το πόσο επηρέασε το σύνολο της αντήχησης την καταληπτότητα των αρχείων και ως εκ τούτου τις επιδόσεις των ακροατών στα σύνολα των σωστών και των λάθος απαντήσεων. Εφαρμόζοντας τους αλγόριθμους αναμόρφωσης, καταπολεμήθηκαν οι ηχηρές συλλαβές και η ασάφεια αυτών και τελικώς, επιτεύχθηκε κάποια βελτίωση στην αύξηση του αριθμού των σωστών απαντήσεων και στη μείωση του αριθμού των λάθος απαντήσεων.

Εν τέλει, στα γραφήματα της μη-καταληπτής ομιλίας ως προς το συναίσθημα, σημειώνεται πως ακόμη και σε συνθήκες μεγάλης ασάφειας που έχει προκληθεί από την απουσία καταληπτότητας λόγω της επίδρασης της αντηχητικής ουράς, των πρωίμων ανακλάσεων, και του ευρυζωνικού θορύβου, οι ακροατές καταφέρνουν και ευστοχούν στα συναισθήματα επειδή κατανοούν και βασίζονται επάνω σε ορισμένα προσωδιακά χαρακτηριστικά της φωνής όπως η θεμέλιος συχνότητα F_0 , η ένταση και η διάρκεια τα οποία δεν είναι τόσο ευαίσθητα στον επηρεασμό από τον θόρυβο και την αντήχηση, όπως επίσης διαπιστώνεται ότι η στην επεξεργασμένη εκδοχή των αρχείων στο στοχευόμενο συναίσθημα του φόβου, η επιλογή του φόβου ως απάντηση εκτιμήθηκε περισσότερο από ότι στην μη επεξεργασμένη και σε αυτό συντέλεσε σημαντικά η επεξεργασία βελτίωσης.

Κεφάλαιο 5^ο

Συμπεράσματα

Η καταγραφή ήχου εξ αποστάσεως με την χρήση οπτοηλεκτρονικών διατάξεων laser και συγκεκριμένα της τεχνικής της εκτρεπόμενης δέσμης laser από δονούμενη επιφάνεια, βρίσκει εφαρμογή στην σύγχρονη εποχή σε συνθήκες και καταστάσεις κατά τις οποίες τα συμβατικά μικρόφωνα, που λειτουργούν μηχανικά με την χρήση διαφράγματος, δεν μπορούν να ανταπεξέλθουν επαρκώς (ή και καθόλου) στην σύλληψη του απομακρυσμένου ήχου. Η χρήση του φωτός ως μέσο μετάφρασης των δονήσεων μιας στοχευόμενης επιφάνειας σε μεγάλη απόσταση, επιτρέπει στην βέλτιστη ανίχνευση ακόμη και των μικρότερων διαταράξεων που συμβαίνουν επάνω σε αυτήν. Μια τέτοια διάταξη υλοποιήθηκε για της ανάγκες αυτής της εργασίας. Σκοπός ήταν η καταγραφή μιας βάσης δεδομένων συναισθηματικής ομιλίας η οποία εκπεμπόταν από ένα μεγάφωνο κοντά σε μια επιφάνεια και εν συνεχεία ανιχνεύονταν οι μικρές μετατοπίσεις τις επιφάνειας από μία φωτοαντίσταση ως χωρικές εκτροπές του φωτός, οι οποίες «μεταφράζονταν» σε αυξομειώσεις σήματος ανάλογες του καταγεγραμμένου ηχητικού υλικού. Με την ολοκλήρωση της καταγραφής, το ηχητικό υλικό υποβλήθηκε σε διεργασίες βελτίωσης και χρησιμοποιήθηκε σε προσεκτικά σχεδιασμένα πειράματα ακρόασης από ένα σύνολο ακροατών, με σκοπό την εξαγωγή στατιστικών αποτελεσμάτων που αφορούσαν την σύγκριση βελτιωμένης και μη βελτιωμένης βάσης δεδομένων ομιλίας ως προς την καταληπτότητα του λεξιλογικού και του συναισθηματικού περιεχομένου των φράσεων που εκπεμπόταν.

Μετά την εξαγωγή του καταγεγραμμένου ηχητικού υλικού διαπιστώθηκε πως τα καταγεγραμμένα αρχεία της ομιλίας περιείχαν σχετικά υψηλά επίπεδα ανεπιθύμητης αντήχησης, λόγω της φύσεως του χώρου που χρησιμοποιήθηκε για την διεξαγωγή των πειραμάτων, καθώς και θόρυβο που προέρχονταν, κατά κύριο λόγο, από τις παρεμβολές όλων εκείνων των ηλεκτρονικών κυκλωμάτων που χρησιμοποιήθηκαν τόσο στην εκπομπή όσο και στην καταγραφή. Για τους παραπάνω λόγους μειώθηκε αισθητά η καταληπτότητα των αρχείων και έτσι εφαρμόστηκαν διαδικασίες βελτίωσης με την χρήση αλγορίθμων αφαίρεσης της αντήχησης και αποθορυβοποίησης.

Εδώ αποδεικνύεται η σημαντικότητα του θορύβου γενικότερα, και ιδιαιτέρως της αντήχησης, στην ευκρίνεια του λόγου, όπου εάν υφίσταται σε μεγάλα επίπεδα και εμφανίζεται

ιδιαίτερα στις πρώιμες ανακλάσεις, αφαιρεί από τον παραγόμενο λόγο τα χαρακτηριστικά γνωρίσματα των συλλαβών και τον λέξεων, μετατρέποντάς τον σε μία ακατανόητη ακολουθία ήχων, όπου κρατά μονάχα των ρυθμό της εκπομπής και μερικά ποιοτικά χαρακτηριστικά της φωνής που δεν εξυπηρετούν σχεδόν καθόλου στην αποσαφήνιση των λέξεων και του νοήματος.

Σημαντικό μέρος στην καταληπτότητα των φράσεων κατείχε το συχνοτικό περιεχόμενο των συναισθημάτων. Διαπιστώθηκε πως σε συναισθήματα με χαμηλότερο συχνοτικό περιεχόμενο και μέτρια ένταση, όπως αυτό του φόβου και της λύπης, ήταν δυσκολότερη η αποσαφήνιση του λεξιλογικού περιεχομένου και λόγω της επιπρόσθετης ηχηρότητας από την αντήχηση γινόταν ακόμα πιο δυσνόητο το τελικό αποτέλεσμα. Σε συναισθήματα με υψηλή ένταση, ξεκάθαρη ποιότητα φωνής και έντονη παρουσία υψηλού συχνοτικού περιεχομένου, παρατηρήθηκε μια σχετική καταληπτότητα. Θα πρέπει εδώ να σημειωθεί πως τα χαρακτηριστικά του κάθε συναισθήματος συμπεριφέρονται διαφορετικά σε κλειστούς χώρους. Σύμφωνα με τις ιδιαιτερότητες του κάθε αντηχητικού χώρου θα επηρεαστεί αναλόγως και το φάσμα του κάθε συναισθήματος, άλλοτε ενισχύοντας περαιτέρω ήδη κάποιες ηχηρές συλλαβές και άλλοτε διατηρώντας ανέπαφη την καταληπτότητα των λέξεων.

Υπήρξαν ορισμένες περιπτώσεις κατά τις οποίες επιτεύχθηκε πλήρης αναγνώριση ολόκληρων των προτάσεων ενώ ήταν εμφανής η απουσία καταληπτότητας. Εδώ αποδεικνύεται πως οι ακροατές ήταν σε θέση να αποσαφηνίσουν όλες τις λέξεις μιας μη καταληπτής πρότασης βασιζόμενοι στην λογική των συμφραζόμενων. Κατανοώντας μόνο λιγοστές λέξεις της πρότασης, κατάφεραν να εξάγουν ολόκληρο το λεξιλογικό περιεχόμενο της χρησιμοποιώντας την κοινή λογική για την επίτευξη ενός γενικότερου νοήματος.

Όσο αναφορά την αναγνώριση των συναισθημάτων, παρατηρήθηκε πως ένα συντριπτικό ποσοστό των ακροατών ήταν σε θέση να αναγνωρίσει το σωστό συναίσθημα στην εκπεμπόμενη ομιλία ακόμη και αν δεν κατανοούσε καθόλου το λεξιλογικό περιεχόμενο της πρότασης. Ακόμη και σε συνθήκες έντονης ηχηρότητας και μη καταληπτότητας, το συναίσθημα εμφανιζόταν τις περισσότερες φορές ξεκάθαρο ιδιαίτερα στις κατηγορίες του θυμού, της χαράς και του ουδέτερου συναισθήματος. Οι ακροατές λόγω της υποσυνείδητης ταξινόμησης που είχαν πραγματοποιήσει στα προσωδιακά και ποιοτικά χαρακτηριστικά της φωνής, αντιλαμβάνοντουσαν και αναλογούσαν άμεσα αυτό που άκουγαν στο πιο αρμόζων συναίσθημα που τους παρεχόταν χωρίς να επηρεάζονται από την απουσία της ευκρίνειας.

Ένα ακόμη χαρακτηριστικό συμπέρασμα αυτών των πειραμάτων ακρόασης, είναι ότι ένα μεγάλο ποσοστό των ακροατών εξέλαβε το συναίσθημα του φόβου ως ουδέτερο συναίσθημα και, σε μικρότερο βαθμό, ως συναίσθημα λύπης. Αυτό συνέβη λόγω της μεγάλης υποβάθμισης ενός από τα σημαντικότερα χαρακτηριστικά του συναισθήματος του φόβου, δηλαδή αυτό του vibrato, το οποίο είχε ως συνέπεια στα αντιπροσωπευτικά αρχεία του να ακούγονται πιο «επίπεδα» συχνοτικά χωρίς έντονη διακύμανση του pitch και με την μέτρια ένταση που είχε εξαρχής ο φόβος ως συναίσθημα, γινόταν άμεση σύνδεση με το ουδέτερο συναίσθημα που έχει παρόμοια χαρακτηριστικά. Επίσης θα πρέπει να σημειωθεί πως σε ένα γενικότερο βαθμό, οποιοδήποτε συλλαβή, λέξη, φράση, ή πρόταση εμπειρείχε έντονα το στοιχείο της άμεσης απόκρισης («ατάκα»), αυτόματα κατευθυνόταν από τους ακροατές περισσότερο προς τα πιο έντονα συναισθήματα όπως είναι αυτό της χαράς και του θυμού ανεξάρτητα από το που πραγματικά άνηκε.

Ως τελικό συμπέρασμα θα πρέπει να επισημανθεί και να τονιστεί το γεγονός ότι οι αλγόριθμοι αφαίρεσης της αντήχησης και αποθορυβοποίησης λειτούργησαν ικανοποιητικά σε όλες τις κατηγορίες των αρχείων που εφαρμόστηκαν. Στις καταληπτές ομάδες ανέβασαν τα ήδη ικανοποιητικά αποτελέσματα σε ακόμα μεγαλύτερα ποσοστά επιτυχίας, και κατάφεραν να κάνουν λίγο περισσότερο κατανοητές της μη-καταληπτές ομάδες, ιδιαίτερα στην κατηγορία του λεξιλογικού περιεχομένου, έχοντας ως αποτέλεσμα την αναγνώριση προτάσεων που δεν ήταν προηγουμένως καταληπτές, και συνεπώς την άνοδο των εύστοχων απαντήσεων και την ανάλογη μείωση των άστοχων απαντήσεων.

Όσο αναφορά την οπτοηλεκτρονική διάταξη που χρησιμοποιήθηκε, αποδείχθηκε ότι με απλά μέσα μπορεί κανείς να έχει πολύ αξιοπρεπή αποτελέσματα, μπορούν όμως να υπάρξουν και σημαντικά περιθώρια βελτίωσης. Η επιλογή ενός χώρου με μικρούς χρόνους αντήχησης παρουσιάζεται ως ιδανικό σενάριο στην εξ' αποστάσεως καταγραφή ήχου με οπτοηλεκτρονικές διατάξεις αφού δεν θα επιβάρυνε τις ηχογραφήσεις με επί πρόσθετο θόρυβο υπό την μορφή πρώιμων ανακλάσεων και αντηχητικής ουράς. Βελτίωση θα μπορούσε να επιτευχθεί επίσης και με την χρήση καλύτερων μέσων μετατροπής σήματος όπως η χρησιμοποιηθείσα φωτοαντίσταση, η καλωδίωση και η κάρτα ήχου του υπολογιστή. Τελικώς, μια μεγάλη πιθανότητα βελτίωσης θα παρουσιαζόταν κατά την τοποθέτηση του μεγαφώνου σε διαφορετικές

θέσης και αποστάσεις απέναντι στην επιφάνεια δόνησης έως ότου βρεθεί το βέλτιστο σημείο εκπομπής του ηχητικού υλικού.

Βιβλιογραφία

- Γρ. Ζέρβας Παναγιώτης Διδακτορική Διατριβή : *Μοντελοποίηση και ψηφιακή επεξεργασία προσωδιακών φαινομένων της ελληνικής γλώσσας με εφαρμογή στην σύνθεση ομιλίας* (2007)
- Ιωάννου Χαράλαμπος Διπλωματική Εργασία *Μελέτη Αλγόριθμων Αποθορυβοποίησης Σήματος Ομιλίας* (2010)
- Καπετανόγλου Νικόλαος Παγίδας Ιωάννης Πτυχιακή Εργασία: *Μελέτη για ηχητική εγκατάσταση προδιαγραφών κινηματογραφικής αίθουσας στον χώρο του αμφιθεάτρου* (2007)
- Καρώνης Δημήτρης Πτυχιακή Εργασία :*Κατασκευή και έλεγχος λειτουργίας οπτικών διατάξεων ανίχνευσης ήχων* (2009)
- Μιχαλέτου Ελένη Διπλωματική εργασία :*Παραμετροποίηση σήματος ομιλίας για αναγνώριση συναισθήματος ομιλητή* (2008)
- Χ. Σπυρίδης Χαράλαμπος *Η φυσική των μουσικών οργάνων* (2008)
- Χαδέλλης Λουκάς *ΤΕΧΝΟΛΟΓΙΑ ΗΧΟΥ* (2010)
- F.Alton Everest, Ken C.Pohlmann *Master Handbook of Acoustics* 5th Edition (2009)
- <http://www.genelec.com/products/previous-models/8030a/>
- Eugene Hecht *Optics* London Reading : Addison-Wesley fourth edition (2002)
- iZotope RX 3 Reference Manual (2013)
- Philipos C. Loizou *Speech Enhancement: Theory and Practice* (2007)
- C B Scruby and L E Drain *Laser Ultrasonics-Techniques and Applications* (1990)
- Siegman, Anthony E. *Lasers* /Anthony E. Siegman.. Sausalito, California : University Science Books (1986)