ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Μη-Γραμμική Υπολογιστική Μοντελοποίηση Φωνής με Στοιχεία Αεροδυναμικής του Φωνητικού Σωλήνα

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

ΑΘΑΝΑΣΙΟΥ Α. ΚΑΤΣΑΜΑΝΗ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &
Μηχανικού Υπολογιστών Ε.Μ.Π.

Αθήνα, Ιούνιος 2009

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
Σχολή Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής

Μη-Γραμμική Υπολογιστική Μοντελοποίηση Φωνής με Στοιχεία Αεροδυναμικής του Φωνητικού Σωλήνα

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

ΑΘΑΝΑΣΙΟΥ Α. ΚΑΤΣΑΜΑΝΗ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &
Μηχανικού Υπολογιστών Ε.Μ.Π.

Συμβουλευτική Επιτροπή:

Καθ. Πέτρος Μαραγκός (Επιβλέπων)
Καθ. Γεώργιος Καραγιάννης
Καθ. Τρύφων Κουσιουρής

Επταμελής εξεταστική επιτροπή:

...	...	...
Π. Μαραγκός Καθηγητής Ε.Μ.Π. (Επιβλέπων)	Γ. Καραγιάννης Καθηγητής Ε.Μ.Π.	Τ. Κουσιουρής Καθηγητής Ε.Μ.Π.
...	...	
Σ. Τσαγγάρης Καθηγητής Ε.Μ.Π.	Γ. Παπαβασιλόπουλος Καθηγητής Ε.Μ.Π.	
...	...	
Α. Ποταμιάνος Αν. Καθ. Πολ. Κρήτης	Ι. Στυλιανού Αν. Καθ. Παν. Κρήτης	

Αθήνα, Ιούνιος 2009

Η παρούσα διδακτορική διατριβή πραγματοποιήθηκε στα πλαίσια του προγράμματος ΠΕΝΕΔ-2003, της Γενικής Γραμματείας Έρευνας και Τεχνολογίας. Το πρόγραμμα συγχρηματοδοτήθηκε κατά 80% από την Ευρωπαϊκή Ένωση και κατά 20% από το Ελληνικό Δημόσιο.

This Ph.D. thesis was supported by grant PENED-2003 of the Greek Ministry of Development-GSRT. It is co-financed by E.U.-European Social Fund (80%) and National Resources (20%).

. . .

ΑΘΑΝΑΣΙΟΣ Α. ΚΑΤΣΑΜΑΝΗΣ
Υποψήφιος Διδάκτωρ Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Γλωσσάρι

ανάλυση κανονικής συσχέτισης canonical correlation analysis. 91

αντιστοίχιση registration,alignment. 114, 116, 120

αχλαδόσχημες κοιλότητες piriform fossae/sinuses. 6

γλωττίδα glottis. 10

δίνη, στρόβιλος vortex. 20

διατμητικό στρώμα shear layer. 23

ενεργά μοντέλα εμφάνισης active appearance models. 95

ηλεκτρομαγνητική καταγραφή των αρθρωτών electromagnetic articulography (EMA). 113, 114

μεσο-οβελιαίο midsagittal. 54

μοντέλο άρθρωσης articulatory model. 121

συνοριακό στρώμα boundary layer. 14

σύμμειξη fusion. 5, 94, 95, 101, 104–106

υπερώα velum. 10, 11

υπογλωττιδική subglottal. 1

φλέβα ροής jet. 25

ΠΡΟΛΟΓΟΣ

Ήρθε η στιγμή να εξηγηθώ. Να εξηγηθώ στους ανυποψίαστους φίλους, στην οικογένειά μου, στον επίδοξο αναγνώστη που θα βρει τυχαία ένα αντίτυπο της διατριβής στη βιβλιοθήκη μου και θα απορήσει με το μέγεθος και το ακατανόητο του τίτλου. Να προσπαθήσω να δώσω μια απλή απάντηση στην προβοκατόρικη ερώτηση φίλου μου, 'Μετά από τόσα χρόνια προσπάθειας για το διδακτορικό σου, πιστεύεις ότι τελικά άξιζε;'.

Και θα μιλήσω κυρίως για τους ανθρώπους του διδακτορικού. Γιατί πιστεύω ότι η συναναστροφή μαζί τους ήταν αυτή από την οποία αποκόμισα τα μεγαλύτερα προσωπικά οφέλη ως μαθητευόμενος διδάκτορας. Η δυναμική συνεργασία, οι έντονες συζητήσεις, οι χαλαρωτικοί περίπατοι, τα αγχωτικά, δημιουργικά ξενύχτια και τα όμορφα ταξίδια ήταν καταλυτικής σημασίας. Και θα εκφράσω την αμέριστη ευγνωμοσύνη μου στους ανθρώπους αυτούς που συμπρωταγωνίστησαν στην προσπάθειά μου. Ως σκηνοθέτης επίσης, ο καθηγητής μου κ. Μαραγκός. Εμπνευστής της κεντρικής ιδέας, ενθουσιώδης περιπατητής δύσβατων μονοπατιών και απαιτητικός συνεργάτης ήταν εκεί για να κινητοποιήσει, να ακούσει, να συμβουλεύσει. Τα παιδιά στο εργαστήριο. Ό,τι και να πω θα είναι λίγο. Ο Γιώργος, φίλος και συζητητής σε όλες τις φάσεις. Μεθοδικός, υπομονετικός και με αυτοπεποίθηση, ικανότατος συνεργάτης. Ξεκινήσαμε και συνεχίζουμε μαζί. Ο Βασίλης, ακούραστος, ενθουσιώδης, ανθρώπινος, στη γωνία του ανεξαρτήτως ώρας, ημέρας ή εποχής, πρόθυμος και έτοιμος να συνδράμει. Ο Ιάσονας, αφοσιωμένος και ασυμβίβαστος, σχολιάζει, αστειεύεται, προτείνει. Ο Σταμάτης και ο Τάσος. Συνταξιδιώτες μου και οι δύο, ο πρώτος ακολουθεί την αναλυτική προσέγγιση και ο δεύτερος βρίσκει τη βέλτιστη λύση, αρκεί να έχει πρώτα αποδείξει την ύπαρξη και τη μοναδικότητά της. Η επιτυχία είναι εξασφαλισμένη. Ο άλλος Γιώργος και η Νατάσα, σκεπτικοί και ανήσυχοι, αναλύουν, προβληματίζονται, συζητούν. Ο Δημήτρης, ο παλαιότερος, συγκάτοικος και συνερευνητής, αναζητά διαμορφώσεις. Ο Σταύρος, ο νεότερος, αναζητά νοήματα μέσα σε χειρονομίες. Η Νάνσυ παρατηρεί, προβληματίζεται, αναζητεί. Η Αγγελική, η Όλγα, η Χάρις, η Δέσποινα, η Βίκυ. Μοναδικός συνδυασμός ανθρώπων.

Οι υπόλοιποι συμμετέχοντες στο ΠΕΝΕΔ, το ερευνητικό πρόγραμμα στο οποίο ήταν ενταγμένο το διδακτορικό μου. Οι καθηγητές κ. Τσαγγάρης και κ. Ποταμιάνος. Ο Πύρρος και ο Γιάννης. Συγκερασμός προσεγγίσεων με στόχο την κατανόηση της ανθρώπινης φωνής. Οι συμμετέχοντες στο ευρωπαϊκό ερευνητικό πρόγραμμα ASPI για την αντιστροφή της φωνής. Ευχαριστώ για τις συζητήσεις και για τη συνεργασία.

Τα μέλη της τριμελούς συμβουλευτικής επιτροπής μου. Ο καθηγητής κ. Καραγιάννης, που πρώτος μου μίλησε για την παραγωγή της ανθρώπινης φωνής και ήταν πάντα πρόθυμος και πολύτιμος σύμβουλος. Ο καθηγητής κ. Κουσιουρής και τα υπόλοιπα μέλη της επταμελούς συμβουλευτικής επιτροπής μου. Οι καθηγητές κ. Παπαβασιλόπουλος και κ. Στυλιανού. Ήταν τιμή μου η συμμετοχή τους στην εξέτασή μου και τους ευχαριστώ ιδιαίτερα για τα σχόλιά τους. Εκφράζω επίσης την ευγνωμοσύνη μου στο κοινωφελές ίδρυμα Αλέξανδρος Σ. Ωνάσης για την οικονομική υποστήριξη που μου παρείχε με τη μορφή υποτροφίας.

Οι άνθρωποι του διδακτορικού μου.

Η οικογένειά μου. Οι γονείς, η Αγγελική και η γιαγιά μου. Μου προσέφεραν και μου προσφέρουν τα πάντα και θα τους είμαι για πάντα ευγνώμων. Οι φίλοι μου. Ήταν εκεί για να συζητήσουμε και κάτι άλλο ή για να προσπαθήσω να τα πω κάπως αλλιώς. Οι πιο

υποψιασμένοι, οι Γιάννηδες, ο Νίκος, ο Παναγιώτης, ο Κώστας, ο Φίλιππος και ο Κωστής. Ο Νίκος από το Κέμπριτζ και ο Νίκος στο Λονδίνο. Γνώριζαν πρόσωπα και καταστάσεις. Και οι ανυποψίαστοι, οι Γιάννηδες, ο Γιώργος, η Ευαγγελία, ο Στράτος, η Βαρβάρα και ο Κώστας. Η Κατερίνα. Άνθρωποι που ήταν και είναι μαζί μου.

Και ναι, τελικά άξιζε!

Αθανάσιος Κατσαμάνης

ΠΕΡΙΛΗΨΗ

Πολλά συμβατικά υπολογιστικά μοντέλα φωνής συνήθως παρακάμπτουν την αεροδυναμική μοντελοποίηση ακολουθώντας φαινομενολογική προσέγγιση για τον προσδιορισμό των ακουστικών πηγών στη φωνητική οδό. Αξιοποιώντας την επικρατούσα θεώρηση για το πεδίο ροής στο φωνητικό σωλήνα και συνδυάζοντας συμπεράσματα που προκύπτουν από τη μελέτη της αεροδυναμικής τόσο στη γλωττίδα όσο και στο υπερλαρύγγειο τμήμα, στα πλαίσια της διδακτορικής διατριβής αναπτύχθηκε ένα μοντέλο που επιτρέπει την υπολογιστική προσομοίωση σημαντικών αεροδυναμικών χαρακτηριστικών που επιδρούν στον παραγόμενο ήχο. Το αεροδυναμικό μοντέλο συνδυάστηκε με ένα βελτιωμένο σύστημα προσομοίωσης του ακουστικού πεδίου μέσα στη φωνητική οδό για σύνθεση φωνής με τη χρήση αρθρωτών. Ο συνδυασμός επιτεύχθηκε μέσω κατάλληλης αεροακουστικής μοντελοποίησης στη γλωττίδα και σε ενδεχόμενες στενώσεις της φωνητικής οδού.

Για τον έλεγχο του συνθέτη φωνής, αναπτύχθηκε σύστημα ταυτοποίησης του ανθρώπινου φωνητικού συστήματος με βάση ένα παρατηρούμενο σήμα φωνής. Το εν λόγω πρόβλημα συχνά αναφέρεται ως αντιστροφή φωνής. Αναπτύχθηκε ένα σύστημα αντιστροφής φωνής το οποίο βασίζεται σε οπτικακουστική θεώρηση της φωνής. Η σύνθετη σχέση μεταξύ της οπτικοακουστικής πληροφορίας και των χαρακτηριστικών της φωνητικής οδού προσεγγίζεται μέσω ενός διακοπτόμενου γραμμικού δυναμικού μοντέλου. Κάθε επιμέρους τμηματικό μοντέλο υπολογίζεται αποδοτικά μέσω στατιστικών τεχνικών όπως είναι η μεγιστοποίηση της πιθανοφάνειας και η ανάλυση κανονικής συσχέτισης. Η εναλλαγή μεταξύ των επιμέρους μοντέλων καθορίζεται από μια διακριτή διαδικασία Markov. Μελετήθηκαν εναλλακτικά συνδυαστικά σχήματα που επιτρέπουν αλληλεπίδραση μεταξύ της ακουστικής και της οπτικής ροής πληροφορίας σε διάφορα επίπεδα συγχρονισμού. Χρησιμοποιώντας τα οπτικά σε συνδυασμό με τα ακουστικά χαρακτηριστικά επιτυγχάνεται η αποδοτική εκτίμηση των τροχιών που ακολουθούνται από διάφορα σημεία ενδιαφέροντος του συστήματος παραγωγής φωνής. Τα πειραματικά αποτελέσματα δείχνουν ότι με την αξιοποίηση της πολυτροπικής πληροφορίας στο προτεινόμενο σύστημα βελτιώνεται η αποτελεσματικότητα της αντιστροφής της φωνής σε σχέση με αντίστοιχα συστήματα που χρησιμοποιούν αποκλειστικά τη μία ή την άλλη πηγή πληροφορίας.

Με βάση το προτεινόμενο υπολογιστικό μοντέλο φωνής και πληθώρα δεδομένων άρθρωσης γίνεται δυνατή η μίμηση του ανθρώπινου φωνητικού συστήματος. Συγκεκριμένα, η ακολουθία καταστάσεων άρθρωσης μοντελοποιείται ως διαδικασία Markov και τα χαρακτηριστικά της ταυτοποιούνται μέσω οπτικοακουστικής αντιστροφής της φωνής. Σε κάθε κατάσταση άρθρωσης, με δεδομένη την αντίστοιχη περιγραφή της γεωμετρίας της φωνητικής οδού είναι δυνατή η σύνθεση φωνής με τη συνδυασμένη εφαρμογή των μοντέλων αεροδυναμικής και ακουστικής. Η γεωμετρία της φωνητικής οδού περιγράφεται μέσω παραμετρικού μοντέλου άρθρωσης που εκπαιδεύεται με την αξιοποίηση δεδομένων άρθρωσης από εικόνες ακτίνων-Χ και προσαρμόζεται κατάλληλα στο ορατό τμήμα της γλώσσας σε εικόνες υπερήχων της στοματικής κοιλότητας. Το προτεινόμενο πλαίσιο επιτρέπει την ευρύτερη εφαρμογή και αξιολόγηση του συστήματος αεροδυναμικής και ακουστικής προσομοίωσης αλλά και της διαδικασίας αντιστροφής φωνής.

ABSTRACT

Conventional computational speech models usually avoid detailed aerodynamic modeling and determine sound sources in the vocal tract in a phenomenological manner. In this dissertation, a model is developed that allows the computational simulation of important aerodynamic properties that could affect the produced sound. The model exploits recent theoretical and experimental results concerning the flow field in the vocal tract and combines conclusions related to aerodynamics and aeroacoustics of the glottis and the supralaryngeal parts. The aerodynamic-aeroacoustic model is combined with an improved vocal tract acoustics simulation module to achieve articulatory synthesis.

To control the articulatory synthesizer, an inversion system was developed that can identify the hidden vocal tract properties given an observed speech signal. The speech inversion system treats speech as essentially an audiovisual process and approximates the complex mapping between the observed information and the vocal tract by means of a switching model. Each submodel is trained via maximum likelihood and canonical correlation analysis. Switching between the submodels is determined by a discrete Markov process. Various alternative audiovisual fusion schemes were investigated that allow interaction between acoustic and optical information at various levels of synchronization. The goal is to recover the underlying vocal tract geometry. Experimental results demonstrate that exploitation of multimodal information by the proposed model clearly benefits inversion results, compared to approaches that exclusively use the one or the other modality.

Based on the proposed computational speech model and various articulatory data, mimicing the human speech system becomes possible. More specifically, the articulatory state sequence is modeled as a Markov process and its properties are identified via audiovisual speech inversion. At each articulatory state, given the corresponding description of the vocal tract, we can resynthesize speech via articylatory synthesis. Vocal tract geometry is described by an articulatory model that is trained on X-ray vocal tract data and is properly fitted to the visible part of the tongue in ultrasound data of the mouth cavity. The proposed framework allows broader application and evaluation of the speech acoustics and aerodynamics simulation system and the speech inversion process.

Κεφάλαιο 1

Εισαγωγή

Σκεφτείτε ότι θέλετε να εκφέρετε κάτι, έστω τη λέξη 'καλημέρα'. Και στη συνέχεια, πείτε τη, έστω χαρούμενα.

Μπορεί να αλλάζει η λέξη, η προσωδία, η ένταση αλλά ο βασικός μηχανισμός παραγωγής φωνής παραμένει ο ίδιος. Για αυτόν που βρίσκεται απέναντι η 'καλημέρα' μας είναι ένα ηχητικό σήμα και η εικόνα του χαρούμενου ομιλητή. Οτιδήποτε πούμε τελικά εκφράζεται ως σήμα ήχου και εικόνα ενός ομιλούντος προσώπου. Και όποιος είναι απέναντί μας γενικά μπορεί να το επαναλάβει.

Αντικείμενο της διδακτορικής διατριβής που παρουσιάζεται είναι η ανάπτυξη ενός μοντέλου για την παραγωγή φωνής που επιτρέπει σε έναν υπολογιστή να μιμηθεί μια ανθρώπινη εκφώνηση. Πέρα από τις ενδεχόμενες εφαρμογές ενός τέτοιου εγχειρήματος, η όλη προσπάθεια υποκινήθηκε βασικά από καθαρό επιστημονικό ενδιαφέρον για το μηχανισμό παραγωγής φωνής και κυρίως για τα εμπλεκόμενα αεροδυναμικά φαινόμενα μέσα στη φωνητική οδό. Η ανάγκη μελέτης τέτοιων φαινομένων σε ένα ευρύτερο πλαίσιο επιβάλλεται από το γεγονός ότι γενικά σχετίζονται σημαντικά με χρονικές μεταβολές της κατάστασης της φωνητικής οδού που κατά κανόνα επιβάλλονται από μηχανισμούς εξωτερικούς του φυσικού αεροδυναμικού συστήματος. Εξάλλου, με την προσπάθεια μίμησης του ανθρώπινου φωνητικού συστήματος επιτρέπεται και η καλύτερη αξιολόγηση διάφορων στρατηγικών μοντελοποίησης. Δίνεται ιδιαίτερη βαρύτητα στη μελέτη των επιδράσεων της μοντελοποίησης των αεροδυναμικών-αεροακουστικών φαινομένων στο συνθετικό σήμα σε σχέση με το παρατηρούμενο αρχικό σήμα φωνής.

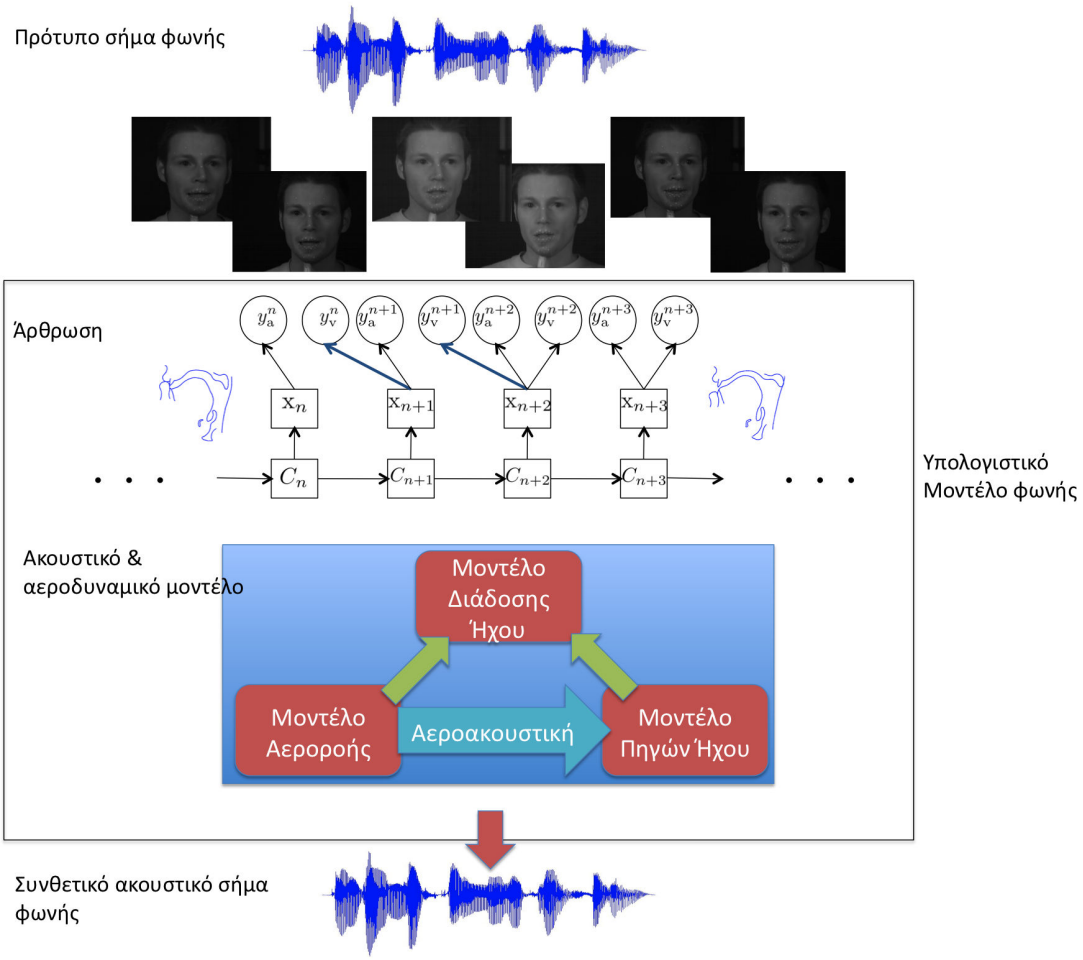
Το προτεινόμενο υπολογιστικό μοντέλο που απεικονίζεται στο Σχ. 1.1 μπορεί να θεωρηθεί ότι περιλαμβάνει δύο επιμέρους επίπεδα, κατ' αναλογία με το ανθρώπινο φωνητικό σύστημα :

Φυσικό επίπεδο Στο επίπεδο αυτό, δεδομένης μιας κατάλληλης χρονικής ακολουθίας ενός συνόλου παραμέτρων φυσιολογίας, όπως είναι η γεωμετρία της φωνητικής οδού, τα χαρακτηριστικά της γλωττίδας και η υπογλωττιδική πίεση, γίνεται δυνατή η σύνθεση φωνής μέσω αριθμητικής επίλυσης του ακουστικού πεδίου (βλ. Κεφάλαιο 3). Οι ακουστικές πηγές ήχου προσδιορίζονται μέσω αεροακουστικής μοντελοποίησης με βάση παράλληλη προσομοίωση του αεροδυναμικού πεδίου (βλ. Κεφάλαιο 4).

Επίπεδο άρθρωσης Στο επίπεδο της άρθρωσης προσομοιώνεται ο ανθρώπινος μηχανισμός ελέγχου της παραγωγής φωνής. Πιο συγκεκριμένα, προσδιορίζεται η ακολουθία των καταστάσεων της φωνητικής οδού που απαιτείται για τη σύνθεση στο φυσικό επίπεδο. Η φωνητική οδός περιγράφεται με κατάλληλο αρθρωτικό μοντέλο. Η μίμηση μιας συγκεκριμένης εκφώνησης γίνεται μέσω μιας διαδικασίας αντιστροφής με βάση το παρατηρούμενο οπτικοακουστικό σήμα φωνής (βλ. Κεφάλαιο 5).

Το εν λόγω μοντέλο παρουσιάζει δύο βασικά καινοτομικά χαρακτηριστικά :

Μη γραμμική αεροδυναμική-αεροακουστική μοντελοποίηση Βασίζεται στην πλέον σύγχρονη περιγραφή της φυσικής του ανθρώπινου φωνητικού συστήματος ενώ έχει



Σχήμα 1.1: *Υπολογιστικό μοντέλο για μίμηση φωνής με αξιοποίηση οπτικοακουστικής πληροφορίας και βελτιωμένη αεροδυναμική μοντελοποίηση*

ληφθεί μέριμνα ώστε να αποφευχθεί σημαντική αύξηση της υπολογιστικής πολυπλοκότητας κατά την αριθμητική προσομοίωση. Εφαρμόζονται σύγχρονα μοντέλα αεροδυναμικής και αεροακουστικής τόσο για τη γλωττίδα και τη φώνηση όσο και για τις στενώσεις και την παραγωγή άφωνων ήχων.

Οπτικοακουστική θεώρηση φωνής Κατ' αναλογία με τον ανθρώπινο μηχανισμό αντίληψης, συνδύαζει το ακουστικό με το οπτικό σήμα φωνής ώστε να εξάγει όσο το δυνατόν ακριβέστερη περιγραφή της άρθρωσης που απαιτείται για τη μίμηση φωνής.

1.1 Γενική περιγραφή και σημασία των ερευνητικών περιοχών

«Υπάρχουν μάλλον τρεις πρωταρχικοί λόγοι για την εισαγωγή εναλλακτικών φυσικών μοντέλων στη θέση των καθιερωμένων: χρησιμότητα, φυσική πιστότητα και ενδιαφέρον. Από αυτά, η χρησιμότητα είναι μάλλον το αντικειμενικότερο και σημαντικότερο», [167]. Είναι λόγια του Teager που πριν από τριάντα περίπου χρόνια προσπαθούσε να στρέψει το ενδιαφέρον της ερευνητικής κοινότητας στη μελέτη της αερορροής μέσα στο φωνητικό σωλήνα. Η επιχειρηματολογία του, ενισχυμένη και από τη σχετικά περιορισμένη επιτυχία των τεχνολογιών φωνής εκείνη την εποχή, αμφισβητούσε σημαντικές υποθέσεις της κλασσικής θεώρησης για την παραγωγή φωνής όπως καθιερώθηκε με τις ερευνητικές εργασίες των Chiba & Kajiyama, Fant και Flanagan [31, 52, 55]. Τα πειραματικά δεδομένα που παρουσίασε παρείχαν

ενδείξεις για την ανάπτυξη μη ομοιόμορφης αερορροής μέσα στο φωνητικό σωλήνα. Με βάση αυτά αλλά και πληθώρα θεωρητικών αναφορών υποστήριξε ότι υπάρχει ενεργός συμμετοχή της μη-ακουστικής αερορροής στην παραγωγή φωνής. Εκτενέστερη επισκόπηση της εν λόγω έρευνας σε αντιδιαστολή με την κλασσική προσέγγιση γίνεται στην Ενότητα 2.5.1.

1.1.1 Αεροδυναμική στο φωνητικό σωλήνα

Οι προτάσεις του Teager μπορεί να μη βρήκαν άμεση αποδοχή αλλά η περαιτέρω μελέτη της αεροδυναμικής συμπεριφοράς του φωνητικού συστήματος τα επόμενα χρόνια ενίσχυσε κάποιες από τις παρατηρήσεις του. Ιδιαίτερη ερευνητική προσπάθεια εστιάστηκε στην ανάλυση του πεδίου ροής στη γλωττίδα, που εμφανίζεται ως ο κύριος μηχανισμός ακουστικής διέγερσης. Οι αρχικές προσεγγίσεις του Van Den Berg και των συνεργατών του [172] και των Ishizaka & Flanagan [73] βελτιώθηκαν σημαντικά στη συνέχεια με την πραγματοποίηση επιπλέον πειραματικών μετρήσεων, όπως π.χ. στο [2], αριθμητικών προσομοιώσεων, π.χ. [3, 183], και με την εφαρμογή πιστότερων φυσικών μοντέλων για τη γλωττίδα, τόσο αεροδυναμικών όσο και μηχανικών [77, 97, 125]. Πέρα από τη γλωττίδα, όσον αφορά στο εσωτερικό της φωνητικής οδού, στα [19, 149] οι Barney, Shadle & Davies έδειξαν πειραματικά την εμφάνιση στροβιλώδους ροής σε ένα μηχανικό μοντέλο της φωνητικής οδού και μελέτησαν τη σημασία της στα πλαίσια της παραγωγής φωνής ενώ ο Thomas ασχολήθηκε με την αριθμητική επίλυση μιας απλοποιημένης εκδοχής του αεροδυναμικού πεδίου για σύνθεση φωνής [169].

1.1.2 Αεροακουστική στο φωνητικό σωλήνα

Δεν ήταν άμεσα ξεκάθαρο βέβαια πώς οι όποιες ανομοιομορφίες της ροής επιδρούν τελικά στον ήχο. Πιθανές συνέπειές τους μοντελοποιήθηκαν σε φαινομενολογικό επίπεδο ως διαμορφώσεις πλάτους και συχνότητας των συντονισμών της φωνής και μελετήθηκαν με τη χρήση μη γραμμικών αλγορίθμων επεξεργασίας σήματος [104, 105, 128]. Η πιο ακριβής φυσική περιγραφή της αλληλεπίδρασης του πεδίου ροής με το ακουστικό πεδίο έγινε στη συνέχεια δυνατή με την κατάλληλη εφαρμογή της αεροακουστικής θεωρίας. Πιο συγκεκριμένα, ως αεροακουστική προσέγγιση για την παραγωγή φωνής θεωρείται η προσπάθεια σύνδεσης του ήχου που παρατηρείται στο μακρινό ακουστικό πεδίο με την αερορροή μέσα στη φωνητική οδό. Η σχετική αναζήτηση συνήθως επικεντρώνεται στον προσδιορισμό των ηχητικών πηγών που αναπτύσσονται είτε λόγω ανομοιογένειας της ροής είτε λόγω αλληλεπίδρασής της με τις περιβάλλουσες συμπαγείς δομές που σχηματίζουν το φωνητικό αεραγωγό. Με βάση την αεροακουστική προσέγγιση μπορεί να θεωρηθεί ότι στη συνέχεια οι πηγές αυτές διεγείρουν το ακουστικό μέσο σαν να είναι εξωτερικές.

Παρά την πολυπλοκότητα του συγκεκριμένου εγχειρήματος, τα τελευταία χρόνια έχουν γίνει σημαντικά βήματα που συμβάλλουν σε μεγάλο βαθμό στην περαιτέρω κατανόηση του υποκείμενου μηχανισμού [65, 70, 71, 91, 107, 109]. Ο McGowan στο [109] προσπάθησε να εξηγήσει, τουλάχιστον σε τάξη μεγέθους, την παραγωγή έμφωνων ήχων στο αεροακουστικό πλαίσιο, για το οποίο έδωσε περαιτέρω διαίσθηση ο Hirschberg στο [65]. Με την εφαρμογή της προσέγγισης του Howe [68], ο Krane στο [91] δίνει ένα πλαίσιο αεροακουστικής μοντελοποίησης των τυρβώδων ήχων. Οι Howe & McGowan στο [70] εξετάζουν πιο συγκεκριμένα την παραγωγή του τυρβώδους [s] ενώ στο [71] ερμηνεύουν την παραγωγή ήχου στη γλωττίδα ως αποτέλεσμα αλληλεπίδρασης της αερορροής με ένα συμπαγές αντικείμενο μεταβαλλόμενο χρονικά και χωρικά κατά τυχαίο τρόπο. Στα [15, 161, 183] μέσω αριθμητικής προσομοίωσης και αεροακουστικών αναλογιών μελετώνται οι αεροδυναμικές πηγές ήχου που εμφανίζονται στη γλωττίδα για διάφορες περιπτώσεις και εξάγονται σημαντικά συμπεράσματα για τα χαρακτηριστικά τους. Σχετικές λεπτομέρειες δίνονται στο Κεφάλαιο 2.

1.1.3 Σύνθεση φωνής με αριθμητική προσομοίωση

Παρ' όλ' αυτά, οι συμβατικοί συνθέτες φωνής με τη χρήση αρθρωτών [24, 26, 43, 114, 146, 150, 151] σε μεγάλο βαθμό δεν έχουν ακόμα ενσωματώσει σημαντικές πτυχές του συστήματος της ανθρώπινης παραγωγής φωνής όπως αυτές έχουν αναδειχθεί μέσω της αεροδυναμικής - αεροακουστικής προσέγγισης. Χαρακτηριστικό παράδειγμα είναι η μοντελοποίηση της ακουστικής διέγερσης στη γλωττίδα αλλά και σε στενώσεις της φωνητικής οδού για την παραγωγή τυρβώδων ήχων που συνήθως πραγματοποιείται σε φαινομενολογικό επίπεδο. Αυτό έχει ως αποτέλεσμα να είναι συχνά απαραίτητος ο ευριστικός προσδιορισμός διάφορων χαρακτηριστικών των πηγών με βάση τις επιθυμητές ιδιότητες του παραγόμενου σήματος. Μια προσπάθεια αεροδυναμικής μοντελοποίησης αξιοποιείται συνήθως αποσπασματικά στις στενώσεις ή στη γλωττίδα, ενώ στην περίπτωση που έχει υιοθετηθεί ενδεχόμενα λεπτομερέστερη αεροδυναμική περιγραφή [26] η υπολογιστική πολυπλοκότητα αυξάνεται σημαντικά, γεγονός που αποτρέπει την πρακτική εφαρμογή της εν λόγω προσέγγισης για σύνθεση. Ιδιαίτερη αναφορά στα επιμέρους ακουστικά και αεροδυναμικά χαρακτηριστικά των συνθετών φωνής γίνεται στα Κεφάλαια 3 και 4 αντίστοιχα. Για τον έλεγχο του συνθέτη φωνής είναι απαραίτητος ο προσδιορισμός της ακολουθίας παραμέτρων φυσιολογίας του φωνητικού συστήματος όπως είναι η γεωμετρία της φωνητικής οδού ή η υπογλωττιδική πίεση. Στο προτεινόμενο μοντέλο, αυτό πραγματοποιείται μέσω αντιστροφής φωνής.

1.1.4 Αντιστροφή φωνής με αξιοποίηση πολυτροπικών δεδομένων

Με την αντιστροφή φωνής επιτυγχάνεται ο προσδιορισμός χαρακτηριστικών του συστήματος της φωνητικής οδού μόνο με εξωτερική παρατήρηση. Ιδανικά, είναι επιθυμητή η ανάκτηση των χαρακτηριστικών που θα επιτρέπουν την ανασύνθεση της παρατηρούμενης φωνής με τη χρήση ενός υπολογιστικού μοντέλου σύνθεσης. Στα πλαίσια των μοντέλων που συνήθως χρησιμοποιούνται το ζητούμενο με την αντιστροφή είναι τελικά ο προσδιορισμός του σχήματος της φωνητικής οδού και οι πηγές ήχου μέσα σε αυτή, κυρίως στη γλωττίδα αλλά και σε στενώσεις της οδού σε περίπτωση τυρβώδων ήχων. Εκτός από το θεωρητικό ενδιαφέρον που παρουσιάζει το πρόβλημα, σχετικές συμπαγείς αναπαραστάσεις που μπορεί να προκύψουν είναι δυνατόν να εφαρμοστούν σε σύνθεση [145], αναγνώριση [87], κωδικοποίηση φωνής [144] ή γλωσσική εκπαίδευση [50].

Η απαίτηση να υπάρχει αντιστοιχία μεταξύ του ανακτώμενου συστήματος παραγωγής φωνής και του πραγματικού είναι προαιρετική και η σχετική επιλογή εξαρτάται από τη συγκεκριμένη εφαρμογή του συστήματος αντιστροφής ή τα χαρακτηριστικά του υπολογιστικού μοντέλου σύνθεσης, αν συμπεριλαμβάνεται. Για παράδειγμα, σε μια εφαρμογή κωδικοποίησης δεν υπάρχει τόσο μεγάλο ενδιαφέρον για τα φυσικά χαρακτηριστικά της φωνητικής οδού αντίθετα με την περίπτωση της εφαρμογής της αντιστροφής σε διδασκαλία άρθρωσης. Στην παρούσα εργασία, βασικό κίνητρο για την αντιστροφή αποτελεί η λεπτομερέστερη μελέτη της φυσικής και η υπολογιστική μοντελοποίηση του ανθρώπινου μηχανισμού παραγωγής φωνής οπότε η απαίτηση για αντιστοιχία του αποτελέσματος αντιστροφής με την πραγματικότητα είναι σχετικά πιο ισχυρή. Με αυτήν την απαίτηση είναι σύμφωνες τόσο η εφαρμογή του προτεινόμενου αεροδυναμικού-αεροακουστικού πλαισίου σύνθεσης όσο και η προσπάθεια αξιοποίησης πληθώρας πολυτροπικών πραγματικών δεδομένων άρθρωσης, όπως είναι εικόνες μαγνητικής τομογραφίας του κρανίου του ομιλητή, βίντεο της γλώσσας καταγεγραμμένο με υπερήχους ή οι τροχιές αισθητήρων πάνω στη γλώσσα όπως παρακολουθούνται από σύστημα ηλεκτρομαγνητικής καταγραφής (EMA). Τα δεδομένα αυτά επιτρέπουν εκτός των άλλων και τη λεπτομερή περιγραφή της γεωμετρίας της οδού που είναι ιδιαίτερα σημαντική για τη μελέτη αεροδυναμικών φαινομένων όπως είναι η αποκόλληση της ροής από τα τοιχώματα ή η γένεση τύρβης. Στα πλαίσια της παρούσας εργασίας, η αντιστροφή φωνής πραγματοποιείται επωφελώς μέσω οπτικοακουστικής φωνητικής μοντελοποίησης, όπως περιγράφεται στο Κεφάλαιο 5.

1.1.5 Οπτικοακουστική μοντελοποίηση φωνής

Η οπτικοακουστική θεώρηση της φωνής, ότι δηλαδή η φωνή είναι η συνισταμένη τόσο ακουστικής όσο και οπτικής πληροφορίας, έχει ωθήσει σημαντικά τις τεχνολογίες φωνής τα τελευταία χρόνια. Η εισαγωγή κατάλληλων οπτικών χαρακτηριστικών από το πρόσωπο του ομιλητή έχει αυξήσει την ευρωστία των συστημάτων αναγνώρισης φωνής [129] ενώ με τη χρήση ομιλούντων προσώπων οι συνθέτες φωνής βελτιώθηκαν σε φυσικότητα και καταληπτότητα [18]. Γενικά, η συμπερίληψη της οπτικής συνιστώσας της φωνής με τρόπο σχετικό με το ανθρώπινο σύστημα παραγωγής [178] και αντίληψης φωνής [111] μπορεί να ωφελήσει σημαντικά την επεξεργασία φωνής και τις διεπαφές ανθρώπου-υπολογιστή. Παρ' όλ' αυτά, δεν υπάρχει ακόμα ένα ολιστικό μοντέλο για την επεξεργασία οπτικοακουστικής φωνής. Στις περισσότερες περιπτώσεις, η σύμμειξη του ήχου και της εικόνας είναι περισσότερο καθοδηγούμενη από τα δεδομένα και λιγότερο βασισμένη σε κάποιο μοντέλο. Βασικό κίνητρο της σύνδεσης που επιχειρείται με την παρούσα εργασία μεταξύ μιας λεπτομερέστερης περιγραφής της παραγωγής φωνής και της οπτικοακουστικής αντιστροφής φωνής είναι η πεποίθηση ότι ο υποκείμενος μηχανισμός παραγωγής φωνής μπορεί να αποτελέσει το συνδετικό κρίκο μεταξύ των παρατηρούμενων οπτικών και ακουστικών εκφράσεων της φωνής. Σε ένα ευρύτερο πλαίσιο, το ζητούμενο είναι ένα οπτικοακουστικό μοντέλο φωνής που να ενσωματώνει γνώση σχετική με την κατάσταση και τη δυναμική της φωνητικής οδού. Αυτό το μοντέλο θα μπορούσε να επιτρέψει την ακριβέστερη ερμηνεία φωνητικών ιδιοτήτων που δεν μπορούν να εξηγηθούν ούτε με την ακουστική του σήματος φωνής μόνο αλλά ούτε και με την εικόνα του ομιλητή, όπως για παράδειγμα το φαινόμενο McGurk [111].

1.2 Ερευνητικές συνεισφορές

Οι κύριες ερευνητικές συνεισφορές που προέκυψαν στα πλαίσια της διδακτορικής διατριβής είναι σε δύο βασικούς άξονες που σχετίζονται με τα δύο επίπεδα του προτεινόμενου υπολογιστικού μοντέλου φωνής.

1.2.1 Αεροδυναμική και αεροακουστική για σύνθεση φωνής με αρθρωτές

Πολλά συμβατικά υπολογιστικά μοντέλα φωνής συνήθως παρακάμπτουν την αεροδυναμική μοντελοποίηση ακολουθώντας ευριστικές τεχνικές που βασίζονται κυρίως σε φαινομενολογικά συμπεράσματα. Έτσι, αποτρέπεται σε μεγάλο βαθμό η λεπτομερής μελέτη της συσχέτισης μεταξύ της αεροροής μέσα στο φωνητικό σωλήνα και του παρατηρούμενου ακουστικού πεδίου. Αξιοποιώντας την επικρατούσα θεώρηση για το πεδίο ροής στο φωνητικό σωλήνα και συνδυάζοντας συμπεράσματα που προκύπτουν από τη μελέτη της αεροδυναμικής τόσο στη γλωττίδα όσο και στο υπερλαρύγγειο τμήμα, αναπτύχθηκε ένα μοντέλο που επιτρέπει την υπολογιστική προσομοίωση σημαντικών αεροδυναμικών χαρακτηριστικών που επιδρούν στον παραγόμενο ήχο.

Παρακάμπτοντας την αριθμητική επίλυση των εξισώσεων Navier-Stokes που θα είχε σημαντικά αυξημένες υπολογιστικές απαιτήσεις, το προτεινόμενο μοντέλο προσομοιώνει βασικά χαρακτηριστικά τόσο του αστρόβιλου (δυναμικού) όσο και του στροβιλώδους πεδίου ροής με τρόπο ώστε να μπορούν τα εν λόγω χαρακτηριστικά να αξιοποιηθούν για σύνθεση φωνής. Έχει δοθεί βαρύτητα στις ιδιότητες της ροής που αναμένεται να είναι σημαντικές για το ακουστικό πεδίο, με βάση τη θεωρία της αεροακουστικής. Η αεροδυναμική προσομοίωση για την παροχή όγκου της μη ακουστικής ροής στα χείλη ερμηνεύει ικανοποιητικά αντίστοιχες πειραματικές μετρήσεις.

Το προτεινόμενο αεροδυναμικό μοντέλο συνδυάστηκε με ένα βελτιωμένο σύστημα προσομοίωσης του ακουστικού πεδίου μέσα στη φωνητική οδό για τη σύνθεση φωνής. Ο συνδυασμός επιτεύχθηκε κυρίως με βάση συμπεράσματα που προκύπτουν από την εφαρμογή της αεροακουστικής θεωρίας στη φωνητική οδό. Το ακουστικό πεδίο υποτίθεται μονοδιάστατο

και περιγράφεται από την ακουστική πίεση και την ακουστική ογκική ταχύτητα. Ο έλεγχος του συνολικού συστήματος γίνεται με τη χρήση παραμέτρων άρθρωσης με ερμηνεία σχετιζόμενη με την ανθρώπινη φυσιολογία, όπως είναι η υπογλωττιδική πίεση ή η ελαστικότητα των φωνητικών χορδών.

Καταβάλλεται ιδιαίτερη φροντίδα για την ακουστική σύζευξη της φωνητικής οδού με επιμέρους ακουστικές κοιλότητες όπως είναι η ρινική ή οι αχλαδόσχημες κοιλότητες (piriform fossae) στο πάνω άκρο του λάρυγγα, που θεωρείται ότι προσδίδουν μεγαλύτερη φυσικότητα στην παραγόμενη φωνή. Προβλέπεται η ύπαρξη μη ακουστικής μέσης ροής μέσα στην οδό και γίνεται εφικτή η σύνθεση ακολουθιών φωνημάτων με τη χρήση δεδομένων άρθρωσης. Χρησιμοποιούνται πραγματικά δεδομένα για τη γεωμετρία της φωνητικής οδού όπως αυτά προκύπτουν με τη χρήση αξονικής τομογραφίας ή ακτίνων-Χ. Η σύγκριση των αποτελεσμάτων της προσομοίωσης με τα πραγματικά σήματα φωνής είναι ικανοποιητική.

1.2.2 Οπτικοακουστική αντιστροφή φωνής

Αναπτύχθηκε σύστημα ταυτοποίησης του ανθρώπινου φωνητικού συστήματος με βάση το παρατηρούμενο σήμα φωνής. Το εν λόγω πρόβλημα συχνά αναφέρεται ως αντιστροφή φωνής. Οι παραδοσιακές τεχνικές για αντιστροφή βασίζονται μόνο στο ακουστικό σήμα φωνής και συνήθως αντιμετωπίζουν προβλήματα λόγω του ότι το ίδιο ακουστικό αποτέλεσμα μπορεί να προκύψει από διαφορετικές καταστάσεις του συστήματος. Για να περιοριστούν τέτοιου είδους προβλήματα αναπτύχθηκε ένα σύστημα αντιστροφής φωνής το οποίο αξιοποιεί επίσης οπτική πληροφορία από το πρόσωπο του ομιλητή. Η σύνθετη σχέση μεταξύ της οπτικοακουστικής πληροφορίας και των χαρακτηριστικών της φωνητικής οδού προσεγγίζεται μέσω ενός διακοπτόμενου γραμμικού δυναμικού μοντέλου. Κάθε επιμέρους τμηματικό μοντέλο υπολογίζεται αποδοτικά μέσω στατιστικών τεχνικών όπως είναι η μεγιστοποίηση της πιθανοφάνειας και η ανάλυση κανονικής συσχέτισης. Η εναλλαγή μεταξύ των επιμέρους μοντέλων καθορίζεται από μια διακριτή διαδικασία Markov.

Μελετήθηκαν εναλλακτικά συνδυαστικά σχήματα που επιτρέπουν αλληλεπίδραση μεταξύ της ακουστικής και της οπτικής ροής πληροφορίας σε διάφορα επίπεδα συγχρονισμού. Για την ανάλυση του προσώπου χρησιμοποιήθηκαν ενεργά μοντέλα εμφάνισης και επιδείχτηκε πλήρως αυτόματη ανίχνευση προσώπου και εξαγωγή οπτικών χαρακτηριστικών. Χρησιμοποιώντας τα οπτικά χαρακτηριστικά όπως προκύπτουν από τα ενεργά μοντέλα εμφάνισης σε συνδυασμό με κατάλληλα ακουστικά χαρακτηριστικά επιτυγχάνεται η αποδοτική εκτίμηση των τροχιών που ακολουθούνται από διάφορα σημεία ενδιαφέροντος του συστήματος παραγωγής φωνής. Πειράματα πραγματοποιήθηκαν στις βάσεις QSMT και MOCHA που περιέχουν ήχο, εικόνα και πληροφορία για την κίνηση των αρθρωτών καταγεγραμμένη ηλεκτρομαγνητικά κατά την ομιλία. Τα αποτελέσματα δείχνουν ότι με την αξιοποίηση τόσο της ακουστικής όσο και της οπτικής πληροφορίας στο προτεινόμενο σύστημα βελτιώνεται η αποτελεσματικότητα της αντιστροφής της φωνής σε σχέση με αντίστοιχα συστήματα που χρησιμοποιούν αποκλειστικά τη μία ή την άλλη πηγή πληροφορίας.

Με βάση το προτεινόμενο υπολογιστικό μοντέλο φωνής και πληθώρα δεδομένων άρθρωσης γίνεται δυνατή η μίμηση του ανθρώπινου φωνητικού συστήματος. Συγκεκριμένα, η ακολουθία καταστάσεων άρθρωσης μοντελοποιείται ως διαδικασία Markov μέσω κρυφών μαρκοβιανών μοντέλων και τα χαρακτηριστικά της ταυτοποιούνται μέσω οπτικοακουστικής αντιστροφής της φωνής. Σε κάθε κατάσταση άρθρωσης, με δεδομένη μια περιγραφή της γεωμετρίας της φωνητικής οδού αλλά και των παραμέτρων άρθρωσης που μπορούν να εκτιμηθούν από το παρατηρούμενο σήμα, είναι δυνατή η σύνθεση φωνής με τη συνδυασμένη εφαρμογή των μοντέλων αεροδυναμικής και ακουστικής που αναπτύχθηκαν στα πλαίσια της διατριβής. Η γεωμετρία της φωνητικής οδού περιγράφεται μέσω παραμετρικού μοντέλου άρθρωσης που εκπαιδεύεται με την αξιοποίηση δεδομένων άρθρωσης από εικόνες ακτίνων-Χ. Με αξιοποίηση και 3Δ δεδομένων της φωνητικής οδού που έχουν συγκεντρωθεί με τη χρήση αξονικής τομογραφίας, το παραμετρικό αυτό μοντέλο προσαρμόζεται κατάλληλα στο

ορατό τμήμα της γλώσσας σε εικόνες υπερήχων της στοματικής κοιλότητας. Παράλληλα με τις εικόνες υπερήχων έχουν καταγραφεί οπτικοακουστικά δεδομένα φωνής και με αυτόν τον τρόπο γίνεται δυνατή η εκπαίδευση των στατιστικών μοντέλων για την αντιστροφή φωνής. Το προτεινόμενο πλαίσιο που εφαρμόστηκε σε δεδομένα της βάσης ASPI επιτρέπει την ευρύτερη εφαρμογή και αξιολόγηση του συστήματος αεροδυναμικής και ακουστικής προσομοίωσης αλλά και της διαδικασίας αντιστροφής φωνής.

1.3 Διάρθρωση της διδακτορικής διατριβής

Συνοψίζοντας, στα κεφάλαια που ακολουθούν παρουσιάζεται αρχικά σε συντομία η σύγχρονη θεώρηση του συστήματος παραγωγής φωνής και πιο συγκεκριμένα οι αεροδυναμικές και αεροακουστικές του συνιστώσες, Κεφάλαιο 2. Στη συνέχεια, δίνονται λεπτομέρειες για την προσομοίωση του ακουστικού πεδίου μέσα στη φωνητική οδό στο Κεφάλαιο 3. Στο Κεφάλαιο 4 περιγράφονται τα αεροακουστικά και αεροδυναμικά μοντέλα που εισάγονται στο σύστημα σύνθεσης φωνής και μελετάται η σημασία τους. Ακολουθεί η περιγραφή του στοχαστικού πλαισίου οπτικοακουστικής αντιστροφής φωνής στο Κεφάλαιο 5, όπου επίσης παρουσιάζεται η ανάπτυξη του μοντέλου άρθρωσης για τη φωνητική οδό από τα πολυτροπικά δεδομένα άρθρωσης και περιγράφεται το συνολικό πλαίσιο μίμησης φωνής. Στο Κεφάλαιο 6 δίνονται τα συμπεράσματα της διδακτορικής διατριβής και συζητώνται μελλοντικές ερευνητικές κατευθύνσεις.

1.4 Ερευνητικό πρόγραμμα ΠΕΝΕΔ για την αεροδυναμική μελέτη και μοντελοποίηση της φωνητικής οδού

Η διδακτορική διατριβή πραγματοποιήθηκε στα πλαίσια ερευνητικού προγράμματος ΠΕΝΕΔ με τίτλο: 'Μοντελοποίηση, σύνθεση και αναγνώριση φωνής με υπολογιστική αεροδυναμική ανάλυση του ανθρώπινου ηχητικού σωλήνα'. Το πρόγραμμα ξεκίνησε το 2005 και αναμένεται να ολοκληρωθεί το 2009. Επιστημονικός υπεύθυνος του προγράμματος ήταν ο Αν. Καθ. του Πολυτεχνείου Κρήτης Αλ. Ποταμιάνος ενώ επίσης συμμετείχαν οι Καθ. του ΕΜΠ Π. Μαραγκός και Σ. Τσαγγάρης καθώς και τρεις υποψήφιοι διδάκτορες. Η σχετική έρευνα κινήθηκε σε τρεις βασικούς άξονες, καθένας από τους οποίους αντιστοιχεί και σε μία ξεχωριστή διδακτορική διατριβή. Ο υποψήφιος διδάκτορας Ι. Παπαγεωργακόπουλος στη σχολή Μηχανολόγων Μηχανικών του ΕΜΠ ασχολήθηκε με την αεροδυναμική αριθμητική ανάλυση της φωνητικής οδού με βάση τις εξισώσεις Navier-Stokes. Στο δεύτερο ερευνητικό άξονα, που αντιστοιχεί στην παρούσα διατριβή, στόχος είναι η υπολογιστική μοντελοποίηση της φωνής με αξιοποίηση στοιχείων αεροδυναμικής. Το τρίτο ζητούμενο είναι η εφαρμογή των συμπερασμάτων της αεροδυναμικής φωνητικής ανάλυσης σε σύνθεση και αναγνώριση φωνής. Αυτό είναι το βασικό αντικείμενο της διδακτορικής διατριβής του Π. Τσιάκουλη. Το ερευνητικό πρόγραμμα, φέρνοντας κοντά ανθρώπους με διαφορετικό υπόβαθρο αλλά και από διαφορετική οπτική γωνία, αποτέλεσε ένα γόνιμο πλαίσιο συνεργασίας, ανταλλαγής και σύνθεσης ιδεών. Συνέβαλε σημαντικά στην ανάπτυξη και ωρίμανση της έρευνας που παρουσιάζεται στην παρούσα διατριβή.

Κεφάλαιο 2

Αεροδυναμική και Αεροακουστική για το Φωνητικό Σωλήνα

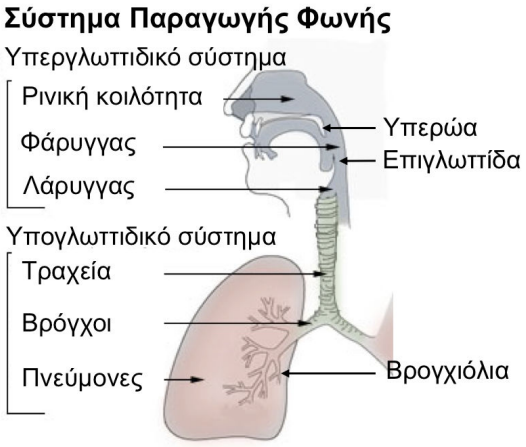
2.1 Εισαγωγή

Παρουσιάζονται σύντομα οι αρχές που διέπουν την επικρατούσα θεώρηση για την παραγωγή φωνής από τον ανθρώπινο φωνητικό σωλήνα. Η συζήτηση ξεκινάει με την παρουσίαση των βασικών χαρακτηριστικών των πρωτευουσών φυσιολογικών δομών που συμμετέχουν στη δημιουργία φωνής. Σκιαγραφείται η βασική αεροδυναμική μέσα στη φωνητική οδό και γίνεται προσπάθεια να δοθεί σε συντομία ο πυρήνας της αεροακουστικής θεωρίας που είναι σχετικός με την παραγωγή φωνής. Ανακεφαλαιώνονται οι παρατηρήσεις του Teager και παρουσιάζονται οι προβληματισμοί του [79, 164–168]. Συνοψίζονται τα πιο σύγχρονα αποτελέσματα που αφορούν στην αερορροή μέσα στο φωνητικό σωλήνα και έχουν προκύψει είτε πειραματικά με μηχανικά ανάλογα [19, 149], είτε μέσω αριθμητικής προσομοίωσης [161, 162, 180, 182, 183]. Η συζήτηση επικεντρώνεται κυρίως στις εργασίες των Krane, Howe, McGowan, Hirschberg που σχετίζουν την αερορροή με την παραγωγή ήχου μέσω αεροακουστικής [65, 70, 71, 90–92, 107, 109].

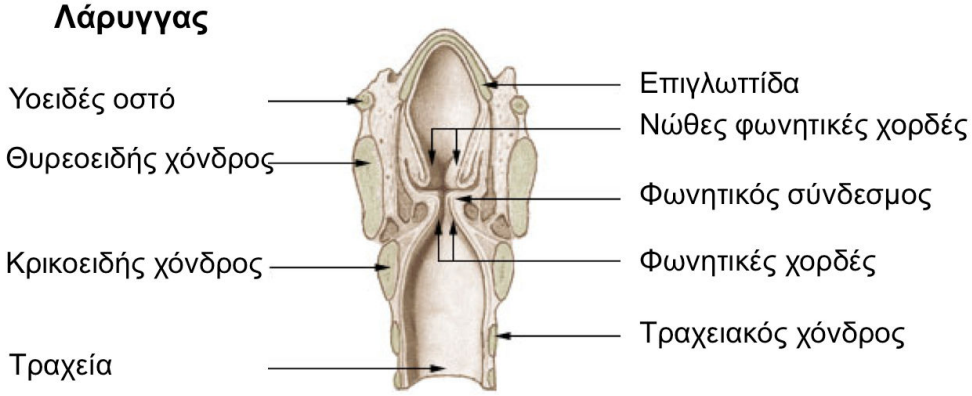
2.2 Ανατομία και φυσιολογία του φωνητικού σωλήνα

Για τη διευκόλυνση της μελέτης του μηχανισμού παραγωγής φωνής είναι συνήθης ο διαχωρισμός του εμπλεκόμενου ανθρώπινου σύστηματος σε τρία μέρη [156, Κεφ. 1], Σχήμα 2.1. Το υπογλωττιδικό σύστημα, ο λάρυγγας και οι δομές και οι αεραγωγοί πάνω από το λάρυγγα έχουν ιδιαίτερα φυσιολογικά χαρακτηριστικά και φαινομενικά διακριτούς ρόλους κατά τη φώνηση, τουλάχιστον σύμφωνα με την κλασσική θεώρηση πηγής - φίλτρου [52, Κεφ. 1], [55, Κεφ. III]. Σε τυπικές περιπτώσεις η φωνή εκλαμβάνεται ως το ακουστικό αποτέλεσμα κατάλληλης διαμόρφωσης της αερορροής που διέρχεται από το λάρυγγα και τις υπερλαρύγγειες δομές. Η απαιτούμενη ενέργεια θεωρείται ότι δίνεται στη ροή από το υπογλωττιδικό σύστημα.

Ως οι βασικότερες δομές στο υπογλωττιδικό σύστημα διακρίνονται οι πνεύμονες. Οι πνευμόνες εσωκλείουν ένα πολύπλοκο σύμπλεγμα κυψελίδων και μικρών διακλαδιζόμενων αεραγωγών, γνωστών και ως βρογχιολίων, όπου είναι δυνατή η αποθήκευση μέχρι και πέντε λίτρων αέρα [156, Ενότητα 1.1.1]. Αλλαγές στον όγκο των πνευμόνων, συνήθως με εύρος ίσο με μισό εώς ένα λίτρο για έναν ενήλικα κατά την παραγωγή φωνής, αντανακλώνται σε αύξηση ή μείωση της πίεσης του αποθηκευμένου αέρα και, αν είναι ανοιχτός ο αεραγωγός προς τα πάνω, σε δημιουργία αερορροής μέσω των βρόγχων (δύο μεγαλύτεροι αγωγοί όπου καταλήγουν τα βρογχιόλια) και της τραχείας (αγωγός - απόληξη των βρόγχων που καταλήγει στη γλωττίδα) προς ή από το λάρυγγα. Κατά τη διάρκεια μιας εκφώνησης που περιλαμβάνει ένα μικρό αριθμό λέξεων, η πίεση στους πνεύμονες διατηρείται μεταξύ 5 και 10 cm H_2O



Σχήμα 2.1: *Ανατομία του συστήματος παραγωγής φωνής*



Σχήμα 2.2: *Στεφανιαία όψη του λάρυγγα*

για κάποιον που μιλάει σε κανονικά επίπεδα εντάσης. Σημειώνεται πως ενδέχεται να εμφανίζονται κάποιες μικρές μεταβολές, εύρους το πολύ ίσου με 30% της μέσης τιμής, ανάλογα με την έμφαση που δίνεται σε συγκεκριμένες λέξεις [156, Ενότητα 1.1.1].

Στο λάρυγγα, εξέχουσα θέση κατέχουν οι λεγόμενες φωνητικές χορδές, δύο αντικριστά τμήματα υμένα σε σχήμα χορδής μήκους 1 με 1.5 cm και πλάτους 2 με 3 mm, Σχήμα 2.2. Η σχισμή που σχηματίζεται μεταξύ τους είναι γνωστή ως γλωττίδα. Ακριβώς από πάνω, παράλληλα με τις φωνητικές χορδές, διακρίνονται οι νώθες φωνητικές χορδές που διαμορφώνουν και το χωνοειδές σχήμα της εισόδου του λάρυγγα. Η επιγλωττίδα, λίγο ψηλότερα, είναι προσαρτημένη στο κάτω μέρος της γλώσσας και κατά την κατάποση, με κατάλληλη κίνηση της γλώσσας προς τα πίσω και ανύψωση του λάρυγγα, σκεπάζει το λάρυγγα για να τον προστατέψει από την ανεπιθύμητη είσοδο τροφής ή σάλιου. Υπό ορισμένες συνθήκες, σύνθετες αλληλεπιδράσεις της διερχόμενης ροής με τις λαρυγγικές δομές έχουν ως αποτέλεσμα την παραγωγή ακουστικής ενέργειας μέρος της οποίας διαδίδεται προς τις ανώτερες δομές του φωνητικού συστήματος.

Αμέσως πάνω από το λάρυγγα, ο αεραγωγός σχηματίζεται από το φάρυγγα και, όταν η υπερώα δεν είναι χαμηλωμένη, στρίβει περίπου 90 μοίρες για να συνεχιστεί με τη στοματική κοιλότητα, Σχήμα 2.3. Το μπροστινό πάνω μέρος της επιφάνειας της στοματικής κοιλότητας είναι γνωστό ως (σκληρός) ουρανίσκος που συμπληρώνεται προς τα πίσω από την υπερώα που έχει τη δυνατότητα με κατάλληλο χαμήλωμα ή ύψωση να ελέγχει το πέρασμα του αέρα από τη στοματική στη ρινική κοιλότητα. Το εμβαδό της εγκάρσιας επιφάνειας του περάσματος μεταξύ των δύο κοιλοτήτων μπορεί να είναι ίσο με 1 cm^2 όταν η υπερώα είναι τελείως

χαμηλωμένη, όπως κατά την αναπνοή, ενώ προσαρμόζεται σε τιμές ανάμεσα σε 0.2 cm^2 και 0.8 cm^2 κατά την παραγωγή ήχων με τη συμμετοχή της ρινικής κοιλότητας [156, Ενότητα 1.1.3.2]. Η γνώση της συνάρτησης εμβαδού, ή, αλλιώς, η γνώση των εμβαδών των επιφανειών των εγκαρσίων τομών σε διάφορα σημεία του φωνητικού συστήματος ανάμεσα στη γλωττίδα και τα χείλη είναι σημαντική για τον προσδιορισμό των ακουστικών και των αεροδυναμικών ιδιοτήτων του. Συνήθως, η επιφάνεια A της εγκάρσιας διατομής σε κάποιο σημείο μοντελοποιείται ως συνάρτηση της απόστασης d ανάμεσα στο εξωτερικό και το εσωτερικό τοίχωμα της οδού στο μέσο οβελιαίο επίπεδο [16]:

$$A = Kd^{\alpha}, \tag{2.1}$$

όπου K και α σταθερές που εξαρτώνται από το συγκεκριμένο ομιλητή και την απόσταση του εν λόγω σημείου από τη γλωττίδα. Το εξωτερικό τοίχωμα ορίζεται από το πίσω τοίχωμα του φάρυγγα, την υπερώα, τον ουρανίσκο, τον πάνω κόφτη και το πάνω χείλος ενώ το εσωτερικό από τη γλώσσα, τον κάτω κόφτη και το κάτω χείλος. Η κίνηση του κορμού της γλώσσας κατά την παραγωγή φωνής, που, ενώ μπορεί να είναι και ανεξάρτητη, σχετίζεται συνήθως σε μεγάλο βαθμό με την κίνηση του σαγονιού, προκαλεί σημαντικές διαφοροποιήσεις στη συνάρτηση επιφανείας από ήχο σε ήχο. Το ίδιο συμβαίνει και με στένωση ή διεύρυνση του φάρυγγα.

Ως φωνητικό σωλήνα ή φωνητική οδό θα θεωρήσουμε τον αεραγωγό που σχηματίζεται από το λάρυγγα και τις δομές πάνω από αυτόν. Στη συνέχεια η μελέτη επικεντρώνεται στις αεροδυναμικές και αεροακουστικές ιδιότητες της φωνητικής οδού.

2.3 Στοιχεία αεροδυναμικής θεωρίας

Οι αρχές διατήρησης της μάζας, της ορμής και της ενέργειας για τη ροή ενός ρευστού είναι ευρέως γνωστές. Παρ' όλ' αυτά, επειδή οι εξισώσεις είναι μη γραμμικές είναι γενικά αδύνατο να πάρουμε μια ακριβή λύση τους στη γενική περίπτωση. Σε αυτή την ενότητα, θα θεωρήσουμε κάποιες βασικές προσεγγίσεις που μπορούν να χρησιμοποιηθούν ώστε να κερδίσουμε διαίσθηση για τη συμπεριφορά της ροής στη φωνητική οδό.

2.3.1 Σωληνοειδές και αστρόβιλο πεδίο

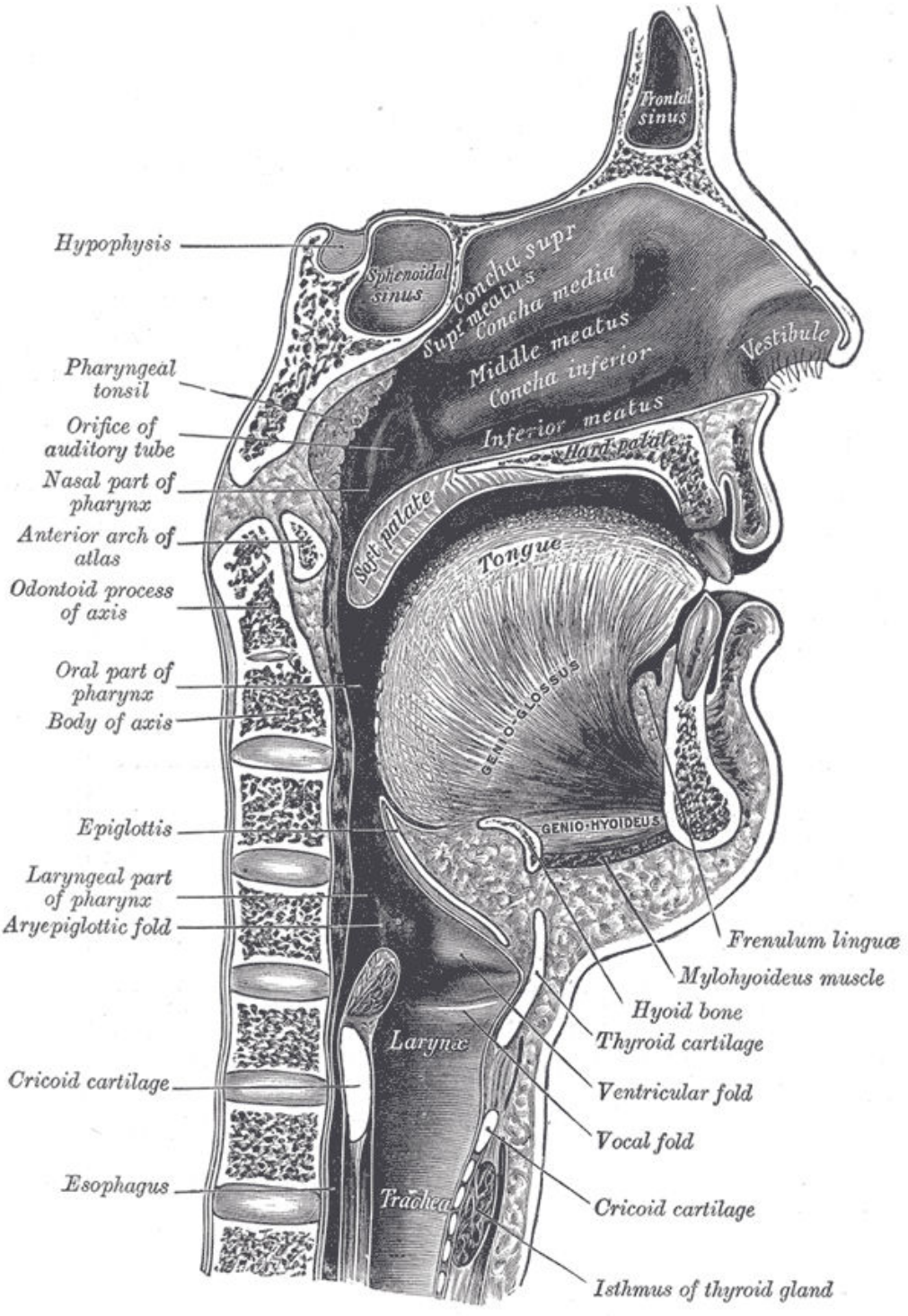
Γενικά, στη ρευστομηχανική χρησιμοποιείται το διανυσματικό πεδίο της ταχύτητας των σωματιδίων του ρευστού αντί για το βαθμωτό πεδίο της παροχής όγκου ή ογκικής ταχύτητας, όπως ισοδύναμα καλείται στην ακουστική. Αυτό σημαίνει ότι τα αεροδυναμικά μοντέλα γενικά έχουν σημαντικούς βαθμούς ελευθερίας που δεν έχουν τα μονοδιάστατα ακουστικά μοντέλα. Υπό μη περιοριστικές συνθήκες, ένα τρισδιάστατο διανυσματικό πεδίο μπορεί να διαχωριστεί σε δύο επιμέρους τρισδιάστατα διανυσματικά πεδία (θέωρημα Helmholtz) [7, σελ. 97-100]:

$$v = v_I + v_S \tag{2.2}$$

Το ένα πεδίο, v_S καλείται σωληνοειδές και το άλλο, v_I καλείται αστρόβιλο. Το διάνυσμα της ταχύτητας σε κάθε σημείο του χώρου γράφεται ως το άθροισμα των δύο επιμέρους διανυσμάτων. Το σωληνοειδές πεδίο μπορεί να υποστηρίξει περιστροφική κίνηση του ρευστού, αλλά όχι συμπίεση και διόγκωση. Το αστρόβιλο πεδίο υποστηρίζει συμπίεση και διόγκωση αλλά όχι περιστροφή. Γι' αυτό, το αστρόβιλο πεδίο είναι απαραίτητο για τη διάδοση ακουστικών διαταραχών. Οι ιδιότητες αυτές διατυπώνονται μαθηματικά με τις παρακάτω ταυτότητες. Η απόκλιση του σωληνοειδούς πεδίου είναι μηδενική όπως και η περιστροφή του αστρόβιλου πεδίου:

$$\nabla \cdot v_S = 0, \quad \nabla \times v_I = 0. \tag{2.3}$$

Η αναλυτική αυτή θεώρηση του πεδίου ταχύτητας είναι χρήσιμη για την περιγραφή της παραγωγής και της διάδοσης του ήχου καθώς και για την αλληλεπίδρασή του με άλλες



Σχήμα 2.3: *Μεσο-οβελιαία όψη της φωνητικής οδού [154]*

συνιστώσες της κίνησης του ρευστού. Ο διαχωρισμός βέβαια είναι στην ουσία μαθηματικό τέχνασμα και στην πράξη δεν είναι δυνατόν να μετρήσουμε ξεχωριστά τα επιμέρους πεδία, ακόμα και στο μακρινό ακουστικό πεδίο. Παρ' όλ' αυτά, μπορεί να δειχτεί ότι διαταραχές της πίεσης που σχετίζονται με αλλαγές της πίεσης στο σωληνοειδές πεδίο διαδίδονται ως ακουστικές διαταραχές στο αστρόβιλο πεδίο.

2.3.2 Δυναμική ροή

Όταν η ροή είναι αστρόβιλη μπορούμε να ορίσουμε ένα δυναμικό έτσι ώστε $v = \nabla\phi$. Σε μια τέτοια περίπτωση μπορούμε να γράψουμε τη συνάρτηση ορμής για ένα ρευστό χωρίς ιξώδες στην ολοκληρωτική της μορφή [20, σελ. 156-164]:

$$\partial\phi/\partial t + |v|^2/2 + I = g(t)$$

όπου I είναι η ειδική ενθαλπία που μπορεί να υπολογιστεί από την εξίσωση $I = \int dp/\rho$ και $g(t)$ είναι μια συνάρτηση του χρόνου που χωρίς απώλεια της γενικότητας μπορεί να συμπεριληφθεί στο δυναμικό (αφού δεν επηρεάζεται το πεδίο ταχύτητας). Η τελευταία εξίσωση είναι η εξίσωση Bernoulli για μια μη μόνιμη, συμπιεστή ισεντροπική, δυναμική (potential) ροή. Όταν θεωρήσουμε συμπαγή (με ομοιόμορφη πυκνότητα) ροή μπορούμε να χρησιμοποιήσουμε την προσέγγιση ασυμπιεστότητας :

$$\partial\phi/\partial t + |v|^2/2 + p/\rho_0 = g(t).$$

Γενικά μια προσέγγιση στην αεροδυναμική λαμβάνεται θεωρώντας την αδιάστατη μορφή των εξισώσεων κίνησης. Σε αυτή τη μορφή εμφανίζεται μπροστά από κάθε όρο ένα αδιάστατο νούμερο που είναι ένα μέτρο της σχετικής σημασίας του όρου αυτού. Υπό συγκεκριμένες συνθήκες κάποιοι μικροί όροι μπορούν να αμεληθούν.

2.3.3 Χαρακτηριστικοί αδιάστατοι αριθμοί

Κατά τη φώνηση οι πιο σημαντικές παράμετροι είναι ο αριθμός Strouhal St, ο αριθμός Reynolds Re, ο αριθμός Helmholtz He και ο αριθμός Mach M [65]. Ο αριθμός Strouhal $St = fD/\mathrm{v}_0$ είναι ένα μέτρο του λόγου της επιτάχυνσης λόγω μη μονιμότητας της ροής και μεταφορικής επιτάχυνσης λόγω ανομοιομορφίας της ροής. Ο αριθμός Helmholtz $He = D/\lambda$, όπου λ το χαρακτηριστικό μήκος κύματος, δίνει πληροφορία για τη συμπάγεια της ροής (ομοιομορφία της πυκνότητας). Ο αριθμός Reynolds $Re = D\mathrm{v}_0/\nu$ όπου ν είναι ο συντελεστής κινηματικής συνεκτικότητας είναι ένα μέτρο του λόγου μεταφορικών δυνάμεων προς τις ιξώδεις δυνάμεις ή δυνάμεις τριβής. Ο αριθμός Mach $M = \mathrm{v}_0/c$ δίνει πληροφορία για τη μεταβολή της πυκνότητας σε μια σταθερή ροή (για $M \ll 1, \Delta\rho/\rho = M^2/2$).

Η σημασία του αδιάστατου αριθμού εξαρτάται σε μεγάλο βαθμό από τη σωστή επιλογή της χαρακτηριστικής συχνότητας f, του χαρακτηριστικού μήκους D της εξεταζόμενης γεωμετρίας και της ταχύτητας v_0 του ρευστού. Γι' αυτό απαιτείται κάποια εμπειρική γνώση της ροής. Αυτή η διαίσθηση μπορεί να δοθεί με πειράματα σαν αυτά που περιγράφονται στα [19, 168]. Επιπλέον, διαφορετικές επιλογές μπορεί να είναι κατάλληλες για τη μελέτη διαφορετικών χαρακτηριστικών της ροής [65, 125].

Όταν ο αριθμός Reynolds είναι πολύ μικρός, $Re \ll 1$, επικρατούν οι δυνάμεις ιξώδους και η μη γραμμικότητα των εξισώσεων δεν είναι σημαντική. Στη φωνητική οδό έχουμε συνήθως αριθμούς Reynolds τάξης μεγέθους 10^3 με αποτέλεσμα η μη γραμμικότητα να είναι ένα σημαντικό χαρακτηριστικό της ροής. Σε πρώτη προσέγγιση όταν $Re \gg 1$ μπορούμε να αγνοήσουμε την τριβή στο κύριο σώμα της ροής. Όταν η ροή είναι αστρόβιλη αυτό δίνει μια δυναμική ροή που υπολογίζεται σχετικά εύκολα. Παρ' όλ' αυτά δεν μπορούμε ποτέ να αμελήσουμε την τριβή στα τοιχώματα.

2.3.4 Συνοριακό στρώμα και αποκόλληση ροής

Υπάρχει πάντα τουλάχιστον μια λεπτή περιοχή κατά μήκος των τοιχωμάτων όπου η τριβή είναι τόσο σημαντική όσο είναι οι δυνάμεις αδράνειας. Αυτή η περιοχή καλείται συνοριακό στρώμα. Συνήθως σε αυτή την περιοχή η πίεση προσδιορίζεται από τον κύριο κορμό της ροής όπου δεν υπάρχει τριβή. Επιπλέον, το συνοριακό στρώμα πάντα περιέχει περιστροφή γιατί πρόκειται για μια σχεδόν παράλληλη ροή στην οποία κυριαρχεί η συνιστώσα που είναι παράλληλη στο τοίχωμα αλλά αλλάζει από την ταχύτητα v_0 του κύριου κορμού της ροής σε μηδέν στο τοίχωμα. Στην ιδανική περίπτωση το συνοριακό στρώμα παραμένει λεπτό και η τριβή είναι υπεύθυνη μόνο για μια μικρή διόρθωση στην ιδανική χωρίς τριβή δυναμική ροή.

Ακόμα και σε μια περιορισμένη περιοχή με γρήγορα μεταβαλλόμενη γεωμετρία η προσέγγιση δυναμικής ροής (potential flow) δεν ισχύει συνήθως. Η πιο θεαματική απόκλιση από τη δυναμική ροή είναι λόγω της αποκόλλησης του συνοριακού στρώματος από τα τοιχώματα. Στο σημείο αποκόλλησης η στροβιλότητα που περιέχεται στο συνοριακό επίπεδο εισάγεται στην κύρια ροή. Αν θεωρήσουμε υψηλό αριθμού Reynolds, η στροβιλότητα παραμένει προσκολλημένη στο σωματίδιο του ρευστού. Η εξέλιξη της κατανομής της στροβιλότητας οδηγεί στο σχηματισμό μιας ελεύθερης φλέβας (στην περίπτωση μόνιμης ροής) ή σε περιοδική εκπομπή στροβίλων (περιοδική ροή).

Η αποκόλληση του συνοριακού στρώματος μπορεί να γίνει κατανοητή ποιοτικά όταν αρχίσουμε με τη μελέτη ενός σωματιδίου ρευστού στην κύρια ροή. Όπως δηλώνεται από την αρχή διατήρησης της ορμής, απούσας τριβής (στην κύρια ροή), η σωματιδιακή ταχύτητα καθορίζεται από ένα σημείο ισορροπίας μεταξύ της μεταφορικής δύναμης και του διαφορικού της πίεσης ∇p. Η χωρική παράγωγος της πίεσης κάθετη στα τοιχώματα εξαφανίζεται σε ένα συνοριακό στρώμα οπότε η πίεση προσδιορίζεται από την εξωτερική ροή, το χωρικό διαφορικό της πίεσης που είναι εφαπτομενικό στα τοιχώματα είναι το ίδιο όπως στο κυρίως σώμα της ροής όπου δεν υπάρχει τριβή. Στην εξωτερική ροή μεταφορικές δυνάμεις είναι σε ισορροπία με το διαφορικό της πίεσης. Καθώς η τριβή στο συνοριακό επίπεδο υποδηλώνει μια απώλεια κινητικής ενέργειας, η μεταφορική δύναμη στο συνοριακό στρώμα μπορεί να μην είναι πάντα αρκετά μεγάλη ώστε να εξισορροπήσει το διαφορικό της πίεσης. Όταν το αντίθετο διαφορικό πίεσης είναι πολύ μεγάλο όπως σε μια απότομη ακμή (δόντια) ή αν το αποκλίνον τμήμα ενός καναλιού είναι πολύ μεγάλο, το συνοριακό στρώμα θα αποκολληθεί.

2.4 Στοιχεία αεροακουστικής θεωρίας

Αντικείμενο της αεροακουστικής είναι η μελέτη της παραγωγής ήχου λόγω αυτεπιδράσεων της αερορροής ή αλληλεπιδράσεών της με επιφάνειες στερεών. Διαφορετικά, πρόκειται για τη μελέτη της ανταλλαγής ενέργειας μεταξύ της αερορροής και του ακουστικού πεδίου ή αλλιώς για τη μελέτη της παραγωγής αεροδυναμικού ήχου [109]. Είναι γενικά αποδεκτό ότι κάθε μηχανισμός παραγωγής ήχου, όπως ο ήχος από μουσικά όργανα, από δονούμενες επιφάνειες ή από συμβατικά ήχεια, μπορεί στην ουσία να προσεγγιστεί ως πρόβλημα αεροδυναμικού ήχου [69]. Η σχετική θεωρία θεμελιώθηκε από τον Lighthill [94]. Βασίζεται στην υπόθεση ότι είναι δυνατή η αποσύμπλεξη του λεπτομερούς υπολογισμού της αερορροής από τον προσδιορισμό του ακουστικού πεδίου. Υποτίθεται ότι ο παραγόμενος ήχος δεν μπορεί να επηρεάσει την αερορροή, δεν υπάρχει δηλαδή ανάδραση. Αυτή η προσέγγιση θεωρείται αποδεκτή για πολλές σημαντικές ροές με χαμηλό αριθμό Mach, όπως είναι για παράδειγμα η αερορροή μέσα στο φωνητικό σωλήνα. Στη συνέχεια, για λόγους πληρότητας, παρουσιάζεται σε συντομία η ακουστική αναλογία του Lighthill και κάποιες σχετικές προεκτάσεις ακολουθώντας κατά βάση την ανάλυση στα [69, 94, 109].

2.4.1 Αρχή διατήρησης ορμής και ακουστική εξίσωση

Η αρχή διατήρησης της ορμής για ένα ιξώδες ρευστό είναι γνωστή και ως εξίσωση Navier-Stokes και εκφράζει το ρυθμό μεταβολής της ορμής ενός στοιχειώδους τμήματος του ρευστού ως προς την πίεση p, τον τανυστή ιξώδους τάσης σ_{ij} και τις δυνάμεις σώματος F (όπως η βαρύτητα) ανά στοιχειώδη όγκο. Οι τελευταίες θεωρούνται αμελητέες και η Reynolds μορφή της εξίσωσης (σε συμβολισμό Einstein [1]) είναι:

$$\frac{d\rho u_i}{dt} = -\frac{\partial \pi_{ij}}{\partial x_j} \tag{2.4}$$

όπου

$$\pi_{ij} = \rho u_i u_j + (p - p_0)\delta_{ij} - \sigma_{ij} \tag{2.5}$$

είναι ο τανυστής ρευστότητας ορμής (momentum flux tensor) και σ_{ij} είναι ο τανυστής ιξώδους τάσης (viscous stress tensor) [69]. Σε ένα γραμμικό ακουστικό μέσο ο τανυστής π_{ij} περιλαμβάνει μόνο την πίεση και θεωρούμε ότι οι ηχητικές διαταραχές είναι αδιαβατικές ($p - p_0 = c_0^2(\rho - \rho_0)$, c_0 η ταχύτητα του ήχου) οπότε τελικά η εξίσωση της ορμής παίρνει τη μορφή:

$$\frac{\partial \rho u_i}{\partial t} + \frac{\partial}{\partial x_i}[c_0^2(\rho - \rho_0)] = 0 \tag{2.6}$$

και η γραμμική ακουστική εξίσωση ως προς τις διαταραχές της πίεσης τελικά προσδιορίζεται ως :

$$\left(\frac{1}{c_0^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)[c_0^2(\rho - \rho_0)] = 0. \tag{2.7}$$

2.4.2 Ακουστική αναλογία

Σύμφωνα με την ακουστική αναλογία του Lighthill, ο ήχος που παράγεται λόγω της τύρβης σε ένα πραγματικό ρευστό είναι ακριβώς ισοδύναμος με αυτόν σε ένα ιδανικό γραμμικό ακουστικό μέσο όταν αυτό διεγείρεται από την κατανομή τάσης :

$$T_{ij} = \pi_{ij} - \pi_{ij}^0 = \rho u_i u_j + \left((p - p_0) - c_0^2(\rho - \rho_0)\right)\delta_{ij} - \sigma_{ij} \tag{2.8}$$

όπου T_{ij} είναι ο Lighthill τανυστής τάσης. Πράγματι, με κατάλληλη αναδιατύπωση της εξίσωσης ορμής στη γενική περίπτωση μπορούμε τελικά να καταλήξουμε στην ακριβή μη γραμμική κυματική εξίσωση:

$$\left(\frac{1}{c_0^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)[c_0^2(\rho - \rho_0)] = -\frac{\partial^2 T_{ij}}{\partial x_i \partial x_j}. \tag{2.9}$$

Αν ο όρος στο δεξί σκέλος έχει φραγμένο πεδίο ορισμού, δηλαδή αν η τυρβώδης περιοχή της ροής είναι περιορισμένη, τότε μπορεί να θεωρηθεί ότι παρέχει τις απαραίτητες διαφορές τάσεων για τη διάδοση ήχου στον εναπομείνοντα χώρο, δρα δηλαδή ως πηγή. Το ρευστό στον εναπομείνοντα χώρο θεωρείται ιδανικό. Αυτή ακριβώς είναι η ακουστική αναλογία.

Με αφορμή διάφορα προβλήματα μηχανικής, έχουν γίνει προσπάθειες κατανόησης της ακουστικής αναλογίας, επαναδιατύπωσης και επίλυσης της Εξ. (2.9) σε διάφορες γεωμετρίες. Οι όροι που σχετίζονται με τις πηγές μπορούν να απλοποιηθούν αν αμεληθούν οι ιξώδεις τάσεις και αν θεωρηθεί ότι η πίεση και η πυκνότητα συνδέονται ισεντροπικά, που είναι μια πολύ καλή προσέγγιση στην περίπτωση ροών με χαμηλό αριθμό Mach. Σε αυτή την περίπτωση ο τανυστής του Lighthill γράφεται ως :

$$T_{ij} \approx \rho_0 u_i u_j. \tag{2.10}$$

[1] Δείκτης που εμφανίζεται δύο φορές σημαίνει άθροιση ως προς το δείκτη αυτό. Ο συμβολισμός εισήχθη από τον Einstein [46]

και κατ' επέκταση:

$$\frac{\partial^2 T_{ij}}{\partial x_i \partial x_j} \approx \rho_0 \mathrm{div}(\boldsymbol{\omega} \times \mathbf{v}) + \frac{\rho_0}{2}\nabla^2 |\mathbf{v}|^2, \tag{2.11}$$

όπου $\mathbf{v}$ είναι η ταχύτητα του ρευστού και έχει αμεληθεί ένας όρος που σχετίζεται με συμπιεστότητα [109]. Διαισθητικά περιμένει κανείς κατά την παραγωγή ήχου να είναι σημαντικότερος ο όρος που περιέχει τη στροβιλότητα $\boldsymbol{\omega}$. Η στροβιλότητα

$$\boldsymbol{\omega} = \boldsymbol{\nabla} \times \mathbf{v} \tag{2.12}$$

είναι ένα μέτρο της περιστροφής του ρευστού. Ο δεύτερος όρος στο δεξί μέλος της Εξ. (2.11) φαίνεται κάποιες περιπτώσεις να έχει μικρή σχέση με την παραγωγή ήχου [130]. Αυτό μπορεί να γίνει πιο συγκεκριμένο αν θεωρηθεί ότι το ακουστικό μέγεθος είναι η ακουστική ολική ενθαλπία αντί για την πυκνότητα ή την πίεση. Η κυματική εξίσωση για την ενθαλπία αποτελμάτωσης ισεντροπικού αερίου $B = \int dp/\rho + 1/2|\mathbf{v}|^2$ μπορεί τότε να διατυπωθεί ως :

$$\left(\frac{1}{c_0^2}\frac{D^2}{Dt^2} - \nabla^2\right) B = \mathrm{div}(\boldsymbol{\omega} \times \mathbf{v}) \tag{2.13}$$

όπου $D/Dt = \partial/\partial t + \mathbf{v} \cdot \nabla$ είναι ο τελεστής υλικής παραγώγισης. Με τον τρόπο αυτό, στο δεξί μέλος της εξίσωσης μένει μόνο ο όρος που έχει σχέση με τη στροβιλότητα.

2.4.3 Επίλυση με εφαρμογή της συνάρτησης Green

Για τη γενίκευση της λύσης σε πολλά προβλήματα, όπως για παράδειγμα για την παραγωγή τυρβωδών ήχων σε μια στένωση, είναι δυνατή η εφαρμογή της συνάρτησης Green. Η συνάρτηση αυτή ικανοποιεί την εξίσωση

$$\Box_{c_0} G(\mathbf{x}, t|\mathbf{y}, \tau) = \delta(\mathbf{x} - \mathbf{y})\delta(t - \tau) \tag{2.14}$$

όπου $\Box_{c_0} = \frac{\partial^2}{\partial t^2} - c_0^2\nabla^2$ είναι ο κυματικός τελεστής (D' Alembertian) στην Εξίσωση (2.13). Όποτε η συνάρτηση Green είναι η απόκριση του συστήματος ως συνάρτηση του χρόνου t όπως μετράται στη θέση $\mathbf{x}$, αν η πηγή είναι τοποθετημένη χωρικά και χρονικά στη θέση $\mathbf{y}$ και στιγμή τ αντίστοιχα.

Με κατάλληλη εφαρμογή της συνάρτησης Green στην περίπτωση της στένωσης σε έναν αεραγωγό απείρου μήκους με μόνιμη ροή, η λύση για τις διαταραχές της ενθαλπίας που σχετίζονται άμεσα με τις ακουστικές διαταραχές δίνεται από τη σχέση [91]:

$$B'(x,t) = \frac{c_0}{2A}\int \nabla \cdot (\omega \times \mathbf{v})(\mathbf{y}, \tau) H\left\{t - \tau - \frac{x}{c_0(1+M)} + \frac{\phi^*(\mathbf{y})}{c_0(1+M)}\right\} d^3\mathbf{y}d\tau \tag{2.15}$$

όπου A είναι το εμβαδό της εγκάρσιας διατομής του αγωγού εκτός της στένωσης, M είναι ο αντίστοιχος αριθμός Mach, H είναι η συνάρτηση Heaviside ενώ $\phi^*(\mathbf{y})$ είναι το δυναμικό ταχύτητας ενός μοναδιαίου μόνιμου, αστρόβιλου πεδίου ροής στον αγωγό. Στο μακρινό ακουστικό πεδίο, η λύση μπορεί να εκφραστεί και ως προς την ακουστική πίεση ως εξής [91]:

$$p(x,t) = -\rho_0 \frac{sgn(x-y)}{2A(1+M)}\int [\omega \times \mathbf{v} \cdot \nabla\phi^*] d^3\mathbf{y} \tag{2.16}$$

Η ολοκληρώσιμη ποσότητα είναι μέσα σε αγκύλες που σημαίνει ότι είναι συνάρτηση της θέσης y της πηγής και του καθυστερημένου χρόνου $t - |x-y|/(c_0(1+M))$ που χρειάζεται για να φτάσει το σήμα από την πηγή στη θέση του παρατηρητή. Ανάλογα μπορούν να διατυπωθούν και οι εξισώσεις για την περίπτωση αγωγού μεταβλητής διατομής [150]. Όπως αναπτύσσεται και στο [150] προκύπτει ότι αν μπορεί να περιγραφτει ο σχηματισμός και η εξέλιξη της στροβιλότητας για ένα ρεύμα αέρα και είναι γνωστό το σχήμα του αγωγού, τότε

μπορεί να θεωρηθεί ένα μοντέλο για τη γένεση ήχου από τη ροή λαμβάνοντας την ένταση της ακουστικής πηγής ανάλογη του

$$\boldsymbol{\omega} \times \mathbf{v} \cdot \hat{\mathbf{U}} \tag{2.17}$$

όπου $\hat{\mathbf{U}} = \nabla\phi^*$ είναι το μοναδιαίο διάνυσμα στη διεύθυνση της ταχύτητας του μέσου πεδίου ροής $\mathbf{U}$.

2.5 Αεροδυναμική μέσα στο φωνητικό σωλήνα

2.5.1 Ιστορικά

2.5.1.1 Μετρήσεις της αερορροής στη στοματική κοιλότητα

Στο [164] ήταν η πρώτη φορά που αναφέρθηκαν μετρήσεις της αερορροής μέσα στη φωνητική οδό. Πιο συγκεκριμένα, το πείραμα που περιγράφεται περιλαμβάνει τη χρήση μιας διάταξης που αποτελείται από τρία ανεμόμετρα για τη μέτρηση της ταχύτητας της ροής μέσα στο στόμα κατά την παραγωγή του φωνήματος /ι/ παρατεταμένα. Οι μετρήσεις ελήφθησαν περίπου 10 mm μακριά από την άκρη της γλώσσας και περίπου στην ίδια απόσταση από τον ουρανίσκο. Οι δύο αισθητήρες που βρίσκονταν ψηλότερα έδωσαν πρακτικά την ίδια μέτρηση ενώ η μέτρηση από τον αισθητήρα που βρισκόταν πιο κάτω ήταν αρκετά διαφορετική. Συγκρίνοντας τις κυματομορφές με αυτές της ακουστικής πίεσης εκτός του στόματος διαπιστώνουμε ότι είναι αρκετά διαφορετικές, εκτός φάσης και με διαφορετικούς συντονισμούς. Σύμφωνα με τον ερευνητή οι συγκεκριμένες μετρήσεις και άλλες για διαφορετικές γεωμετρίες που αναφέρει ότι έχει πραγματοποιήσει αποτελούν ένδειξη του ότι η ροή στο στόμα δεν είναι ομοιόμορφη και έχει διαχωριστεί από τα τοιχώματα. Ισχυρίζεται ότι οι παρατηρήσεις του θα μπορούσαν να προκύψουν με την υπόθεση ύπαρξης ροής που εναλλάσεται από τον ουρανίσκο στη γλώσσα με κάποια ορισμένη συχνότητα. Συμπεραίνει έτσι ότι έχουμε ενεργή παραγωγή ήχου στη στοματική κοιλότητα αφού η ύπαρξη ρυθμού μεταβολής της διανυσματικής ροής προκαλεί ήχο.

Στην προσπάθειά τους να ενθαρρύνουν επανεξέταση της θεωρίας παραγωγής φωνής οι Teager & Teager στο [166] εμφανίζουν παρόμοιες μετρήσεις για τον ήχο /ε/ μέσα στη φωνητική οδό και κάνουν ανάλογες διαπιστώσεις. Περιγράφουν με λεπτομέρεια την πειραματική τους μεθοδολογία και συζητούν ενδεχόμενα προβλήματα που μπορεί να παρουσιάζει και πώς προσπάθησαν να τα αντιμετωπίσουν. Επιπλέον, συζητούν τα προβλήματα της κλασσικής θεωρίας ακουστικής που βασίζεται στη δυναμική ροή και παρουσιάζουν διάφορα παράδοξα που προκύπτουν και δεν μπορούν να εξηγηθούν αν δεν εμπλουτιστεί περαιτέρω η αρχική θεώρηση. Για το διαχωρισμό της ροής που παρατηρούν στο στόμα επισημαίνουν ότι πολλές από τις μετρήσεις τους είναι ενδεικτικές της ύπαρξης στροβιλώδους ροής. Με βάση την αρχή λειτουργίας της σφυρίχτρας για την οποία εμφανίζονται κάποια πειραματικά αποτελέσματα μετρήσεων με ανεμόμετρα και όπου έχουμε ενεργή παραγωγή ήχου και δεν μπορεί να εφαρμοστεί η αρχή πηγής φίλτρου προτείνουν ότι θα πρέπει να αναπτυχθεί ένα ανάλογο μοντέλο αλληλεπίδρασης φλέβας αέρα-κοιλότητας για τη φωνητική οδό. Κάτι τέτοιο, όπως σημειώνουν θα μπορούσε να φέρει πιο κοντά τη θεωρία στα πειραματικά δεδομένα και στην καθημερινή εμπειρία. Για τις πηγές ήχου μέσα στη φωνητική οδό αναφέρεται ότι η γλωττίδα είναι μόνο μία από αυτές ενώ θα έπρεπε ενδεχόμενα να ληφθούν υπόψη μηχανισμοί παραγωγής ήχου όπως αυτοί στις σφυρίχτρες (απλές ή στροβιλώδεις) ή στους αιολιανούς τόνους.

2.5.1.2 Προσπάθειες για φυσική μοντελοποίηση

Στην ίδια κατεύθυνση, στο [167] μετά την παρουσίαση ενός συνόλου από διαπιστωμένες ατέλειες της κλασσικής γραμμικής θεωρίας για την παραγωγή φωνής γίνεται μια πρώτη

προσπάθεια να δοθούν οι βασικές αρχές ενός φαινομενολογικού μοντέλου της φωνητικής οδού για την παραγωγή έμφωνων ήχων. Ως σημαντικά προβλήματα του γραμμικού μοντέλου αναφέρονται μεταξύ άλλων η αδυναμία του να εξηγήσει ήχους που θα μπορούσε να βγάλει το ανθρώπινο φωνητικό σύστημα όπως κλικς, σφυρίγματα, χασμουρητά ή ροχαλητά αλλά και το γεγονός ότι υποβαθμίζει χρονο-συχνοτικά σημαντικά γεγονότα επιμένοντας στη στατική συχνοτική προσέγγιση. Περιγράφοντας τη φυσική της παραγωγής της φωνής παρουσιάζεται η έννοια του κύματος ορμής που είναι σχετική με το φαινόμενο του διαχωρισμού της ροής που σύμφωνα με τους ερευνητές είναι πρωτεύουσας σημασίας μέσα στη φωνητική οδό. Γίνεται αναφορά στη διάδοση αλλαγών στην ταχύτητα ενός ρευστού που σε ένα τζετ πραγματοποιείται με ταχύτητα κοντά στην ακουστική. Αυτή η διάδοση δε σχετίζεται με διάδοση παλμών πίεσης και ροής. Συμπεραίνεται ότι η διάδοση γίνεται με τη μορφή σολιτονίων (παλμών) [136] χωρίς αντίστοιχη μεταβολή στην πίεση. Το χαρακτηριστικό τους είναι ότι όταν συγκρούονται μεταξύ τους ή με στροβίλους η συμπεριφορά τους δεν είναι γραμμική. Όταν συμβαίνουν τέτοιες συγκρούσεις παράγονται μεταβολές στην πίεση. Ο αντίστοιχος ήχος μπορεί να έχει αρμονικές που δεν είναι εμφανείς στις αρχικές ροές (για δυο ροές με συχνότητες f_1, f_2 μετά τη σύγκρουση ο ήχος θα μπορούσε να περιέχει τις συχνότητες $2f_1, 2f_2, f_1 - f_2, f_1 + f_2$ και υψηλότερης τάξης αρμονικές για παράδειγμα. Παρά του ότι δίδεται μια τυπική ανάλυση του φαινομένου, είναι αρκετά εξιδανικευμένη και δεν είναι άμεσες οι συσχετίσεις με την παραγωγή φωνής.

Τα μοντέλα που προτείνονται στη συνέχεια για τα διάφορα μέρη του φωνητικού συστήματος είναι αρκετά αφηρημένα. Περιγράφονται κάποιες γενικές αρχές τους αλλά δεν προτείνεται κάποια συγκεκριμένη υπολογιστική προσέγγιση ώστε να είναι αξιοποιήσιμα. Για το υπογλωττιδικό σύστημα, έχει ενδιαφέρον η διαπίστωση ότι είναι συζευγμένο με το υπόλοιπο σύστημα και ότι συμμετέχει στην παραγωγή φωνής με ένα σύνθετο μηχανισμό διαχωρισμού ροής, στροβιλισμών και σχετικών φαινομένων και όχι απλά ως μια δεξαμενή πίεσης. Για τη γλωττίδα και το λάρυγγα τονίζεται η σημασία του διαχωρισμού της ροής αλλά και της τρισδιάστατης γεωμετρίας για τη διάδοση των οποιωνδήποτε διαταραχών. Για το φάρυγγα και το στόμα τονίζεται η σημασία της συμπεριφοράς των διαφόρων τζετ που μπορεί να δημιουργούνται και το πώς οι συγκρούσεις τους, οι παγιδεύσεις τους ή η εξασθένησή τους μπορεί να είναι σημαντική για τον παραγόμενο ήχο. Οι τέσσερεις ρευστοδυναμικοί μηχανισμοί που κατά τους ερευνητές πρέπει να διερευνηθούν περαιτέρω και είναι σημαντικοί για τη φώνηση είναι: (1) Αποκόλληση ροής, (2) Αξονικές και ακτινικές γρήγορα εναλλασσόμενες ροές μέσα σε κοιλότητες που σχηματίζονται είτε από τα τοιχώματα είτε από τοπικές ροές, (3) Μη γραμμική σύζευξη πιέσεων και ροών που έχει ως αποτέλεσμα ενεργή ανάδραση μεταξύ τζετ και στροβίλων στα διάφορα τμήματα της φωνητικής οδού και (4) μη γραμμική διαμόρφωση και γέννηση αρμονικών από την αλληλεπίδραση των κυμάτων ορμής. Η συμπεριφορά της ροής που παρατηρείται είναι χαρακτηριστική για κάθε ξεχωριστό ήχο.

Στο άρθρο τους [168], οι Teager & Teager επανεξετάζουν το θέμα της ροής μέσα στο φωνητικό σωλήνα, παρουσιάζουν μετρήσεις με ανεμόμετρα μέσα στο στόμα στο ίδιο πνεύμα και συζητούν πιθανές εξηγήσεις για τις παρατηρήσεις τους. Αναφέρονται και στις διαδικασίες ακουής και αντίληψης της φωνής ενώ τελικά προτείνουν μια μεθοδολογία ανάλυσης που θα μπορούσε να αναδείξει ενδιαφέρουσες πλευρές του συστήματος παραγωγής φωνής. Πιο συγκεκριμένα, ως εξήγηση για τη μη μονιμότητας του πεδίου ροής εμφανίζεται ο όρος μη γραμμικής μεταγωγής στις εξισώσεις της ακουστικής μέσα σε κινούμενο μέσο. Στην κλασσική ακουστική προσέγγιση μέσα στο φωνητικό σωλήνα ο όρος αυτός αμελείται, υποθέτοντας μηδενική μέση ροή. Ως πιθανούς μηχανισμούς ενεργής παραγωγής ήχου αναφέρουν (1) την αρχή λειτουργίας της σφυρίχτρας, που έχει να κάνει με ένα τζετ που διεγείρει εφαπτομενικά μια κοιλότητα, (2) την κίνηση ενός τζετ κατά μήκος του εσωτερικού τοιχώματος της κοιλότητας, (3) την ύπαρξη ενός τζετ με περιδίνιση μέσα σε μια κοιλότητα, (4) την ύπαρξη τζετ με ακτινικούς στροβίλους και (5) την Αιολιανή αστάθεια (παραγωγή ήχου σε ένα εμπόδιο σε μια κατά τα άλλα ομοιόμορφη ροή).

Η παρουσίαση του συστήματος παραγωγής φωνής ως ένα ενεργό σύστημα και όχι απλώς ως ένα παθητικό φίλτρο στην ουσία εμπνέει και διαφορετικό τρόπο θεώρησης για τις διαδικασίες αντίληψης της φωνής. Αμφισβητείται ότι το αυτί είναι ένας απλός αναλυτής τόνων. Όπως αναφέρεται χαρακτηριστικά δεν υπάρχουν απομονωμένοι ακουστικοί τόνοι στη φύση και αυτοί σχετίζονται μόνο με ανθρώπινες κατασκευές. Η μελέτη των φυσικών ήχων επιβάλλει να δοθεί βαρύτητα σε μεταβατικά φαινόμενα και στην ενδεχόμενη περιοδικότητά τους. Η όλη συζήτηση είναι πάνω στις βασικές αρχές της φασματικής ανάλυσης της φωνής. Στην ουσία υποστηρίζεται ότι θα πρέπει να αντικατασταθεί αυτή με ανάλυση ζωνοπερατών διαμορφώσεων. Όπως αναφέρεται θα πρέπει να είναι δυνατή η ανίχνευση στο σήμα το κατά πόσον αυτό ήταν προϊόν παθητικής ή ενεργητικής παραγωγής. Υποστηρίζεται ότι αυτό θα μπορούσε να είναι δυνατό αν μπορούσαμε να διακρίνουμε χαρακτηριστικά της ενέργειας της πηγής που παρήγαγε το σήμα. Η ενέργεια αυτή σχετίζεται με το γινόμενο του τετραγώνου του πλάτους επί το τετράγωνο της συχνότητας. Η προτεινόμενη μεθοδολογία αποτέλεσε πηγή έμπνευσης για μετέπειτα αξιόλογες ερευνητικές προσπάθειες ανάλυσης φωνής με χρήση του λεγόμενου ενεργειακού τελεστή Teager-Kaiser και διαδικασιών αποδιαμόρφωσης πλάτους και συχνότητας με καινοτομικούς αλγορίθμους με μεγάλη χρονική ευκρίνεια [42, 81, 104, 105, 128].

Στην ιδιαίτερα κατατοπιστική παρουσίαση του Kaiser [79] που ακολούθησε τις πρώτες δημοσιεύσεις του Teager πρακτικά παρουσιάζεται εμπεριστατωμένα και σε σχέση με τις επικρατούσες τότε αντιλήψεις η πρόταση για μια καινούρια θεωρητική και υπολογιστική προσέγγιση για τη μελέτη της παραγωγής της φωνής. Ως βασικό κίνητρο για την όλη προσπάθεια δίνεται πρακτικά η συχνά διαπιστούμενη απόσταση μεταξύ θεωρίας και πράξης όσον αφορά στη φωνή. Οι ερευνητές της φωνής βρίσκονται μακριά από τους ανθρώπους που έχουν πρακτική γνώση σχετική με τη φωνή όπως είναι θεραπευτές, ηθοποιοί, τραγουδιστές και είναι κοινή πεποίθηση ότι τα συμβατικά μοντέλα για τη φωνή δεν μπορούν εύκολα να αποδειχτούν ουσιαστικά χρήσιμα γιατί υπάρχει μεγάλη απόκλιση των προβλέψεων τους από τις πραγματικές παρατηρήσεις. Το μοντέλο πηγής φίλτρου που επικρατεί εμφανίζεται κατά κάποιο τρόπο ως ένας συμβιβασμός. Οι βασικές υποθέσεις πάνω στις οποίες στηρίζεται περιλαμβάνουν ότι το ρευστό είναι αστρόβιλο και ισοτροπικό. Η πίεση υποτίθεται ότι συνδέεται με την ογκική ταχύτητα μέσω της ακουστικής αντίστασης. Τελικά, συνήθως η δυσκολία στον προσδιορισμό μιας κατάλληλης μοντελοποίησης για έναν συγκεκριμένο ήχο ανάγεται στον κατάλληλο προσδιορισμό μιας κατάλληλης πηγής που θα διεγείρει το γραμμικό σύστημα.

Στον ίδιο άξονα με τον Teager, στο [79] παρουσιάζεται μια σειρά παρατηρήσεων που αφορούν στη φωνή και φαινομενικά δεν μπορούν να εξηγηθούν από την κλασσική θεώρηση. Εκτός των άλλων, ενδιαφέρον έχει η παρατήρηση ότι το πουλί μάινα μπορεί να παράγει φωνή που δύσκολα μπορεί να διακριθεί από ανθρώπινη ενώ ο φωνητικός του μηχανισμός φαίνεται να είναι αρκετά διαφορετικός από τον ανθρώπινο. Το ίδιο ισχύει και για τον εγκέφαλο ή το ακουστικό του σύστημα. Για να διερευνηθούν αυτά τα ζητήματα προτείνεται η μελέτη της ρευστοδυναμικής μέσα στο φωνητικό σωλήνα και η ανάπτυξη μοντέλων που θα λαμβάνουν υπόψη τους τα χαρακτηριστικά της ροής. Σημαντικό ρόλο εμφανίζεται να έχει το φαινόμενο διαχωρισμού της ροής κατά το σχηματισμό ενός τζετ αλλά και η εμφάνιση, διάδοση στροβίλων καθώς και η αλληλεπίδραση μεταξύ τους αλλά και με τα τοιχώματα. Τέλος, γίνεται και κάποια νύξη για αντικατάσταση της φασματικής ανάλυσης της φωνής με κάποια ανάλυση που θα μπορούσε καλύτερα να αναδείξει τα μεταβατικά φαινόμενα στη φωνή που φαίνεται να είναι και τα πιο σημαντικά. Συζητείται η κατανομή Wiener αλλά στην ουσία προτείνεται γενικότερα ότι η χρονοσυχνοτική ανάλυση μπορεί να έχει πολλά πλεονεκτήματα [33].

2.5.1.3 Μετά τους Teager και Kaiser

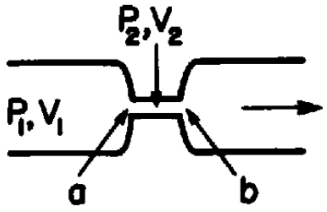
Δυναμικά μηχανικά μοντέλα με αρχικά απλοποιημένες γεωμετρίες προσφέρουν τη δυνατότητα να μελετηθούν μη μόνιμες κινήσεις του ρευστού διατηρώντας την ελεγξιμότητα και την επαναληψιμότητα των πειραμάτων. Στο [19] αναφέρονται πειράματα με ένα δυναμικό μηχανικό μοντέλο που προσομοιώνει την κίνηση των φωνητικών χορδών και την ίδια τη φωνητική

οδό. Στόχος των πειραμάτων είναι η σύγκριση των μετρήσεων των ακουστικών διαταραχών που προκαλούνται στο μακρινό πεδίο από την περιοδικό ανοιγοκλείσιμο του μοντέλου των χορδών με τις αντίστοιχες προβλέψεις από τη σχετική θεωρία διάδοσης επίπεδων ακουστικών κυμάτων μέσα στη φωνητική οδό. Αρχικά χρησιμοποιήθηκε απλά η θεωρία του ακουστικού πεδίου σε κινούμενο μέσο όπως έχει συζητηθεί στο [38]. Οι προβλεπόμενες ακουστικές διαταραχές σε απόσταση 60 cm από την άκρη του μηχανικού μοντέλου ήταν σημαντικά ασύμφωνες με τις αντίστοιχες μετρήσεις (περίπου κατά 10 dB). Στη συνέχεια, τα αποτελέσματα βελτιώθηκαν σημαντικά όταν ενσωματώθηκε στη θεώρηση κι ένα αεροακουστικό μοντέλο. Η ιδέα που περιγράφεται είναι πρακτικά ότι στην έξοδο του μοντέλου της γλωττίδας δημιουργούνται δίνες που διαδίδονται μέσα στο σωλήνα με ταχύτητα συγκρίσιμη με τη ταχύτητα της μέσης ροής του σωλήνα. Οι στρόβιλοι αυτοί είναι υπεύθυνοι για την παραγωγή ήχου όταν συναντήσουν μια σημαντική αλλαγή στη γεωμετρία του σωλήνα. Η πρώτη τέτοια σημαντική αλλαγή είναι το τελείωμα του σωλήνα, οπότε και παράγεται ήχος στο σημείο αυτό σύμφωνα με την προσέγγιση του Howe. Οι βελτιωμένες προβλέψεις του ακουστικού πεδίου συμφώνησαν αρκετά καλύτερα με το μετρούμενο πεδίο. Οι ενδεχόμενες διαφορές αποδόθηκαν στο απλοποιημένο μοντέλο για τους στροβίλους που απαιτούσε αξονική συμμετρία για παράδειγμα. Το πείραμα φαίνεται ότι επιβεβαιώνει τις διαπιστώσεις του [65] για τη σημασία της αεροακουστικής στα χείλια. Έχει ενδιαφέρον πάντως το γεγονός ότι εμφανίζονται οι στρόβιλοι να επιβιώνουν για τόσο μεγάλη απόσταση ενώ σύμφωνα με τον [91] θα έπρεπε να έχουν εξασθενήσει.

Στο [149] γίνεται προσπάθεια να αξιολογηθούν τα αποτελέσματα των πειραμάτων του δυναμικού μηχανικού μοντέλου και να εκτιμηθεί κατά πόσο μπορούν να χρησιμοποιηθούν για να εξαχθούν συμπεράσματα για την ανθρώπινη παραγωγή φωνής. Για το σκοπό αυτό μετρήθηκαν οι πιέσεις πριν και μετά το μοντέλο της γλωττίδας, η σωματιδιακή ταχύτητα του αέρα λίγο πριν το άκρο του μοντέλου και της φωνητικής οδού και η ακουστική πίεση στο μακρινό πεδίο. Οι μετρήσεις έγιναν από τη μία στο δυναμικό μηχανικό μοντέλο και από την άλλη για τέσσερεις άντρες που προέφεραν ένα ουδέτερο φωνήεν, ώστε η γεωμετρία της φωνητικής τους οδού να είναι όσο το δυνατόν πλησιέστερη στη γεωμετρία του μοντέλου. Οι μετρήσεις της πίεσης πριν και μετά το μοντέλο της γλωττίδας σε σύγκριση με πραγματικές μετρήσεις φαίνεται να διαφέρουν σημαντικά ως προς το πλάτος αλλά όχι τόσο ως προς τα γενικά χαρακτηριστικά της κυματομορφής. Η διαφορά στο πλάτος οφείλεται σύμφωνα με τους ερευνητές στο ότι η επιφάνεια διατομής της ανθρώπινης γλωττίδας εξελίσσεται χρονικά περισσότερο ως πριονωτή κυματομορφή και όχι ως ημιτονοειδής, όπως η επιφάνεια διατομής του μοντέλου. Αυτό έχει ως αποτέλεσμα η παράγωγος της κυματομορφής που σχετίζεται με τις μετρήσεις πίεσης να είναι μικρότερη για το μοντέλο. Επιπλέον υπάρχει διαφορά και στη διάρκεια του κύκλου που το μοντέλο της γλωττίδας παραμένει κλειστό, με αποτέλεσμα να είναι περιορισμένη η μέγιστη πίεση που παρατηρείται. Οι κυματομορφές της ταχύτητας είναι αρκετά παρόμοιες, γεγονός που ενισχύει την υπόθεση ότι και στη φωνητική οδό έχουμε τη διάδοση στροβίλων. Το ακουστικό μακρινό πεδίο ήταν 10-20 dB υψηλότερα για τις μετρήσεις σε ανθρώπους. Αυτό θα μπορούσε να εξηγηθεί από το γεγονός ότι το μη ακουστικό πεδίο εξαρτάται γενικά έντονα από τη γεωμετρία οπότε αναμένεται να είναι πολύ διαφορετικό τελικά για τη φωνητική οδό και να συνεισφέρει με διαφορετική βαρύτητα στο μακρινό ακουστικό πεδίο.

2.5.2 Περιγραφή του πεδίου ροής

Κατά την παρουσίαση της βασικής αεροδυναμικής στο φωνητικό σωλήνα από τον Stevens [156], το ενδιαφέρον επικεντρώνεται στη μελέτη της πίεσης και της ροής του αέρα που παρατηρούνται για διάφορες καταστάσεις του σωλήνα κατά την παραγωγή φωνής. Συγκεκριμένα, γίνεται προσπάθεια σύνδεσης της διαφοράς πίεσης μεταξύ δύο σημείων μέσα σε έναν αεραγωγό και της αντίστοιχης αερορροής για απλοποιημένες γεωμετρίες αγωγών που προσομοιώνουν το γενικό σχήμα του φωνητικού σωλήνα κατά την παραγωγή ήχων όπως είναι



Σχήμα 2.4: *Στένωση σε αεραγωγό*

τα φωνήεντα ή τα τυρβώδη σύμφωνα. Η ροή θεωρείται μόνιμη και ασυμπίεστη (Αυτό ισχύει προσεγγιστικά. Η διάδοση ήχου απαιτεί συμπιεστό ρευστό.). Η ύπαρξη στενώσεων θεωρείται το σημείο κλειδί και αυτό γιατί σε έναν αγωγό σταθερής διατομής, η μεταβολή της πίεσης κατά μήκος του, ακόμα και στην περίπτωση τυρβώδους ροής, είναι σχετικά αμελητέα.

Σε έναν αγωγό μεταβλητής διατομής, αμελώντας αρχικά απώλειες, ισχύει ο νόμος Bernoulli, δηλαδή

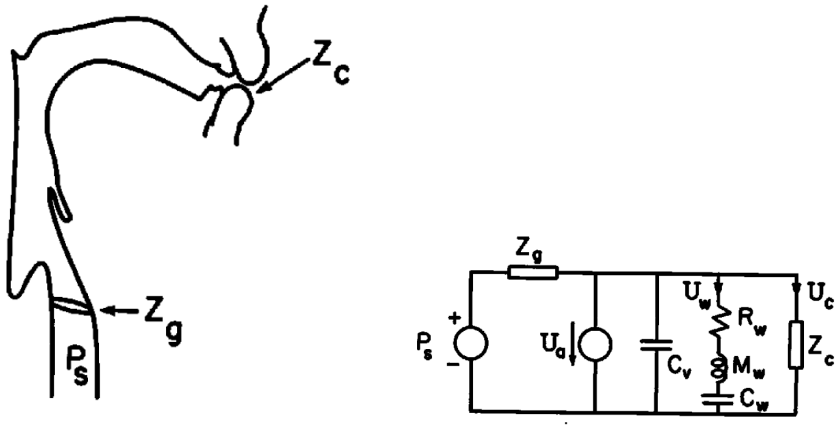
$$p + \frac{\rho \mathrm{v}^2}{2} = \text{const.} \tag{2.18}$$

όπου ρ είναι η πυκνότητα του αέρα και v είναι η ταχύτητα του ρευστού, ενώ $U = A\mathrm{v}$ είναι η παροχή όγκου μέσα στον αγωγό εμβαδού διατομής A. Οι διαφορές στην πίεση p πριν και μετά τη στένωση λόγω του νόμου Bernoulli αμελούνται γιατί θεωρούμε την κινητική ενέργεια πριν και μετά σχετικά αμελητέα (σταθερά μεγάλο εμβαδό διατομής A εκτός της στένωσης). Όπως έχει προκύψει από πειραματικές μετρήσεις [172], ένα φαινόμενο που παίζει σημαντικό ρόλο στον καθορισμό των απωλειών, όπως αυτές εκφράζονται μέσω της αντίστασης που ορίζεται ως το πηλίκο της διαφοράς πίεσης προς την αεροροή $R = \Delta P/U$, είναι η δημιουργία στροβίλων στα άκρα της στένωσης [156, Ενότητα 1.2.1.2]. Οι στρόβιλοι μπορούν να θεωρηθούν ως στοιχεία του αέρα που περιστρέφονται. Οι απώλειες που εμφανίζονται οφείλονται στην θερμική ανάλωση της ενέργειας των στροβίλων εξαιτίας του ιξώδους του αέρα. Η πτώση πίεσης λόγω των απωλειών αυτών εκφράζεται τελικά κατ' αναλογία της πίεσης Bernoulli [10, 103] και δίνεται ως

$$\Delta P = k_L \underbrace{\frac{1}{2}\rho \frac{U^2}{A^2}}_{\text{Bernoulli}} . \tag{2.19}$$

Έχει θεωρηθεί ότι η ταχύτητα του ρευστού, $\mathrm{v} = AU$ μέσα στη στένωση εμβαδού διατομής A, είναι αρκετά μεγαλύτερη σε σύγκριση με την ταχύτητα εκτός. Η σταθερά k_L γενικά εξαρτάται από τη γεωμετρία της στένωσης και κατά μέσο όρο μπορεί να θεωρηθεί, όπως προκύπτει από σχετικές πειραματικές μετρήσεις, ότι $k_L \approx 1$ [156], $k_L = 0.875$ [172] ή $k_L = 1.1$ [4].

Ως ένα απλό μοντέλο ροής για την περίπτωση της φωνητικής οδού που φαίνεται στα αριστερά του Σχήματος 2.5, θα μπορούσε να θεωρηθεί το ισοδύναμο ηλεκτρικό κύκλωμα δεξιά. Οι πιέσεις μπορούν γενικά να θεωρηθούν ισοδύναμες τάσεων ενώ οι παροχές όγκου ισοδύναμες ρευμάτων. Ως P_s συμβολίζεται η υπογλωττιδική πίεση που λαμβάνεται ίση με την πίεση στους πνεύμονες. Η πηγή ροής U_a μοντελοποιεί τη ροή που δημιουργείται με την ενεργή μεταβολή του όγκου της φωνητικής οδού μεταξύ των στενώσεων στη γλωττίδα και στα χείλια. Οι σύνθετες αντιστάσεις Z_g και Z_c βασικά εκφράζουν τις απώλειες που εμφανίζονται αντίστοιχα στις στενώσεις αυτές και στις οποίες έγινε σύντομη αναφορά προηγουμένως. Στη γενική περίπτωση, εκφράζουν την πτώση πίεσης μέσα στη στένωση λόγω μόνιμης ροής, λόγω τυρβώδους ροής (απώλειες λόγω ιξώδους) και τις απώλειες λόγω των στροβιλισμών στα άκρα (θερμικές). Με τη χωρητικότητα C_v μοντελοποιείται η ελαστικότητα του αέρα ενώ τα στοιχεία R_w, M_w και C_w αντιπροσωπεύουν τη σύνθετη αντίσταση των τοιχωμάτων του φωνητικού σωλήνα. Ένα τέτοιο στατικό μοντέλο ροής θεωρείται ότι ισχύει σε χαμηλές συχνότητες, μέχρι 200Hz [103, 156].



(α') φωνητική οδός (β') ισοδύναμο κύκλωμα ροής

Σχήμα 2.5: *Ισοδύναμο απλό μοντέλο ροής για φωνητική οδό με δύο στενώσεις, μία στα χείλη και μία στη γλωττίδα [156]*

Η πτώση πίεσης στη γλωττίδα υπό αυτές τις συνθήκες μπορεί τελικά να εκφραστεί ως

$$\Delta P_g = R_g U + \frac{d}{dt}(M_g U), \tag{2.20}$$

όπου M_g είναι η ακουστική μάζα του αέρα μέσα στη γλωττίδα. Το σχήμα της γλωττίδας θεωρείται ορθογώνιο παραλληλεπίπεδο με πάχος h και εγκάρσιες διαστάσεις, μήκος και πλάτος, l και d αντίστοιχα. Για ένα τμήμα σωλήνα μήκους l και διατομής A η ακουστική μάζα είναι $M_g = \rho l/A$ κι εκφράζει την αδράνεια του αέρα (η μάζα ενός ταλαντωτή στον οποίο εφαρμόζεται κάποια δύναμη για να τον θέσει σε κίνηση). Από την άλλη, για την αντίσταση R_g έχουμε

$$R_g U_g = \frac{12\mu h}{l d^3} + k_g \frac{\rho U^2}{2(ld)^2}. \tag{2.21}$$

Ο πρώτος όρος εκφράζει απώλειες λόγω ιξώδους, με δείκτη ιξώδους μ, για στρωτή ροή. Ο δεύτερος όρος σχετίζεται με τη δημιουργία στροβίλων όπως περιγράφτηκε προηγουμένως, Εξίσωση (2.19), και είναι γενικά ο επικρατών εκτός της περίπτωσης πολύ μικρών γλωττιδικών διατομών.

Ανάλογα, για την πτώση πίεσης στη στένωση στα χείλια (ή οπουδήποτε αλλού εκτός της γλωττίδας), έχουμε

$$\Delta P_c = R_c U + \frac{d}{dt}(M_c U). \tag{2.22}$$

Συνήθως η στένωση μοντελοποιείται ως ένας κύλινδρος μήκους l_c και εγκάρσιας διατομής A_c με ακουστική μάζα $M_c = \rho l_c/A_c$ Η αντίσταση R_{cv} που εκφράζει τις απώλειες λόγω του ιξώδους, παίρνει τη μορφή $R_{cv} = 8\pi\mu l_c/A_c^2$ (νόμος Hagen-Poiseuille) [10, 103]. Επιπλέον αυτών, όπως και στη γλωττίδα, υπάρχουν και οι απώλειες που σχετίζονται με τους στροβίλους στα άκρα. Τέλος, θα μπορούσε να συμπεριληφθεί και η επιπλέον πτώση πίεσης P_{c_tur} [10] που οφείλεται στην παραγωγή ήχου με αεροακουστικούς μηχανισμούς στους οποίους θα γίνει αναφορά αργότερα. Στο πλαίσιο μελέτης της γενικότερης αεροδυναμικής μια τέτοια πτώση πίεσης θεωρείται μάλλον αμελητέα. Γενικότερα, φαινόμενα αλληλεπίδρασης με το ακουστικό πεδίο δε συμπεριλαμβάνονται στη μέχρι τώρα ανάλυση.

Η ροή του άερα μέσα στη φωνητική οδό περιγράφεται σε γενικές γραμμές στο Σχήμα 2.7. Αυτή η γεωμετρία έχει τα γενικά χαρακτηριστικά όλων των ήχων της φωνής, ανεξαρτήτως της πηγής τους [91]. Αρχικά, οι πνεύμονες στα αριστέρα ωθούν τον αέρα μέσα στην οδό. Μετά την επιτάχυνση της ροής κατά την εισαγωγή μέσα στην αριστερή στένωση, ακολουθεί επιβράδυνσή και αποκόλληση της από τα τοιχώματα ώστε να σχηματιστεί εκτοξευόμενη φλέβα.

Αυτή η φλέβα περιλαμβάνει περιοχή εστιασμένης μεγάλης ορμής περικυκλωμένη από περιοχή αποτελματωμένου αέρα. Οι δύο αυτές περιοχές οριοθετούνται από διατμητικό στρώμα όπου τα σωματίδια του αέρα εκτελούν όχι μόνο μεταφορική αλλά και περιστροφική κίνηση. Η στροβιλότητα, ως ένα μέτρο της περιστροφικής αυτής κίνησης, τείνει να συγκεντρώνεται σε συνεκτικές δομές οι οποίες μπορεί να είναι τυρβώδεις και οι οποίες μετακινούνται από αριστερά προς τα δεξιά. Το διατμητικό στρώμα διαχέεται υπό την επίδραση δυνάμεων τριβής (δυνάμεων συνεκτικότητας), όπως αυτές ενισχύονται και λόγω της ανάμειξης των περιστρεφόμενων συμπαγών δομών. Έτσι, η φλέβα απλώνεται στην εγκάρσια διάσταση και τελικά η ορμή της εξαπλώνεται σε ολόκληρη τη διατομή της φωνητικής οδού. Η εμφάνιση της στροβιλότητας είναι συνέπεια της αποκόλλησης του συνοριακού στρώματος της ροής στην έξοδο της γλωττίδας. Η αποκόλληση εισάγει στον κύριο όγκο της ροής την περιστροφική κίνηση που δημιουργείται στο λεπτό ιξώδες στρώμα αέρα κοντά στα τοιχώματα. Σημειώνεται ότι μέχρι το σημείο αποκόλλησης της ροής, η στροβιλότητα είναι περιορισμένη στα συνοριακά στρώματα. Αν η φλέβα συναντήσει κάποια αλλαγή στη γεωμετρία της φωνητικής οδού, όπως για παράδειγμα μια στένωση, τότε οι στροβιλώδεις δομές θα παράξουν ασταθείς δυνάμεις στα τοιχώματα του εμποδίου καθώς περνούν από αυτό. Αυτές οι δυνάμεις είναι υπεύθυνες για τη μεταβίβαση ενέργειας στο ακουστικό πεδίο [91].

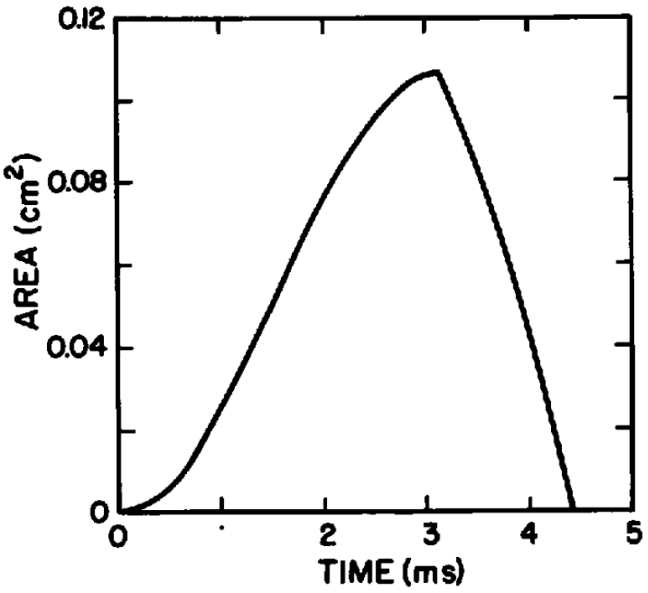
Οι ήχοι φωνής μπορεί να θεωρηθούν ότι παράγονται μέσω τριών μηχανισμών που είναι: (1) μετακίνηση όγκου αέρα εξαιτίας της κίνησης της φωνητικής οδού, (2) ασταθείς δυνάμεις σε ένα εμπόδιο, μια στένωση ή γενικότερα σε μια μεταβολή της γεωμετρίας του σωλήνα, που προκαλούνται από την ασταθή κίνηση των δομών στροβιλότητας της φλέβας του αέρα και (3) άμεση εκπομπή ήχου εξαιτίας της ασταθούς κίνησης των δομών της φλέβας [91].

2.5.3 Αεροδυναμική στη γλωττίδα

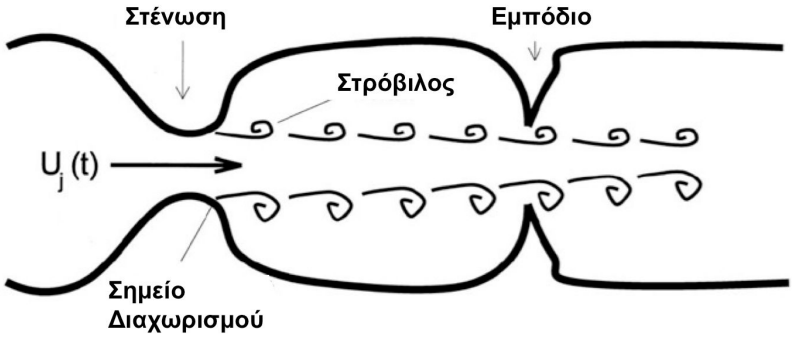
Η ακριβής κίνηση των φωνητικών χορδών μπορεί να παρατηρηθεί με χρήση γλωττογραφίας, που γίνεται δυνατή με διάφορα μέσα, όπως κάμερα υψηλής ταχύτητας, οπτική ίνα ή με τη χρήση διόδου εκπομπής φωτός [66]. Είναι συνέπεια πολύπλοκων αλληλεπιδράσεων των ελαστικών υμένων στη γλωττίδα και της αεροδυναμικής κατάστασης στο φωνητικό σωλήνα. Για πιο λεπτομερή ανάλυση, βλέπε [73, 77, 78, 156, 173]. Ο μηχανισμός που επιτρέπει την παλλόμενη κίνηση των χορδών σχετίζεται με το γεγονός ότι το πάνω και το κάτω μέρος τους μπορούν να κινηθούν το ένα σε σχέση με το άλλο. Χρησιμοποιώντας το μηχανικό μοντέλο των δύο μαζών, [73, 125] μπορεί κάποιος να προσομοιώσει το μηχανισμό αυτό και να προσδιορίσει το εμβαδό διατομής της γλωττίδας ως συνάρτηση του χρόνου για σταθερή υπογλωττιδική πίεση, Σχήμα 2.6.

Δεδομένου αυτού του εμβαδού μπορεί να υπολογιστεί η σύνθετη αντίσταση στη γλωττίδα με χρήση των Εξισώσεων (2.21) και (2.22). Με βάση το ισοδύναμο κύκλωμα του Σχήματος 2.5 ή παραλλαγές του, ανάλογα με το αν η σύνθετη αντίσταση της στένωσης πάνω από τη γλωττίδα θεωρείται σημαντική ή όχι, υπολογίζεται και η ροή μέσα στο φωνητικό σωλήνα. Στην περίπτωση που η σύνθετη αντίσταση της γλωττίδας είναι σχετικά μεγάλη, όταν δηλαδή δεν υπάρχει άλλη σημαντική στένωση και η γλωττιδική διατομή δεν έχει το μέγιστο εμβαδόν, μπορεί να θεωρηθεί ότι η διέγερση στο ισοδύναμο κύκλωμα είναι μια πηγή ογκικής ταχύτητας στη γλωττίδα [156].

Η κλασσική αυτή θεώρηση που βασίζεται μόνο στον υπολογισμό της ογκικής ταχύτητας που διαπερνάει τη γλωττίδα είναι ανεπαρκής σύμφωνα με τον McGowan [71, 109] κι αυτό γιατί παραβλέπει σημαντικά αεροδυναμικά φαινόμενα που λαμβάνουν χώρα κατά τη φώνηση. Για τη μελέτη των σχετικών φαινομένων χρησιμοποιήθηκε μια εξιδανικευμένη γεωμετρία του φωνητικού σωλήνα που όμως έχει το απαραίτητο χαρακτηριστικό της απότομης μεταβολής εγκάρσιας επιφάνειας αμέσως μετά τη γλωττίδα και επιπλέον έχει πεπερασμένο μήκος. Στο σημείο όπου υπάρχει η απότομη μεταβολή της επιφάνειας, αναμένεται η εμφάνιση περιστροφικής ροής αέρα η οποία δεν είναι ακουστική, με την έννοια ότι δεν σχετίζεται απαραίτητα με συμπιεστότητα και το σχετικό πεδίο ταχύτητας δεν υπακούει στην ακουστική



Σχήμα 2.6: *Εμβαδό της επιφάνειας διατομής της γλωττίδας με χρήση του μοντέλου δύο μαζών.*



Σχήμα 2.7: *Σκίτσο της κίνησης του αέρα κατά την παραγωγή φωνής [91]. Ο αέρας που βγαίνει από τους πνεύμονες από τα αριστερά προς τα δεξιά περνάει μέσα από τη στένωση στα δεξιά με ταχύτητα $U_j(t)$. Το ρεύμα του αέρα διαχωρίζεται σε κάποιο σημείο μετά το σημείο μέγιστης στένωσης, σχηματίζοντας μια φλέβα. Στο διατμητικό στρώμα της φλέβας μεταφέρονται δίνες που πριν το διαχωρισμό ήταν περιορισμένες στο οριακό στρώμα της ροής. Αλληλεπίδραση των δινών με τα τοιχώματα προκαλεί ακουστικές διαταραχές.*

κυματική εξίσωση.

2.6 Αεροακουστική στη φωνητική οδό

Η παραγωγή ήχου μέσω της ροής στη φωνητική οδό είναι γνωστό ότι αποτελεί τον πρωταρχικό μηχανισμό για την εκφορά άφωνων ήχων αλλά και ένα δευτερεύοντα μηχανισμό για φώνηση. Από τη σκοπιά της αεροακουστικής ιδιαίτερα σημαντική είναι η εμφάνιση στροβιλότητας. Η στροβιλότητα είναι ουσιαστικά η ποσότητα της ροής που είναι απαραίτητη για την κατανόηση όχι μόνο της δυναμικής της τυρβώδους κίνησης της αερορροής αλλά και της παραγωγής ήχου μέσω της αερορροής. Ακόμα και μελέτες της παραγωγής φωνής που χρησιμοποιούν πιο λεπτομερή μοντέλα ροής [1], δεν έχουν επικεντρωθεί στο ρόλο της στροβιλότητας και στη σημασία της για την αεροακουστική θεωρία. Ο McGowan ήταν ο πρώτος που ενσωμάτωσε αρχές της αεροακουστικής στη μελέτη της παραγωγής φωνής αλλά περιόρισε την αναζήτησή του μόνο σε έμφωνους ήχους. Ενώ οι επόμενες, πιο προσιτές συνεισφορές [65, 97, 125] κατάφεραν να αξιοποιήσουν πολλές σχετικές ιδέες, ασχολήθηκαν και αυτές κυρίως με έμφωνους ήχους. Ο Davies [39] ασχολήθηκε με την επίδραση της μετακίνησης του αέρα μέσα

στην φωνητική οδό στη διάδοση του ήχου εφαρμόζοντας ακουστική για κινούμενο μέσο στη φωνή, αλλά δεν διερεύνησε άμεσα την παραγωγή φωνής από την αερορροή. Πιο πρόσφατα, στην [183] αναφέρθηκε εφαρμογή του αεροακουστικού φορμαλισμού των Ffowcs-Williams και Hawkings στην παραγωγή έμφωνων ήχων άλλα δεν έγινε καμία προσπάθεια για άφωνους ήχους.

Για την παραγωγή άφωνων ήχων έχει δημιουγηθεί μια ποιοτική εικόνα της φυσικής μέσα στα χρόνια. Από την αρχή (π.χ., [52]) είχε γίνει εμφανής η ανάγκη ύπαρξης τυρβώδους αερορροής αλλά εκτός από την επισήμανση του τυχαίου και ευρυζωνικού συχνοτικού χαρακτήρα της πηγής δεν δίνονταν παραπάνω λεπτομέρειες για το σχετικό μηχανισμό. Ο Stevens [156] αξιοποίησε πολλές ιδέες από την αεροακουστική θεωρία, και πιο συγκεκριμένα τη μορφή του ακουστικού φάσματος από μία τυρβώδ φλέβα ροής και την ιδέα ότι η αερορροή παράγει το θόρυβο πιο αποδοτικά όταν παρευρίσκεται ένα εμπόδιο, οπότε ο ήχος παράγεται όχι στο σημείο όπου σχηματίζεται η τυρβώδης ροή αλλά εκεί όπου αλληλεπιδρά με ένα εμπόδιο όπως είναι τα δόντια. Με άλλα λόγια, το κύριο μέρος του εκπεμπόμενου ήχου δεν πρόερχεται από την τυρβώδη φλέβα αέρα άμεσα αλλά από την αλληλεπίδρασή της με το περιβάλλον. Η ιδέα αυτή επαληθεύτηκε σε μια σειρά από πειράματα από τη Shadle [147] χωρίς όμως να εξηγείται περαιτέρω ο μηχανισμός παραγωγής ήχου στην περίπτωση μη ύπαρξης διακριτού εμποδίου τριγύρω. Αργότερα προτάθηκε ότι ο ήχος που παράγεται από την αερορροή μέσα στη φωνητική οδό εξαρτάται από το τρισδιάστατες λεπτομέρειες της γεωμετρίας της φωνητικής οδού, οπότε μια απλή αξονική κατανομή διατομών μπορεί να μην είναι αρκετή για το χαρακτηρισμό της φωνητικής οδού κατά την παραγωγή άφωνων ήχων.

Η δυναμική της αερορροής είναι εκ των πραγμάτων ιδιαίτερα σύνθετη και για αυτό γεννάται η ανάγκη για διάφορες προσεγγίσεις ώστε το πρόβλημα να γίνει περισσότερο προσιτό. Η αεροακουστική θεωρία παρέχει τα μέσα για την εισαγωγή λογικών προσεγγίσεων παρέχοντας τυπικές διατυπώσεις στις οποίες η μη ακουστική κίνηση εμφανίζεται ως ακουστική πηγή. Η σχετική διατύπωση στην ουσία παρέχει ένα φιλτράρισμα πληροφορίας [91]. Δεν είναι ανάγκη η ακριβής αναπαράσταση της αερορροής και η μορφή του όρου που αντιστοιχεί στην πηγή παρέχει στην ουσία καθοδήγηση σχετικά με το πώς μπορούν να εφαρμοστούν ικανοποιητικές προσεγγίσεις.

Κεφάλαιο 3

Σύνθεση Φωνής με Αριθμητική Προσομοίωση

3.1 Εισαγωγή

Μελετώνται οι εξισώσεις που διέπουν τη διάδοση του ήχου μέσα στη φωνητική οδό σύμφωνα με τη γραμμική ακουστική. Συγκεκριμένα, προσδιορίζονται οι εξισώσεις όπως αυτές προκύπτουν με κατάλληλη εφαρμογή των αρχών διατήρησης της μάζας και της ορμής μέσα στη φωνητική οδό. Στη συνέχεια διακριτοποιούνται κατάλληλα τόσο στη συχνότητα όσο και στο χρόνο. Η αριθμητική επίλυση πραγματοποιείται αρχικά με δύο διαφορετικά σχήματα, του Portnoff [127] και του Maeda [101] που στην ουσία διέπονται από τη φιλοσοφία που διέπει πολλούς από τους συνθέτες φωνής που βασίζονται σε αριθμητική προσομοίωση [24, 43]. Μελετάται η αποδοτικότητα των δύο σχημάτων και η ακρίβειά τους για την προσομοίωση του ακουστικού πεδίου μέσα σε ένα σωλήνα ομοιόμορφης διατομής. Υιοθετείται το βασικό σχήμα του Maeda για τη συνέχεια με κάποιες τροποποιήσεις που αφορούν στη μοντελοποίηση των δονούμενων τοιχωμάτων της φωνητικής οδού και που τελικά προσδίδουν περαιτέρω ευστάθεια.

Προσομοιώνεται η παραγωγή φωνηέντων χρησιμοποιώντας ενδεικτικές γεωμετρίες όπως δίνονται από τον Fant [52] αλλά και γεωμετρίες που έχουν προκύψει από δεδομένα μαγνητικής τομογραφίας και συνοδεύονται από μετρήσεις του παρατηρούμενου ακουστικού φάσματος [159, 160]. Υπολογίζονται οι αντίστοιχες αποκρίσεις συχνότητας, εντοπίζονται οι συχνότητες συντονισμού και συγκρίνονται με αυτές που έχουν μετρηθεί από τα πραγματικά σήματα φωνής. Προκειμένου να γίνει δυνατή η ενσωμάτωση και ακουστικών κοιλοτήτων όπως είναι η ρινική κοιλότητα και οι αχλαδόσχημες κοιλότητες (piriform fossa) αναπτύσσεται μια βελτιωμένη εκδοχή του αριθμητικού σχήματος ακολουθώντας σε γενικές γραμμές το [114]. Για τη γλωττίδα εφαρμόζεται το μοντέλο των δύο μαζών που προτείνεται στο [73] και το συνολικό σύστημα προσομοίωσης τροποποιείται κατάλληλα ώστε να είναι δυνατή η ζεύξη της γλωττίδας με τη φωνητική οδό. Παρουσιάζεται τέλος αποτέλεσμα σύνθεσης ακολουθίας φωνηέντων στο χρόνο με βάση δεδομένα γεωμετρίας της φωνητικής οδού όπως έχουν καταγραφεί με τη χρήση ακτίνων Χ. Δίνονται λεπτομέρειες για τον υπολογισμό των αντίστοιχων συναρτήσεων εμβαδού από τα μεσο-οβελιαία σχήματα της φωνητικής οδού και περιγράφονται οι μεταβλητές ελέγχου του συστήματος προσομοίωσης. Αναφέρονται και άλλες προσπάθειες για σύνθεση φωνής με προσομοίωση όπως στα [95, 146, 151].

3.2 Διάδοση ήχου στη φωνητική οδό

Ακολουθώντας σε γενικές γραμμές την ανάλυση των [126, 127, 132] εξάγονται οι εξισώσεις κίνησης για ακουστική κυματική διάδοση μέσα στο φωνητικό σωλήνα. Η συνήθης πρακτική για την εξαγωγή των αντίστοιχων εξισώσεων περιλαμβάνει τη γραμμικοποίηση των εξισώσεων

που εκφράζουν την αρχή διατήρησης της μάζας και την αρχή διατήρησης της ορμής καθώς και τη σχέση μεταξύ πίεσης και πυκνότητας γραμμικής ακουστικής [126]. Η φωνητική οδός μοντελοποιείται ως ένας ανομοιόμορφος χρονομεταβλητός σωλήνας με ελαστικά τοιχώματα και η κυματική κίνηση θεωρείται ως διάδοση επίπεδου κύματος με διαταραχές που εισάγονται από τα ελαστικά τοιχώματα. Το σχεδόν μονοδιάστατο μοντέλο θεωρείται ότι είναι ακριβές όσο το μήκος κύματος του ήχου είναι μεγάλο σε σύγκριση με τις εγκάρσιες διαστάσεις της φωνητικής οδού. Η ορθότητα ενός τέτοιου ισχυρισμού είναι οριακή στις υψηλότερες ακουστικές συχνότητες, μεγαλύτερες από 5 kHz. Προσπάθειες για προσομοίωση του τρισδιάστατου ακουστικού πεδίου για την παραγωγή φωνηέντων περιγράφονται στα [47, 116, 117, 163]. Στο [163] χρησιμοποιούνται πεπερασμένες διαφορές, ενώ στο [47] εφαρμόζεται η μέθοδος του πίνακα γραμμών μεταφοράς. Μέθοδοι πεπερασμένων στοιχείων [116] και τεχνικές προσαρμογής υψηλότερων ακουστικών ρυθμών έχουν επίσης εφαρμοστεί [117]. Συστηματική αξιολόγηση και σύγκριση της τρισδιάστατης και της μονοδιάστατης προσέγγισης δεν έχει ακόμα πραγματοποιηθεί.

3.2.1 Αρχές διατήρησης μάζας και ορμής

Ένα επαρκές σύνολο ποσοτήτων για την περιγραφή της ακουστικής κυματικής διάδοσης σε ένα ρευστό όπως είναι ο άερας είναι η πυκνότητα ρ, η πίεση p, η θερμοκρασία T, η εντροπία και η τοπική ή σωματιδιακή ταχύτητα $\mathbf{v}$ του ρευστού. Αυτές οι ποσότητες είναι αλληλεξαρτημένες και συνδέονται με εξισώσεις κατάστασης. Δεδομένου του ότι χρειάζονται πολλές προσεγγίσεις για να μοντελοποιηθεί η φωνητική οδός ως ένα πρακτικά μονοδιάστατο σύστημα, οι εξισώσεις αρχικά εξάγονται στη γενική τους μορφή ως ένα σύστημα μη γραμμικών μερικών διαφορικών εξισώσεων σε τρεις χωρικές διαστάσεις και μετά περιορίζονται σε ένα μονοδιάστατο σύστημα με τρόπο ώστε κάθε προσέγγιση και η λογική της να δηλώνονται ξεκάθαρα.

Διατυπώνεται η αρχή διατήρησης της συνολικής μάζας του ρευστού σε ολοκληρωτική μορφή. Θεωρείται μια παραμορφώσιμη μακροσκοπική περιοχή ρευστού V που περικλείεται από την κλειστή επιφάνεια Σ. Αναπαριστάνουμε ένα στοιχειώδη όγκο σε αυτή την περιοχή με dV και μια στοιχειώδη επιφάνεια του συνόρου με $d\mathbf{\Sigma} = \hat{\mathbf{n}} d\Sigma$, που ορίζεται από το μοναδιαίο κάθετο διάνυσμα $\hat{\mathbf{n}}$ με φορά προς τα έξω. Αν δεν υπάρχουν πηγές ή καταβόθρες στην περιοχή V, ρυθμός μεταβολής της συνολικής μάζας που εσωκλείεται από την επιφάνεια ισούται με τον καθαρό ρυθμό εισόδου ροής στην περιοχή :

$$\frac{d}{dt} \iiint_V \rho dV = - \iint_\Sigma \rho \mathbf{v} \cdot d\mathbf{\Sigma}. \tag{3.1}$$

Αυτή είναι η εξίσωση συνέχειας σε ολοκληρωτική μορφή για τη μάζα του ρευστού.

Χρησιμοποιώντας το δεύτερο νόμο του Νεύτωνα μπορούμε να γράψουμε την αρχή διατήρησης της ορμής του ρευστού. Θεωρείται ένα στοιχείο του ρευστού με μάζα $dm = \rho dV$, όπου με dV αναπαρίσταται ένας στοιχειώδης όγκος. Αυτό το στοιχείο μπορεί να μετασχηματίζεται με το χρόνο αλλά η μάζα του παραμένει σταθερή. Αν του εφαρμοστεί η δύναμη $\mathbf{f} = -\boldsymbol{\nabla} p dV$ και η ταχύτητά του είναι $\mathbf{v}$ τότε η ορμή του είναι $\rho \mathbf{v} dV$ και ο δεύτερος νόμος του Νέυτωνα δίνει

$$\frac{D}{Dt}(\rho \mathbf{v} dV) = -\boldsymbol{\nabla} p dV \tag{3.2}$$

όπου ο τελεστής $\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla$ υποδηλώνει τη μεταφορική παράγωγο και δείχνει ότι η ποσότητα που παραγωγίζεται μετράται στο ακίνητο σύστημα συντεταγμένων του ρευστού. Αφού η μάζα dm εξ ορισμού παραμένει σταθερή στο χρόνο,

$$\frac{D}{Dt}(\rho dV) = 0$$

η Εξ. (3.2) γίνεται

$$\rho \frac{D\mathbf{v}}{Dt} dV = -\boldsymbol{\nabla} p dV$$

και γι΄ αυτό

$$\rho \frac{D\mathbf{v}}{Dt} = -\boldsymbol{\nabla} p$$

ή, στο ακίνητο σύστημα συντεταγμένων :

$$-\boldsymbol{\nabla} p = \rho \frac{\partial \mathbf{v}}{\partial t} + \rho \mathbf{v} \cdot \boldsymbol{\nabla} \cdot \mathbf{v}. \tag{3.3}$$

3.2.2 Γραμμική ακουστική προσέγγιση

Οι εξισώσεις συνέχειας μάζας και ορμής είναι μη γραμμικές εξισώσεις για τις ποσότητες $p, \mathbf{v}$ και ρ. Οι ακουστικές διαταραχές μπορούν να θεωρηθούν ως μικρές διαταραχές πλάτους γύρω από μια καθολική κατάσταση ισορροπίας. Για ένα ρευστό, η καθολική αυτή κατάσταση περιγράφεται από τις τιμές $(p_0, \rho_0, \mathbf{v}_0)$ της πίεσης, πυκνότητας και ταχύτητας του ρευστού αντίστοιχα κατά την απουσία διαταραχών [126]. Τα καθολικά αυτά μεγέθη ικανοποιούν τις εξισώσεις του πεδίου (3.1) και (3.1) στην ισορροπία αλλά όταν υπάρχουν διαταραχές ισχύει :

$$p = p_0 + p' \ , \ \rho = \rho_0 + \rho', \ \mathbf{v} = \mathbf{v_0} + \mathbf{v}' \tag{3.4}$$

Δεδομένων των υποθέσεων γραμμικού ακουστικού πεδίου είναι δυνατή η γραμμικοποίηση γύρω από τη θέση ισορροπίας. Έτσι, χρησιμοποιώντας τις σχέσεις (3.4) και αμελώντας όρους δεύτερης ή μεγαλύτερης τάξης ως προς τις διαταραχές, η αρχή διατήρησης της ορμής (3.3) γράφεται :

$$\rho_0 \left(\frac{\partial \mathbf{v}'}{\partial t} + \mathbf{v}_0 \boldsymbol{\nabla} \cdot \mathbf{v}' + \big(\boldsymbol{\nabla} \cdot \mathbf{v}_0\big)\mathbf{v}' + \left[\frac{\partial \mathbf{v}_0}{\partial t} + \mathbf{v}_0 \boldsymbol{\nabla} \cdot \mathbf{v}_0 \right] \frac{\rho'}{\rho_0} \right) = -\boldsymbol{\nabla} p' \tag{3.5}$$

Η συμπίεση ενός ρευστού λόγω του περάσματος ενός ακουστικού κύματος είναι περίπου αδιαβατική, δηλαδή η εντροπία του αέρα μένει σχεδόν σταθερή κατά τη συμπίεση [126, σελ. 11-13]. Γι΄ αυτό μπορούμε να γράψουμε

$$\rho' = \rho_0 \kappa_s p'$$

όπου κ_s η αδιαβατική συμπιεστότητα του ρευστού. Κατά συνέπεια, η αρχή διατήρηση της μάζας (3.1) γράφεται :

$$\frac{\partial}{\partial t} \iiint_V \rho_0 (1 + \kappa_s p') dV + \iint_\Sigma \rho_0 [\mathbf{v_0} + \mathbf{v}' + \kappa_s p' \mathbf{v_0}] d\boldsymbol{\Sigma} = 0 \tag{3.6}$$

και λαμβάνοντας υπόψη ότι :

$$\frac{\partial}{\partial t} \iiint_V \rho_0 dV + \iint_\Sigma \rho_0 \mathbf{v_0} d\boldsymbol{\Sigma} = 0 \tag{3.7}$$

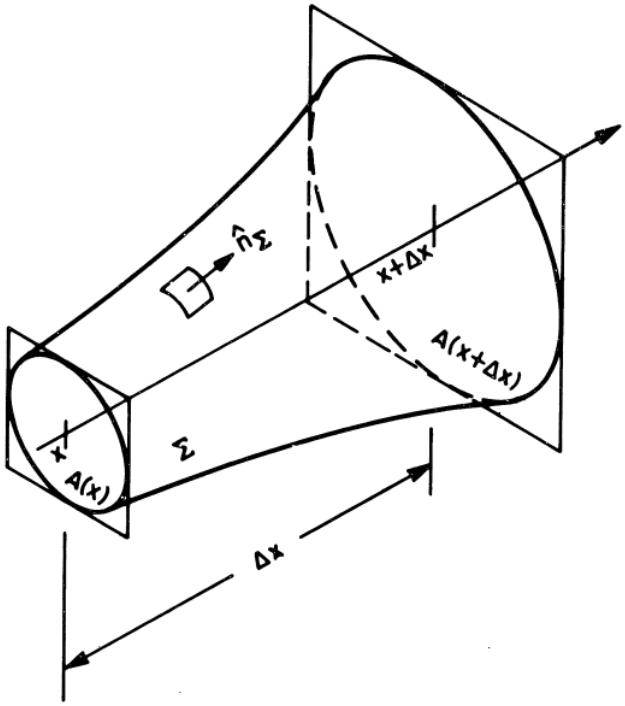
η Εξ. (3.6) απλοποιείται ως :

$$\frac{\partial}{\partial t} \iiint_V \kappa_s p' dV + \iint_\Sigma [\mathbf{v}' + \kappa_s p' \mathbf{v_0}] d\boldsymbol{\Sigma} = 0 \tag{3.8}$$

Στο [127, σελ.25, Εξ.(2.7)] παρ΄ όλ΄ αυτά γίνεται δεκτό ότι:

$$\frac{\partial}{\partial t} \iiint_V (1 + \kappa_s p') dV + \iint_\Sigma [\mathbf{v}' + \kappa_s p' \mathbf{v_0}] d\boldsymbol{\Sigma} = 0 \tag{3.9}$$

που είναι και η μορφή της εξίσωσης που έχει επικρατήσει.



Σχήμα 3.1: *Στοιχειώδης όγκος ελέγχου στη φωνητική οδό για τη μελέτη της ροής του άερα [127]*

3.2.3 Εφαρμογή σε σωλήνα χρονομεταβλητής διατομής

Οι γραμμικοποιημένες εξισώσεις (3.5) και (3.9) εφαρμόζονται στη συνέχεια σε έναν ανομοιόμορφο χρονομεταβλητό σωλήνα με ελαστικά τοιχώματα. Για λόγους απλότητας, παραλείπονται οι τονισμοί των συμβόλων όπου δεν υπάρχει κίνδυνος αμφισημίας. Θεωρείται το στοιχειώδες μήκος Δx του σωλήνα που φαίνεται στο Σχήμα 3.1. Η επιφάνεια που αντιστοιχεί στο τοίχωμα του σωλήνα μεταξύ των παράλληλων επιπέδων x και Δx σημειώνεται με Σ και η επιφάνεια της εγκάρσιας διατομής στο x σημειώνεται ως $A(x)$. Παρά του ότι η επιφάνεια μπορεί να είναι μεταβαλλόμενη χρονικά, εδώ σημειώνεται μόνο η χωρική της εξάρτηση. Εφαρμόζοντας την αρχή διατήρησης της μάζας έχουμε

$$\iint\limits_{A(x+\Delta x)} [\mathrm{v}_x + \kappa_s p \mathrm{v}_{0x}] dA - \iint\limits_{A(x)} [\mathrm{v}_x + \kappa_s p \mathrm{v}_{0x}] dA + \iint_{\Sigma} \rho v \cdot \hat{\mathbf{n}}_{\Sigma} d\boldsymbol{\Sigma} = \frac{d}{dt} \int\limits_{x}^{x+\Delta x} dx \iint\limits_{A(x)} (1 + \kappa_s p) dA \quad (3.10)$$

Αφού δεν περνάει μάζα από τα τοιχώματα το τελευταίο ολοκλήρωμα στο αριστερό σκέλος εξαφανίζεται. Διαιρώντας και τα δύο μέλη της εξίσωσης με Δx και παίρνοντας το όριο καθώς $\Delta_x \to 0$ παίρνουμε:

$$-\lim_{\Delta x \to 0} \frac{1}{\Delta_x} \left\{ \iint\limits_{A(x+\Delta x)} [\mathrm{v}_x + \kappa_s p \mathrm{v}_{0x}] dA - \iint\limits_{A(x)} [\mathrm{v}_x + \kappa_s p \mathrm{v}_{0x}] dA \right\} = \lim_{\Delta_x \to 0} \frac{1}{\Delta_x} \frac{d}{dt} \int\limits_{x}^{x+\Delta x} dx \iint\limits_{A(x)} (1 + \kappa_s p) dA. \quad (3.11)$$

Εφαρμόζοντας τον ορισμό της παραγώγου στο αριστερό μέλος και το θεμελιώδες θεώρημα ολοληρωτικού λογισμού στο δεξί μέλος της τελευταίας εξίσωσης παίρνουμε:

$$-\frac{\partial}{\partial x}\iint A(x)[\mathrm{v}_x + \kappa_s p \mathrm{v}_{0x}]dA = \frac{\partial}{\partial t}\iint\limits_{A(x)}(1+\kappa_s p)dA. \tag{3.12}$$

Ορίζοντας την ακουστική ογκική ταχύτητα στο x ως

$$U(x,t) = \iint_{A(x)} \mathrm{v}_x dA$$

και προσεγγίζοντας

$$\iint_{A(x)} p dA \approx pA$$

παίρνουμε :

$$-\frac{\partial U}{\partial x} - \kappa_x \mathrm{v}_{0x}\frac{\partial (pA)}{\partial x} = \frac{\partial A}{\partial t} + \kappa_x \frac{\partial (pA)}{\partial t}. \tag{3.13}$$

Η αρχή διατήρησης της μάζας εφαρμόζεται απλά όπως φαίνεται στη συνέχεια. Αμελούμε σε πρώτη τάξη τα αποτελέσματα των ακτινικών και των περιμετρικών συνιστωσών της χωρικής παραγώγου της πίεσης στο ρευστό. Γι΄ αυτό η αξονική συνιστώσα της αρχής διατήρησης της ορμής γράφεται

$$-\frac{\partial p}{\partial x} = \rho_0 \frac{\partial}{\partial t}\left(\frac{U}{A}\right) + \rho_0 \mathrm{v}_{0x}\frac{\partial}{\partial x}\left(\frac{U}{A}\right) \tag{3.14}$$

με την προσέγγιση

$$\mathrm{v}_x \approx \frac{U}{A}.$$

Και οι δύο εξισώσεις (3.13), (3.14) περιέχουν την ποσότητα A που δεν είναι γνωστή εκ των προτέρων. Παρά του ότι η επιφάνεια διατομής μπορεί να προσδιοριστεί στην ισορροπία οπότε δεν υπάρχει διάδοση ήχου, η αλλαγή στην πίεση του ρευστού στο σωλήνα η οποία συνοδεύει ένα ακουστικό κύμα αναμένεται να προκαλέσει μια μικρή αλλαγή στην επιφάνεια εγκάρσιας διατομής. Θεωρώντας αυτή τη μικρή αλλαγή ως γραμμική διαταραχή, γράφουμε :

$$A = A_0 + \delta A$$

όπου A_0 είναι η επιφάνεια ισορροπίας της εγκάρσιας διατομής. Αφού $\delta A \ll A_0$, γράφουμε [127] :

$$\frac{U}{A} = \frac{U}{A_0+\delta A} = \frac{U}{A_0}[1 - \frac{\delta A}{A_0} + (\frac{\delta A}{A_0})^2 - \ldots] \approx \frac{U}{A_0} \tag{3.15}$$

και

$$pA = p(A_0 + \delta A) \approx A_0.$$

Έτσι, οι Εξ. (3.13), (3.14) γράφονται αντίστοιχα

$$-\frac{\partial U}{\partial x} - \kappa_s \mathrm{v}_{0x}\frac{\partial (pA_0)}{\partial x} = \frac{\partial (A_0+\delta A)}{\partial t} + \kappa_s \frac{\partial (pA_0)}{\partial t}. \tag{3.16}$$

και

$$-\frac{\partial p}{\partial x} = \rho_0 \frac{\partial}{\partial t}\left(\frac{U}{A_0}\right) + \rho_0 \mathrm{v}_{0x}\frac{\partial}{\partial x}\left(\frac{U}{A_0}\right) \tag{3.17}$$

3.2.3.1 Δονούμενα τοιχώματα

Η διαταραχή δA μπορεί να προσδιοριστεί σε συνάρτηση της πίεσης του ρευστού στα τοιχώματα. Η κάθετη μετατόπιση του τοιχώματος από τη θέση ισορροπίας του σημειώνεται ως $y(x,t)$. Αν $S_0(x,t)$ είναι η περίμετρος του σωλήνα στην ισορροπία ($y=0$) τότε η προσεγγιστική αλλαγή στην επιφάνεια, δA, λόγω της μετατόπισης ξ είναι

$$\delta A = S_0 y.$$

Ένα στοιχειώδες μοναδιαίο τμήμα της επιφάνειας του τοιχώματος μοντελοποιείται ως μάζα $M_w(x)$ προσδεδεμένη στο άκρο ελατηρίου με σταθερά $K_w(x)$ και σταθερά απόσβεσης $b_w(x)$. Για μικρές ταλαντώσεις γύρω από τη θέση ισορροπίας για ένα στοιχειώδες μήκος dx του σωλήνα με επιφάνεια $d\Sigma$ η εξίσωση κίνησης θα είναι

$$pd\Sigma - (K_w dx)y - (b_w dx)\dot{y} = (M_w dx)\ddot{y}$$

ή, αν θεωρήσουμε ότι $d\Sigma = S_0 dx$:

$$p = \frac{M_w}{S_0}\ddot{y} + \frac{b_w}{S_0}\dot{y} + \frac{K_w}{S_0}y \tag{3.18}$$

Η λύση σε αυτή την εξίσωση μπορεί να εκφραστεί ως το ολοκλήρωμα συνέλιξης [127] :

$$y(x,t) = \int_{-\infty}^{t} p(x,a)h(x,t-a)da$$

όπου $h(x,t)$ είναι η κρουστική απόκριση της Εξ. (3.18) στη θέση x κατά μήκος του σωλήνα.

3.2.3.2 Ύπαρξη μέσης ροής

Μπορεί τώρα να υποστηριχτεί ότι οι μεταφορικοί όροι της μορφής $\mathrm{v}_{0x}\frac{\partial}{\partial x}$ στις Εξ. (3.16), (3.17) μπορούν να αμεληθούν σε καλή προσέγγιση όταν η συνεχής ταχύτητα της ροής v_{0x} είναι μικρή σε σύγκριση με την ταχύτητα του ήχου c. Γνωρίζουμε ότι στην περίπτωση ενός σωλήνα με σταθερά τοιχώματα και κατά τμήματα σταθερή εγκάρσια επιφάνεια διατομής, όταν δεν υπάρχει κάποια μέση ροή, το σύστημα των εξισώσεων (3.16), (3.17) ικανοποιείται από μια υπέρθεση κυμάτων σε κάθε τμήμα του σωλήνα τα οποία έχουν τη μορφή:

$$p(x,t) = p_+(x-ct) + p_-(x+ct)$$

και

$$U(x,t) = U_+(x-ct) + U_-(x+ct).$$

Παρατηρώντας ότι για κάθε συνάρτηση της μορφής $f(x \pm ct)$ ισχύει $\frac{\partial f(x\pm ct)}{\partial t} = \pm c f'(x \pm ct)$ ενώ $\mathrm{v}_{0x}\frac{\partial f(x\pm ct)}{\partial x} = \mathrm{v}_{0x} f'(x \pm ct)$, μπορεί να θεωρηθεί ότι το λάθος που εισάγεται αμελώντας τους μεταφορικούς όρους είναι μικρό όταν $\mathrm{v}_{0x} \ll c$. Η ισχύς αυτής της υπόθεσης εξετάζεται διεξοδικά στην Ενότητα 4.3 ώστε να διαφανεί ενδεχόμενη επίδραση της μέσης ροής στη διάδοση του ακουστικού πεδίου.

3.2.4 Εξισώσεις γραμμικού ακουστικού πεδίου στη φωνητική οδό, χωρίς μέση ροή

Το τελικό σύνολο εξισώσεων για τη διάδοση ήχου στη φωνητική οδό είναι το παρακάτω:

$$-\frac{\partial U}{\partial x} = \frac{1}{c_0^2\rho_0}\frac{\partial(pA_0)}{\partial t} + \frac{\partial A}{\partial t} \tag{3.19}$$

$$-\frac{\partial p}{\partial x} = \rho_0\frac{\partial}{\partial t}(\frac{U}{A_0}) \tag{3.20}$$

όπου

$$A(x,t) = A_0(x,t) + S_0(x,t)y(x,t) \tag{3.21}$$

και η μετατόπιση $y(x,t)$ των τοιχωμάτων ικανοποιεί τη διαφορική εξίσωση:

$$p(x,t) = M_w(x)\ddot{y}(x,t) + b_w\dot{y}(x,t) + K_w(x)y(x,t). \tag{3.22}$$

Οι παράμετροι που πρέπει να προσδιοριστούν είναι: $A_0(x,t)$ η εγκάρσια επιφάνεια διατομής της φωνητικής οδού (στην ισορροπία), $S_0(x,t)$ η περίμετρος της φωνητικής οδού (στην ισορροπία), $M_w(x)$ η επιφανειακή πυκνότητα ή μάζα ανά μοναδιαία επιφάνεια της φωνητικής οδού, $b_w(x)$ η απόσβεση/μοναδιαία επιφάνεια του τοιχώματος της φωνητικής οδού, $K_w(x)$ η σταθερά ελατηρίου ανα μοναδιαία επιφάνεια του τοιχώματος της φωνητικής οδού, ρ_0 η πυκνότητα του αέρα. Η ταχύτητα του ήχου στη θερμοκρασία σώματος είναι $c_0 = 3.5 \times 10^4 \text{cm/sec}$.

Η εξαγωγή των εξισώσεων βασίστηκε στις παρακάτω υποθέσεις :

Μονοδιάστατη προσέγγιση Οι εγκάρσιοι ρυθμοί διάδοσης είναι μικροί. Η ισχύς της υπόθεσης είναι ικανοποιητική για συχνότητες μέχρι 4kHz [116].

Γραμμική προσέγγιση Το ακουστικό κύμα μπορεί να θεωρηθεί ως γραμμική διαταραχή του ακουστικού μέσου. Η υπόθεση ισχύει όσο το πλάτος του κύματος πίεσης είναι μικρό σε σύγκριση με την ατμοσφαιρική πίεση.

Αμελητέα μέση ροή Η ταχύτητα της μέσης ροής είναι μικρή σε σχέση με την ταχύτητα του ήχου. Ενδεχόμενη άρση της εν λόγω υπόθεσης θα αξιολογηθεί στη συνέχεια, στο Κεφάλαιο 4.

Αμελητέο ιξώδες και θερμικές απώλειες Οι απώλειες λόγω ιξώδους και θερμικής αγωγιμότητας στο κύριο κομμάτι του μέσου και στο συνοριακό στρώμα είναι μικρές, ειδικά όταν συγκρίνονται με τις απώλειες λόγω της δόνησης των τοιχωμάτων και της εκπομπής ήχου στα χείλη. Θα εξεταστεί ενδεχόμενη σημασία τους κατά την προσομοίωση της ακουστικής διάδοσης στη συχνότητα όπως παρουσιάζεται στην Ενότητα 3.3.1.

Ομαλή συνάρτηση εμβαδού Η εγκάρσια επιφάνεια διατομής της φωνητικής οδού δεν αλλάζει υπερβολικά γρήγορα.

Τοιχώματα τοπικά και ελάχιστα αντιδρώντα Τα στοιχεία του τοιχώματος της φωνητικής οδού είναι τοπικά αντιδρώντα δηλαδή γειτονικά στοιχεία δεν είναι συζευγμένα. Επιπλέον, οι δονήσεις των τοιχωμάτων της φωνητικής οδού μπορούν να θεωρηθούν ως μικρές ταλαντώσεις γύρω από ένα σημείο ισορροπίας.

Οι μερικές διαφορικές εξισώσεις που εξάχθηκαν περιγράφουν τη φωνητική οδό ως μια ακουστική γραμμή μεταφοράς. Από τη μελέτη προβλημάτων συνοριακών τιμών είναι γνωστό ότι οι ιδιορυθμοί ενός τέτοιου συστήματος καθορίζονται τόσο από μια διαφορική εξίσωση που περιγράφει το σύστημα όσο και από τον καθορισμό ενός κατάλληλου συνόλου συνοριακών συνθηκών.

3.2.5 Συνοριακές συνθήκες

Εξάγονται οι συνοριακές συνθήκες που αντιστοιχούν στην εκπομπή ήχου στο στόμα και στα ρουθούνια. Ένα λογικό μοντέλο για το φορτίο εκπομπής στο στόμα (ή ρουθούνια), συμβατό με την υπόθεση για μονοδιάστατη διάδοση είναι η σύνθετη αντίσταση εκπομπής για τη διαταραχή που προκαλείται από ένα πιστόνι που βρίσκεται μέσα σε μια σταθερή σφαίρα [55]. Η σφαίρα θα μπορούσε να προσομοίωσει την επιφάνεια του κεφαλιού ενώ η έξοδος του πιστονιού αντιστοιχεί στο στόμα. Η έκφραση γι' αυτή την αντίσταση έχει εξαχθεί

από τους Morse & Ingard ως μια άπειρη σειρά από ποικίλλες σφαιρικές αρμονικές και δεν μπορεί να γραφτεί σε κλειστή μορφή [115].

Ως δεύτερη προσέγγιση υποτίθεται ότι η διάμετρος του πιστονιού είναι μικρή σε σύγκριση με τη διάμετρο της σφαίρας. Σε αυτή την περίπτωση επιλέγουμε για φορτίο εκπομπής τη σύνθετη αντίσταση εκπομπής ενός πιστονιού με άπειρη επίπεδη επιφάνεια στην έξοδο. Η σύνθετη αυτή αντίσταση μπορεί να εκφραστεί σε κλειστή μορφή. Μια σύγκριση μεταξύ των δύο διαφορετικών προσεγγίσεων δείχνει ότι η αλλαγή από σφαιρική σε επίπεδη εξωτερική επιφάνεια αλλάζει ελάχιστα το μέσο φορτίο σύνθετης αντίστσης στο πιστόνι παρά του ότι η κατευθυντικότητα της εκπομπής διαφέρει ουσιαστικά [55]. Τελικά, με τη δεύτερη προσέγγιση, η σύνθετη αντίσταση εκπομπής (εκφρασμένη ως κανονικοποιημένη ακουστική σύνθετη αντίσταση $z_n = Z_A/Z_0 = (p/U)/(\rho c/A)$) προκύπτει ότι μπορεί να θεωρηθεί ανάλογη της σύνθετης αντίστασης της παράλληλης σύνδεσης μιας αντίστασης και μιας επαγωγής [55, 127] :

$$z_n = \frac{j\omega L}{1 + j\omega \frac{L}{R}} \tag{3.23}$$

όπου $L_{rad} = 8a/3\pi c_0$ και $R_{rad} = 128/9\pi^2$ και A, a είναι η επιφάνεια και η ακτίνα του πιστονιού (στόματος) αντίστοιχα. Για να χρησιμοποιηθεί ως συνοριακή συνθήκη για την κυματική εξίσωση που εξάχθηκε η Εξίσωση (3.23) πρέπει να μετασχηματιστεί στο χρόνο. Χρησιμοποιώντας τη συνήθη αναλογία μεταξύ ακουστικής πίεσης και ηλεκτρικής τάσης και ακουστικής ογκικής ταχύτητας με το ηλεκτρικό ρεύμα, έχουμε στο πεδίο του χρόνου:

$$U(t) = \frac{A}{\rho_0 c_0} \left[\frac{1}{L_{rad}} \int_{-\infty}^{t} p(\tau) d\tau + \frac{p(t)}{R_{rad}} \right] \tag{3.24}$$

Η τελευταία εξίσωση είναι αυτή που χρησιμοποιείται ως συνοριακή συνθήκη για το φορτίο εκπομπής στα χείλια ή στα ρουθούνια.

3.3 Προσομοίωση στη συχνότητα

Στο πεδίο της συχνότητας, η προσομοίωση γίνεται ακολουθώντας την προσέγγιση που προτείνεται από τον Portnoff [127]. Συγκεκριμένα, οι εξισώσεις μετασχηματίζονται στο πεδίο της συχνότητας, θεωρώντας λύση μόνιμης κατάστασης της μορφής :

$$p(x,t) = \hat{p}(x,\omega)e^{j\omega t} \tag{3.25}$$

$$U(x,t) = \hat{U}(x,\omega)e^{j\omega t} \tag{3.26}$$

$$y(x,t) = \hat{y}(x,\omega)e^{j\omega t}. \tag{3.27}$$

Τελικά, προκύπτουν οι εξισώσεις :

$$\frac{\partial \hat{p}}{\partial x} + Z\hat{U} = 0 \tag{3.28}$$

$$\frac{\partial \hat{U}}{\partial x} + Y\hat{p} = 0 \tag{3.29}$$

όπου η σύνθετη αντίσταση και η σύνθετη αγωγιμότητα είναι

$$Z = j\omega_0\rho_0/A_0 \tag{3.30}$$

$$Y = j\omega[\kappa_s A_0 + \frac{(K_w - \omega^2 M_w) - j\omega b_w}{(K_w - \omega^2 M_w)^2 + \omega^2 b_w^2} S_0]. \tag{3.31}$$

ενώ για να λάβουμε υπόψη μας και απώλειες λόγω ιξώδους η μεταγωγής θερμότητας μπορούμε να συμπεριλάβουμε τις εκφράσεις αυτών των απωλειών στο πεδίο συχνοτήτων. Τα

αποτελέσματα ιξώδους τριβής στα τοιχώματα του σωλήνα μπορούν να συμπεριληφθούν με την πρόσθεση ενός όρου αντίστασης ενώ για τη θερμική αγωγιμότητα μπορεί να προστεθεί ένας όρος αγωγιμότητας [55]:

$$\frac{\partial \hat{p}}{\partial x} + Z\hat{U} + R_a\hat{U} = 0 \tag{3.32}$$

$$\frac{\partial \hat{U}}{\partial x} + Y\hat{p} + G_a\hat{p} = 0 \tag{3.33}$$

όπου

$$R_a = \frac{S}{A^2}\sqrt{\frac{\omega\rho_0\mu}{2}} \tag{3.34}$$

$$G_a = S\frac{n-1}{\rho_0 c^2}\sqrt{\frac{\lambda\omega}{2C_p\rho_0}} \tag{3.35}$$

με μ το συντελεστή συνεκτικότητας του ρευστού, λ το συντελεστή θερμικής αγωγιμότητας του ρευστού, n την αδιαβατική σταθερά και C_p τη θερμική χωρητικότητα του ρευστού σε σταθερή πίεση.

Οι οριακές συνθήκες είναι:

$$\hat{p} - Z_{rad}\hat{U} = 0, \tag{3.36}$$

$$Y_g\hat{p} + \hat{U} = \hat{U_g} \tag{3.37}$$

για τα χείλια και τη γλωττίδα αντίστοιχα. Για τη γλωττίδα έχει θεωρηθεί ένα ισοδύναμο δίκτυο Norton και Y_g είναι η αντίστοιχη αγωγιμότητα της γλωττίδας. Αν αυτή θεωρηθεί πολύ μεγαλύτερη από την σύνθετη αντίσταση εισόδου της φωνητικής οδού, τότε απλά έχουμε $\hat{U} = \hat{U_g}$.

Οι εξισώσεις που προκύπτουν διακριτοποιούνται με βάση το παρακάτω σχήμα (central-difference with averaging):

$$[\frac{df}{dx} + g]_{x=(i-1/2)\Delta x} = 0 \rightarrow \frac{1}{\Delta x}(f_i - f_{i-1}) + \frac{1}{2}(g_i + g_{i-1}) = 0 \tag{3.38}$$

όπου Δx είναι το βήμα της χωρικής διακριτοποίησης. Για το γραμμικό σύστημα $AX = B$ που είναι τελικά προς επίλυση, δείτε το [127, σελ. 46-47]. Για την επίλυση αξιοποιείται η τετραγωνικότητα και η αραιότητα του πίνακα A.[1]

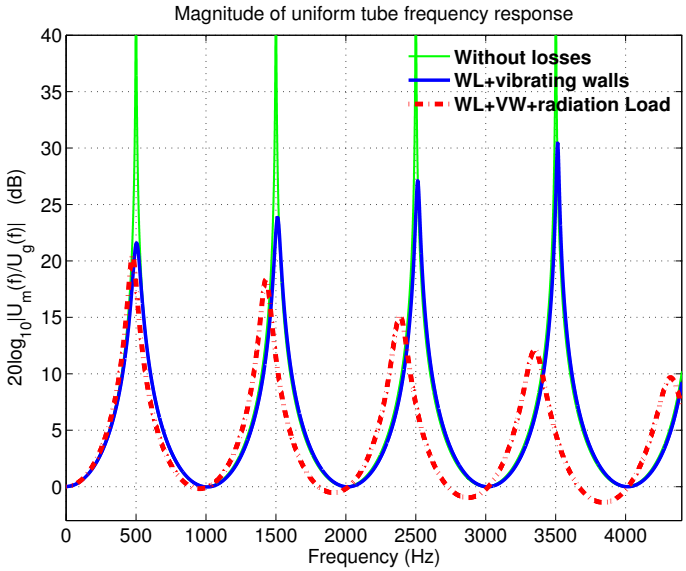
3.3.1 Απόκριση συχνότητας ομοιόμορφου σωλήνα

Η απόκριση συχνότητας του σωλήνα δίνεται ως $H(f) = U_m(f)/U_g(f)$ όπου U_m είναι η ογκική ταχύτητα στα χείλη και U_g είναι η ογκική ταχύτητα στη γλωττίδα. Στο Σχήμα 3.2 δίνεται η ειδική περίπτωση του μέτρου της απόκρισης συχνότητας ($|H(f)|$) για έναν κυλινδρικό σωλήνα σταθερού εμβαδού εγκάρσιας διατομής $A = 5 \times 10^{-}4\ \mathrm{m}^2$ με μήκος $l = 0,175$ m. Αν θεωρήσουμε σταθερά τοιχώματα, μηδενικές απώλειες και ανοιχτό το άκρο που αντιστοιχεί στα χείλια (η συνολική πίεση είναι ίση με την ατμοσφαιρική στα χείλια, οπότε η ακουστική πίεση είναι $p_{lips} = 0$), τότε αναμένονται συντονισμοί (απειρισμοί του μέτρου της απόκρισης συχνότητας) στις συχνότητες [134]:

$$f_{0k} = k\frac{c}{4l},\ k = 1, 3, 5\ldots \tag{3.39}$$

δηλαδή στις συχνότητες $f = 500, 1500, 2500\ldots$Hz για ταχύτητα ήχου $c = 350$ m/sec και για τον σωλήνα υπό εξέταση. Στο αριστερό άκρο της φωνητικής οδού ως οριακή συνθήκη

[1] Η επίλυση γίνεται με τη συνάρτηση mldivide του MATLAB.



Σχήμα 3.2: *Πλάτος της απόκρισης συχνότητας ομοιόμορφου σωλήνα όπως προκύπτει από προσομοίωση του ακουστικού πεδίου στο πεδίο της συχνότητας. Φαίνεται η επίδραση του φορτίου εκπομπής και των δονούμενων τοιχωμάτων. Οι συχνότητες συντονισμού για την περίπτωση χωρίς απώλειες είναι πολύ κοντά στις θεωρητικά αναμενόμενες.*

	F_1	F_2	F_3	F_4	F_5
Χωρίς απώλειες	500.10 Hz	1500.20 Hz	2500.50 Hz	3501.30 Hz	4502.60 Hz
+φορτίο εκπομπής	-5.74%	-5.49%	-5.11%	-4.69%	-4.29%
+δονούμενα τοιχώματα	+0.78%	+0.80%	+0.57%	+0.40%	+0.28%
+απώλειες λόγω ιξώδους	-0.02%	-0.01%	-0.00%	-0.01%	-0.00%
+θερμικές απώλειες	-0.02%	-0.01%	-0.00%	-0.00%	-0.00%

Πίνακας 3.1: *Συχνότητες συντονισμών για τον ομοιόμορφο σωλήνα καθώς γίνεται σταδιακά συνθετότερη η μοντελοποίηση του ακουστικού πεδίου. Η προσθήκη του φορτίου εκπομπής έχει ως αποτέλεσμα τη μείωση των συχνοτήτων συντονισμού ενώ η πρόσθετη επίδραση των δονούμενων τοιχωμάτων, των απωλειών συνεκτικότητας και των θερμικών απωλειών είναι σχετικά μικρή.*

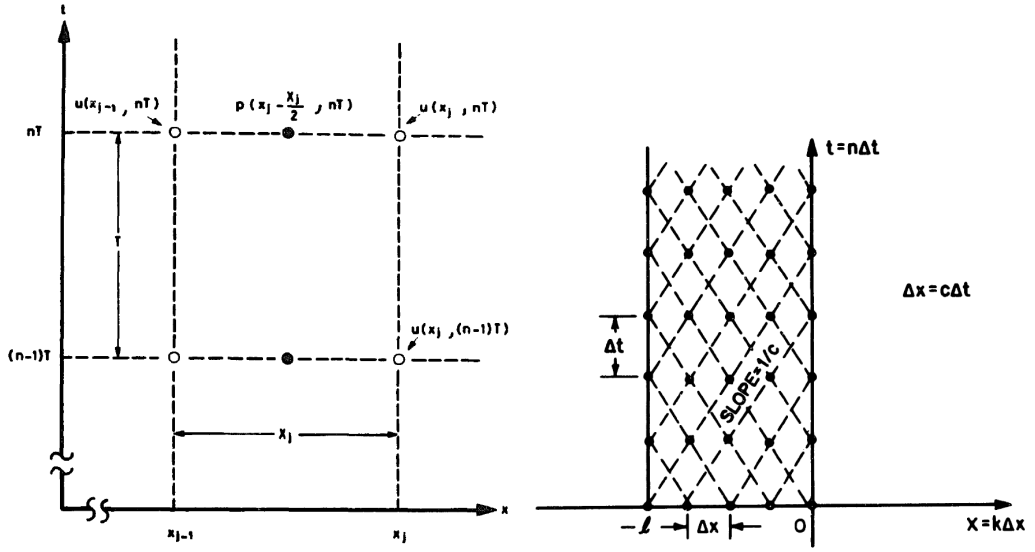
λαμβάνεται ότι $U_g = 1$ για όλες τις συχνότητες. Σημειώνεται ότι η απόκριση συχνότητας υπολογίζεται μέχρι τη συχνότητα $F_{max} = 5$ kHz με βήμα $\Delta f = 0.1$ Hz. Το χωρικό βήμα διακριτοποίησης είναι $\Delta x = 0.001$ m. Για την περίπτωση που δεν υπάρχουν απώλειες οπότε και είναι γνωστές οι αναμενόμενες συχνότητες συντονισμού ήταν δυνατή η εκτίμηση του λάθους προσέγγισης λόγω της διακριτοποίησης, $e \sim O(\Delta x^2)$.

Στους Πίνακες 3.2, 3.2 παρουσιάζονται οι συχνότητες των συντονισμών και τα αντίστοιχα εύρη ζώνης καθώς η μοντελοποίηση του ακουστικού πεδίου στο σωλήνα γίνεται συνθετότερη. Τα 3 dB εύρη ζώνης που δίνονται έχουν υπολογιστεί με βάση την προσαρμογή κατάλληλης παραβολής στα σημεία των συντονισμών [2]. Όπως φαίνεται και στο Σχ. 3.2 η επίδραση του φορτίου εκπομπής είναι κυρίως σημαντική στις υψηλές συχνότητες ενώ τα δονούμενα τοιχώματα έχουν ως αποτέλεσμα την αύξηση των ευρών ζώνης των χαμηλόσυχνων συντονισμών [13, 53, 55, 127]. Οι θερμικές απώλειες και οι απώλειες λόγω συνεκτικότητας είναι λιγότερο σημαντικές για την εξεταζόμενη γεωμετρία.

[2] Έγινε προσαρμογή μοντέλου της μορφής $ax^2 + b$ με χρήση της συνάρτησης fit του MATLAB στα σημεία τοπικού μεγίστου του πλάτους της απόκρισης συχνότητας

	B_1	B_2	B_3	B_4	B_5
Χωρίς απώλειες	0.26 Hz	0.27 Hz	0.28 Hz	0.30 Hz	0.28 Hz
+φορτίο εκπομπής	$\times$12.91	$\times$105.39	$\times$242.80	$\times$389.13	$\times$572.65
+δονούμενα τοιχώματα	$\times$16.86	$\times$2.52	$\times$1.44	$\times$1.18	$\times$1.09
+απώλειες λόγω ιξώδους	$\times$1.06	$\times$1.09	$\times$1.08	$\times$1.07	$\times$1.07
+θερμικές απώλειες	$\times$1.03	$\times$1.04	$\times$1.04	$\times$1.03	$\times$1.03

Πίνακας 3.2: *Εύρη ζώνης των συντονισμών για τον ομοιόμορφο σωλήνα καθώς γίνεται σταδιακά συνθετότερη η μοντελοποίηση του ακουστικού πεδίου. Φαίνεται ότι η επίδραση του φορτίου εκπομπής γίνεται κυρίως σημαντική για τους υψίσυχνους συντονισμούς σε αντίθεση με την επίδραση των δονούμενων τοιχώματων.*



Σχήμα 3.3: *Πλέγμα για την αριθμητική προσομοίωση της ακουστικής διάδοσης μέσα στη φωνητική οδό όπως χρησιμοποιήθηκε από του Maeda (αριστερά) και του Portnoff (δεξιά).*

3.4 Προσομοίωση στο χρόνο

Οι εξισώσεις που περιγράφουν τη διάδοση των ακουστικών κυμάτων μέσα στη φωνητική οδό είναι ένα σύστημα γραμμικών μερικών διαφορικών εξισώσεων. Το πρόβλημα της προσομοίωσης του μονοδιάστατου μοντέλου της φωνητικής οδού έχει διδιάστατο χαρακτήρα: υπάρχει μια χωρική και μια χρονική διάσταση. Για να επιλυθεί αριθμητικά διατυπώνεται με βάση διακριτές μεταβλητές και πεπερασμένες διαφορές. Είναι σημαντικός ο προσδιορισμός του πλέγματος των σημείων στα οποία θα υπολογιστεί τελικά η λύση. Τα πλέγματα που χρησιμοποιήθηκαν από τους Portnoff και Maeda φαίνονται στο Σχ. 3.3. Ο χώρος διακριτοποιείται στα σημεία $x_n = \sum_{i=1}^{n} X_i, n = 0, 1, \ldots N$ όπου $X_i = \Delta x$ στο σχήμα του Portnoff ενώ το πλέγμα του Maeda είναι χωρικά ανομοιόμορφο. Επιπλέον, στο τελευταίο η πίεση δειγματοληπτείται στο μέσο των ανομοιόμορφων διαστημάτων.

Για τη διακριτοποίηση, μελετήσαμε τα αριθμητικά σχήματα που προτάθηκαν από τους Portnoff [127] και Maeda [101] καθώς και την επέκταση που προτάθηκε στο [114] και που αφορά στη σύζευξη της φωνητικής οδού με κοιλότητες όπως είναι η ρινική και άλλες. Τα σχήματα που παρουσιάζονται στα [24, 43] είναι βασισμένα σε παρόμοια φιλοσοφία. Τα εν λόγω σχήματα καλούνται implicit επειδή για τον προσδιορισμό της κατάστασης $X[n]$ κάθε χρονική στιγμή χρειάζεται να επιλυθεί ένα σύστημα εξισώσεων που εμπλέκει και την κατάσταση την προηγούμενη χρονική στιγμή:

$$G(X[n], X[n-1]) = 0.$$

Αντίθετα, τα λεγόμενα explicit σχήματα δίνουν την κατάσταση άμεσα ως :

$$X[n] = F(X[n-1]).$$

Για τα τελευταία, υπάρχει ο περιορισμός ότι τα διαστήματα διακριτοποίησης πρέπει να ικανοποιούν τη σχέση $\Delta x \geq c\Delta t$, όπως προκύπτει από τη συνθήκη Courant-Friedrichs-Lewy για να υπάρχει ευστάθεια. Πρακτικά, με ένα implicit σχήμα είναι δυνατόν να πάρουμε την ίδια ακρίβεια με αρκετά μεγαλύτερα διαστήματα διακριτοποίησης. Το μειονέκτημα είναι βέβαια ότι η υλοποίηση των implicit σχημάτων είναι συνήθως δυσκολότερη.

Το σχήμα του Portnoff για τη διακριτοποίηση των χωρικών και χρονικών παραγώγων καλείται σχήμα κεντρικών διαφορών, και λαμβάνεται ως επέκταση του κανόνα του τραπεζίου σε δύο διαστάσεις. Αντίθετα ο Maeda εφαρμόζει τον κανόνα του μέσου σημείου για τη χωρική διακριτοποίηση και τον κανόνα του τραπεζίου για τη χρονική διακριτοποίηση. Η χωρική διακριτοποίηση του Maeda έχει το χαρακτηριστικό ότι οδηγεί σε αναπαράσταση της φωνητικής οδού με τη μορφή ηλεκτρικού κυκλώματος. Και οι δύο κανόνες που χρησιμοποιούνται αφορούν αρχικά στον προσεγγιστικό υπολογισμό ολοκληρωμάτων. Ο κανόνας του μέσου σημείου διατυπώνεται ως :

$$\int_a^b g(x)dt \approx (b-a)g(a+b/2)$$

με λάθος προσέγγισης $-\frac{1}{24}(b-a)^3 f''(a) + O((b-a)^4)$. Ο κανόνας του τραπεζίου από την άλλη έχει ως εξής :

$$\int_a^b g(\alpha)d\alpha \approx (b-a)(g(a)+g(b))/2$$

Το λάθος της προσέγγισης είναι $\frac{1}{12}(b-a)^3 f''(a) + O((b-a)^4)$. Αν $g(\alpha) = df(\alpha)/d\alpha$ τότε παίρνουμε $f(b) - f(a) = (b-a)(g(a)+g(b))/2$ και με $b = x_n$ και $a = x_{n-1}$ προκύπτει η διακριτοποίηση της χωρικής παραγώγου $g(x) = \frac{\partial f}{\partial x}$ (central difference with averaging) :

$$g(x_n) = \frac{2}{X_n}(f(x_n) - f(x_{n-1}) - g(x_{n-1}).$$

Σε περίπτωση ομοιόμορφης χωρικής διακριτοποίησης είναι $X_n = x_n - x_{n-1} = \Delta x$.

3.4.1 Χωρική διακριτοποίηση

3.4.1.1 Αρχή διατήρησης της μάζας

Με βάση τα παραπάνω, η εξίσωση συνέχειας :

$$\frac{\partial U(x,t)}{\partial x} + \frac{1}{\rho_0 c_0^2}\frac{\partial}{\partial t}[A_0(x,t)p(x,t)] + \frac{\partial A(x,t)}{\partial t} = 0$$

διακριτοποιείται χωρικά από τον Portnoff ως :

$$\frac{2}{X_n}(U(x_n,t) - U(x_{n-1},t)) - \frac{\partial U(x,t)}{\partial x}|_{x=x_{n-1}} + \frac{1}{\rho_0 c^2}\frac{d}{dt}[A_0(x_n,t)p(x_n,t)] + \frac{dA(x_n,t)}{dt} = 0 \quad (3.40)$$

ή, αντικαθιστώντας την παράγωγο στο σημείο x_{n-1} :

$$U(x_n,t) - U(x_{n-1},t) + \frac{X_n}{2}\frac{1}{\rho_0 c_0^2}\frac{d}{dt}[A_0(x_n,t)p(x_n,t) + A(x_{n-1},t)p(x_{n-1},t)] + \frac{X_n}{2}\frac{d}{dt}[A(x_n,t) + A(x_{n-1},t)] = 0 \quad (3.41)$$

Με τον κανόνα του μέσου σημείου και το ανομοιόμορφο πλέγμα του Maeda η εξίσωση, μετά από ολοκλήρωση, διακριτοποιείται χωρικά ως :

$$U(x_n,t) - U(x_{n-1},t) + X_n \frac{1}{\rho_0 c_0^2}\frac{d}{dt}[A_0(x_n,t)P(x_n - X_n/2,t)] + X_n\frac{d}{dt}A_0(x_n,t) + X_n\frac{d}{dt}S_0(x_n,t)y(x_n,t) = 0. \quad (3.42)$$

όπου έχουμε θεωρήσει ότι $A_0(x,t) = A_0(x_n,t), S_0(x,t) = S_0(x_n,t)$ για $x_{n-1} < x \leq x_n$. Επίσης, έχει χρησιμοποιηθεί η Εξ. (3.21) και $y(x_n,t)$ είναι η μετατόπιση των δονούμενων τοιχωμάτων. Το μέγεθος

$$U_{vw}(x_n,t) = X_n\frac{d}{dt}S_0(x_n,t)y(x_n,t)$$

αντιπροσωπεύει τη μεταβολή της ακουστικής ογκικής ταχύτητας λόγω της δόνησης αυτής.

3.4.1.2 Αρχή διατήρησης της ορμής

Η εξίσωση κίνησης τροποποιείται κατάλληλα με την προσθήκη ενός όρου αντίστασης της ροής ώστε να συμπεριλάβει και απώλειες λόγω ιξώδους σε στενώσεις [26, 101]:

$$\frac{\partial p(x,t)}{\partial x} + \rho_0\frac{\partial}{\partial t}\frac{U(x,t)}{A_0(x,t)} + r(x,t)U(x,t) = 0, \quad (3.43)$$

όπου $r(x,t)$ είναι γνωστή ως αντίσταση Hagen-Poiseuille :

$$r(x,t) = \frac{8\pi\mu}{A_0(x,t)^2}. \quad (3.44)$$

Η συγκεκριμένη αντίσταση γίνεται σημαντική μόνο σε στενώσεις. Για διατομές μεγαλύτερου σχετικά πλάτους οι αντιστάσεις λόγω ιξώδους συγκεντρώνονται στο συνοριακό στρώμα και εξαρτώνται από τη συχνότητα. Προσπάθεια να ληφθούν και αυτές υπόψη σε μια αριθμητική προσομοίωση στο χρόνο έγινε στο [26, σελ. 72].

Η χωρική διακριτοποίηση επιτυγχάνεται με τον κανόνα του τραπεζίου ως :

$$p(x_n,t) - p(x_{n-1},t) + \frac{X_n}{2}\rho_0\frac{d}{dt}\left[\frac{U(x_n,t)}{A_0(x_n,t)} + \frac{U(x_{n-1},t)}{A_0(x_{n-1},t)}\right] + \frac{X_n}{2}[r(x_n,t)U(x_n,t) + r(x_{n-1},t)U(x_{n-1},t)] = 0 \quad (3.45)$$

ενώ με τον κανόνα του μέσου σημείου είναι :

$$p(x_n - X_n/2,t) - p(x_{n-1} - X_{n-1}/2,t) + \frac{X_{n-1}}{2}\rho_0\frac{d}{dt}\left[\frac{U(x_n,t)}{A_0(x_{n-1},t)}\right] + \frac{X_n}{2}\rho_0\frac{d}{dt}\left[\frac{U(x_n,t)}{A_0(x_n,t)}\right] + \frac{X_{n-1}}{2}\frac{r(x_{n-1},t)}{A_0(x_{n-1},t)}U(x_n,t) + \frac{X_n}{2}\frac{r(x_n,t)}{A_0(x_n,t)}U(x_n,t) = 0 \quad (3.46)$$

3.4.2 Χρονική διακριτοποίηση

Μετά τη χρονική διακριτοποίηση ($t = k\Delta t$), για το σχήμα του Portnoff και για κάθε χρονική στιγμή έχουμε το σύστημα $2N-2$ εξισώσεων με $2N$ αγνώστους ($n = 1 \ldots N$) (θεωρείται ότι

$f_n^k = f(x_{n-1}, k\Delta t), f_{n+\frac{1}{2}}^k = f(x_n - X_n/2, k\Delta t)$):

$$-U_n^k + (Yp)_n^k + U_{n+1}^k + (Yp)_{n+1}^k = U_n^{k-1} + (Yp)_n^{k-1} - U_{n+1}^{k-1} + (Yp)_{n+1}^{k-1} - \frac{X_n}{\Delta t}[A_{n+1}^k - A_{n+1}^{k-1} + A_n^k - A_n^{k-1}] \quad (3.47)$$

$$((Z+R)U)_n^k - p_n^k + ((Z+R)U)_{n+1}^k + p_{n+1}^k = ((Z-R)U)_n^{k-1} + p_n^{k-1} + ((Z-R)U)_{n+1}^{k-1} - p_{n+1}^{k-1} \quad (3.48)$$

όπου

$$Z_n^k = \rho_0 \frac{X_n}{\Delta t} \frac{1}{A_{0n}^k} \quad (3.49)$$

$$Y_n^k = \frac{1}{\rho_0 c_0^2} \frac{X_n}{\Delta_t} A_{0n}^k \quad (3.50)$$

$$R_n^k = \frac{X_n}{2} r_n^k. \quad (3.51)$$

Το ίδιο σύστημα θα είχαμε πάρει αν εφαρμόζαμε τους μετασχηματισμούς που ισοδυναμούν με λήψη κεντρικής διαφοράς στη διάσταση της παραγώγισης και μέσης τιμής στην άλλη [127]:

$$\frac{\partial f}{\partial x}|_{x=(i+1/2)\Delta x}^{t=(n+1/2)\Delta t} = \frac{1}{2\Delta x}[f_{i+1}^{n+1} - f_i^{n+1} + f_{i+1}^n - f_i^n] \quad (3.52)$$

$$\frac{\partial f}{\partial t}|_{x=(i+1/2)\Delta x}^{t=(n+1/2)\Delta t} = \frac{1}{2\Delta t}[f_{i+1}^{n+1} - f_{i+1}^n + f_i^{n+1} - f_i^n], \quad (3.53)$$

ενώ για το σχήμα του Maeda έχουμε :

$$-U_n^k + 2Y_n^k p_{n+\frac{1}{2}}^k + U_{n+1}^k = + U_n^{k-1} + 2Y_n^{k-1} p_{n+\frac{1}{2}}^{k-1} - U_{n+1}^{k-1} - \frac{2X_n}{\Delta t}[A_{0n}^k - A_{0n}^{k-1}] - (U_{vwn}^k + U_{vwn}^{k-1}) \quad (3.54)$$

$$(Z_n^k + Z_{n-1}^k + R_n^k + R_{n-1}^k)u_n^k + p_{n+\frac{1}{2}}^k - p_{n-\frac{1}{2}}^k = (Z_n^{k-1} + Z_{n-1}^{k-1} - R_n^{k-1} - R_{n-1}^{k-1})u_n^{k-1} - p_{n+\frac{1}{2}}^{k-1} + p_{n-\frac{1}{2}}^{k-1}. \quad (3.55)$$

3.4.3 Δονούμενα τοιχώματα

Μένει να προσδιοριστεί το εμβαδό A_n^k της εγκάρσιας διατομής των δονούμενων τοιχωμάτων κάθε χρονική στιγμή. Επειδή ισχύει η Εξ. (3.21) και θεωρείται ότι η εγκάρσια επιφάνεια και η περίμετρος είναι γνωστές στην ηρεμία ως A_{0n}^k, S_{0n}^k, αρκεί να προσδιορίσουμε την απομάκρυνση y. Αν θεωρήσουμε την απόκλιση των τοιχωμάτων σταθερή για ένα διάστημα χωρικής διακριτοποίησης, διακριτοποιώντας χωρικά την (3.22) παίρνουμε:

$$p = M_w \frac{d^2 y(x_n, t)}{dt^2} + b_w \frac{dy(x_n, t)}{dt} + K_w y(x_n, t) \quad (3.56)$$

Έχει αποφευχθεί να προσδιοριστεί χωρικός δείκτης για την πίεση ώστε να συμπεριληφθούν οι δύο διαφορετικές διακριτοποιήσεις των αριθμητικών σχημάτων.

3.4.3.1 Διακριτοποίηση με τη μέθοδο της αμετάβλητης κρουστικής απόκρισης

Για τον Portnoff η απόκλιση προσδιορίζεται ως έξοδος του συστήματος ταλαντωτή που μοντελοποιεί ένα στοιχειώδες τμήμα της επιφάνειας του τοιχώματος με είσοδο την πίεση p. Η συνάρτηση μεταφοράς της Εξ. (3.57) υπολογίζεται στη συχνότητα και στη συνέχεια διακριτοποιείται με τη μέθοδο της αμετάβλητης κρουστικής απόκρισης (impulse invariance). Τελικά προκύπτει ότι:

$$y_n^k = c_{0n} p_n^{k-1} + c_{1n} y_n^{k-1} + c_{2n} y_n^{k-2} \tag{3.57}$$

όπου

$$c_{\nu n} = c_\nu(x), \nu = 0, 1, 2 \text{ και } x_{n-1} < x \leq x_n \tag{3.58}$$

$$c_0 = \begin{cases} \frac{\Delta t}{\omega_0 M_w} e^{\sigma_0 \Delta t} \sin \omega_0 \Delta t & \text{αν } b_w^2 - 4M_w K_w < 0 \\ \frac{\Delta t}{\alpha_0 M_w} e^{\sigma_0 \Delta t} \sinh \alpha_0 \Delta t & \text{αν } b_w^2 - 4M_w K_w > 0 \\ \frac{\Delta t^2}{M_w} e^{\sigma_0 \Delta t} & \text{αν } b_w^2 - 4M_w K_w = 0 \end{cases} \tag{3.59}$$

$$c_1 = \begin{cases} 2e^{\sigma_0 \Delta t} \cos \omega_0 \Delta t & \text{αν } b_w^2 - 4M_w K_w < 0 \\ 2e^{\sigma_0 \Delta t} \cosh \alpha_0 \Delta t & \text{αν } b_w^2 - 4M_w K_w > 0 \\ 2e^{\sigma_0 \Delta t} & \text{αν } b_w^2 - 4M_w K_w = 0 \end{cases} \tag{3.60}$$

$$c_2 = -e^{2\sigma_0 \Delta t} \tag{3.61}$$

και

$$\sigma_0 = -b_w / 2M_w \tag{3.62}$$

$$\omega_0 = \sqrt{K_w/M_w - (b_w/2M_w)^2} \tag{3.63}$$

$$\alpha_0 = \sqrt{(b_w/2M_w)^2 - K_w/M_w} \tag{3.64}$$

Το πλεονέκτημα με αυτή τη διακριτοποίηση είναι ότι δεν είναι αναγκαίο να λυθεί η Εξ. (3.57) ταυτόχρονα με τις Εξ. (3.47) και (3.48). Πράγματι, η απομάκρυνση των τοιχωμάτων κάθε χρονική στιγμή μπορεί να προσδιοριστεί με βάση ποσότητες σε προηγούμενες χρονικές στιγμές μόνο.

3.4.3.2 Διακριτοποίηση με τον κανόνα του τραπεζίου

Διαφορετική είναι η πρακτική που ακολουθεί ο Maeda και τελικά οδηγείται σε ταυτόχρονη επίλυση όλων των εξισώσεων. Αρχικά εκφράζει τη χωρικά διακριτοποιημένη εξίσωση κίνησης των τοιχωμάτων με βάση την ογκική ταχύτητα U_{vw}:

$$p(x_n - X_n/2, t) = M_w \frac{d}{dt} \frac{U_{vw}(x_n, t)}{S_0(x_n, t) X_n} + \frac{b_w}{X_n S_0(x_n, t)} U_{vw}(x_n, t) + \frac{K_w}{S_0(x_n, t)} \int_0^t \frac{U_{vw}(x_n, t)}{X_n} \tag{3.65}$$

Στη συνέχεια, με χρονική διακριτοποίηση χρησιμοποιώντας τον κανόνα του τραπεζίου και λύνοντας ως προς U_{vw} παίρνουμε:

$$U_{vwn}^k = Y_{wn}^k (p_{n+\frac{1}{2}}^k + \underbrace{M_w Q_n^{k-1} - \frac{K_w}{S_{0n}^k} V_n^{k-1}}_{\Psi_n^{k-1}}) \tag{3.66}$$

όπου

$$Y_{wn}^k = 1/(\frac{2M_w}{\Delta_t X_n S_{0n}^k} + \frac{b_w}{X_n S_{0n}^k} + \frac{\Delta t K_w}{2 X_n S_{0n}^k}). \tag{3.67}$$

Για τους όρους Q, V που προκύπτουν κατά τη διακριτοποίηση ισχύει:

$$Q_n^{k-1} = \frac{4}{\Delta t}\frac{U_{vwn}^{k-1}}{S_{0n}^{k-1}X_n} - Q_n^{k-2} \tag{3.68}$$

$$V_n^{k-1} = \Delta t\frac{U_{vwn}^{k-1}}{X_n} + V_n^{k-2}. \tag{3.69}$$

Είναι φανερό ότι η ογκική ταχύτητα λόγω της δόνησης των τοιχωμάτων εξαρτάται από την πίεση την ίδια χρονική στιγμή οπότε οι Εξ. (3.54), (3.55) και (3.66) πρέπει να λυθούν ταυτόχρονα. Αν και στην πράξη δε χρησιμοποιείται άμεσα, η μετατόπιση των τοιχωμάτων προκύπτει ως :

$$y_n^k = \frac{\Delta t}{2X_nS_{0n}^k}(U_{vwn}^k + U_{vwn}^{k-1}) + \frac{S_{0n}^{k-1}}{S_{0n}^k}y_n^{k-1}. \tag{3.70}$$

Πρακτικά, αφού αντικατασταθεί η U_{vwn}^k στην Εξ. (3.54) με την Εξ. (3.66) λύνεται η Εξ. (3.54) ως προς $p_{n+\frac{1}{2}}^k$:

$$p_{n+\frac{1}{2}}^k = b_n^k(U_n^k - U_{n+1}^k - \Phi_n^{k,k-1}) \tag{3.71}$$

όπου

$$b_n^k = 1/(2Y_n^k + Y_{wn}^k) \tag{3.72}$$

και

$$\Phi_n^{k,k-1} = \frac{2X_n}{\Delta t}[A_{0n}^k - A_{0n}^{k-1}] + Y_{wn}^k\Psi_n^{k-1} - 2Y_n^{k-1}p_{n+\frac{1}{2}}^{k-1} + U_{n+1}^{k-1} - U_n^{k-1} + U_{vwn}^{k-1} \tag{3.73}$$

Αντικαθιστώντας στην Εξ. (3.55) τελικά παίρνουμε για $n = 2, \dots N$ το σύστημα των $N-1$ εξισώσεων:

$$\begin{aligned} &- b_{n-1}^kU_{n-1}^k + \underbrace{(Z_n^k + Z_{n-1}^k + R_n^k + R_{n-1}^k + b_n^k + b_{n-1}^k)}_{H_{n,n-1}^k}U_n^k - b_n^kU_{n+1}^k = \\ &\underbrace{(Z_n^{k-1} + Z_{n-1}^{k-1} - R_n^k - R_{n-1}^{k-1})U_n^{k-1} - p_{n+\frac{1}{2}}^{k-1} + p_{n-\frac{1}{2}}^{k-1} + b_n^k\Phi_n^{k,k-1} - b_{n-1}^k\Phi_{n-1}^{k,k-1}}_{F_{n,n-1}^{k,k-1}}. \end{aligned} \tag{3.74}$$

Αξίζει να σημειωθεί ότι αυτή η προσέγγιση για τη μοντελοποίηση των δονούμενων τοιχωμάτων δεν προκύπτει από κάποια εγγενή απαίτηση της διακριτοποίησης του Maeda. Είναι επίσης δυνατός ο προσδιορισμός της ογκικής ταχύτητας U_{vw} ως :

$$U_{vwn}^k = 2\frac{X_n}{\Delta t}(S_{0n}^ky_n^k - S_{0n}^{k-1}y_n^{k-1}) \tag{3.75}$$

οπότε μπορεί να βρεθεί ως συνάρτηση μεγεθών της προηγούμενης χρονικής στιγμής αφού η μετατόπιση y_n^k μπορεί να προσδιοριστεί με τη μέθοδο της αμετάβλητης κρουστικής απόκρισης όπως περιγράφτηκε. Αυτό το υβριδικό σχήμα μοντελοποίησης των τοιχωμάτων είναι κιόλας που χρησιμοποιήθηκε τελικά ώστε να αποφευχθούν συγκεκριμένες περιπτώσεις ασταθούς συμπεριφοράς [3].

[3]Κατά τον υπολογισμό της απόκρισης συχνότητας για τη συνάρτηση επιφανείας που αντιστοιχεί στο φώνημα /u/ όπως δίνεται στο [52] το σύστημα προέκυπτε ασταθές μετά την προσθήκη των δονούμενων τοιχωμάτων με τη μέθοδο της διακριτοποίησης του τραπεζίου.

3.4.4 Οριακή συνθήκη εκπομπής

Για την οριακή συνθήκη εκπομπής, χρησιμοποιείται ο κανόνας του τραπεζίου για τη διακριτοποίηση οπότε προκύπτει:

$$U_{N+1}^k - [G_{rad}^k + \frac{\Delta t}{2} S_{rad}^k] p_{N+1}^k = U_{N+1}^{k-1} - [G_{rad}^{k-1} - \frac{\Delta t}{2} S_{rad}^{k-1}] p_{N+1}^{k-1} \tag{3.76}$$

όπου

$$G_{rad}^k = \frac{9\pi^2 A_{0N}^k}{128 \rho_0 c} \tag{3.77}$$

$$S_{rad}^k = \frac{3\pi \sqrt{\pi A_{0N}^k}}{8\rho_0}. \tag{3.78}$$

Οπότε για το σχήμα του Portnoff μένει ο προσδιορισμός άλλης μίας εξίσωσης που θα προέλθει από την οριακή συνθήκη στη γλωττίδα.

Για το σχήμα του Maeda παίρνουμε την πίεση στα χείλη:

$$p_{lips}^k = b_{N+1}^k (U_{N+1}^k - U_{N+1}^{k-1} + [G_{rad}^{k-1} - \frac{\Delta t}{2} S_{rad}^{k-1}] p_{lips}^{k-1}) \tag{3.79}$$

με

$$b_{N+1}^k = \frac{1}{G_{rad}^k + \frac{\Delta t}{2} S_{rad}^k} \tag{3.80}$$

και παρόμοια άλλη μία εξίσωση για το σύστημα προς επίλυση (εξίσωση κίνησης στα χείλη):

$$\underbrace{(Z_N^k + b_{N+1}^k + b_N^k)}_{H_{N,N+1}^k} U_{N+1}^k - b_N^k U_N^k = \underbrace{(Z_N^{k-1} + b_{N+1}^k) U_{N+1}^{k-1} + p_{N+\frac{1}{2}}^{k-1} - (1 + b_{N+1}^k [G_{rad} - \frac{\Delta t}{2} S_{rad}]) p_{lips}^{k-1} - b_N^k \Phi_N^{k,k-1}}_{F_{N,N+1}^{k,k-1}}. \tag{3.81}$$

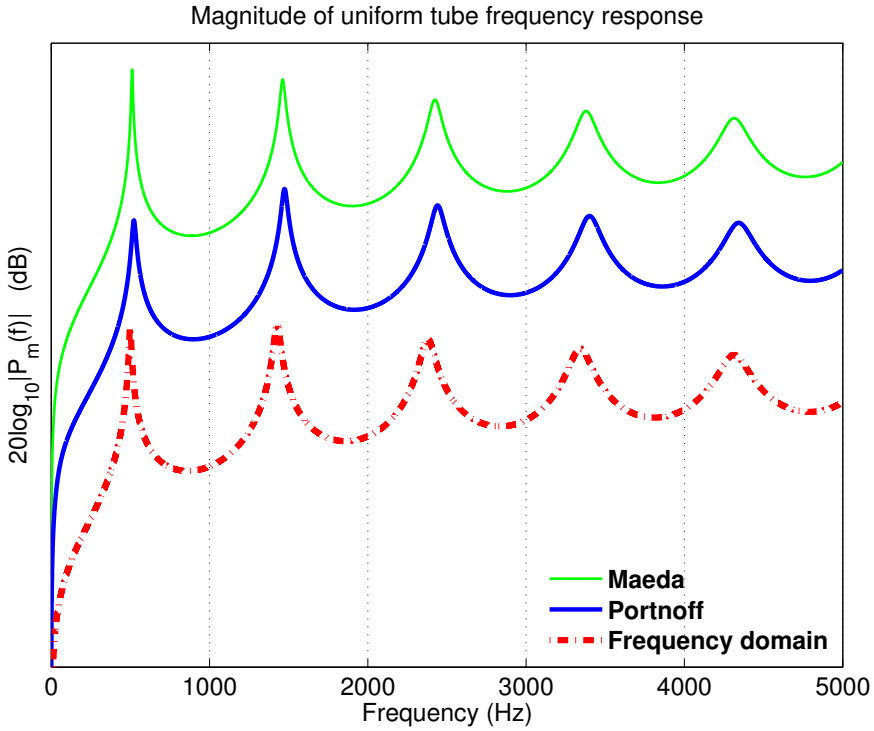
3.4.5 Απόκριση συχνότητας ομοιόμορφου σωλήνα

Πριν την παρουσίαση του συνθέτη φωνής που επίσης περιλαμβάνει ένα μοντέλο για τη γλωττίδα και συζευγμένες ακουστικές κοιλότητες είναι σκόπιμο να γίνει μια αξιολόγηση των επιμέρους προσεγγίσεων που έχουν περιγραφτεί ώστε να δικαιολογηθούν οι επιλογές που έγιναν στη συνέχεια. Για το σκοπό αυτό, υπολογίζεται η απόκριση συχνότητας ενός σωλήνα ομοιόμορφης εγκάρσιας διατομής σε διάφορες περιπτώσεις. Χρησιμοποιείται η ίδια χωρική και χρονική συχνότητα διακριτοποίησης για τα δύο σχήματα. Συγκεκριμένα, η χρονική συχνότητα διακριτοποίησης είναι ίση με $F_{sim} = 48$ kHz ενώ είναι $X_n = \Delta x = 0.001$ m. Το μήκος του σωλήνα είναι ίσο με $l = 17.5$ cm ενώ η ταχύτητα του ήχου λαμβάνεται ίση με $c = 350$ m/s. Γίνεται σύγκριση με την απόκριση συχνότητας όπως αυτή υπολογίζεται με προσομοίωση στο πεδίο της συχνότητας.

Πρακτικά, το σύστημα διεγείρεται από έναν μοναδιαίο διακριτό παλμό ογκικής ταχύτητας ενώ η σύνθετη αντίσταση στη γλωττίδα θεωρείται άπειρη, που σημαίνει ότι η εναπομείνουσα οριακή συνθήκη στο αριστερό άκρο της φωνητικής οδού λαμβάνεται ως:

$$U_g^k = \delta[k]. \tag{3.82}$$

Το πλάτος της απόκρισης συχνότητας λαμβάνεται ως το πλάτος του διακριτού μετασχηματισμού Fourier της πίεσης στα χείλη. Το εύρος συχνοτήτων που μας ενδιαφέρει είναι μέχρι $F_{max} = 5$ kHz ώστε να θεωρείται αποδεκτή η προσέγγιση διάδοσης επίπεδου ακουστικού



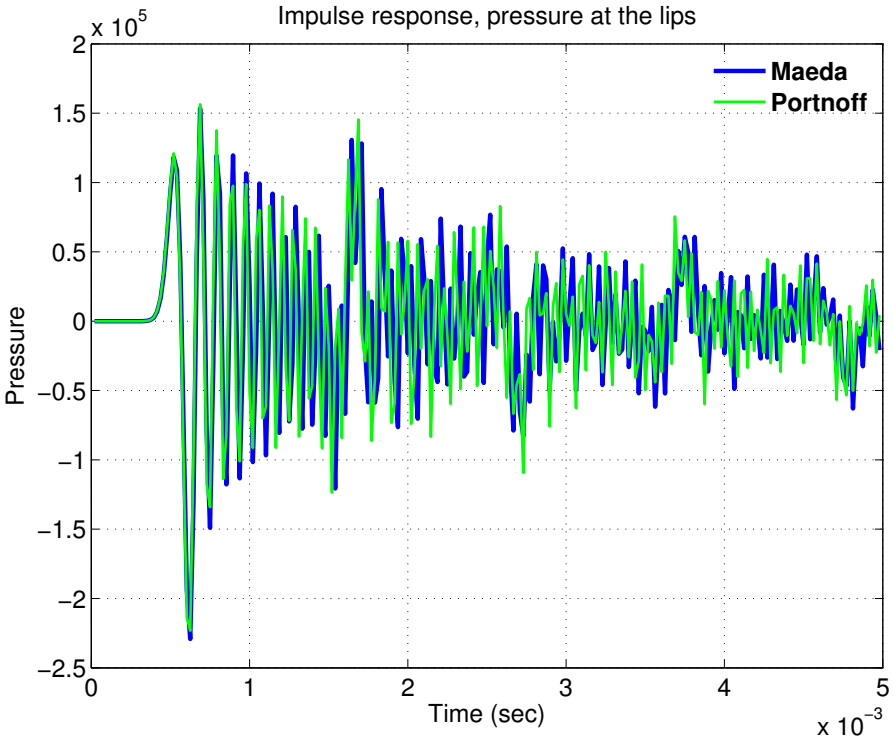
Σχήμα 3.4: *Πλάτος της απόκρισης συχνότητας του ομοιόμορφου σωλήνα για τα τρία διαφορετικά σχήματα προσομοίωσης, όπως έχουν παρουσιαστεί : σχήμα Maeda, σχήμα Portnoff και προσομοίωση στο πεδίο της συχνότητας. Τα τρία γραφήματα έχουν διαχωριστεί τεχνητά ως προς τον άξονα της τεταγμένης, για λόγους οπτικοποίησης.*

κύματος. Προσομοιώνεται η απόκριση για χρόνο $t = 1$ sec ώστε το σήμα στην έξοδο να έχει στην ουσία αποσβεστεί στο πέρας της προσομοίωσης [4]. Για τη μοντελοποίηση των τοιχωμάτων στην προσομοίωσή μας επιλέξαμε τις τιμές που μετρήθηκαν στο [74] για το χαλαρό μάγουλο, δηλαδή $M_w = 21kg/m^2$, $b_w = 8000kg/m^2$, $K_w = 845000kg/m^2s^2$.

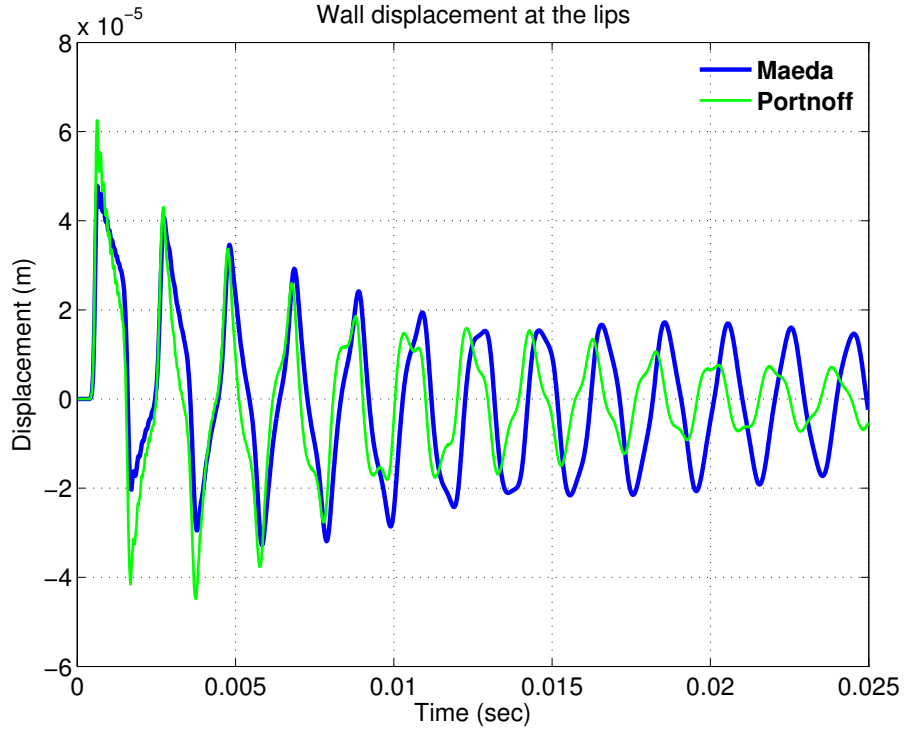
Το σχήμα του Portnoff με βάση την ανάλυση στο [127] είναι αλάνθαστο για την περίπτωση που ο σωλήνας δεν έχει απώλειες και $\Delta x = c\Delta t$. Αντίθετα, το σχήμα του Maeda αναμένεται να μεταβάλλει το συχνοτικό περιεχόμενο της εξόδου για υψηλές συχνότητες λόγω αναδίπλωσης αν η διακριτοποίηση δεν είναι αρκετά λεπτομερής [101]. Στην περίπτωση μας, όπως φαίνεται και στο Σχ. 3.4 οι διαφορές στην προκύπτουσα απόκριση συχνότητας είναι πρακτικά αμελητέες δεδομένου του ότι η χωρική ανάλυση είναι αρκετά μεγάλη. Στο Σχ. 3.5 δίνεται ένα τμήμα της κρουστικής απόκρισης υπολογισμένο με τα δύο διαφορετικά σχήματα, ενώ στο Σχ. 3.6 δίνεται η απόκλιση του τοιχώματος του σωλήνα όπως αυτή υπολογίζεται στην άκρη του. Σημειώνεται ότι η μέθοδος διακριτοποίησης του Maeda για τη μετατόπιση των τοιχωμάτων εμφάνισε αστάθειες σε κάποιες περιπτώσεις και γι' αυτό, όπως αναφέρθηκε και προηγουμένως, εγκαταλείφθηκε. Για τη συνέχεια, για την προσομοίωση του ακουστικού πεδίου υιοθετείται το αριθμητικό σχήμα του Maeda που έχει το πλεονέκτημα να είναι διαισθητικά πλησιέστερα στο καθιερωμένο ισοδύναμο ηλεκτρικό κύκλωμα για τη φωνητική οδό και επιπλέον να είναι σημαντικά πιο αποδοτικό [5] δεδομένου του ότι περιλαμβάνει την επίλυση ενός συστήματος που έχει το μισό μέγεθος από ότι το σύστημα που προκύπτει από το σχήμα του Portnoff.

[4] Σημειώνεται ότι τα μεγέθη λαμβάνονται στην πράξη σε μονάδες του διεθνούς συστήματος μέτρησης S.I. οπότε ένας μοναδιαίος παλμός ογκικής ταχύτητας αντιστοιχεί σε 1 m^3/sec που είναι αφύσικα μεγάλη τιμή για την ογκική ταχύτητα στη φωνητική οδό που συνήθως λαμβάνει τιμές μικρότερες του 0.01 m^3/sec [10].

[5] Στα πειράματα μας, το σχήμα του Maeda ήταν τουλάχιστον 5 φορές αποδοτικότερο, με βάση τη συνολική διάρκεια προσομοίωσης.



Σχήμα 3.5: *Κρουστική απόκριση του ομοιόμορφου σωλήνα για τα δύο διαφορετικά σχήματα προσομοίωσης στο χρόνο : σχήμα Maeda (σκούρα γραμμή) και σχήμα Portnoff (ανοιχτόχρωμη γραμμή). Η απόκριση δίνεται για τα πρώτα 5 ms.*



Σχήμα 3.6: *Μετατόπιση των τοιχωμάτων ομοιόμορφου σωλήνα κατά τον υπολογισμό της κρουστικής απόκρισης. Δίνονται τα αποτελέσματα από δύο αριθμητικά σχήματα : του Portnoff που διακριτοποιεί την εξίσωση κίνησης των τοιχωμάτων με τη μέθοδο της αμετάβλητης κρουστικής απόκρισης και του Maeda που χρησιμοποιεί του κανόνα του τραπεζίου.*

	Σχετικό λάθος υπολογισμού (%)					
	Στο χρόνο			Στη συχνότητα		
	F_1	F_2	F_3	F_1	F_2	F_3
i	-3.28	-2.47	-9.96	-5.99	-3.62	0.21
e	-3.51	-3.50	-5.58	-8.22	-5.98	-8.82
a	-14.52	-6.74	-3.35	-16.48	-6.39	-0.99
o	-13.60	-6.68	-2.54	-11.08	0.03	-0.21
u	-8.95	-3.44	-1.42	-2.01	0.41	-0.00
i_	-4.89	-7.32	-5.27	-7.71	-6.94	-5.69
Μέση τιμή	-8.12	-5.02	-4.69	-8.58	-3.75	-2.59

Πίνακας 3.3: *Σχετικά λάθη στον υπολογισμό των συχνοτήτων συντονισμού για τις συναρτήσεις εμβαδού του [52] με την προσομοίωση του ακουστικού πεδίου στο χρόνο και στη συχνότητα.*

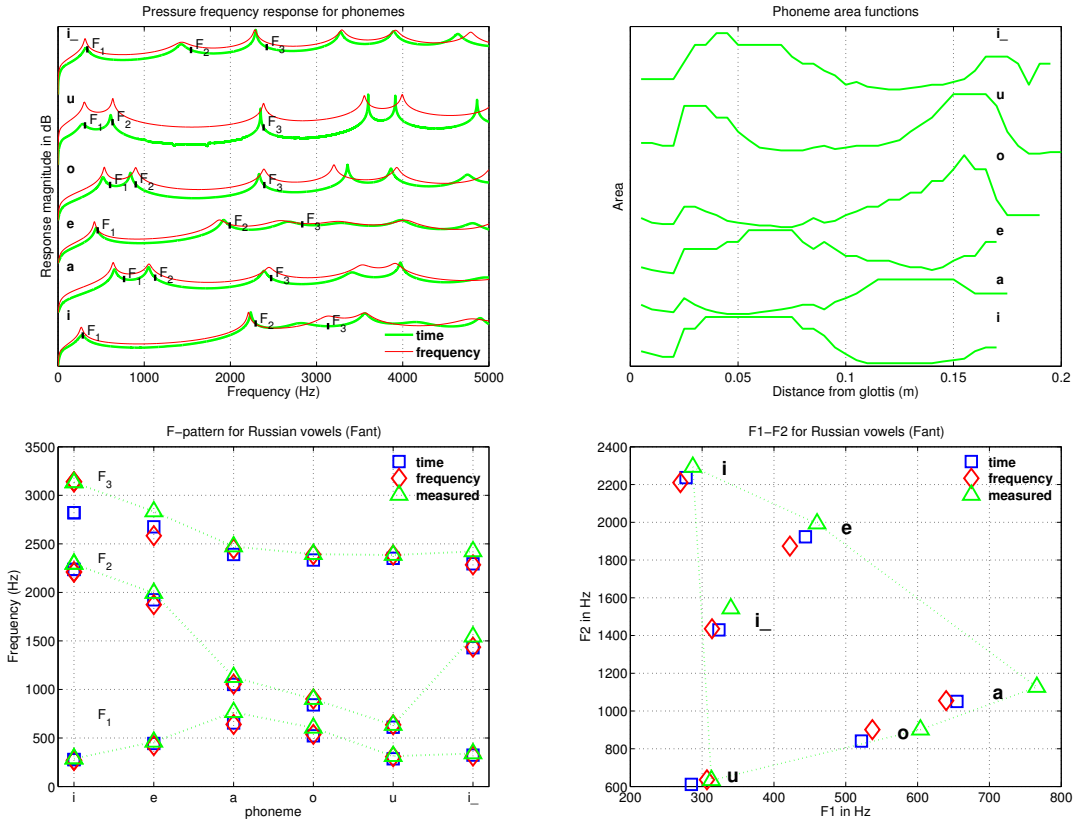
3.5 Απόκριση συχνότητας φωνητικής οδού με βάση πραγματικά δεδομένα

Έχει ενδιαφέρον να μελετηθεί η συμπεριφορά της αριθμητικής προσομοίωσης για την περίπτωση πραγματικών δεδομένων. Για το σκοπό αυτό υπολογίστηκε η απόκριση συχνότητας της φωνητικής οδού με συνάρτηση εμβαδού που να αντιστοιχεί στην εκφώνηση πραγματικών φωνημάτων. Δεδομένα αυτής της μορφής έχουν δημοσιευτεί για τρεις ομιλητές, μια γυναίκα [160] και δύο άντρες [52, 159]. Τα δεδομένα που δίνονται στα [159, 160] έχουν υπολογιστεί με χρήση ογκομετρικών εικόνων όπως έχουν προκύψει από μαγνητικές τομογραφίες της φωνητικής οδού για την εκφώνηση 10 διαφορετικών φωνηέντων ενώ τα δεδομένα στο [52] είναι για 7 φωνηέντα και έχουν εκτιμηθεί περισσότερο ευριστικά. Έχουν επίσης μετρηθεί και δημοσιευτεί οι συχνότητες συντονισμών των πραγματικών σημάτων φωνής που υποτίθεται ότι αντιστοιχούν στις συναρτήσεις επιφανείας. Η καταγραφή του σήματος φωνής για την περίπτωση των μαγνητικών τομογραφιών δε γίνεται παράλληλα με τις τομογραφίες αφού το υψηλό επίπεδο θορύβου στον χώρο καταγραφής κάνει κάτι τέτοιο πρακτικά δύσκολο και θα απαιτούσε την εφαρμογή μιας μεθόδου αποθορυβοποίησης [28]. Αντίθετα, η καταγραφή γίνεται εκ των υστέρων και γίνεται προσπάθεια να προσομοιωθούν οι ίδιες συνθήκες εκφώνησης, π.χ. ο ομιλητής είναι ξαπλωμένος [160].

Ακολουθώντας τη μέθοδο που περιγράφεται στην Ενότητα 3.4.5 προσομοιώνουμε την απόκριση συχνότητας χρησιμοποιώντας τα πραγματικά δεδομένα της συνάρτησης εμβαδού και υπολογίζουμε τις αντίστοιχες συχνότητες συντονισμού και τα εύρη ζώνης. Η προσομοίωση γίνεται στο χρόνο και στη συχνότητα και λαμβάνονται υπόψη απώλειες λόγω φορτίου εκπομπής, δονούμενων τοιχωμάτων και συνεκτικότητας. Οι συχνότητες συντονισμού στην έξοδο υπολογίζονται με τη χρήση αλγορίθμου επιλογής κορυφών χρησιμοποιώντας κάποια πρότερη γνώση σχετικά με τη ζώνη συχνοτήτων στην οποία μπορεί να βρίσκεται ο κάθε συντονισμός. Τα αποτελέσματα της προσομοίωσης στο χρόνο που παρουσιάζονται στα Σχήματα 3.7, 3.8, 3.9 εμφανίζονται καλύτερα από αυτά της προσομοίωσης στη συχνότητα και δείχνουν αρκετά μεγάλη συμφωνία με τους συντονισμούς που έχουν μετρηθεί από τα πραγματικά σήματα φωνής (απόκλιση της τάξης του 5%). Στους Πίνακες 3.3, 3.4, 3.5 δίνονται λεπτομερώς τα σχετικά λάθη στον υπολογισμό των συντονισμών. Τα σύμβολα των φωνηέντων που προσομοιώνονται εξηγούνται στον Πίνακα 3.6.

3.6 Μοντέλο δύο μαζών, οριακή συνθήκη στη γλωττίδα

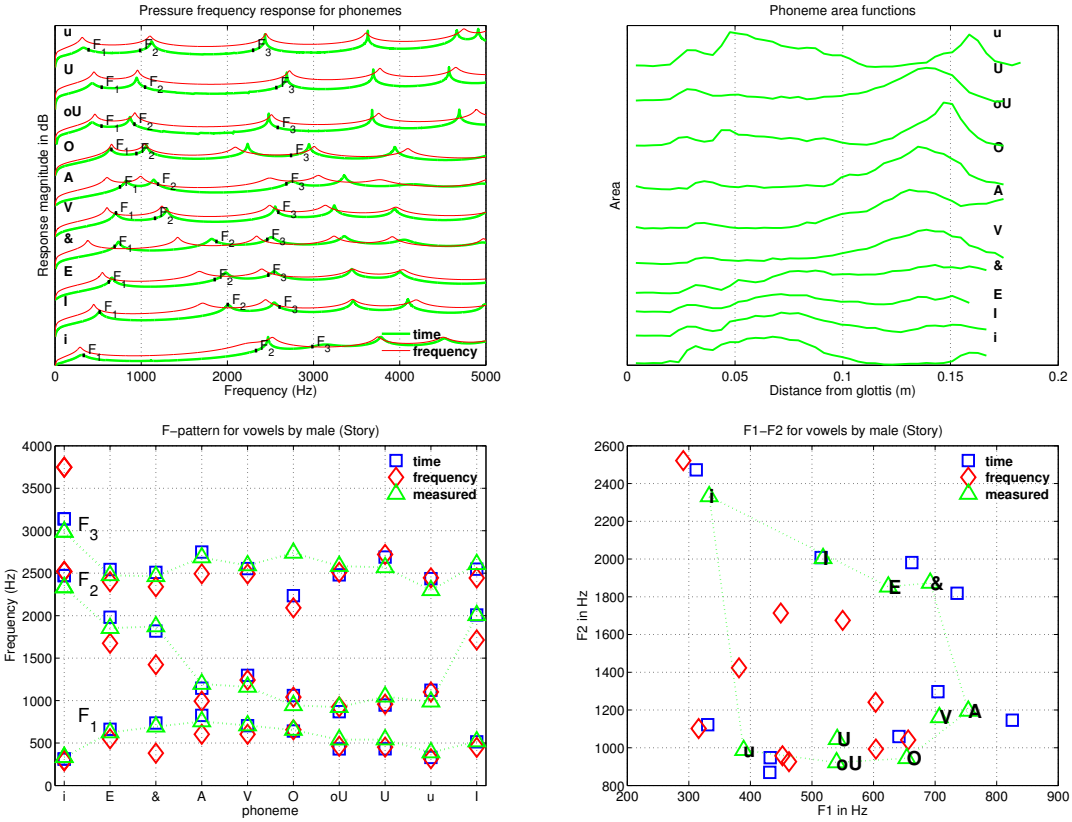
Για οριακή συνθήκη στη γλωττίδα από τον Portnoff χρησιμοποιείται το μοντέλο των δύο μαζών των Ishizaka-Flanagan [73] για τη γλωττίδα που φαίνεται στο Σχήμα 3.10. Υποτίθεται ότι οι φωνητικές χορδές είναι διπλευρικά συμμετρικές. Συζητώνται γι' αυτό οι ιδιότητες της



Σχήμα 3.7: *Πειράματα προσομοίωσης χρησιμοποιώντας πραγματικά δεδομένα, όπως έχουν δημοσιευτεί στο [52] για 6 ρώσικα φωνήεντα. Απόκριση συχνότητας (πλάτος) όπως υπολογίστηκε με προσομοίωση στο χρόνο και στη συχνότητα. Χρησιμοποιήθηκαν οι συναρτήσεις εμβαδού που δίνονται στο γράφημα.*

	Σχετικό λάθος υπολογισμού (%)					
	Στο χρόνο			Στη συχνότητα		
	F_1	F_2	F_3	F_1	F_2	F_3
i	-6.30	6.02	5.17	-12.61	8.13	25.51
E	6.11	6.92	2.78	-11.88	-9.60	-3.06
&	6.32	-2.90	1.91	-44.88	-24.01	-5.01
A	9.47	-4.11	2.38	-19.92	-16.89	-7.16
V	-0.34	11.76	-1.42	-14.65	6.90	-3.86
O	-1.95	12.27	-18.40	0.35	10.29	-23.60
oU	-20.11	-5.67	-4.03	-14.31	0.48	-2.83
U	-20.12	-9.34	4.73	-16.43	-8.38	5.94
u	-14.99	13.68	5.96	-18.79	11.61	6.39
I	-0.60	0.20	-2.21	-13.26	-14.47	-6.20
Μέση τιμή	-4.25	2.88	-0.31	-16.64	-3.59	-1.39

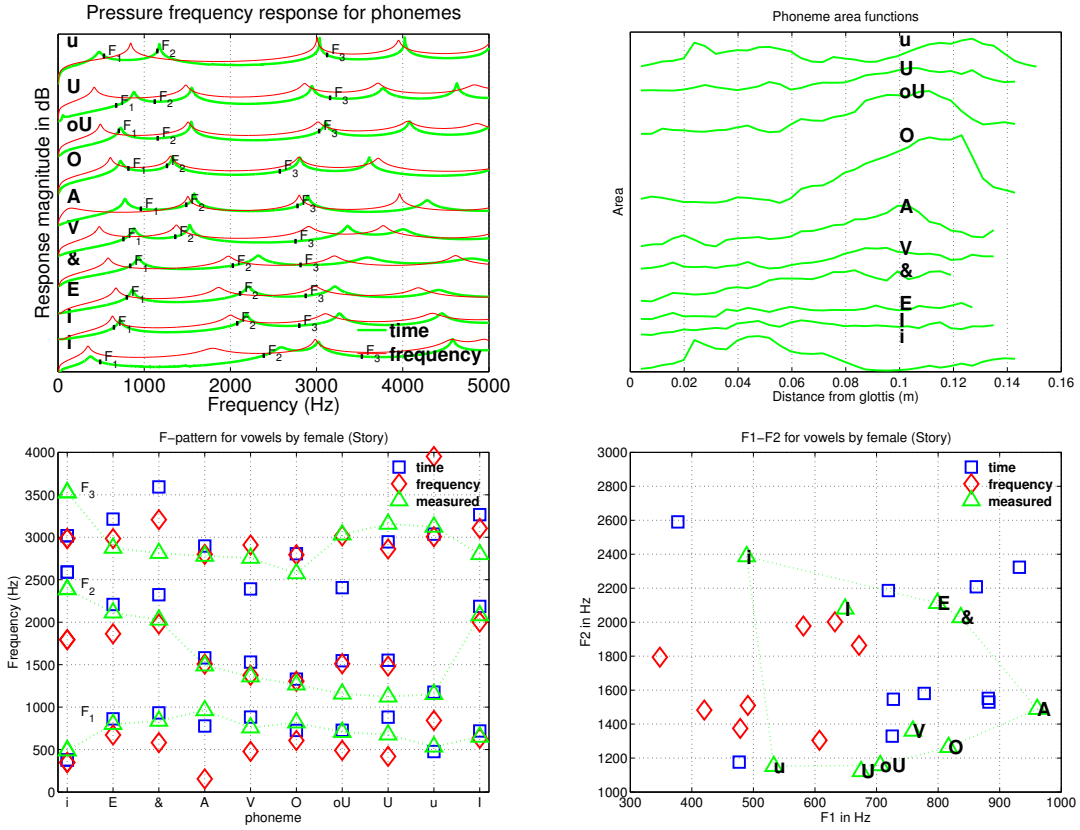
Πίνακας 3.4: *Σχετικά λάθη στον υπολογισμό των συχνοτήτων συντονισμού για τις συναρτήσεις εμβαδού του [159] με την προσομοίωση του ακουστικού πεδίου στο χρόνο και στη συχνότητα.*



Σχήμα 3.8: *Πειράματα προσομοίωσης χρησιμοποιώντας πραγματικά δεδομένα, όπως έχουν δημοσιευτεί στο [159] για 10 φωνήεντα της αμερικάνικης γλώσσας από άνδρα ομιλητή. Απόκριση συχνότητας (πλάτος) όπως υπολογίστηκε με προσομοίωση στο χρόνο και στη συχνότητα. Χρησιμοποιήθηκαν οι συναρτήσεις εμβαδού που δίνονται στο γράφημα.*

	Σχετικό λάθος υπολογισμού (%)					
	Στο χρόνο			Στη συχνότητα		
	F_1	F_2	F_3	F_1	F_2	F_3
i	-22.86	8.53	-14.40	-28.83	-24.86	-15.35
E	7.89	4.56	11.81	-15.94	-11.76	3.81
&	11.35	14.58	27.65	-30.56	-2.50	13.99
A	-19.10	6.22	4.22	-83.92	1.27	0.69
V	16.28	12.50	-13.23	-36.94	1.07	5.55
O	-11.20	5.14	9.02	-25.67	3.22	8.55
oU	3.02	33.67	-20.58	-30.45	30.51	-0.46
U	30.59	38.29	-6.68	-37.75	32.15	-9.34
u	-10.54	2.11	-2.72	58.18	160.90	26.55
I	10.82	5.16	16.68	-2.56	-3.72	10.92
Μέση τιμή	1.62	13.08	1.18	-23.45	18.63	4.49

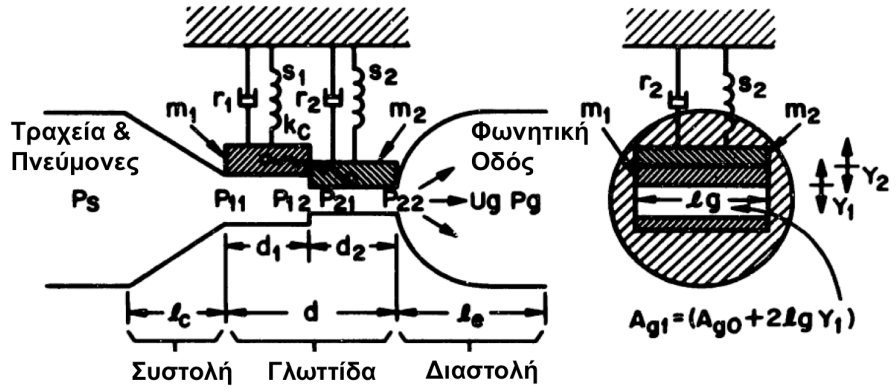
Πίνακας 3.5: *Σχετικά λάθη στον υπολογισμό των συχνοτήτων συντονισμού για τις συναρτήσεις εμβαδού του [160] με την προσομοίωση του ακουστικού πεδίου στο χρόνο και στη συχνότητα.*



Σχήμα 3.9: *Πειράματα προσομοίωσης χρησιμοποιώντας πραγματικά δεδομένα, όπως έχουν δημοσιευτεί στο [160] για 10 φωνήεντα της αμερικάνικης γλώσσας από γυναίκα ομιλήτρια. Απόκριση συχνότητας (πλάτος) όπως υπολογίστηκε με προσομοίωση στο χρόνο και στη συχνότητα. Χρησιμοποιήθηκαν οι συναρτήσεις εμβαδού που δίνονται στο γράφημα.*

Φώνημα	Περιβάλλον
i	beet
I	bit
E	bet
&	at
V	but
A	cot
O	paw, cause
U	book
u	who, boot

Πίνακας 3.6: *Φωνήματα της αμερικάνικης γλώσσας [160] που προσομοιώθηκαν στο πεδίο του χρόνου και στο πεδίο της συχνότητας.*



Σχήμα 3.10: *Μουτέλο δύο μαζών για τη γλωττίδα όπως προτείνεται στο [73]. Αριστερά είναι η μεσοοβελιαία και δεξιά η εγκάρσια όψη του.*

μιας μόνο χορδής και τα ίδια ισχύουν και για την άλλη. Ένα σχηματικό διάγραμμα του συστήματος της γλωττίδας φαίνεται στο Σχ. 3.10. Η τραχεία που οδηγεί στους πνεύμονες αναπαρίσταται με το σωλήνα στα αριστερά. Ο λάρυγγας, που οδηγεί στο φωνητικό σωλήνα, είναι στα δεξιά. Αυτοί οι σωλήνες θεωρούνται κυλινδρικού σχήματος και σταθερού μεγέθους. Η γλωττίδα είναι μια στένωση ανάμεσά τους και το μέγεθός της εξαρτάται από τη μετατόπιση των χορδών. Η είσοδος στη γλωττιδική στένωση λαμβάνει χώρα κατά μήκος της απόστασης συστολής l_c. Το άνοιγμα προς τη φωνητική οδό έχει μήκος l_e. Στο Σχ. 3.10 φαίνονται και οι αντίστοιχες αεροδυναμικές πιέσεις $P_{11}, P_{12}, P_{21}, P_{22}$ σε κάθε τμήμα του μοντέλου της γλωττίδας.

Στο μοντέλο των δύο μαζών η φωνητική χορδή χωρίζεται κατά μήκος (πάχος) σε πάνω και κάτω μέρος . Κάθε μέρος περιλαμβάνει έναν απλό μηχανικό ταλαντωτή με μάζα ελατήριο και απόσβεση (m, s και r). Οι δύο μάζες μιας χορδής m_1 και m_2 επιτρέπεται να κινηθούν μόνο εγκάρσια κατά x_1 και x_2 και είναι συζευγμένες με ένα γραμμικό ελατήριο συντελεστή δυσκαμψίας k_c όπως φαίνεται στο Σχήμα. Άλλα μεγέθη που εμφανίζονται είναι : l_g το ουσιαστικό μήκος των χορδών ή της γλωττιδικής σχισμής, d_1, d_2 τα πάχη των μαζών m_1, m_2 αντίστοιχα s_1, s_2 τα ισοδύναμα ελατήρια, r_1, r_2 οι ισοδύναμες αποσβέσεις, A_{g01}, A_{g02} τα εμβαδά των εγκάρσιων επιφανειών διατομής της γλωττιδικής σχισμής όταν οι δύο μάζες είναι σε ηρεμία και U_g η μέση ογκική ταχύτητα κατά μήκος της επιφάνειας της γλωττίδας.

Λόγω της υπόθεσης συμμετρίας, οι μεταβολές στις εγκάρσιες επιφάνειες λόγω των μετατοπίσεων x_1, x_2 διπλασιάζονται για να δώσουν τη συνολική μεταβολή της γλωττιδικής επιφάνειας, δηλαδή ισχύει:

$$A_{g1} = A_{g01} + 2l_g x_1 \tag{3.83}$$

$$A_{g2} = A_{g02} + 2l_g x_2 \tag{3.84}$$

Λόγω των μικρών διαστάσεων της γλωττίδας (σε σύγκριση με ένα μήκος κύματος στις συχνότητες που μας ενδιαφέρουν) και λόγω της υψηλής ταχύτητας της γλωττιδικής ροής σε σύγκριση με την ταχύτητα των φωνητικών χορδών, μπορούμε να υποθέσουμε ότι η γλωττιδική ροή είναι σχεδόν σταθερή. Χρησιμοποιείται η εξίσωση Bernoulli για μονοδιάστατη ροή ώστε να εκτιμηθεί η κατανομή της πίεσης κατά μήκος της ροής. Το απότομο στένεμα στην είσοδο της γλωττίδας προκαλεί μια vena contracta που περιβάλλεται από τελματωμένο αέρα. Αυτό το φαινόμενο κάνει την επιφάνεια εισόδου A_{g1} να εμφανίζεται μικρότερη από την πραγματικότητα και η πτώση πίεσης είναι μεγαλύτερη από αυτή που υπαγορεύεται από μια ιδανική μεταβολή της επιφάνειας διατομής. Το φαινόμενο αυτό θεωρήθηκε αμελητέο σε πιο σύγχρονες μελέτες της γλωττίδας [125] με βάση το επιχείρημα ότι η είσοδος στη

γλωττίδα δεν είναι αρκετά απότομη ώστε να έχουμε αποκόλληση της ροής από το τοίχωμα. Στο μοντέλο της γλωττίδας που τελικά παρουσιάζεται στο Κεφάλαιο 4 το φαινόμενο αυτό αμελείται. Στην κλασσική θεώρηση, ο παράγοντας απωλειών για ένα τέτοιο στένεμα έχει μελετηθεί σε πειράματα ρευστοδυναμικής και έχει βρεθεί ότι είναι μεταξύ του 0.4 και του 0.5. Στην πράξη θεωρείται κατ' αρχάς η τιμή που δίνεται στο [172] και είναι ίση με 0.37. Οπότε η πτώση πίεσης στην είσοδο λαμβάνεται ως

$$P_{B1}(1+0.37) \text{ ή } 0.69\rho(U_{g1}^2/A_{g1}^2)$$

όπου $P_{B1} = \frac{1}{2}\rho u_{g1}^2$ είναι η πίεση Bernoulli, ρ είναι η πυκνότητα του αέρα και U_{g1} η σωματιδιακή ταχύτητα στη χαμηλότερη άκρη των χορδών.

Μέσα στη στένωση που σχηματίζεται από το χαμηλότερο τμήμα της χορδής η πτώση πίεσης θεωρείται ότι κυρίως εξουσιάζεται από απώλειες λόγω συνεκτικότητας, που είναι επίσης συμβατό με τις παρατηρήσεις στο [172]. Σε αυτή την περιοχή η πίεση πέφτει γραμμικά με την απόσταση σύμφωνα με μια αντίσταση στην ογκική ροή που ισούται με $12\mu d_1 l_g^2/A_{g1}^3$, όπου μ είναι ο συντελεστής διατμητικής συνεκτικότητας.

Στη σημείο συνένωσης μεταξύ των μαζών m_1, m_2, η ογκική ταχύτητα της ροής U_g είναι συνεχής αλλά η σωματιδιακή ταχύτητα αλλάζει. Εμφανίζεται μια αντίστοιχη απότομη αλλαγή στην πίεση που ισούται με την αλλαγή κινητικής ενέργειας ανά μονάδα όγκου του ρευστού. Αυτή η αλλαγή στην πίεση στον σύνδεσμο μεταξύ των μαζών είναι :

$$\Delta p = 1/2\rho(U_{g1}^2 - U_{g2}^2) = 1/2\rho_0 U_g^2(1/A_{g2}^2 - 1/A_{g1}^2).$$

Κατά μήκος της στένωσης που δημιουργείται στη ανώτερη άκρη των χορδών, δηλαδή στη μάζα m_2, θεωρείται πάλι ότι επικρατούν οι απώλειες λόγω συνεκτικότητας και όπως στο χαμηλότερο τμήμα, η αντίσταση λαμβάνεται ως $12\mu d_2 l_g^2/A_{g2}^3$. Στο απότομο άνοιγμα στην έξοδο της γλωττίδας, η πίεση ανακάμπτει προς την ατμοσφαιρική (υποθέτοντας έλλειψη στενώσεων στη σχετικά μεγάλη φωνητική οδό). Η εκτίμηση του ποσού της ανάκαμψης βασίζεται σε θεωρήσεις που αφορούν στην ορμή της ροής, όπως υπαγορεύονται από τη θεωρία ρευστών [73]. Σε πιο σύγχρονες μελέτες της γλωττίδας η ανάκαμψη αυτή της πίεσης θεωρείται αμελητέα δεδομένου του ότι η εγκάρσια επιφάνεια εισόδου της φωνητικής οδού είναι σημαντικά μεγαλύτερη του ανοίγματος της γλωττίδας.

Στη βάση αυτής της ανάλυσης για την κατανομή της πίεσης, τα στοιχεία της ακουστικής σύνθετης αντίστασης της γλωττιδικής σχισμής σχηματίζουν ένα ισοδύναμο ηλεκτρικό κύκλωμα που όπου η ροή U_g μπορεί να θεωρηθεί ότι αντιστοιχεί σε συνεχές ρεύμα. Τα στοιχεία του κυκλώματος συγκεκριμένα δίνονται ως :

$$R_c = 1.37\frac{\rho_0}{2}\frac{|U_g|}{A_{g1}^2}, L_c = \int_0^{l_c}\frac{dx}{A_c(x)}$$

$$R_{v1} = 12\frac{\mu l_g^2 d_1}{A_{g1}^3}, L_{g1} = \frac{\rho_0 d_1}{A_{g1}}$$

$$R_{12} = \frac{\rho_0}{2}(\frac{1}{A_{g2}^2} - \frac{1}{A_{g1}^2})|U_g|$$

$$R_{v2} = 12\frac{\mu l_g^2 d_2}{A_{g1}^3}, L_{g1} = \frac{\rho_0 d_2}{A_{g2}}$$

$$R_e = -\frac{\rho_0}{2}\frac{2}{A_{g2}A_1}(1 - \frac{A_{g2}}{A_1}|U_g|).$$

Η οριακή συνθήκη που προκύπτει, ακολουθώντας το συμβολισμό και τη διακριτοποίηση του Maeda όπου $U(x_0,t) = U_g(t)$ είναι η ογκική ταχύτητα στη γλωττίδα, είναι :

$$R_m(t)|U(x_0,t)|U(x_0,t) + R_v(t)U(x_0,t) + L_g(t)\frac{dU(x_1,t)}{dt} + L_1(t)\frac{dU(x_1,t)}{dt} + R_1(t)U(x_0,t) + p(x_1 - X_1/2) - p_{sub}(t) = 0 \quad (3.85)$$

όπου $R_m(t) = R_c(t) + R_{12}(t) + R_e(t)$, $R_g(t) = R_{v1}(t) + R_{v2}(t)$, $L_g(t) = L_{g1}(t) + L_{g2}(t)$ και $p_{sub}(t)$ είναι η υπογλωττιδική πίεση που δίνεται ως είσοδος για την προσομοίωση. Διακριτοποιώντας χρονικά με βάση τον κανόνα του τραπεζίου παίρνουμε:

$$R_m^k|U_1^k|U_1^k + (Z_g^k + Z_1^k + R_g^k + R_1^k)U_1^k + p_{1+\frac{1}{2}}^k = - R_m^{k-1}|U_1^{k-1}|U_1^{k-1} + (Z_g^{k-1} + Z_1^{k-1} - R_g^{k-1} - R_1^{k-1})U_1^{k-1} - p_{1+\frac{1}{2}}^{k-1} + p_{sub}^k + p_{sub}^{k-1}, \quad (3.86)$$

με

$$Z_g^k = \frac{2}{\Delta t}\rho_0(\frac{d_1}{A_{g1}^k} + \frac{d_2}{A_{g2}^k}), \quad (3.87)$$

$$R_g^k = 12\mu l_g^2(\frac{d_1}{(A_{g1}^k)^3} + \frac{d_2}{(A_{g2}^k)^3}) \quad (3.88)$$

$$R_m^k = \underbrace{\frac{0.19\rho}{(A_{g_1}^k)^2}}_{R_{k_1}} + \underbrace{\frac{\rho[0.5 - \frac{A_{g2}}{A_1^k}(1 - \frac{A_{g2}}{A_1^k})]}{A_{g_2}^2}}_{R_{k_2}}, \quad (3.89)$$

όπου το μ αναπαριστά το συντελεστή ιξώδους του αέρα. Οι επιφάνειες A_{g1}, A_{g2} δίνονται από το μηχανικό μοντέλο για τη γλωττίδα κάθε χρονική στιγμή. Ο Maeda ακολουθεί μια απλούστερη προσέγγιση και προσδιορίζει την έμφωνη διέγερση μέσω του εμβαδού διατομής της γλωττίδας για μια θεμελιώδη περίοδο με βάση το παραμετρικό μοντέλο:

$$A_g(t; A_p, t_0) = \begin{cases} 0.5A_p(1 - \cos(\pi t/t_1) & \text{για } 0 < t \le t_1, \\ A_p(K\cos(\pi(t - t_1)/t_1) - K + 1) & \text{για } t_1 < t \le t_2, \\ 0 & \text{για } t_2 < t \le t_0. \end{cases} \quad (3.90)$$

όπου A_p, t_0 είναι παράμετροι που καθορίζουν την κυματομορφή ενώ $K = 1/(1 - \cos(\pi t_2/t_1))$. Επιπλέον, στην ουσία αμελεί τη μη γραμμικότητα της Εξ. (3.86) θεωρώντας σχετικά αργή μεταβολή της $U_g(t)$ και λαμβάνει τον όρο του γινομένου που είναι σε απόλυτη τιμή από την προηγούμενη χρονική στιγμή. Αντικαθιστώντας την $p_{1+\frac{1}{2}}^k$ στην Εξ. (3.86) τελικά παίρνουμε και την τελευταία εξίσωση του συστήματος προς επίλυση (με βάση το σχήμα του Maeda):

$$(R_m^k|U_1^{k-1}| + \underbrace{Z_g^k + Z_1^k + R_g^k + R_1^k + b_1^k}_{H_{1,g}^k})U_1^k - b_1^k U_2^k = -R_m^k|U_1|^{k-2}U_1^{k-1} + \underbrace{(Z_g^{k-1} + Z_1^{k-1} - R_g^{k-1} - R_1^{k-1})U_1^{k-1} - p_{1+\frac{1}{2}}^{k-1} + b_1^k\Phi_1^{k,k-1} + p_{sub}^k + p_{sub}^{k-1}}_{F_{1,g}^{k,k-1}}. \quad (3.91)$$

Το τελικό σύστημα για τον Maeda είναι γραμμικό της μορφής $AX = B$ και μπορεί να λυθεί αρκετά αποδοτικά δεδομένου του ότι ο πίνακας A είναι τριδιαγώνιος. Από την άλλη ο Portnoff κρατάει τη μη γραμμικότητα αξιοποιώντας τη μορφή του συστήματος. Το τελικό αριθμητικό σχήμα που υιοθετήσαμε περιλαμβάνει τη μη γραμμική εξίσωση και επιλύεται με έναν ανάλογο τρόπο που περιγράφεται στη Παράρτημα 3.Γ.

3.7 Σύζευξη επιμέρους ακουστικών κοιλοτήτων

Η σύζευξη της κύριας φωνητικής οδού με επιμέρους ακουστικές κοιλότητες όπως είναι η ρινική μπορεί να έχει αξιοσημείωτες επιδράσεις στο τελικό ακουστικό αποτέλεσμα [36, 37]. Στην προσπάθεια να προσεγγιστεί όσο το δυνατόν καλύτερα το φυσικό σύστημα, θεωρήθηκε σημαντική η κατάλληλη τροποποίηση του συνθέτη φωνής ώστε να ενσωματωθούν τελικά η ρινική κοιλότητα και οι κοιλότητες που είναι γνωστές ως piriform fossae. Ακολουθήθηκε το πλαίσιο που προτείνεται στο [114].

3.7.1 Ρινική κοιλότητα

Το ακουστικό πεδίο μέσα στη ρινική κοιλότητα υποτίθεται ότι μπορεί να περιγραφεί κατ' αναλογία με αυτό της κύριας φωνητικής οδού. Μένει να περιγραφεί κατάλληλα η συμπεριφορά του πεδίου στο σημείο ζεύξης. Στο σημείο ζεύξης της φωνητικής οδού με τη ρινική κοιλότητα x_K ισχύει η παρακάτω οριακή συνθήκη:

$$u(x_K^-, t) = u(x_K^+, t) + u_{NT}(x_{(NT)0}, t) \tag{3.92}$$

όπου $u(x_K^-, t), u(x_K^+, t)$ είναι οι ακουστικές ογκικές ταχύτητες αμέσως πριν και αμέσως μετά τη ζεύξη αντίστοιχα ενώ $u_{NT}(x_{(NT)0}, t)$ είναι η ογκική ταχύτητα στην είσοδο της ρινικής κοιλότητας. Τα σημεία $x_K, x_{(NT)0}$ ταυτίζονται και η πίεση που επικρατεί εκεί είναι $p(x_K, t)$. Οι τροποποιημένες εξισώσεις για το σημείο ζεύξης δίνονται στο παράρτημα 3.Α'.

Για την οριακή συνθήκη εκπομπής της ρινικής κοιλότητας ισχύουν τα ίδια με τα χείλια ενώ το τελικό ηχητικό σήμα προκύπτει ως άθροισμα των πιέσεων στην έξοδο τόσο της μύτης όσο και των χειλιών. Στο Σχ. 3.11 φαίνεται το αποτέλεσμα της ζεύξης της ρινικής κοιλότητας με τη φωνητική οδό για τα ρώσικα φωνήεντα /i/ και /u/. Ως συνάρτηση εμβαδού της ρινικής κοιλότητας χρησιμοποιήθηκε αυτή που έχει δημοσιευτεί στο [36] που έχει μήκος ίσο με 10.8 cm. Επιλέχτηκε μικρό ρινοφαρυγγικό άνοιγμα (1.42 cm^2) ώστε να παρατηρείται μικρός βαθμός ρινικότητας. Οι αποκρίσεις συχνότητας των ένρινων φωνηέντων εμφανίζουν διάφορα ζεύγη πόλων-μηδενικών με πιο εμφανές αυτό που είναι στις χαμηλές συχνότητες, κοντά στο 1 kHz.

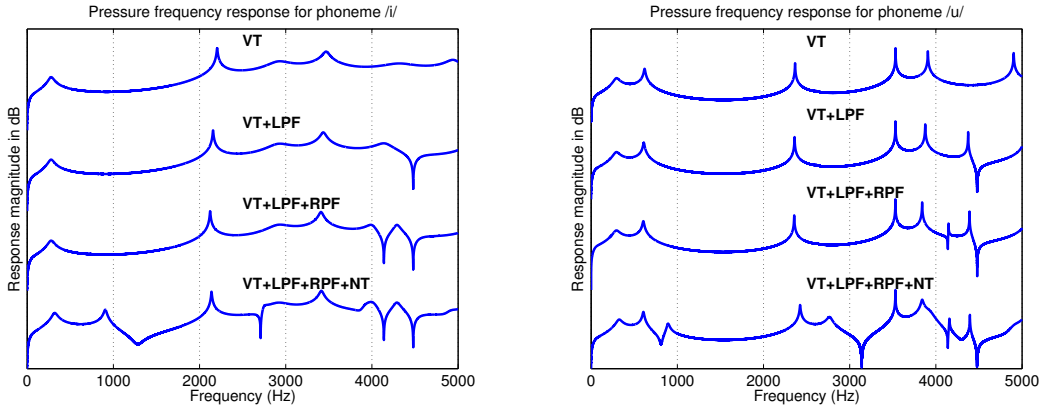
3.7.2 Αχλαδόσχημες κοιλότητες (Piriform fossae)

Ανάλογα τροποποιείται το σύστημα στην περίπτωση που χρειάζεται να συζευχθούν και άλλες κοιλότητες όπως είναι οι piriform fossae [37, 114] που βρίσκονται λίγο πάνω από το λάρυγγα και θεωρούνται υπεύθυνες για την εμφάνιση φασματικών μηδενικών στη φασματική ζώνη 4-5 kHz. Πιο συγκεκριμένα, για τις τελευταίες, η συνθήκη που επικρατεί στο σημείο ζεύξης x_L είναι :

$$u(x_L^-, t) = u(x_L^+, t) + u_{LPF}(x_{(LPF)0}, t) + u_{RPF}(x_{(RPF)0}, t) \tag{3.93}$$

και οι εξισώσεις του συστήματος που τροποποιούνται δίνονται στο παράρτημα 3.Β'. Για τις κοιλότητες που είναι κλειστές από το ένα άκρο [75] και οπότε δεν εκπέμπουν χρησιμοποιείται η ίδια συνθήκη εκπομπής όπως δίνεται από την Εξ. (3.79) με μηδενική αντίστροφη επαγωγή και αντίστροφη αντίσταση όπως προκύπτει για πολύ μικρή (σχεδόν μηδενική) επιφάνεια εκπομπής [114].

Για την προσομοίωση των piriform fossae, δοκιμάσαμε να χρησιμοποιήσουμε τα μετρηθέντα δεδομένα που δίνονται στο [37]. Η ανάγκη όμως της πολύπλοκης εφαρμογής κάποιου συντελεστή διόρθωσης στη ζεύξη λόγω του αχλαδωτού σχήματος των κοιλοτήτων τελικά μας οδήγησε στην επιλογή ενός απλού κυλινδρικού σχήματος για τις κοιλότητες. Πιο συγκεκριμένα θεωρήσαμε ότι η δεξιά και η αριστερή κοιλότητα έχουν ομοιόμορφη διατομή (1 cm^2 και 1.5 cm^2) με μήκη 2.1 cm και 1.9 cm αντίστοιχα. Τα μηδενικά που παρατηρούμε στο φάσμα κατά τη ζεύξη τους είναι αυτά που περίπου αναμένουμε από τη θεωρία (αντιστοιχούν στη συχνότητα $c/(4l)$ όπου l το μήκος της κοιλότητας και c η ταχύτητα του ήχου στην κοιλότητα.



Σχήμα 3.11: *Ακουστικά αποτελέσματα της ζεύξης της κύριας φωνητικής οδού (VT) με επιμέρους ακουστικές κοιλότητες, όπως είναι η ρινική NT και οι λεγόμενες αχλαδόσχημες κοιλότητες (piriform fossae, RPF και LPF, αριστερή και δεξία αντίστοιχα). Δίνονται ενδεικτικά οι αποκρίσεις συχνότητας που αντιστοιχούν σε δύο ρώσικα φωνήεντα, τα /i/ και /u/. Είναι εμφανή τα μηδενικά που εισάγονται στο φάσμα.*

3.8 Σύνθεση ακολουθιών φωνημάτων

Για τη σύνθεση ακολουθιών φωνημάτων και όχι απλά τον υπολογισμό κρουστικών αποκρίσεων είναι απαραίτητο να γίνει η κατάλληλη ζεύξη του μοντέλου της γλωττίδα με την υπόλοιπη φωνητική οδό. Το σύστημα των εξισώσεων που προκύπτει προς επίλυση παρουσιάζει μια ιδιαιτερότητα ως προς το ότι περιλαμβάνει μια μη γραμμική εξίσωση, πολυωνυμική δευτέρου βαθμού. Λόγω των χαρακτηριστικών του συστήματος, η επίλυση γίνεται τελικά δυνατή αποδοτικά με κατάλληλη τριγωνοποίηση όπως περιγράφεται στο παράρτημα 3.Γ. Για επαλήθευση, παρουσιάζεται η σύνθεση μιας ακολουθίας φωνηέντων με δεδομένα που έχουν συγκεντρωθεί μέσω ακτίνων-Χ.

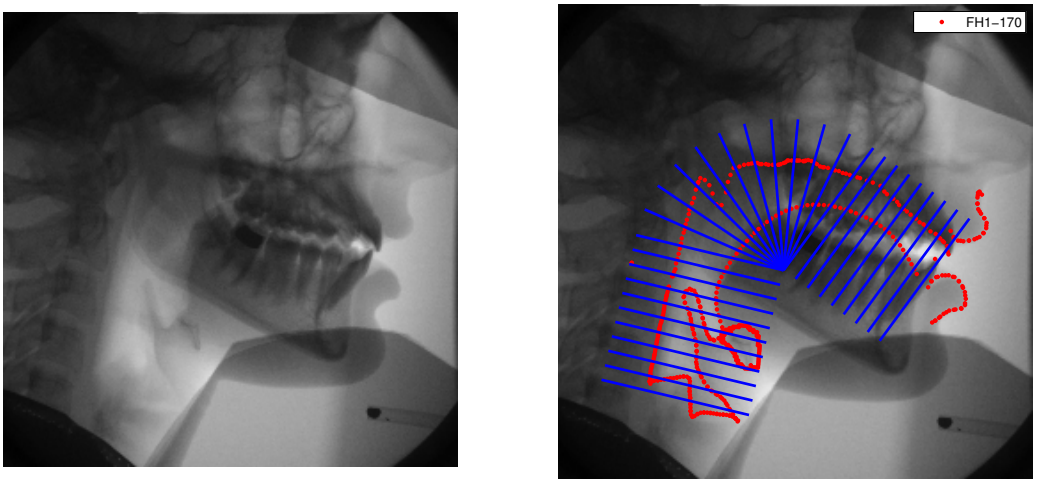
Το προτεινόμενο μοντέλο εφαρμόζεται για τη σύνθεση της ακολουθίας φωνηέντων /iu/ για την οποία υπάρχουν διαθέσιμες εικόνες της φωνητικής οδού ενός ομιλητή όπως έχουν ληφθεί με τη χρήση ακτίνων-Χ (βλ. Σχήμα 3.12(α΄)). Στις εικόνες αυτές έχει περιγραφτεί ημιαυτόματα το σχήμα της φωνητικής οδού στο μέσο οβελιαίο επίπεδο. Για τον υπολογισμό της συνάρτησης επιφανείας χρησιμοποιείται κατάλληλο ημιπολικό πλέγμα με βάση το οποίο υπολογίζεται το άνοιγμα της φωνητικής οδού στο επίπεδο αυτό. Η συνάρτηση επιφανείας προκύπτει μετά την εφαρμογή κατάλληλου μοντέλου.

3.8.1 Ημιπολικό πλέγμα φωνητικής οδού

Το ημιπολικό πλέγμα τοποθετείται στο μεσο-οβελιαίο επίπεδο, όπως φαίνεται στο Σχήμα 3.12(α΄), και βρίσκονται οι συντεταγμένες των σημείων τομής των γραμμών του πλέγματος με τα όρια της φωνητικής οδού, Σχήμα 3.12(β΄). Περιλαμβάνει τρία επιμέρους τμήματα, ένα γραμμικό πλέγμα που αντιστοιχεί στην περιοχή του λάρυγγα και το χαμηλό τμήμα του φάρυγγα, ένα πολικό πλέγμα για την περιοχή του φάρυγγα, της ρινοφαρυγγικής σύζευξης και της οπίσθιας στοματικής κοιλότητας και τέλος ένα γραμμικό πλέγμα για την περιοχή της στοματικής κοιλότητας. Πρακτικά, η περιγραφή του πλέγματος γίνεται με τη βοήθεια 9 παραμέτρων :

z_{orig} Συντεταγμένες του αρχής του πολικού τμήματος. Το συγκεκριμένο σημείο θεωρείται και το σημείο αναφοράς όλου του πλέγματος.

l_l, l_m Μήκη των δύο γραμμικών τμημάτων του πλέγματος. Το ένα αντιστοιχεί στο μήκος από το σημείο αναφοράς ως τη χαμηλότερη γραμμή του κάτω γραμμικού πλέγματος, που ιδανικά βρίσκεται ακριβώς πάνω από τις φωνητικές χορδές στην εικόνα. Το άλλο



(α') (β')

Σχήμα 3.12: *Εικόνες της φωνητικής οδού, με τη χρήση ακτίνων Χ. Παρουσιάζεται και το ημιπολικό πλέγμα όπως τοποθετείται για τον υπολογισμό της συνάρτησης εμβαδού.*

αντιστοιχεί στο μήκος από το σημείο αναφοράς ως την τελευταία γραμμή του πάνω γραμμικού πλέγματος που ιδανικά περνάει από τον πάνω κόφτη.

θ_l, θ_m Οι γωνίες που σχηματίζουν με τον οριζόντιο άξονα οι ακριανές γραμμές του πολικού πλέγματος.

$dl_l, dl_m, d\theta$ Χωρική ανάλυση για τα δύο γραμμικά τμήματα του πλέγματος και το πολικό τμήμα. Η τελευταία αντιστοιχεί στη γωνία μεταξύ δύο γραμμών του πολικού πλέγματος.

l_g Το μήκος της γραμμής του ημιπολικού πλέγματος.

Ο προσδιορισμός του ημιπολικού πλέγματος γίνεται πάνω σε μια εικόνα ακτίνων Χ της φωνητικής οδού με τη βοήθεια κατάλληλης γραφικής διεπαφής. Το πλέγμα θεωρείται ότι κινείται μαζί με το κεφάλι, παραμένοντας σταθερό ως προς τον ουρανίσκο. Η κίνηση του κεφαλιού στις διαθέσιμες εικόνες ακτίνων Χ θεωρείται ούτως ή άλλως αμελητέα.

3.8.2 Εκτίμηση συναρτήσεων εμβαδού

Έχοντας τοποθετήσει το πλέγμα πάνω στις εικόνες της φωνητικής οδού, ακολουθεί αυτόματη εξαγωγή των σημείων τομής των γραμμών του πλέγματος με τις καμπύλες που περιγράφουν το εσωτερικό και το εξωτερικό τοίχωμα της φωνητικής οδού. Το εσωτερικό τοίχωμα θεωρείται ότι σχηματίζεται κυρίως από τη γλώσσα και το υπεργλωττιδικό τοίχωμα του λάρυγγα ενώ καταλήγει στον κάτω κόφτη. Αμελείται η επιγλωττίδα στην παρούσα ανάλυση. Το εξωτερικό τοίχωμα περιλαμβάνει τον πάνω κόφτη, τον ουρανίσκο, τη μαλακή υπερώα (στη θέση όπου το ρινοφαρυγγικό άνοιγμα είναι κλειστό) το φαρυγγικό τοίχωμα και το υποφαρυγγικό τοίχωμα του λάρυγγα. Τα χείλη εξαιρούνται αφού θεωρείται ότι λόγω της διαμήκους μεταβλητότητάς τους δεν περιγράφονται κατάλληλα από το συγκεκριμένο ημιπολικό πλέγμα. Σε κάθε γραμμή του πλέγματος προσδιορίζεται ο άξονας της φωνητικής οδού ως η γραμμή που ισαπέχει σε κάθε σημείο της από το εσωτερικό και το εξωτερικό τοίχωμα και υπολογίζεται τελικά η απόσταση μεταξύ των τοιχωμάτων αυτών. Με βάση την απόσταση αυτή και χρησιμοποιώντας κατάλληλο μοντέλο είναι τελικά δυνατή η εκτίμηση της συνάρτησης επιφανείας πάνω στον άξονα της φωνητικής οδού.

Τέτοια μοντέλα στην ουσία ενσωματώνουν εμμέσως την τρίτη διάσταση που δεν είναι ορατή στις μεσο-οβελιαίες εικόνες της φωνητικής οδού 2.2. Είναι της μορφής $A = Kd^{\alpha}$ όπου A είναι το εμβαδό της εγκάρσιας διατομής της οδού, d το μεσο-οβελιαίο άνοιγμα, η απόσταση δηλαδή που μπορούμε να μετρήσουμε πάνω στο πλέγμα, και K, α παράμετροι που εξαρτώνται από τον ομιλητή και την απόσταση από τη γλωττίδα του σημείου όπου πραγματοποιείται ο υπολογισμός [62]. Στη βιβλιογραφία έχουν εμφανιστεί και διάφορες βελτιώσεις όπως για παράδειγμα στο [21] όπου οι παράμετροι εξαρτώνται και από το μέσο οβελιαίο άνοιγμα και βελτιστοποιούνται με τη χρήση ακουστικών δεδομένων. Στα πειράματά μας, χρησιμοποιούμε τελικά τις παραμέτρους του μοντέλου που δίνονται στο [153]. Δεδομένου του ότι αυτές εξαρτώνται από τον ομιλητή, στην παρούσα φάση, είναι σε εξέλιξη προσπάθεια να εξαχθεί ένα μοντέλο μετατροπής μεσο-οβελιαίας απόστασης σε εγκάρσια διατομή από διαθέσιμα δεδομένα αξονικής τομογραφίας που αφορούν στο συγκεκριμένο ομιλητή στα δεδομένα του οποίου πειραματιζόμαστε. Για τα χείλια, θεωρούμε ότι μπορούν να προσεγγιστούν ως κύλινδρος ελλειπτικής διατομής με μήκος ίσο με την απόσταση του κάτω κόφτη από την άκρη των χειλιών και μικρό άξονα ίσο με την ελάχιστη απόσταση μεταξύ του πάνω και του κάτω χείλους. Ο μεγάλος άξονας εκτιμάται ευριστικά ως ακέραιο πολλαπλάσιο του μικρού.

Στο Σχήμα 3.13 δίνονται τα διαδοχικά σχήματα της φωνητικής οδού για την ακολουθία φωνηέντων προς σύνθεση καθώς και οι εκτιμώμενες συναρτήσεις εμβαδού. Αντιστοιχούν σε εικόνες ακτίνων-Χ που έχουν ληφθεί με συχνότητα ίση με 25 Hz.

3.8.3 Αρθρωτικές παράμετροι και σύνθεση

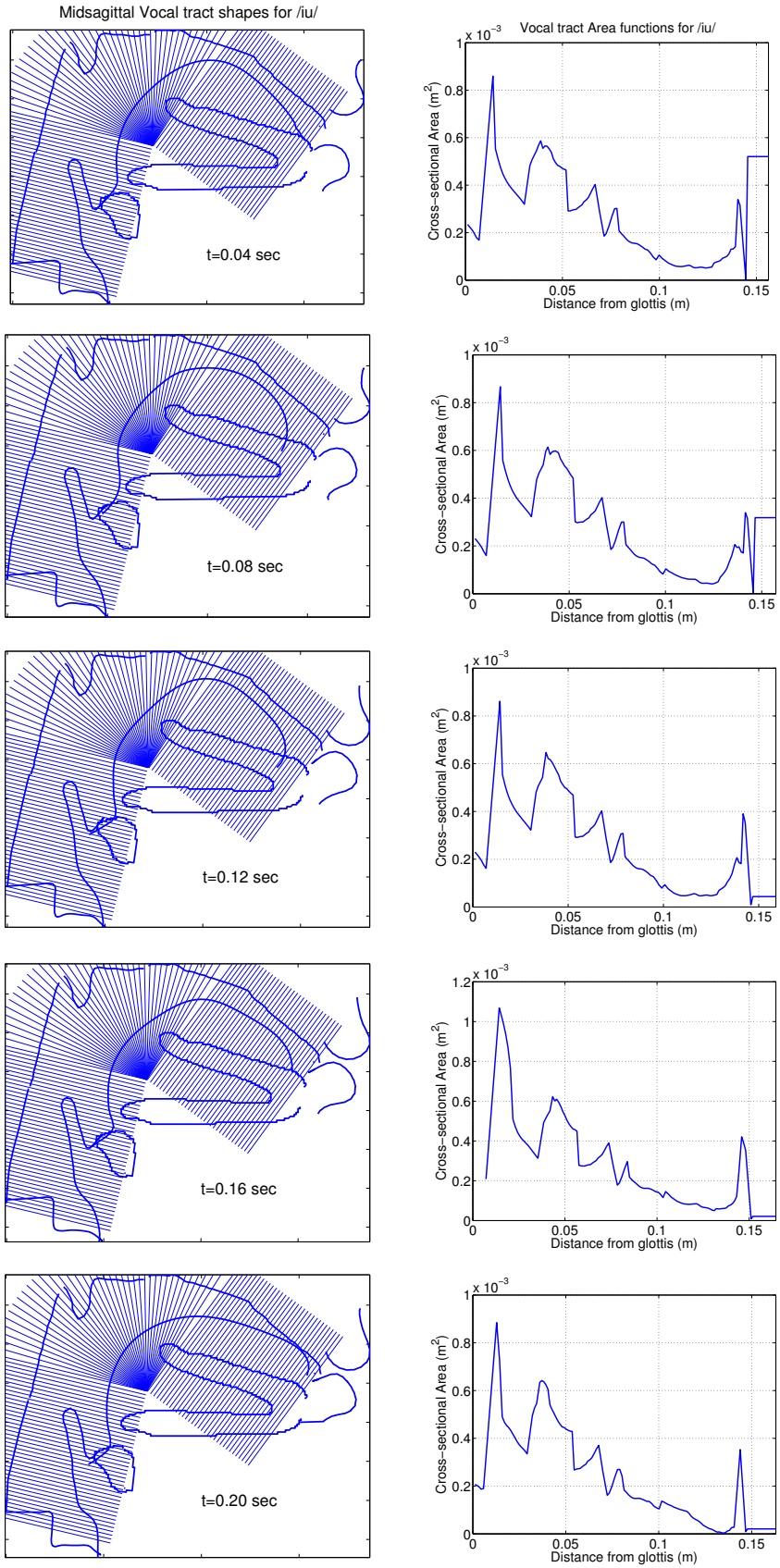
Για τον έλεγχο της προσομοίωσης χρησιμοποιήθηκε η υπογλωττιδική πίεση P_s και η τάση των φωνητικών χορδών Q ακολουθώντας το παράδειγμα του [57]. Με την παράμετρο Q που πολλαπλασιάζει τις μάζες του μοντέλου της γλωττίδας και διαιρεί τις σταθερές των ελατηρίων επιτυγχάνεται εμμέσως ο έλεγχος της θεμελιώδους συχνότητας της πηγής ήχου. Στην προσπάθειά μας να προσομοιώσουμε τις μεταβολές της έντασης και της θεμελιώδους συχνότητας του πραγματικού σήματος οι παράμετροι άρθρωσης θεωρήθηκαν όπως στο Σχήμα. Στην πράξη ορίστηκαν συγκεκριμένες τιμές στόχοι για τις παραμέτρους αυτές, όπως για παράδειγμα $Q = 0.9$ που πρέπει να επιτευχθούν σε κάποιο χρονικό διάστημα, για παράδειγμα από 100 έως 150 ms. Οι ενδιάμεσες τιμές προκύπτουν με κυβική παρεμβολή.

Στο Σχήμα 3.15 δίνεται λεπτομέρεια του ακουστικού σήματος που προκύπτει με την προσομοίωση σε αντιπαράθεση με το ακουστικό σήμα που έχει μετρηθεί ενώ στο Σχήμα 3.16 παρουσιάζοντα τα αντίστοιχα σπεκτρογραφήματα. Στο Σχήμα 3.17 δίνονται επίσης η γλωττιδική ογκική ταχύτητα και το εμβαδόν του ανοίγματος της γλωττίδας όπως αυτό μετράται για την πρώτη μάζα. Αρνητικό εμβαδόν ανοίγματος υποδηλώνει κλειστή γλωττίδα.

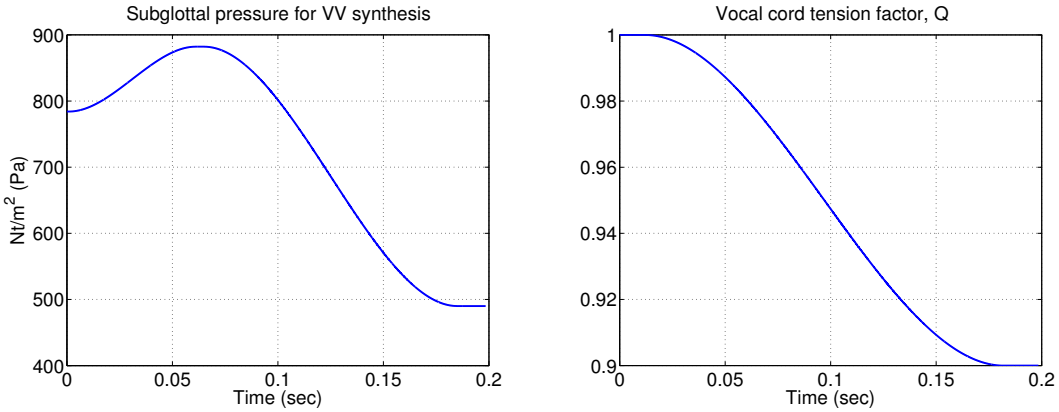
Άτυπος ακουστικός έλεγχος καταδεικνύει ότι το συνθετικό σήμα όντως μιμείται το πραγματικό παρά τις διαφορές που εμφανίζονται στα σπεκτρογραφήματα. Υπάρχουν βέβαια και διαφορές στη θεμελιώδη συχνότητα και στη χροιά οι οποίες θα μπορούσαν να δικαιολογηθούν λόγω ανακριβούς ορισμού των αρθρωτικών παραμέτρων και ατελειών του εφαρμοζόμενου μοντέλου για τη συνάρτηση εμβαδού της φωνητικής οδού το οποίο όπως αναφέρθηκε δεν έχει προσαρμοστεί στον ομιλητή μας.

3.9 Συζήτηση

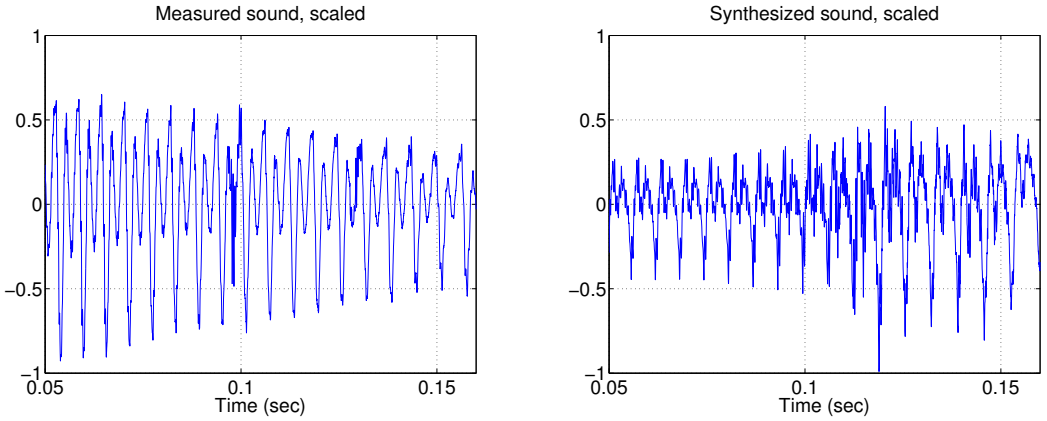
Ακολουθώντας τη βασική φιλοσοφία συμβατικών συνθετών φωνής με αρθρωτές αναπτύχθηκε σύστημα προσομοίωσης του ακουστικού πεδίου μέσα στη φωνητική οδό. Συγκεκριμένα, η φωνητική οδός προσομοιώνεται ως ένας σωλήνας χωρικά και χρονικά μεταβαλλόμενης διατομής. Η εκπομπή ήχου από το σωλήνα θεωρείται ως εκπομπή από ένα μικρό άνοιγμα σε ένα άπειρο επίπεδο ενώ στην είσοδό του είναι συζευγμένο ένα μοντέλο δύο μαζών για τη γλωττίδα που ταλαντώνεται χωρίς εξωτερική διέγερση πέρα από τη διερχόμενη ογκική ταχύ-



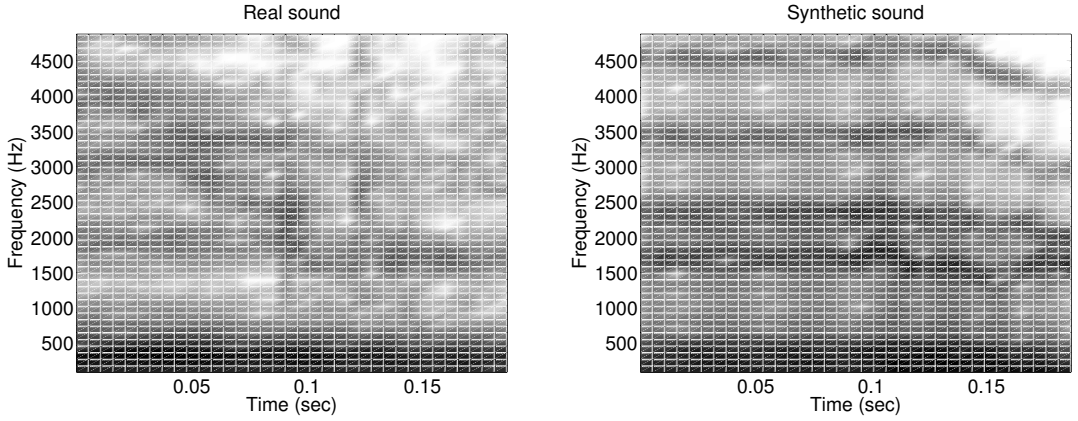
Σχήμα 3.13: *Διαδοχικά μεσο-οβελιαία σχήματα της φωνητικής οδού και συναρτήσεις εμβαδού για την ακολουθία φωνηέντων /iu/. Η συχνότητα με την οποία έχουν ληφθεί οι αντίστοιχες εικόνες ακτίνων X είναι 25 Hz.*



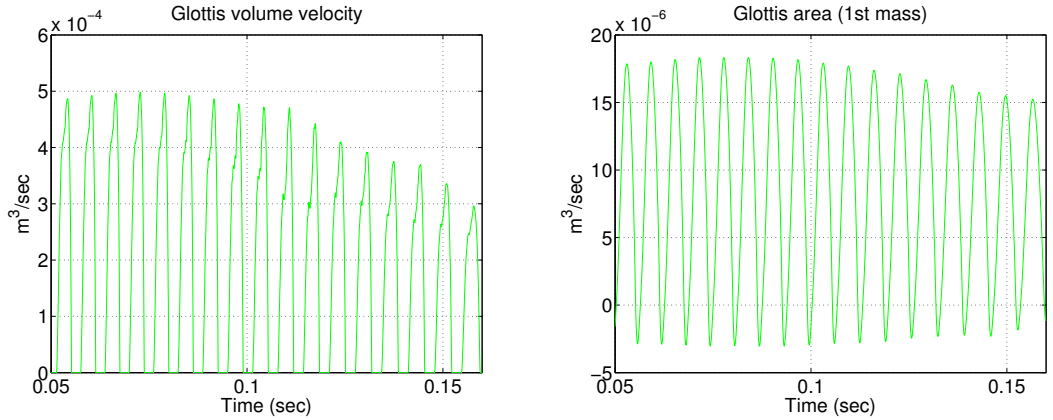
Σχήμα 3.14: *Μεταβολή της υπογλωττιδικής πίεσης και του παράγοντα τάσης των φωνητικών χορδών όπως χρησιμοποιήθηκαν για την προσομοίωση. Εφαρμόζεται κυβική παρεμβολή μεταξύ τιμών-στόχων των παραμέτρων αυτών. Έγινε προσπάθεια να προσομοιωθούν η ένταση και η θεμελιώδης συχνότητα του ηχητικού σήματος που είχε μετρηθεί κατά την καταγραφή των εικόνων ακτίνων X.*



Σχήμα 3.15: *Λεπτομέρεια από το πραγματικό και το συνθετικό σήμα φωνής, μετά την προσομοίωση χρησιμοποιώντας τα μεσο-οβελιαία σχήματα της φωνητικής οδού όπως έχουν μετρηθεί με ακτίνες X.*



Σχήμα 3.16: *Σπεκτρογραφήματα του πραγματικού και του συνθετικού σήματος φωνής, μετά την προσομοίωση χρησιμοποιώντας τα μεσο-οβελιαία σχήματα της φωνητικής οδού όπως έχουν μετρηθεί με ακτίνες X.*



Σχήμα 3.17: *Λεπτομέρεια της ογκικής ταχύτητας και του εμβαδού ανοίγματος της γλωττίδας κατά την προσομοίωση. Αρνητικό εμβαδό ανοίγματος υποδηλώνει ότι η γλωττίδα είναι κλειστή.*

τητα του αέρα. Το μοντέλο αυτό επιτρέπει τη μελέτη φαινομένων αλληλεπίδρασης πηγής - φωνητικής οδού. Στο βασικό τμήμα της οδού είναι επίσης συζευγμένες η ρινική κοιλότητα και οι αχλαδόσχημες κοιλότητες στην κορυφή του λάρυγγα. Οι τελευταίες είναι υπεύθυνες για κάποια μικρή μετακίνηση των συντονισμών του παραγόμενου σήματος και θεωρείται ότι προσδίδουν μεγαλύτερη φυσικότητα στη συνθετική φωνή. Σημειώνεται ότι δεν έγινε προσπάθεια να μοντελοποιηθούν ενδεχόμενες επιδράσεις λόγω σύζευξης με το υπογλωττιδικό τμήμα του ανθρώπινου συστήματος παραγωγής φωνής [30]. Το αριθμητικό σχήμα που εφαρμόζεται συνδυάζει τον κανόνα του μέσου σημείου για τη χωρική διακριτοποίηση και του τραπεζίου για τη χρονική διακριτοποίηση. Η δόνηση των τοιχωμάτων μοντελοποιείται στη συχνότητα και προσομοιώνεται μέσω της μεθόδου αμετάβλητης κρουστικής απόκρισης. Η επίλυση του τελικού συστήματος, που είναι γενικά γραμμικό αλλά περιλαμβάνει μια πολυωνυμική δευτέρου βαθμού εξίσωση για τη γλωττίδα, επιτυγχάνεται με κατάλληλη τριγωνοποίηση.

Για αξιολόγηση παρουσιάστηκαν αποτελέσματα σύνθεσης για συναρτήσεις εμβαδού της φωνητικής οδού που αντιστοιχούν σε διάφορα φωνήεντα όπως έχουν μετρηθεί με αξονική τομογραφία και έχουν δημοσιευτεί στα [52, 159, 160]. Συγκρίνονται οι εκτιμώμενοι συντονισμοί των αντίστοιχων αποκρίσεων συχνότητας με αυτούς που έχουν μετρηθεί από τα πραγματικά σήματα φωνής και τα αποτελέσματα είναι αρκετά ικανοποιητικά. Τέλος, δίνεται παράδειγμα σύνθεσης ακολουθίας φωνηέντων με βάση δεδομένα για το σχήμα της φωνητικής οδού όπως έχουν καταγραφεί με ακτίνες Χ.

Το προτεινόμενο σύστημα προσομοίωσης θα συνδυαστεί στη συνέχεια με κατάλληλο αεροδυναμικό μοντέλο για την περιγραφή και της μη ακουστικής ροής στη φωνητική οδό. Στόχος είναι η μοντελοποίηση αεροακουστικών φαινομένων που θεωρούνται σημαντικά για τη φωνή και για την περιγραφή των οποίων δεν αρκεί η υπόθεση ότι η παραγωγή φωνής διέπεται απλά από τις γραμμικές ακουστικές εξισώσης διάδοσης ενός επιπέδου κύματος σε ένα ήρεμο μέσο.

3.Α΄ Εξισώσεις ακουστικής ζεύξης ρινικής κοιλότητας

Δεδομένης της οριακής συνθήκης στο σημείο ζεύξης με τη ρινική κοιλότητα, εφαρμόζουμε κατάλληλα την Εξίσωση (3.55) και παίρνουμε:

$$-b^k_{K-1}U^k_{K-1} + H^k_{K,K-1}U^k_K - b^k_K U^k_{NC} = F^{k,k-1}_{K,K-1}, \tag{3.94}$$

$$-b_K^k u_K + \underbrace{(Z_K^k + R_K^k + b_K^k)}_{H_{NC}^k} U_{NC}^k + p_{NC}^k = \underbrace{(Z_K^{k-1} - R_K^{k-1})U_{NC}^{k-1} - p_{NC}^{k-1} + p_{K+\frac{1}{2}}^{k-1} - b_K^{k-1}\Phi_K^{k,k-1}}_{F_{K,NC}^{k,k-1}}, \quad (3.95)$$

$$\underbrace{(Z_{K+1}^k + R_{K+1}^k + b_{K+1}^k)}_{H_{K+1}^k} U_{K+1}^k - b_{K+1}^k U_{K+1} - p_{NC}^k = \underbrace{(Z_{K+1}^{k-1} - R_{K+1}^{k-1})U_{K+1}^{k-1} + p_{NC}^{k-1} - p_{K+\frac{3}{2}}^{k-1} + b_{K+1}^{k-1}\Phi_{K+1}^{k,k-1}}_{F_{K+1,NC}^{k,k-1}}. \quad (3.96)$$

$$\underbrace{(Z_{(NT)1}^k + R_{(NT)1}^k + b_{(NT)1}^k)}_{H_{(NT)1}^k} U_{(NT)1}^k - b_{(NT)1}^k U_{(NT)1} - p_{NC}^k = \underbrace{(Z_{(NT)1}^{k-1} - R_{(NT)1}^{k-1})U_{(NT)1}^{k-1} + p_{NC}^{k-1} - p_{(NT)1+\frac{1}{2}}^{k-1} + b_{(NT)1}^{k-1}\Phi_{(NT)1}^{k,k-1}}_{F_{NC,(NT)1}^{k,k-1}}. \quad (3.97)$$

Δηλαδή είναι :

$$p_{NC}^k = b_K^k U_K - H_{NC}^k(U_{K+1}^k + U_{(NT)1}^k) + F_{K,NC}^{k,k-1}. \quad (3.98)$$

Απαλοίφοντας τα U_{NC}^k, p_{NC}^k τελικά παίρνουμε [114] :

$$-b_{K-1}^k U_{K-1}^k + H_{K,K-1}^k U_K^k - b_K^k U_{K+1}^k - b_K^k U_{(NT)1} = F_{K,K-1}^{k,k-1}, \quad (3.99)$$

$$-b_K U_K^k + (H_{K+1}^k + H_{NC}^k)U_{K+1}^k - b_{K+1}^k U_{K+2} + H_{NC}^k U_{(NT)1} = F_{K+1,NC}^{k,k-1} + F_{K,NC}^{k,k-1}, \quad (3.100)$$

$$-b_K U_K^k + H_{NC}^k U_{K+1}^k + (H_{(NT)1}^k + H_{NC}^k)U_{(NT)1}^k - b_{(NT)1}^k U_{(NT)2} = F_{NC,(NT)1}^{k,k-1} + F_{K,NC}^{k,k-1}. \quad (3.101)$$

3.Β΄ Εξισώσεις ακουστικής ζεύξης αχλαδόσχημων κοιλοτήτων

Δεδομένης της οριακής συνθήκης στο σημείο ζεύξης των αχλαδόσχημων κοιλοτήτων, οι εξισώσεις που τροποποιούνται έχουν ως εξής :

$$-b_{L-1}^k U_{L-1}^k + H_{L,L-1}^k U_L^k - b_L^k U_{L+1}^k - b_L^k U_{(LPF)1} - b_L^k U_{(RPF)1} = F_{L,L-1}^{k,k-1}, \quad (3.102)$$

$$-b_L U_L^k + (H_{L+1}^k + H_{PFC}^k)U_{L+1}^k - b_{L+1}^k U_{L+2} + H_{PFC}^k U_{(LPF)1} + H_{PFC}^k U_{(RPF)1} = F_{L+1,PFC}^{k,k-1} + F_{L,PFC}^{k,k-1}, \quad (3.103)$$

$$-b_L u_L^k + H_{PFC}^k U_{L+1} + (H_{(LPF)1}^k + H_{PFC}^k)U_{(LPF)1}^k - b_{(LPF)1}^k U_{(LPF)2} + H_{PFC}^k U_{(LPF)1} = F_{PFC,(LPF)1}^{k,k-1} + F_{L,PFC}^{k,k-1}, \quad (3.104)$$

$$-b_L U_L^k + H_{PFC}^k U_{L+1} + (H_{(RPF)1}^k + H_{PFC}^k)U_{(RPF)1}^k - b_{(RPF)1}^k U_{(RPF)2} + H_{PFC}^k U_{(RPF)1} = F_{PFC,(RPF)1}^{k,k-1} + F_{L,PFC}^{k,k-1}, \quad (3.105)$$

με τα $H_{PFC}, F_{L,PFC}$ να ορίζονται κατ΄ αναλογία με τα $H_{NC}, F_{K,NC}$. Οι υπόλοιπες εξισώσεις του ακουστικού πεδίου μέσα στις όποιες κοιλότητες είναι κατ΄ αναλογία με αυτές που επικρατούν στην κύρια φωνητική οδό.

3.Γ΄ Επίλυση συστήματος ακουστικής προσομοίωσης

Μαζί με τις εξισώσεις στα σημεία ζεύξης με κοιλότητες που παρατέθηκαν προηγουμένως το τελικό σύστημα εξισώσεων που προκύπτει επίσης περιλαμβάνει :

$$R_m^k|U_1^k|U_1^k + H_{1,g}^k U_1^k - b_1^k U_2^k = -R_m^k|U_1|^{k-1}U_1^{k-1} + F_{1,g}^{k,k-1}. \quad (3.106)$$

$$-b_{n-1}^k U_{n-1}^k + H_{n,n-1}^k U_n^k - b_{n+1}^k U_{n+1}^k = F_{n,n-1}^{k,k-1}, n = 2, \dots L-1, L+2, \dots K-1, \dots N \quad (3.107)$$

$$-b_{(NT)n-1}^k U_{(NT)n-1}^k + H_{(NT)n,n-1}^k U_{(NT)n}^k - b_{(NT)n+1}^k U_{(NT)n+1}^k = F_{(NT)n,n-1}^{k,k-1}, n = 2, \dots N_{nt} \quad (3.108)$$

$$-b_{(LPF)n-1}^k U_{(LPF)n-1}^k + H_{(LPF)n,n-1}^k U_{(LPF)n}^k - b_{(LPF)n+1}^k U_{(LPF)n+1}^k = F_{(LPF)n,n-1}^{k,k-1}, n = 2, \dots N_{lpf} \quad (3.109)$$

$$-b_{(RPF)n-1}^k U_{(RPF)n-1}^k + H_{(RPF)n,n-1}^k U_{(RPF)n}^k - b_{(RPF)n+1}^k U_{(RPF)n+1}^k = F_{(RPF)n,n-1}^{k,k-1}, n = 2, \dots N_{rpf} \quad (3.110)$$

όπου N_{nt}, N_{lpf}, N_{rpf} είναι το πλήθος των σημείων διακριτοποίησης της ρινικής κοιλότητας και των piriform fossa αντίστοιχα. Επιπλέον, και οι εξισώσεις που αντιστοιχούν στις οριακές συνθήκες εκπομπής :

$$-b_N^k U_N^k + H_{N+1,N}^k U_{N+1}^k = F_{N+1,N}^{k,k-1}, \quad (3.111)$$

$$-b_{(NT)N_{nt}}^k U_{(NT)N_{nt}}^k + H_{(NT)N_{nt}+1,N_{nt}}^k U_{(NT)N_{nt}+1}^k = F_{(NT)N_{nt}+1,N_{nt}}^{k,k-1}, \quad (3.112)$$

$$-b_{(RPF)N_{rpf}}^k U_{(RPF)N_{rpf}}^k + H_{(RPF)N_{rpf}+1,N_{rpf}}^k U_{N_{rpf}+1}^k = F_{N_{rpf}+1,N_{rpf}}^{k,k-1}, \quad (3.113)$$

$$-b_{(LPF)N_{lpf}}^k U_{(LPF)N_{lpf}}^k + H_{(LPF)N_{lpf}+1,N_{lpf}}^k U_{N_{lpf}+1}^k = F_{N_{lpf}+1,N_{lpf}}^{k,k-1}. \quad (3.114)$$

Όλες οι εξισώσεις είναι γραμμικές εκτός από την Εξ. (3.106). Για την επίλυση του συστήματος το τριγωνοποιούμε αρχικά εφαρμόζοντας απαλοιφή Gauss ξεκινώντας από τις εξισώσεις που εκφράζουν τις συνθήκες εκπομπής. Με αυτόν τον τρόπο επιτυγχάνουμε τελικά να εκφράσουμε τη μη γραμμική εξίσωση μόνο ως συνάρτηση της ογκικής ταχύτητας u_1, οπότε η εξίσωση αυτή μπορεί να επιλυθεί ανεξάρτητα ως τριώνυμο. Ανάλογη προσέγγιση ακολουθείται και στο [127]. Έχοντας προσδιορίσει την ταχύτητα u_1 μπορούμε να βρούμε και τις υπόλοιπες ογκικές ταχύτητες εύκολα με αντικατάσταση προς τα εμπρός. Αν για λόγους παρουσίασης αμελήσουμε τους χρονικούς δείκτες, έχουμε με την απαλοιφή Gauss :

$$\alpha|U_1|U_1 + \beta U_1 + \gamma = 0 \quad (3.115)$$

$$l_{n,n-1}U_{n-1} + l_{n,n}U_n = c_n, n = 2, \dots, N+1 \quad (3.116)$$

$$g_{n,n-1}U_{(LPF)n-1} + g_{n,n}U_{(LPF)n} = d_n, n = 2, \dots, N_{lpf}+1 \quad (3.117)$$

$$f_{n,n-1}U_{(RPF)n-1} + f_{n,n}U_{(RPF)n} = e_n, n = 2, \dots, N_{rpf}+1 \quad (3.118)$$

$$h_{n,n-1}U_{(NT)n-1} + h_{n,n}U_{(NT)n} = w_n, n = 2, \dots, N_{nt}+1 \quad (3.119)$$

$$g_{1,L}U_L + g_{1,L+1}U_{L+1} + g_{1,1}U_{(LPF)1} = d_1 \quad (3.120)$$

$$f_{1,L}U_L + f_{1,L+1}U_{L+1} + f_{LPF,RPF}U_{(LPF)1} + f_{1,1}U_{(RPF)1} = e_1 \quad (3.121)$$

$$h_{1,K}U_K + h_{1,K+1}U_{K+1} + h_{1,1}U_{(NT)1} = w_1 \quad (3.122)$$

όπου

$$l_{n,n-1} = -b_{n-1}, n = 2, \ldots, L-1, L+2, \ldots, K, K+2, \ldots, N+1, \tag{3.123}$$

$$l_{n,n} = H_{n,n-1} + b_n \frac{l_{n+1,n}}{l_{n+1,n+1}}, n = 1, \ldots, L-1, L+2, \ldots, K-1, K+2, \ldots, N, \tag{3.124}$$

$$c_n = F_{n,n-1} + b_n \frac{c_{n+1}}{l_{n+1,n+1}}, n = 1, \ldots, L-1, L+2, \ldots, K-1, K+2, \ldots, N, \tag{3.125}$$

$$g_{n,n-1} = -b_{(LPF)n-1}, n = 2, \ldots, N_{lpf}+1, \tag{3.126}$$

$$g_{n,n} = H_{(LPF)n,n-1} + b_{(LPF)n} \frac{g_{n+1,n}}{g_{n+1,n+1}}, n = 2, \ldots, N_{lpf}, \tag{3.127}$$

$$d_n = F_{(LPF)n,n-1} + b_{(LPF)n} \frac{d_{n+1}}{g_{n+1,n+1}}, n = 2, \ldots, N_{lpf}, \tag{3.128}$$

$$f_{n,n-1} = -b_{(RPF)n-1}, n = 2, \ldots, N_{rpf}+1, \tag{3.129}$$

$$f_{n,n} = H_{(RPF)n,n-1} + b_{(RPF)n} \frac{f_{n+1,n}}{f_{n+1,n+1}}, n = 2, \ldots, N_{rpf}, \tag{3.130}$$

$$e_n = F_{(RPF)n,n-1} + b_{(RPF)n} \frac{e_{n+1}}{f_{n+1,n+1}}, n = 2, \ldots, N_{rpf}, \tag{3.131}$$

$$h_{n,n-1} = -b_{(NT)n-1}, n = 2, \ldots, N_{nt}+1, \tag{3.132}$$

$$h_{n,n} = H_{(NT)n,n-1} + b_{(NT)n} \frac{h_{n+1,n}}{h_{n+1,n+1}}, n = 2, \ldots, N_{nt}, \tag{3.133}$$

$$w_n = F_{(NT)n,n-1} + b_{(NT)n} \frac{w_{n+1}}{h_{n+1,n+1}}, n = 2, \ldots, N_{nt}, \tag{3.134}$$

με ειδικές περιπτώσεις τις :

$$l_{N+1,N+1} = H_{N+1,N}, c_{N+1} = F_{N+1,N} \tag{3.135}$$

$$g_{N_{lpf}+1,N_{lpf}+1} = H_{(LPF)N_{lpf}+1,N_{lpf}}, d_{N_{lpf}+1} = F_{(LPF)N_{lpf}+1,N_{lpf}} \tag{3.136}$$

$$f_{N_{rpf}+1,N_{rpf}+1} = H_{(RPF)N_{rpf}+1,N_{rpf}}, e_{N_{rpf}+1} = F_{(RPF)N_{rpf}+1,N_{rpf}} \tag{3.137}$$

$$h_{N_{nt}+1,N_{nt}+1} = H_{(NT)N_{nt}+1,N_{nt}}, w_{N_{nt}+1} = F_{(NT)N_{nt}+1,N_{nt}} \tag{3.138}$$

και :

$$l_{L,L} = H_L + b_L \frac{f_{1,L}}{f_{1,1}} + b_L(1 - \frac{f_{LPF,RPF}}{f_{1,1}}) \frac{g_{1,L}}{g_{1,1}} + b_L(1 - \frac{f_{1,L+1}}{f_{1,1}} - \frac{g_{1,L+1}}{g_{1,1}} + \frac{f_{LPF,RPF}}{f_{1,1}} \frac{g_{1,L+1}}{g_{1,1}}) \frac{l_{L+1,L}}{l_{L+1,L+1}} \tag{3.139}$$

$$c_L = F_{L,L-1} + b_L \frac{e_1}{f_{1,1}} + b_L(1 - \frac{f_{LPF,RPF}}{f_{1,1}}) \frac{d_1}{g_{1,1}} + b_L(1 - \frac{f_{1,L+1}}{f_{1,1}} - \frac{g_{1,L+1}}{g_{1,1}} + \frac{f_{LPF,RPF}}{f_{1,1}} \frac{g_{1,L+1}}{g_{1,1}}) \frac{c_{L+1}}{l_{K+1,K+1}} \tag{3.140}$$

$$l_{L+1,L+1} = H_{L+1} + H_{PFC} + b_{L+1} \frac{l_{L+2,L+1}}{l_{L+2,L+2}} - H_{PFC} \frac{f_{1,L+1}}{f_{1,1}} - H_{PFC}(1 - \frac{f_{LPF,RPF}}{f_{1,1}}) \frac{g_{1,L+1}}{g_{1,1}}, \tag{3.141}$$

$$l_{L+1,L} = -b_L - H_{PFC} \frac{g_{1,L}}{g_{1,1}} - H_{PFC}(1 - \frac{f_{LPF,RPF}}{f_{1,1}}) \frac{f_{1,L}}{f_{1,1}}, \tag{3.142}$$

$$c_{L+1} = F_{L+1,PFC} + F_{L,PFC} - H_{PFC} \frac{e_1}{f_{1,1}} - H_{PFC}(1 - \frac{f_{LPF,RPF}}{f_{1,1}}) \frac{d_1}{g_{1,1}} + b_{L+1} \frac{c_{L+2}}{l_{L+2,L+2}} \tag{3.143}$$

$$l_{K,K} = H_K + b_K(1 - \frac{h_{1,K+1}}{h_{1,1}})\frac{l_{K+1,K}}{l_{K+1,K+1}} + b_K\frac{h_{1,K}}{h_{1,1}} \tag{3.144}$$

$$c_K = F_{K,K-1} + b_K(1 - \frac{h_{1,K+1}}{h_{1,1}})\frac{c_{K+1}}{l_{K+1,K+1}} + b_K\frac{w_1}{h_{1,1}} \tag{3.145}$$

$$l_{K+1,K+1} = H_{K+1} + H_{NC} + b_{K+1}\frac{l_{K+2,K+1}}{l_{K+2,K+2}} - H_{NC}\frac{h_{1,K+1}}{h_{1,1}} \tag{3.146}$$

$$l_{K+1,K} = -b_K - H_{NC}\frac{h_{1,K}}{h_{1,1}}, \tag{3.147}$$

$$c_{K+1} = F_{K+1,NC} + F_{K,NC} - H_{NC}\frac{w_1}{h_{1,1}} + b_{K+1}\frac{c_{K+2}}{l_{K+2,K+2}} \tag{3.148}$$

$$g_{1,1} = H_{(LPF)1} + H_{PFC} + b_{(LPF)1}\frac{g_{2,1}}{g_{2,2}} - H_{PFC}\frac{f_{LPF,RPF}}{f_{1,1}}, \tag{3.149}$$

$$g_{1,L} = -b_L + H_{(PFC)}\frac{b_L}{f_{1,1}}, g_{1,L+1} = H_{PFC} - H_{PFC}\frac{H_{PFC}}{f_{1,1}}, \tag{3.150}$$

$$d_1 = F_{PFC,(LPF)1} + F_{L,PFC} - H_{PFC}\frac{e_1}{f_{1,1}}, \tag{3.151}$$

$$f_{1,1} = H_{(RPF)1} + H_{PFC} + b_{(RPF)1}\frac{f_{2,1}}{f_{2,2}}, \tag{3.152}$$

$$f_{1,L} = -b_L, f_{1,L+1} = H_{PFC}, f_{LPF,RPF} = H_{PFC}, \tag{3.153}$$

$$h_{1,1} = H_{(NT)1} + H_{NC} + b_{(NT)1}\frac{h_{2,1}}{h_{2,2}}, \tag{3.154}$$

$$h_{1,K} = -b_K, h_{1,K+1} = H_{NC}. \tag{3.155}$$

Η μη γραμμική εξίσωση στη γλωττίδα λαμβάνει τη μορφή :

$$\underbrace{R_m^k}_{\alpha}|U_1^k|U_1^k + \underbrace{(Z_g^k + Z_1^k + R_g^k + R_1^k + R_{vt}^k)}_{\beta}U_1^k + \underbrace{R_m^{k-1}(U_1^{k-1})^2 + p_{vt}^k + p_{1+\frac{1}{2}}^{k-1} + (p_{sub}^k + p_{sub}^{k-1}) + (R_g^{k-1} + R_1^{k-1} - Z_g^{k-1} - Z_1^{k-1})U_1^{k-1}}_{\gamma} = 0 \tag{3.156}$$

που μπορεί να επιλυθεί ως προς την ογκική ταχύτητα U_1. Τα μεγέθη R_{vt}^k, p_{vt}^k αντιστοιχούν στην ουσία σε ένα ισοδύναμο κύκλωμα που 'βλέπει' η γλωττίδα προς τη φωνητική οδό [127]. Οι τιμές τους προκύπτουν μετά την τριγωνοποίηση που περιγράφτηκε προηγουμένως με στόχο την απαλοιφή της ογκικής ταχύτητας U_2 από τη μη γραμμική εξίσωση της γλωττίδας ώστε να μπορεί αυτή να λυθεί ανεξάρτητα :

$$R_{vt}^k = b_1^k - b_1^k\frac{b_1^k}{l_{2,2}} \tag{3.157}$$

$$p_{vt}^k = -b_1^k\frac{c_2}{l_{2,2}} - b_1^k\Phi_1^{k,k-1} \tag{3.158}$$

Στην ουσία η πίεση $p_{1+\frac{1}{2}}^k$ στην Εξ. (3.86) αντικαθίσταται με :

$$p_{1+\frac{1}{2}}^k = b_1^k(U_1^k - U_2^{k-1} - \Phi_1^{k,k-1}) \tag{3.159}$$

$$= b_1^k(U_1^k - b_1^k\frac{b_1^k}{l_{2,2}} - \frac{c_2}{l_{2,2}} - \Phi_1^{k,k-1}) \tag{3.160}$$

$$= R_{vt}^k U_1^k + p_{vt}^k. \tag{3.161}$$

Η επίλυση του συνολικού συστήματος είναι στη συνέχεια δυνατή με αντικατάσταση προς τα εμπρός αφού ο πίνακας του τροποποιημένου συστήματος είναι τριγωνικός κάτω.

Κεφάλαιο 4

Μοντελοποίηση αεροδυναμικών και αεροακουστικών φαινομένων για σύνθεση φωνής

4.1 Εισαγωγή

Στο Κεφάλαιο 3 μελετήθηκε η μοντελοποίηση του ακουστικού πεδίου της φωνητικής οδού. Όπως σημειώθηκε, το ακουστικό αυτό πεδίο αφορά σε μικρές διαταραχές της πίεσης και της ογκικής ταχύτητας του αέρα που διαδίδονται με την ταχύτητα του ήχου και που τελικά εκπέμπονται από τα χείλη ή τα ρουθούνια. Η ύπαρξη μη ακουστικής αερορροής μέσα στη φωνητική οδό θεωρήθηκε αμελητέα μετά την περιοχή της γλωττίδας. Η γλωττίδα εμφανίζεται πρακτικά να είναι ο μηχανισμός ακουστικής διέγερσης της φωνητικής οδού δημιουργώντας τις κατάλληλες διαταραχές ογκικής ταχύτητας. Δεδομένου του ότι με το μοντέλο των δύο μαζών μπορούν να ληφθούν υπόψη και φαινόμενα ανάδρασης του ακουστικού πεδίου στη γλωττίδα η κλασσική αυτή προσέγγιση θεωρείται ικανοποιητική για την προσομοίωση της παραγωγής φωνηέντων παρά του ότι, όπως συζητήθηκε και στο Κεφάλαιο 2, ενδεχόμενα αποκλίνει σημαντικά από την πραγματική φυσική λειτουργία του φωνητικού συστήματος.

Οι αποκλίσεις αυτές πάντως και η παραμέληση της μη ακουστικής αερορροής αποτρέπει τη σύνθεση ήχων όπως είναι οι τυρβώδεις για παράδειγμα, για το σχηματισμό των οποίων είναι απαραίτητη η θεώρηση αεροδυναμικών μηχανισμών όπως είδαμε και στο Κεφάλαιο 2. Δεδομένης της δυσκολίας μοντελοποίησης τέτοιων φαινομένων που θα απαιτούσε τη γνώση της λεπτομερούς τρισδιάστατης γεωμετρίας της φωνητικής οδού αλλά και τη γνώση του συνολικού πεδίου αερορροής, ως λύση ανάγκης γι' αυτές τις περιπτώσεις συνήθως εφαρμόζονται διάφορα ευριστικά φαινομενολογικά μοντέλα πηγών ήχου. Πρακτικά αυτό συνεπάγεται την εμφάνιση πηγών τυχαίας πίεσης ή ογκικής ταχύτητας στο ισοδύναμο ηλεκτρικό κύκλωμα στο σημείο όπου αναμένεται η εμφάνιση τύρβης ή γενικότερα σε κατάλληλο σημείο ώστε να είναι αποδεκτό το ακουστικό αποτέλεσμα. Εκτός από τη θέση βέβαια, και τα υπόλοιπα χαρακτηριστικά των πηγών αυτών, δηλαδή το φάσμα και η έντασή τους, ρυθμίζονται σε μεγάλο βαθμό ευριστικά ώστε ο παραγόμενος ήχος να έχει κατάλληλα φασματικά χαρακτηριστικά και ένταση. Χαρακτηριστικές είναι οι περιπτώσεις που έχουν δημοσιευτεί στα [56, 95, 103, 118, 151]. Ενδεικτικά, στο [151], αναφέρεται η εισαγωγή δύο πηγών θορύβου, μιας πηγής πίεσης σε σειρά στη γλωττίδα για τη μοντελοποίηση της τραχύτητας της φωνής και μιας πηγής ογκικής ταχύτητας παράλληλα στο ισοδύναμο ηλεκτρικό κύκλωμα λίγο μετά από το σημείο της μεγαλύτερης στένωσης της φωνητικής οδού. Η πρώτη πηγή μοντελοποιεί μια φυσική διπολική πηγή ήχου και έχει πλάτος ανάλογο της διαφοράς του τετραγώνου του τοπικού αριθμού Reynolds από ένα κατώφλι. Η δεύτερη μοντελοποιεί ένα φυσικό μονόπολο και προτιμήθηκε γιατί η θέση στην οποία τοποθετείται δεν είναι και τόσο σημαντική. Και οι δύο πηγές συνοδεύονται από εσωτερική αντίσταση για τη μοντελοποίηση ενδεχόμενων απωλειών

στη στένωση όπου εμφανίζεται η πηγή ήχου. Και οι δύο πηγές θορύβου εισάγουν λευκό θόρυβο ομοιόμορφα κατανεμημένο. Εναλλακτικά, στο [103] ο θόρυβος φιλτράρεται κατάλληλα ώστε το τελικό φάσμα να συμφωνεί με μετρήσεις σε πειράματα με μηχανικά ανάλογα [156]. Ενδιαφέρον έχει και η προσέγγιση στο [118] όπου η πηγή θορύβου έχει παραμετρικό φάσμα η μορφή του οποίου βελτιστοποιείται ώστε να είναι δυνατή η σύνθεση του ζητούμενου ήχου. Οι επιμέρους προσεγγίσεις δεν έχουν αξιολογηθεί συστηματικά προς το παρόν και πέρα από το αποδεκτό ακουστικό αποτέλεσμα δεν υπάρχει κάποια άλλη σύγκριση με το αντίστοιχο πραγματικό σήμα φωνής.

Για το προτεινόμενο υπολογιστικό μοντέλο θεωρείται ζητούμενο η κατά το δυνατόν ακριβέστερη προσέγγιση του φυσικού συστήματος και επιπλέον η αποφυγή ευριστικών τεχνικών όπως αυτών που αναφέρθηκαν προηγουμένως. Στόχος είναι να αξιοποιηθεί, όσο αυτό είναι υπολογιστικά δυνατό, η γνώση σχετικά με το συνολικό πεδίο αερορροής μέσα στο φωνητικό σωλήνα. Σε αυτό το πλαίσιο αναπτύσσεται ένα μοντέλο αερορροής μέσα στο σωλήνα, βλ. Ενότητα 4.2. Το μοντέλο αυτό περιλαμβάνει τόσο το στροβιλώδες όσο και το αστρόβιλο πεδίο υιοθετώντας την προσέγγιση που περιγράφεται στα [91, 150]. Για το αστρόβιλο πεδίο συνδυάζει κατάλληλα και το μοντέλο δύο μαζών για τη γλωττίδα ώστε να μοντελοποιηθούν φαινόμενα αλληλεπίδρασης της αερορροής με τη φωνητική οδό [108]. Το μοντέλο για το αστρόβιλο πεδίο είναι χαμηλόσυχνο και θεωρείται ότι δίνει τη λεγόμενη μέση ροή. Οι πηγές ήχου που εισάγονται στο ισοδύναμο ηλεκτρικό κύκλωμα υπαγορεύονται πλήρως από την αλληλεπίδραση του στροβιλώδους με το αστρόβιλο πεδίο όπως υπαγορεύεται από την αεροακουστική θεωρία, βλ. Ενότητα 4.5. Ενδεχόμενη ανάδραση από το ακουστικό πεδίο θεωρείται αμελητέα.

Μελετάται επίσης πιθανή επίδραση της μέσης (αναπνευστικής) ροής στο ακουστικό πεδίο. Τυπικές μετρήσεις της ογκικής ταχύτητας που αντιστοιχεί σε αυτή τη ροή κατά τη διάρκεια της παραγωγής φωνής [10] ενισχύουν την ανάγκη διερεύνησης σχετικών φαινομένων. Προηγούμενες προσεγγίσεις για σύνθεση με αρθρωτές (articulatory synthesis) που λαμβάνουν υπόψη το πλήρες πεδίο ροής του αέρα έχουν αναφερθεί στα [22, 26, 39, 72], όπου προτείνεται η λύση των αντίστοιχων μη-γραμμικών εξισώσεων. Για την μείωση της υπολογιστικής πολυπλοκότητας υποθέτουμε ότι η μέση ροή του αέρα στην φωνητική οδό δεν επηρεάζεται από την αντίστοιχη ακουστική ροή. Έτσι, είναι δυνατόν να αποπλέξουμε τις εξισώσεις που αφορούν στις ακουστικές διαταραχές από τις αντίστοιχες εξισώσεις για την μέση ροή. Η προσέγγιση αυτή υποκινήθηκε από την ανάλυση στο [183], όπου εφαρμόστηκε ευθεία αριθμητική ανάλυση για τον υπολογισμό του πεδίου μέσης ροής, σε δύο διαστάσεις, και ένα ακουστικό ανάλογο για την πρόβλεψη του εκπεμπόμενου ήχου, επίσης σε δύο διαστάσεις. Διερευνούμε την δυνατότητα για παρόμοιες προσεγγίσεις σε ένα αριθμητικά απλούστερο σύστημα σύνθεσης με αρθρωτές.

Για τη γλωττίδα θεωρείται ένα βελτιωμένο μοντέλο δύο μαζών με βάση προτάσεις που έγιναν στα [97, 108, 125]. Συγκεκριμένα, στο νέο μηχανικό μοντέλο η επιφάνεια της κάθε φωνητικής χορδής είναι συνεχής ώστε να μπορεί να δικαιολογηθεί και μετακινούμενο σημείο αποκόλλησης της ροής. Επιπλέον, περιγράφεται κατάλληλα και η τρίτη διάσταση ώστε η συμπεριφορά της γλωττίδας κατά την παραγωγή άφωνων ήχων να είναι περισσότερο φυσική. Ανάλογα έχει βελτιωθεί και το αεροδυναμικό μοντέλο και περιγράφει καλύτερα κάποια φαινόμενα που σχετίζονται με το συνοριακό στρώμα της αερορροής στη γλωττίδα και με το διαχωρισμό της κατά την είσοδο στη φωνητική οδό. Ο προσδιορισμός των ακουστικών πηγών γίνεται με βάση την αεροακουστική όπως περιγράφεται στα [71, 91]. Για τη σύνθεση ακολουθιών φωνηέντων-τυρβώδων ήχων χρησιμοποιήθηκαν πραγματικά δεδομένα όπως αυτά έχουν μετρηθεί με τη χρήση εικόνων αξονικής τομογραφίας και έχουν δημοσιευτεί στα [119, 158].

4.2 Αεροδυναμική μοντελοποίηση για τη φωνητική οδό

Η πολυπλοκότητα της τρισδιάστατης γεωμετρίας της φωνητικής οδού έχει καταστήσει πρακτικά αδύνατη την επίλυση των εξισώσεων αεροδυναμικής μέσα στο φωνητικό σωλήνα στη

γενική περίπτωση [148]. Στις περισσότερες των περιπτώσεων θεωρείται κάποια απλοποιημένη γεωμετρία ή επικεντρώνεται το ενδιαφέρον σε κάποιο συγκεκριμένο τμήμα όπως είναι η γλωττίδα ή κάποια στένωση.

Μια προσπάθεια επίλυσης των διδιάστατων εξισώσεων για μια εξιδανικευμένη γεωμετρία και εξιδανικευμένες οριακές συνθήκες αναφέρεται από τον Thomas στο [169] με τη χρήση πεπερασμένων στοιχείων. Αναφέρονται δύο είδη πειραμάτων, στατικά πειράματα, με στατική δηλαδή ροή στην είσοδο και πειράματα χρονομεταβλητά, όπου η είσοδος θεωρείται ότι είναι ένα χρονομεταβλητό τζετ που προσομοιώνει την παροχή όγκου στη γλωττίδα. Τα τοιχώματα θεωρούνται σταθερά ενώ η έξοδος της φωνητικής οδού θεωρείται ότι είναι ανοιχτή (σταθερή πίεση, δεν έχει γίνει προσπάθεια να προσομοιωθεί κάποια συνθήκη εκπομπής). Σημειώνεται ότι λύνονται οι εξισώσεις για τις δύο διαστάσεις και όχι τις τρεις, αφού όπως αναφέρεται κάτι τέτοιο θα ήταν απαγορευτικό υπολογιστικά. Για τα στατικά πειράματα, αρχικά δίνεται εικόνα των ροϊκών γραμμών του πεδίου για την περίπτωση αξονικά συμμετρικού τζετ στην είσοδο. Ενδιαφέρον παρουσιάζει η αποκόλληση της ροής σε μια πλάτυνση της φωνητικής οδού στην περιοχή της επιγλωττίδας. Η ροή προσκολλάται στη γλώσσα ενώ φαίνεται ότι εμφανίζεται και περιοχή επανακυκλοφόρησης για απόσταση περίπου $2cm$. Περισσότερο ενδιαφέρον παρουσιάζει η εικόνα της ροής για μη συμμετρικό τζετ στην είσοδο και στο σημείο προσομοίωσης των δοντιών οπότε και πάλι έχουμε προσκόλληση της ροής στο εσωτερικό λαρυγγοφαρυγγικό τοίχωμα ενώ στα δόντια σχηματίζεται ένα τζετ. Γίνεται εμφανής η γέννηση στροβίλων. Για την περίπτωση χρονομεταβλητής εισόδου δίνονται αποτελέσματα πειραμάτων με εξιδανικευμένα μοντέλα, χωρίς να προκύπτει κάποιο ιδιαίτερα ενδιαφέρον συμπέρασμα πέρα ίσως από το ότι η οριακή συνθήκη για μηδενική πίεση στην έξοδο ήταν ανεπαρκής. Τέλος, γίνεται προσπάθεια για σύνθεση φωνής, και πιο συγκεκριμένα ενός διφθόγγου, χωρίς όμως να δίνονται πολλές λεπτομέρειες. Ενδιαφέρον έχει το γεγονός ότι για 1 δευτερόλεπτο συνθετικής φωνής χρειάστηκαν περίπου 100 ώρες υπολογισμών. Δεν γίνεται παρ' όλ' αυτά αναφορά στο πώς άλλαξε η οριακή συνθήκη για τη σύνθεση φωνής. Μάλλον τα αποτελέσματα που δίνονται αφορούν σε ογκική ταχύτητα και όχι σε ακουστική πίεση. Σημειώνεται επίσης ότι δεν υπάρχει αλληλεπίδραση ακουστικού-μη ακουστικού πεδίου στη γλωττίδα δεδομένου του ότι το τζετ στην είσοδο θεωρείται συγκεκριμένο. Στο [72] αναφέρεται η αριθμητική επίλυση των εξισώσεων Reynolds Averaged Navier Stokes με στόχο τη σύνθεση φωνής χωρίς όμως να γίνεται ιδιαίτερη αναφορά στο πεδίο της ροής.

Για την περιοχή της γλωττίδας έχουν παρουσιαστεί κατά καιρούς διάφορες ερευνητικές εργασίες με στόχο τον προσδιορισμό του πεδίου ροής. Δεδομένου του ότι γίνονται διάφορες υποθέσεις για τη γεωμετρία αλλά και για τις επικρατούσες συνθήκες, πρόκειται πρακτικά για εναλλακτικές μοντελοποιήσεις. Ιδιαίτερο ενδιαφέρον παρουσιάζουν οι πιο πρόσφατες εργασίες στα [15, 161, 162, 180, 183]. Στα [180, 183] υποτίθεται διδιάστατο αξονικά συμμετρικό πεδίο και εφαρμόζεται άμεσο αριθμητικό σχήμα για την επίλυση των εξισώσεων συμπιεστού ρευστού για διάφορες χρονομεταβλητές γεωμετρίες. Αμελείται η γένεση τύρβης. Στο [15] θεωρείται τρισδιάστατο, επίσης ομαλό, αξονικά συμμετρικό πεδίο και εφαρμόζεται αριθμητικό σχήμα επίλυσης για τις εξισώσεις ασυμπίεστου ρευστού. Αξίζει να σημειωθεί ότι βασικό μέλημα αυτών των ερευνητικών προσπαθειών είναι και ο προσδιορισμός του σχετιζόμενου ακουστικού πεδίου. Για αυτό το σκοπό υιοθετούνται διαφορετικές προσεγγίσεις με παρόμοια σε γενικές γραμμές αποτελέσματα. Στα [161, 162] διερευνάται αριθμητικά η ύπαρξη τύρβης στην περιοχή της γλωττίδας και η ενδεχόμενη επίδρασή της στο ακουστικό πεδίο. Η επίδραση της γεωμετρίας για τη γένεση τύρβης στην περίπτωση στένωσης μελετάται στο [135]. Στα [77, 78] παρουσιάζονται αποτελέσματα χρησιμοποιώντας μοντέλα μικρότερης τάξης για τη γλωττίδα χωρίς όμως να διερευνάται ο παραγόμενος ήχος.

Αποτελέσματα και συμπεράσματα που προκύπτουν μέσω αριθμητικών προσομοιώσεων είναι ιδιαίτερα σημαντικά για την περίπτωση της φωνητικής οδού. Παρά του ότι λόγω της αυξημένης τους πολυπλοκότητας δεν μπορούν να χρησιμοποιηθούν ευρέως για σύνθεση φωνής για παράδειγμα, βοηθούν γενικά στην επισήμανση διάφορων κρίσιμων φαινομένων για την

παραγωγή φωνής. Σημαντική βέβαια είναι και η συνεισφορά άμεσων πειραματικών μετρήσεων που είτε αφορούν στην ίδια τη φωνητική οδό [164] και είναι γενικότερα περιορισμένες λόγω εγγενών δυσκολιών ή σε μηχανικά ανάλογα [19, 41, 140, 149].

Στόχος του προτεινόμενου υπολογιστικού μοντέλου φωνής είναι να ενσωματώσει σε ένα ενιαίο πλαίσιο αεροδυναμικά φαινόμενα που είναι σημαντικά για τη φωνή. Τα συμβατικά συστήματα σύνθεσης φωνής με αρθρωτές συνήθως υιοθετούν ένα απλοϊκό αεροδυναμικό μοντέλο. Για τη γλωττίδα, συχνά εφαρμόζεται το μοντέλο των δύο μαζών των Ishizaka, Flanagan [73] που βασίζεται σε αρκετά απλοποιητικές παραδοχές για το πραγματικό πεδίο. Για το εσωτερικό της φωνητικής οδού, ενδεικτική είναι η προσέγγιση των Flanagan και Cherry [56] που πρακτικά δε διαφοροποιούν το ακουστικό από το μη ακουστικό πεδίο. Για την περιγραφή της ηχητικής πηγής σε στενώσεις, όπου απαιτείται στην ουσία η χρήση της μη ακουστικής ροής, χρησιμοποιούν μια εξομαλυμένη εκδοχή του πεδίου ογκικής ταχύτητας, όπως αυτό δίνεται μέσω της ακουστικής προσομοίωσης. Η πρακτική αυτή με μικρές παραλλαγές ακολουθείται και από μεταγενέστερους συνθέτες φωνής με τη χρήση αρθρωτών [24, 26, 151]. Αντίθετα, στις ερευνητικές εργασίες της Scully [146] και του Maeda [103] αναφέρεται η εφαρμογή ξεχωριστού αεροδυναμικού μοντέλου που όπως περιγράφεται ισχύει για χαμηλές συχνότητες. Πρακτικά, το πεδίο θεωρείται μονοδιάστατο και επιλύεται ένα ανεξάρτητο ανάλογο ηλεκτρικό κύκλωμα για τη λεγόμενη μέση ροή, ή αλλιώς τη μη ακουστική ογκική ταχύτητα (βλ. και Ενότητα 2.5.2). Σε όλες αυτές τις περιπτώσεις το μη ακουστικό πεδίο λαμβάνεται πρακτικά ως μονοδιάστατο και αστρόβιλο. Η μόνη προσπάθεια να αξιοποιηθεί για πρακτική σύνθεση φωνής μια πλουσιότερη περιγραφή του πεδίου παρουσιάζεται στο [150], όπου περιγράφεται μοντελοποίηση και της στροβιλώδους ροής, τουλάχιστον στην περιοχή των στενώσεων.

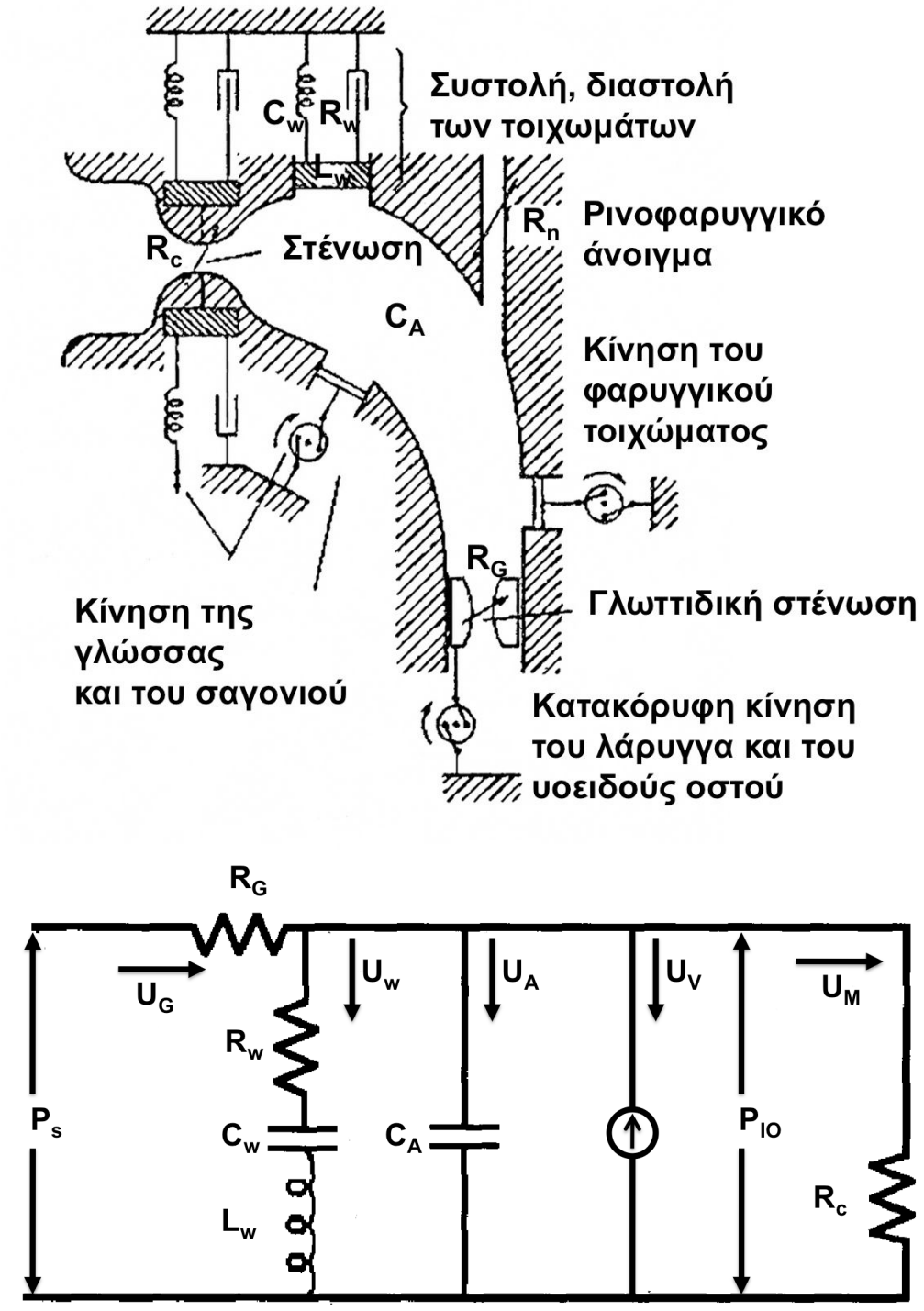
Το αεροδυναμικό μοντέλο που παρουσιάζεται στη συνέχεια, όπως στο [150] και σύμφωνα με το πλαίσιο που τίθεται στο [91], διακρίνει ανάμεσα σε αστρόβιλο και σωληνοειδές πεδίο. Το αστρόβιλο πεδίο είναι συμπιεστό και είναι απαραίτητο για τη μετάδοση της ακουστικής κίνησης, ενώ το στροβιλώδες πεδίο, αλληλεπιδρώντας με το πρώτο μπορεί να προκαλέσει τη γένεση ήχου, όπως σκιαγραφείται στην Ενότητα 2.4.

4.2.1 Αστρόβιλο πεδίο

Κατά την παραγωγή φωνής, οπότε εμφανίζονται μικροί αριθμοί Mach, η αστρόβιλη (δυναμική) ροή μπορεί να θεωρηθεί σχετικά ασυμπίεστη κι έτσι ο λεπτομερής υπολογισμός της απαιτεί την επίλυση ενός προβλήματος Laplace, $\nabla^2\phi = 0$ με κατάλληλες οριακές συνθήκες [6]. Προκειμένου να αποφευχθεί η απαιτούμενη αυξημένη υπολογιστική πολυπλοκότητα υιοθετείται μοντελοποίηση του πεδίου που συνδυάζει μια απλουστευμένη περιγραφή της κατεύθυνσής του σε κάθε σημείο της φωνητικής οδού [150] με προσδιορισμό του μέτρου του μέσω ενός κατάλληλα προσαρμοσμένου ανεξάρτητου ισοδύναμου ηλεκτρικού κυκλώματος [108].

4.2.1.1 Προσδιορισμός της έντασης

Συγκεκριμένα, το ισοδύναμο ηλεκτρικό κύκλωμα έχει τα βασικά χαρακτηριστικά ανάλογων μοντέλων (π.χ. [120, 138]) και περιλαμβάνει συγκεντρωμένα στοιχεία για τη μοντελοποίηση της πτώσης πίεσης (P_{io}, R_c) σε στενώσεις της φωνητικής οδού, των απωλειών λόγω ελαστικών τοιχωμάτων (U_w, C_w, L_w, R_w), της παροχής όγκου U_v λόγω μεταβολής του σχήματος και της ακουστικής συμπιεστότητας του αέρα, C_A (βλ. και Σχήμα 4.1). Η ύπαρξη μιας στένωσης στη φωνητική οδό θεωρείται ότι είναι χαρακτηριστικό της άρθρωσης για την παραγωγή πολλών ήχων, όπως τυρβώδεις ή εκρηκτικοί ενώ στην περίπτωση φωνηέντων, όπου συνήθως η στένωση είναι αρκετά μεγάλης διατομής, η αντίστοιχη αντίσταση είναι αμελητέα.



Σχήμα 4.1: *Διαγραμματική αναπαράσταση της φωνητικής οδού και του λάρυγγα [138]. Το ισοδύναμο ηλεκτρικό κύκλωμα που θεωρήθηκε.*

Ταυτοποίηση του κυκλώματος και επίλυση Για τον προσδιορισμό της αερορροής στη γλωττίδα, ακολουθείται η ανάλυση που προτάθηκε στο [73] για το μοντέλο δύο μαζών. Η υπόθεση είναι ότι εμβαδό διατομής είναι σταθερό για όλο το πάχος της γλωττίδας και ίση με A_g. Στην είσοδο της γλωττίδας, λόγω της απότομης αλλαγής διατομής από την τραχεία θεωρείται ότι εμφανίζεται μια περιοχή vena contracta που στην ουσία επιτείνει τα αποτελέσματα του φαινομένου Bernoulli και αντιστοιχεί σε αντίσταση

$$R_{vc} = 1.37\frac{\rho}{2}\frac{|U_g|}{A_g^2}$$

ενώ στην έξοδο θεωρείται ότι δεν έχουμε πλήρη ανάνηψη της πίεσης στο άνοιγμα προς τη φωνητική οδό με αποτέλεσμα η αντίσταση που εμφανίζεται εκεί να είναι :

$$R_e = -\frac{\rho}{2}\frac{2}{A_g A_{vt}}(1 - \frac{A_g}{A_{vt}})|U_g|,$$

όπου A_{vt} είναι το εμβαδό της εγκάρσιας διατομής του ανοίγματος. Σε όλο το πάχος της γλωττίδας λαμβάνονται απώλειες λόγω ιξώδους που για στένωση ορθογώνιας διατομής όπως υποτίθεται ότι είναι η γλωττίδα μπορούν να αναπαρασταθούν με την αντίσταση :

$$R_v = 12\mu\frac{l_g}{\pi d_g^4}$$

όπου l_g, d_g είναι το μήκος και το πάχος της γλωττίδας αντίστοιχα ενώ μ είναι ο συντελεστής ιξώδους του αέρα. Η συνολική οπότε αντίσταση της γλωττίδας λαμβάνεται ως $R_g = R_{vc} + R_e + R_v$. Ανάλογα, οι απώλειες στη στένωση της φωνητικής οδού, που θεωρείται ότι έχει διατομή κυκλικού σχήματος, εκφράζονται ως [155] :

$$P_{io} = 128\mu\frac{l_c}{\pi d^4}U_m + k_L(\frac{\rho}{2}\frac{U_m^2}{A^2}),$$

όπου U_m είναι η ογκική ταχύτητα της μη ακουστικής ροής στο εσωτερικό της φωνητικής οδού και l_c το μήκος της στένωσης. Η χωρητικότητα C_A εκφράζει τη συμπιεστότητα του αέρα και λαμβάνεται ίση με :

$$C_A = \frac{V_{vt}}{\rho c_0^2}$$

όπου V_{vt} είναι ο όγκος του αέρα στο εσωτερικό της φωνητικής οδού και c_0 είναι η ταχύτητα του ήχου. Για τα χαρακτηριστικά των ελαστικών τοιχωμάτων, δηλαδή τη συγκεντρωμένη αντίστασή τους λόγω συνεκτικότητας, που εκφράζεται μέσω της αντίστασης R_w, τη συγκεντρωμένη υποχωρητικότητά τους, που εκφράζεται μέσω της χωρητικότητας C_w, και τη συγκεντρωμένη μάζα τους, που εκφράζεται μέσω της επαγωγής L_w χρησιμοποιούμε τις τιμές που προτείνονται στο [108]. Είναι $C_W = 2.5458 \times 10^{-9}$ m^5/Nt, $R_W = 18.56 \times 10^5$ $\mathrm{Nt{\cdot}sec/m}^5$ και $L_W = 1.92 \times 10^3$ $\mathrm{Nt{\cdot}sec}^2/\mathrm{m}^5$.

Στη συνέχεια εφαρμόζεται η αρχή διατήρησης της μάζας μέσα στη φωνητική οδό και οι εξισώσεις διακριτοποιούνται κατάλληλα με βάση τον κανόνα του τραπεζίου, βλ. Ενότητα 3.4.3.2, Εξίσωση (3.71). Έτσι, προκύπτει τελικά το παρακάτω σύστημα μη γραμμικών εξισώσεων για κάθε χρονική στιγμή :

$$\frac{\rho}{2}\left(\frac{1.37}{A_g^2} - \frac{2}{A_g A_{vt}}(1 - \frac{A_g}{A_{vt}})\right)|U_g|U_g + R_v U_g - P_s + P_{io} = 0 \tag{4.1}$$

$$k_L(\frac{\rho}{2}\frac{U_m^2}{A^2}) + 128\mu\frac{l}{\pi d^4}U_m - P_{io} = 0 \tag{4.2}$$

$$b_{io}U_g - b_{io}U_m - P_{io} - b_{io}\Phi = 0, \tag{4.3}$$

όπου $b_{io} = 1/(2Y_a + Y_w)$ με

$$Y_a = F_{sim}C_A, Y_w = \frac{1}{2F_{sim}L_W + R_W + \frac{1}{2F_{sim}C_w}}$$

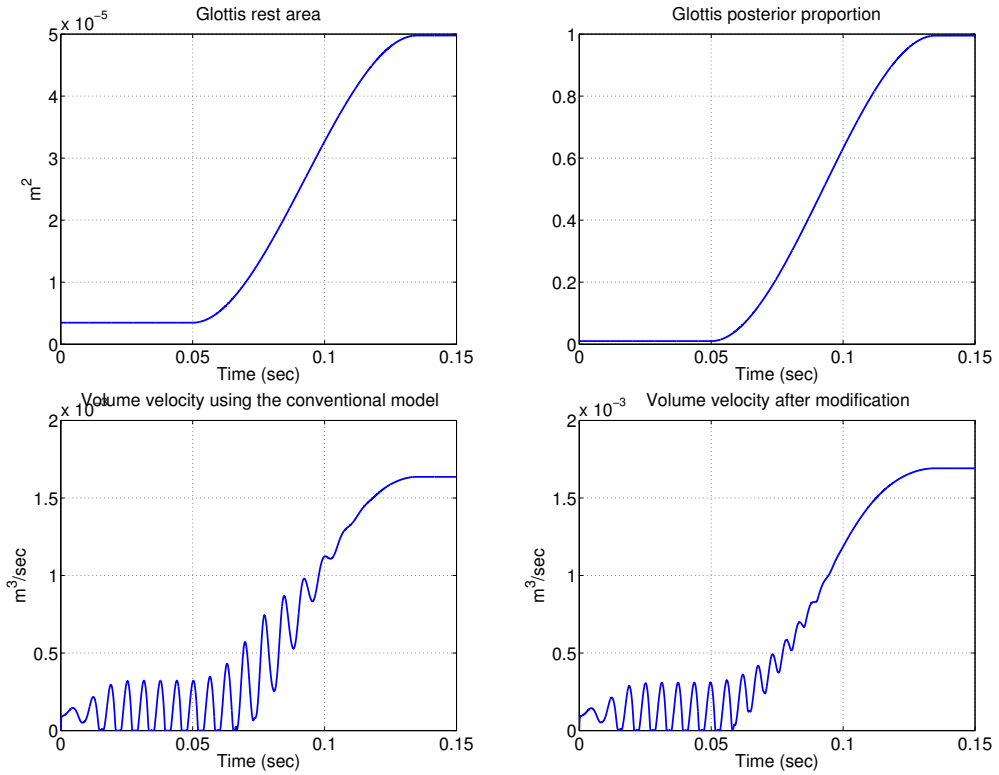
και F_{sim} είναι η συχνότητα της αριθμητικής προσομοίωσης. Ο όρος Φ περιλαμβάνει τη συνεισφορά της ογκικής ταχύτητας λόγω μεταβολής του σχήματος και λόγω της δόνησης των τοιχωμάτων, όπως δίνεται από την Εξίσωση (3.73) και εξαρτάται και από την προηγούμενη χρονική στιγμή. Το σύστημα επιλύεται αριθμητικά με τη μέθοδο Newton, [34, Ενότητα 5.2].

Διεπαφή με το μοντέλο δύο μαζών Για τον προσδιορισμό του ανοίγματος της γλωττίδας κάθε χρονική στιγμή επιστρατεύεται το μοντέλο των δύο μαζών ώστε να συμπεριληφθούν και ενδεχόμενα φαινόμενα αλληλεπίδρασης της φωνητικής οδού με το λάρυγγα. Η διεπαφή μεταξύ των δύο μοντέλων επιτυγχάνεται πρακτικά ακολουθώντας τη στρατηγική που περιγράφεται στο [108]. Δεδομένου του ότι το αεροδυναμικό μοντέλο για τη δυναμική ροή έχει ισχύ μόνο σε χαμηλές συχνότητες, θεωρείται ότι η ροή U_g είναι μια εξομαλυμένη έκδοση της γλωττιδικής ροής που προκύπτει από το μοντέλο των δύο μαζών. Συγκεκριμένα, η απαίτηση είναι η ροή μέσα στο αεροδυναμικό μοντέλο της φωνητικής οδού να μη διαφέρει περισσότερο από 4% ή από 0.04lt/sec από τη ροή του μοντέλου των δύο μαζών. Κάθε χρονική στιγμή η πίεση στο εσωτερικό της φωνητικής οδού, όπως υπολογίζεται από το αεροδυναμικό μοντέλο, δίνεται ως παράμετρος του μοντέλου των δύο μαζών και προσδιορίζεται η ογκική ταχύτητα U_{gmm}. Ενημερώνεται η μέση της τιμή και αν αυτή διαφέρει αρκετά από την προηγούμενη τιμή της γλωττιδικής ταχύτητας στο χαμηλόσυχνο μοντέλο U_g τότε προσδιορίζεται κατάλληλο άνοιγμα γλωττίδας ώστε να αρθεί αυτή η απόκλιση. Η διαδικασία είναι επαναληπτική [1] γιατί στη συνέχεια υπολογίζεται καινούρια πίεση P_{io} και κατ' επέκταση καινούρια τιμή για τη ροή U_{gmm}. Τελικά, η προσομοίωση επαναλαμβάνεται από την αρχή αφού πρώτα εξομαλυνθεί η μεταβολή του γλωττιδικού ανοίγματος που προσδιορίστηκε αρχικά ώστε να αποφευχθεί η εμφάνιση πιθανών υψίσυχνων φαινομένων. Σημειώνεται ότι η εξομάλυνση είναι προαπαιτούμενο για όλες τις εμπλεκόμενες παραμέτρους άρθρωσης, όπως για παράδειγμα για τις συναρτήσεις εμβαδού εγκάρσιας διατομής. Πραγματοποιείται με ένα φίλτρο κινούμενου μέσου με μήκος που αντιστοιχεί σε $40ms$.

Μοντελοποίηση της τρίτης διάστασης του μοντέλου των δύο μαζών Για την προσομοίωση καταστάσεων στις οποίες δεν υπάρχει φώνηση είναι σημαντική η κατάλληλη παραμετροποίηση του μοντέλου των δύο μαζών για τη γλωττίδα ώστε να καθίσταται δυνατή η διακοπή της ταλάντωσης των φωνητικών χορδών και η ανάπτυξη αυξημένης δυναμικής ροής μέσα στη φωνητική οδό. Όπως διαπιστώνεται στο [73] αυτό μπορεί να επιτευχθεί με κατάλληλη επιλογή του ανοίγματος ηρεμίας της γλωττίδας A_{g0}. Όταν το άνοιγμα γίνει αρκετά μεγάλο, λόγω των χαρακτηριστικών απόσβεσης των ελατηρίων του μοντέλου, σταματάει η ταλάντωση. Στο [57] και σε σχετικές ερευνητικές εργασίες χρησιμοποιείται αυτή ακριβώς η παράμετρος άρθρωσης για να ελέγξει τη διακοπή και την έναρξη της φώνησης για παράδειγμα κατά τη σύνθεση ακολουθιών έμφωνων-άφωνων ήχων. Έτσι, το εμβαδό του ανοίγματος ηρεμίας μπορεί να μεταβάλλεται από 0.05 ως 0.5 cm^2 ενώ το σημείο στο οποίο σταματάει η φώνηση εξαρτάται και από τις παραμέτρους απόσβεσης του μοντέλου των δύο μαζών αλλά και την υπογλωττιδική πίεση.

Το πρόβλημα που εντοπίζεται και σχετίζεται με την αεροδυναμική μοντελοποίηση αφορά στην παρατήρηση ότι στο κλασσικό μοντέλο δύο μαζών καθώς αυξάνεται το άνοιγμα ηρεμίας αυξάνεται και το πλάτος των ταλαντώσεων της γλωττιδικής ογκικής ταχύτητας, που είναι ασύμβατο με πειραματικές μετρήσεις της ογκικής ταχύτητας κατά την ανθρώπινη παραγωγή φωνής . Για αυτό το λόγο, εφαρμόστηκε η τροποποίηση του μοντέλου της γλωττίδας που

[1] Ο αλγόριθμος που χρησιμοποιείται αποτελεί μια μικρή παραλλαγή του αλγόριθμου βελτιστοποίησης gradient descent.



Σχήμα 4.2: *Γλωττιδική ογκική ταχύτητα καθώς διακόπτεται η φώνηση με τη χρήση είτε του κλασσικού μοντέλου δύο μαζών είτε του τροποποιημένου. Στην πάνω σειρά δίνονται οι αντίστοιχες παράμετροι άρθρωσης.*

προτείνεται στο [108]. Προτιμήθηκε αυτή η εναλλακτική από αυτή που προτείνεται στο [98] και περιλαμβάνει τροποποίηση των παραμέτρων απόσβεσης. Συγκεκριμένα, η ιδέα είναι να θεωρηθεί ότι η ταλάντωση δεν είναι ομοιόμορφη σε όλο το μήκος l_g. Θεωρείται ότι το πίσω τμήμα της γλωττίδας γενικά έχει μεγαλύτερο άνοιγμα ηρεμίας (θεωρείται ότι $A_p = 0.5$ cm^2) και δεν ταλαντώνεται. Για το υπόλοιπο τμήμα που ταλαντώνεται, η προσομοίωση τροποποιείται κατάλληλα ώστε να ληφθεί υπόψη το περιορισμένο του μήκος. Οι ταλαντούμενες μάζες, για παράδειγμα, θεωρούνται ποσοστό των συνολικών. Σε κατάσταση φώνησης το πίσω τμήμα θεωρείται ότι έχει πολύ μικρό μήκος ίσο με περίπου 0.5% του συνολικού μήκους της γλωττίδας. Για τη διακοπή της φώνησης το μήκος του μη ταλαντώμενου κομματιού αυξάνεται για να φτάσει τελικά το 100% του συνολικού μήκους, οπότε και η ταλάντωση της γλωττίδας σταματάει τελείως. Η συγκεκριμένη τροποποίηση δεν αποτελεί απλά έναν ευριστικό τρόπο αντιμετώπισης του προβλήματος αλλά αντικατοπτρίζει και τη συμπεριφορά της γλωττίδας όπως αυτή παρατηρείται με τη βοήθεια γλωττογραφίας. Η νέα παράμετρος άρθρωσης είναι το ποσοστό α του μήκους του μη ταλαντούμενου τμήματος προς το ταλαντούμενο. Η σχέση της νέας παραμέτρου με την προηγούμενη είναι :

$$A_{g0} = \alpha A_p + (1 - \alpha) A_a$$

όπου A_a είναι το άνοιγμα ηρεμίας του ταλαντώμενου μέρους ($A_a = 0.03$ cm^2).

Στο Σχήμα 4.2 φαίνεται η γλωττιδική ογκική ταχύτητα όπως προκύπτει για τα δύο μοντέλα στην περίπτωση ισοδύναμης μεταβολής της αντίστοιχης παραμέτρου άρθρωσης. Σημειώνεται ότι για λόγους αριθμητικής ευστάθειας η ελάχιστη τιμή της παραμέτρου α είναι 0.005. Διαπιστώνεται ότι πράγματι επιτυγχάνεται η αύξηση του μέσου μέτρου της ροής χωρίς να αυξάνεται και το πλάτος των ταλαντώσεων.

Εφαρμογή του αεροδυναμικού μοντέλου για την ένταση του δυναμικού πεδίου Το αεροδυναμικό μοντέλο για τον προσδιορισμό της έντασης του δυναμικού πεδίου εφαρμόστη-

κε για επαλήθευση στην περίπτωση εκφώνησης μιας ακολουθίας /AsA/. Οι συναρτήσεις εμβαδού είναι αυτές που αντιστοιχούν σε αρσενικό ομιλητή και έχουν δημοσιευτεί στο [159] για το /A/ και στο [119] για το /s/. Οι παράμετροι άρθρωσης που χρησιμοποιήθηκαν είναι η υπογλωττιδική πίεση P_s, το ποσοστό α του μήκους του μη ταλαντούμενου τμήματος της γλωττίδας και ο παράγοντας Q της τάσης της γλωττίδας που σχετίζεται με τη θεμελιώδη συχνότητα ταλάντωσης. Η μεταβολή των παραμέτρων άρθρωσης είναι όπως στο [108] ώστε τα αποτελέσματα να είναι συγκρίσιμα και να μπορούν να χρησιμοποιηθούν ως αναφορά και οι πραγματικές αεροροές που έχουν μετρηθεί και επίσης δημοσιεύονται [96]. Σημειώνεται ότι η αεροροή που μετράται στο στόμα μπορεί να συγκριθεί με αυτή που προκύπτει από την προσομοίωση μετά την εφαρμογή ενός εξομαλυντικού φίλτρου ώστε να αφαιρεθεί η συνιστώσα του ακουστικού πεδίου από τη μέτρηση. Ανάλογες μετρήσεις έχουν δημοσιευτεί και στο [148] για την ακολουθία /isi/. Τα προκύπτοντα αποτελέσματα είναι κοντά στα δημοσιευμένα αποτελέσματα και στις μετρήσεις (βλ. Σχήμα 4.3). Χαρακτηριστική είναι βέβαια η διαφορά που εμφανίζεται κατά τη μετάβαση από το /s/ στο /A/ (βλ. Σχήμα 4.4). Η διαφορά αυτή είναι μικρότερη στην προσομοίωση του McGowan. Φαίνεται από το διάγραμμα του εμβαδού της στένωσης που δίνεται στο [108] για την προσομοίωση ότι έχει ληφθεί υπόψη κάποιο φαινόμενο συνάρθρωσης που εμείς δεν έχουμε συμπεριλάβει. Οι αποκλίσεις θα μπορούσαν να δικαιολογηθούν και από το γεγονός ότι οι συναρτήσεις εμβαδού που χρησιμοποιούνται αν και πραγματικές, δεν είναι αυτές που αντιστοιχούν στις μετρήσεις και μάλιστα είναι από διαφορετικό ομιλητή για κάθε φώνημα. Επιπλέον, και ο συγχρονισμός των παραμέτρων άρθρωσης δεν είναι και ο πλέον ακριβής.

4.2.1.2 Προσδιορισμός της κατεύθυνσης

Για τον προσδιορισμό της κατεύθυνσης του δυναμικού πεδίου ροής, υποτίθεται διδιάστατη αξονικά συμμετρική γεωμετρία για τη φωνητική οδό. Τα σωματίδια του αέρα πρακτικά θεωρείται ότι κινούνται σε διευθύνσεις που ακολουθούν το εξομαλυμένο σχήμα των τοιχωμάτων, οπότε οι ροϊκές γραμμές του πεδίου, εφαπτομενικές στο διάνυσμα ταχύτητας σε κάθε σημείο του, γίνεται δυνατό να προσδιοριστούν κατά προσέγγιση με βάση μόνο τη γεωμετρία [150]. Αυτό μπορεί να γίνει με την υπόθεση ότι για κάθε ξεχωριστό διάστημα διακριτοποίησης, στο οποίο το σχήμα του τοιχώματος μπορεί να θεωρηθεί γραμμικό, οι γραμμές και τα τοίχωματα έχουν το ίδιο εστιακό σημείο πάνω στον άξονα της φωνητικής οδού.

Η χωρική διακριτοποίηση είναι σημαντικό να είναι αρκετά λεπτομερής αφού οι όποιες μεταβολές της γεωμετρίας επηρεάζουν το πεδίο ροής. Αν η διακριτοποίηση του άξονα συμμετρίας είναι

$$x_1, x_2, \ldots x_n, x_{n+1}, \ldots, x_N$$

και σε κυλινδρικές συντεταγμένες οι αντίστοιχες ακτινικές αποστάσεις του ενός τοιχώματος είναι

$$r_1, r_2, \ldots r_n, r_{n+1}, \ldots, r_N, r_i > 0$$

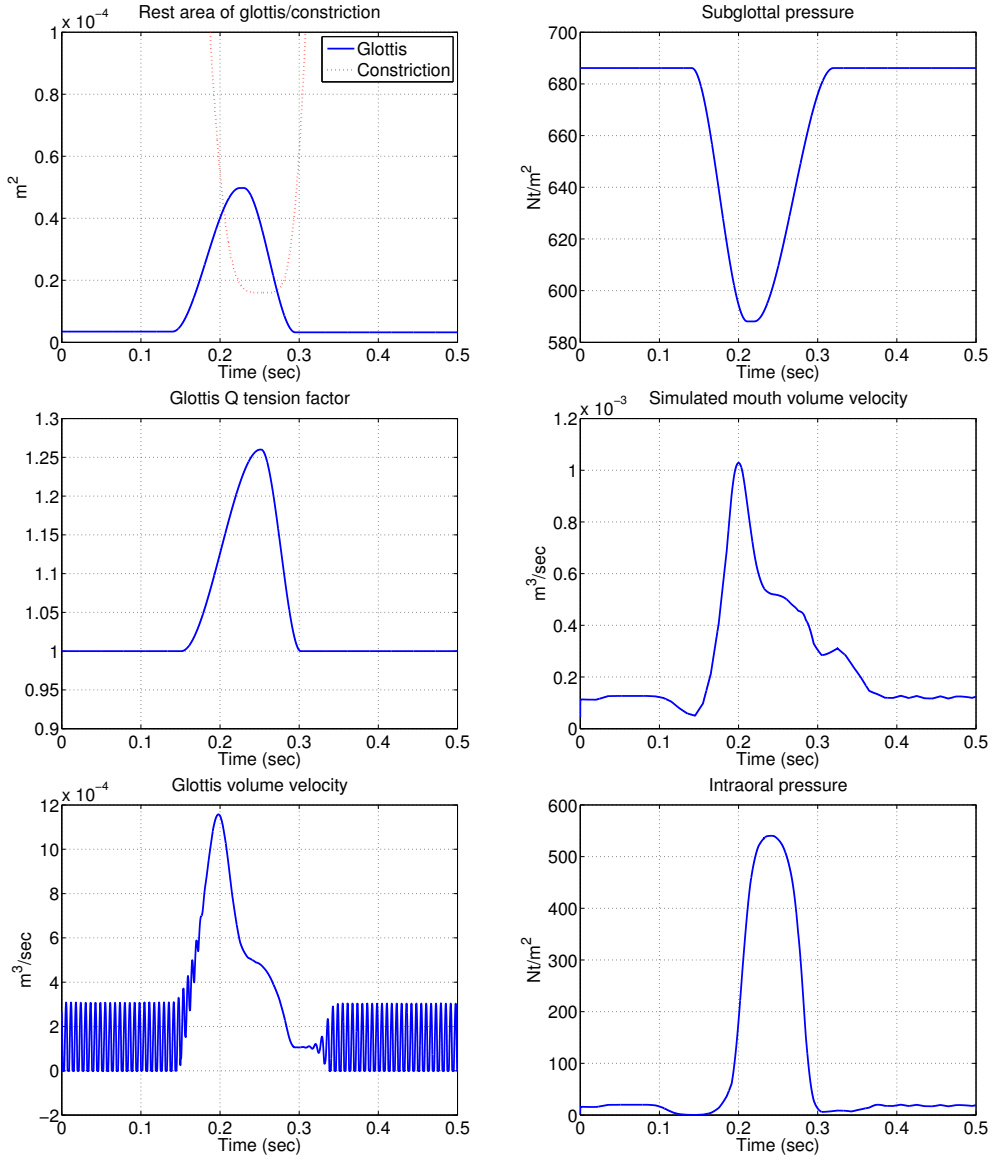
τότε η δυναμική γραμμή που διέρχεται από το σημείο $(x, r), X_n < x < X_{n+1}, r > 0$ έχει ως εστιακό σημείο το $(x_f, 0)$ όπου (βλ. Σχήμα 4.5(α')) :

$$x_f = x_n - r_n \frac{x_{n+1} - x_n}{r_{n+1} - r_n}$$

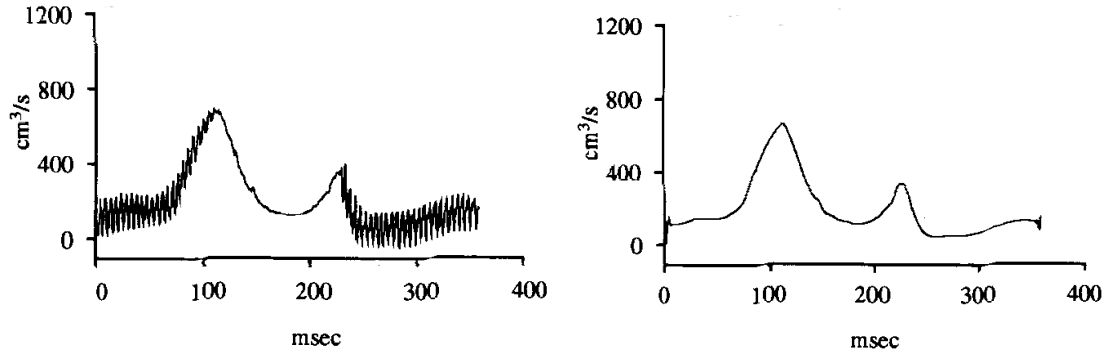
και η γωνία που σχηματίζει το διάνυσμα της ταχύτητας με τον άξονα συμμετρίας είναι :

$$\theta = \arctan\left(\frac{r}{x - x_f}\right).$$

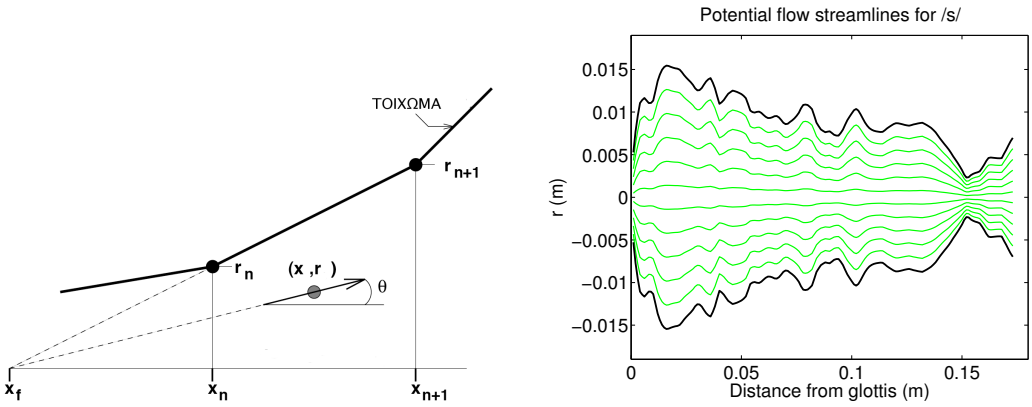
Στο Σχήμα 4.5(β') δίνονται οι δυναμικές γραμμές του δυναμικού πεδίου για μια αξονικά συμμετρική γεωμετρία όπως έχει προκύψει με βάση τη συνάρτηση εμβαδού του φωνήματος /s/ που έχει δημοσιευτεί στο [119]. Σημειώνεται ότι ο αριθμός των δυναμικών γραμμών



Σχήμα 4.3: *Προσομοίωση του πεδίου δυναμικής αεροροής για την ακολουθία φωνημάτων AsA. Χρησιμοποιήθηκαν συναρτήσεις εμβαδού που έχουν μετρηθεί με τη βοήθεια μαγνητικής τομογραφίας. Οι παράμετροι άρθρωσης είναι κατά το [108].*



Σχήμα 4.4: *Αεροροή στο στόμα για την εκφώνηση μιας ακολουθίας /asa/ όπως έχει εκφωνηθεί από γυναίκα ομιλήτρια κι έχει δημοσιευτεί στο [108]. Δίνεται και η εξομαλυμένη αεροροή η οποία προσομοιώνεται από το αεροδυναμικό μοντέλο.*



(α′) (β′)

Σχήμα 4.5: *Αριστερά: Σχηματική αναπαράσταση του τρόπου προσδιορισμού της διεύθυνσης της δυναμικής ταχύτητας σε ένα σημείο της φωνητικής οδού. Δεξιά: Οι δυναμικές γραμμές όπως προσδιορίζονται με χρήση του μοντέλου [150] για την αξονικά συμμετρική γεωμετρία που αντιστοιχεί στο φώνημα /s/.*

που έχουν σχεδιαστεί είναι τυχαίος και η συσχέτιση της πυκνότητάς τους σε κάθε περιοχή με το μέτρο της ταχύτητας δεν είναι ακριβής, όπως θα έπρεπε θεωρητικά να ισχύει . Απλά επιλέχθηκαν κάποια ισοκατανεμημένα σημεία στην αρχή της φωνητικής οδού ως σημεία εκκίνησης. Παρά την απλότητά του, το μοντέλο αυτό για τη διεύθυνση της δυναμικής ροής θεωρείται ότι αναπαριστά ικανοποιητικά την πληροφορία που απαιτείται στα πλαίσια της σύνθεσης φωνής.

4.2.2 Στροβιλώδες πεδίο

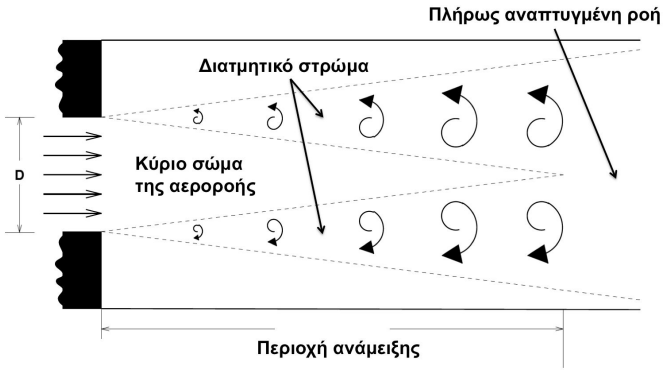
Για το στροβιλώδες πεδίο, η μοντελοποίηση που εφαρμόζεται σημειώνεται πως δεν είναι γενική αλλά αποσκοπεί στην αναπαράσταση των φαινομένων που σχετίζονται με την παραγωγή αεροακουστικού ήχου μέσα στη φωνητική οδό. Σε μια στένωση έχουμε πρακτικά το διαχωρισμό της αερορροής υπό τη μορφή μιας φλέβας (τζετ) που μεταφέρει μια σειρά στροβίλων οι οποίοι μετατοπίζονται με τη μισή ταχύτητα του τζετ και από κάποιο σημείο και μετά εξασθενούν. Αυτή η διαδικασία μοντελοποιείται ακολουθώντας την τεχνική που προτείνεται στο [150].

Συγκεκριμένα, όπως περιγράφεται και στο Κεφάλαιο 2 στην έξοδο μιας στένωσης, υπό τις κατάλληλες προϋποθέσεις, υψηλής ταχύτητας ροή είναι δυνατόν να διαχωριστεί από τα τοιχώματα και να σχηματίσει μια φλέβα που περιβάλλεται από αποτελματωμένο ρευστό (stagnant fluid) στην περιοχή μετά τη στένωση. Η στροβιλότητα που μέσα στη στένωση ήταν περιορισμένη στο συνοριακό στρώμα της ροής μετά την αποκόλλησή της εμφανίζεται σε ένα διατμητικό στρώμα και προκαλεί τη μίξη του δυναμικού πυρήνα του τζετ με το περιβάλλον ρευστό. Αυτή η ανάμειξη έχει ως αποτέλεσμα τελικά την απόσβεση του διατμητικού στρώματος (Σχήμα 4.6).

Το διατμητικό στρώμα του τζετ περιέχει δίνες διαφόρων μεγεθών που καθώς μεταφέρονται αλληλεπιδρούν μεταξύ τους και επιτείνουν την ανάμειξη. Αυτές ακριβώς οι δίνες είναι σημαντικές για την παραγωγή ήχου και η γένεση, διάδοση και απόσβεσή τους είναι που περιγράφεται από το εφαρμοζόμενο μοντέλο [91, 150]. Για διευκόλυνση, η περιστροφική κίνηση θεωρείται μέσω της κυκλοφορίας Γ που ορίζεται ως

$$\Gamma = -\int_{\Sigma} \omega \cdot \mathbf{n} d\Sigma \tag{4.4}$$

για τυχαία επιφάνεια Σ και διάνυσμα $\mathbf{n}$ κάθετο στην επιφάνεια. Η συνολική κυκλοφορία που περνάει από το σημείο διαχωρισμού της ροής σε μια περίοδο dt μπορεί να δειχτεί ότι



Σχήμα 4.6: *Σκαρίφημα του τζετ και του διατμητικού στρώματος που περιλαμβάνει τους στροβίλους που δημιουργούνται μετά την αποκόλληση της ροής [150].*

είναι ίση με :

$$\Gamma = \frac{1}{2}U_j^2 dt$$

όπου U_j είναι η ταχύτητα της φλέβας (τζετ).

Η μεταφορική ταχύτητα των στροβίλων είναι συνήθως μικρότερη αυτής του τζετ. Όπως αναφέρεται στο [150], η θεωρία προβλέπει ότι είναι ίση με τη μισή ταχύτητα του τζετ ενώ η συχνότητα εκπομπής στροβίλων έχει βρεθεί πειραματικά περίπου ίση με $f_{shed} = 0.8U_c/D$, όπου U_c είναι η μεταφορική ταχύτητα των στροβίλων και D είναι η διατομή της στένωσης από όπου γίνεται η εκπομπή. Για τη μοντελοποίηση της στροβιλότητας, οι δίνες αναπαρίστανται ως συγκεντρωμένα στοιχεία και περιγράφεται η γέννηση, η έντασή τους, η μεταφορά και η απόσβεσή τους.

Για τη γέννηση των στροβίλων πρέπει στο σημείο μεγαλύτερης στένωσης της φωνητικής οδού κατ' αρχάς το εμβαδό της εγκάρσιας διατομής να είναι μεταξύ 0.4 και 0.001 cm^2. Επίσης, ο αριθμός Mach απαιτείται να είναι μεγαλύτερος του 0.01 ενώ τέλος θα πρέπει οι συνθήκες αυτές να έχουν διαρκέσει τουλάχιστον 2 ms. Η τελευταία συνθήκη είναι για να προσομοιωθεί η καθυστέρηση που συνήθως ακολουθεί το σχηματισμό του τζετ μέχρι την έναρξη εκπομπής στροβίλων. Το διάστημα που μεσολαβεί μεταξύ της γέννησης δύο διαδοχικών στροβίλων προσδιορίζεται με βάση την ταχύτητα του τζετ, τη διάμετρο της στένωσης και την επιλογή μιας κατάλληλης τιμής για τον αριθμό Strouhal. Στην πράξη, για κάθε στρόβιλο, λαμβάνεται ίσο με ένα κανονικά κατανεμημένο τυχαίο αριθμό με μέση τιμή

$$E\{T_{shed}\} = \frac{D}{StU_j}.$$

Όσον αφορά στην ένταση του στροβίλου, η κυκλοφορία του λαμβάνεται ίση με :

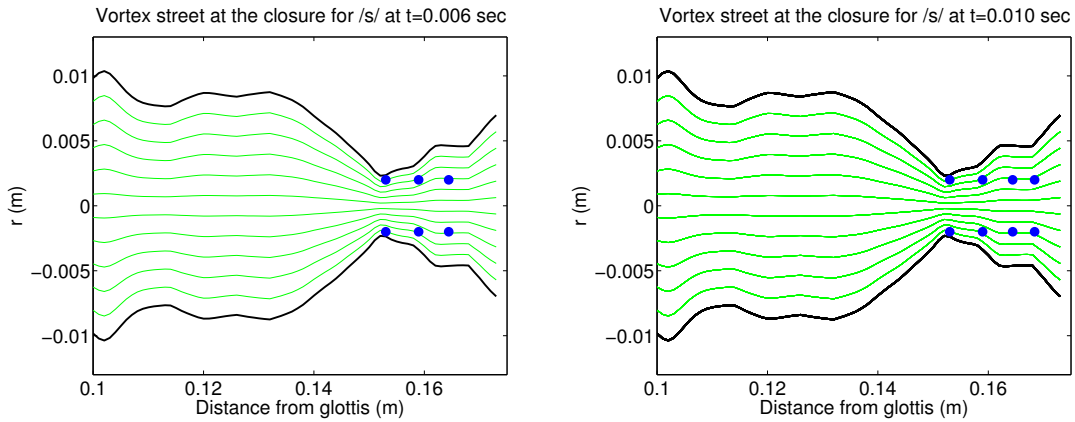
$$\Gamma = \frac{1}{2}U_j^2 T_{shed}$$

Η περιστροφική ενέργεια αυξάνεται καθώς αυξάνεται ο χρόνος μεταξύ της γέννησης διαδοχικών στροβίλων T_{shed}. Οι θέσεις των στροβίλων κάθε χρονική στιγμή δίνονται από τη σχέση :

$$x^k = x^{k-1} + vT_{sim}$$

όπου v είναι η ταχύτητα του στροβίλου και T_{sim} είναι το διάστημα χρονικής διακριτοποίησης που χρησιμοποιείται κατά την προσομοίωση.

Στο Σχήμα 4.7 δίνεται μια αναπαράσταση της προσομοίωσης του στροβιλώδους πεδίου για δύο χρονικές στιγμές για το φώνημα /s/. Το πεδίο έχει υπερτεθεί στις δυναμικές γραμμές του αστρόβιλου (δυναμικού) πεδίου. Η ογκική ταχύτητα (παροχή όγκου) του τελευταίου είναι σταθερή και ίση με 0.8 lt/sec. Οι στρόβιλοι διαδίδονται παράλληλα στον άξονα συμμετρίας. Στην ουσία δύο τελείες σε αξονικά συμμετρικές θέσεις αντιπροσωπεύουν έναν στροβιλώδη δακτύλιο στις τρεις διαστάσεις.



Σχήμα 4.7: *Το στροβιλώδες πεδίο όπως προσεγγίζεται από το εφαρμοζόμενο μοντέλο για δύο χρονικές στιγμές για τη στένωση του φωνήματος /s/. Δυο τελείες σε αξονικά συμμετρικές θέσεις αντιπροσωπεύουν ένα στροβιλώδη δακτύλιο στις τρεις διαστάσεις.*

4.3 Διερεύνηση επίδρασης μέσης ροής

Χαλαρώνοντας την υπόθεση μηδενικής ταχύτητας στην κατάσταση ισορροπίας, ανάλογα με την ερευνητική εργασία που παρουσιάζεται στο [39], και αποδεχόμενοι ότι υπάρχει μη μηδενική μέση παροχή U_0 μέσα στο σωλήνα για την οποία υποθέτουμε ότι

$$\iint_{A(x)} p\nu_{0x} dA \simeq pU_0, \tag{4.5}$$

$$\nu_{0x} \simeq U_0/A(x), \tag{4.6}$$

$$\partial U_0/\partial x = 0. \tag{4.7}$$

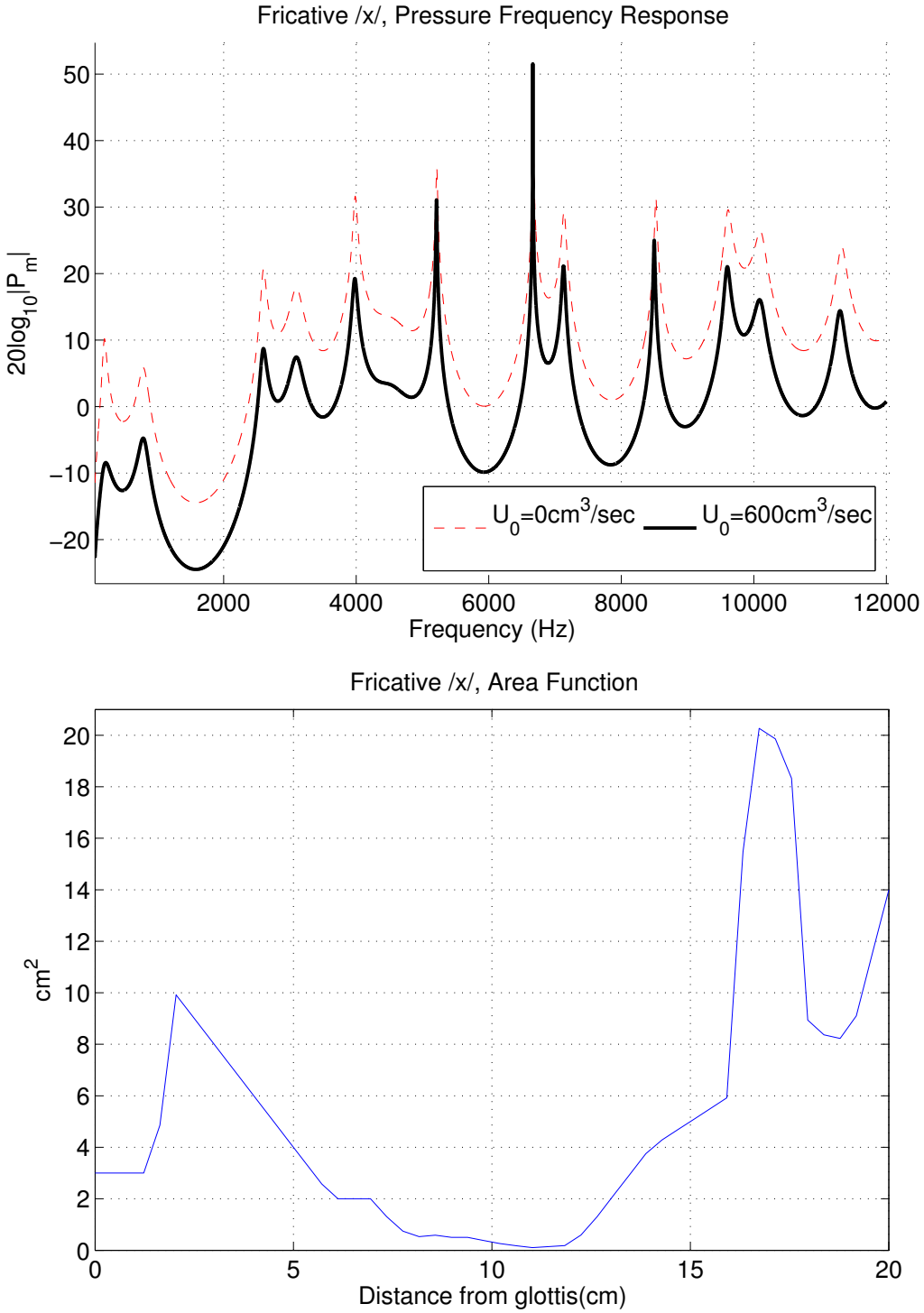
Η συνολική παροχή είναι $U_{tot} = U_0 + U$. Σε αυτή την περίπτωση, μπορεί να δειχτεί ότι οι εξισώσεις διάδοσης του ήχου μπορούν να θεωρηθούν ότι είναι:

$$-\frac{\partial U}{\partial x} = \frac{1}{c_0^2\rho_0}\frac{\partial (pA)}{\partial t} + \frac{\partial A}{\partial t} + \frac{U_0}{\rho_0 c_0^2}\frac{\partial p}{\partial x} \tag{4.8}$$

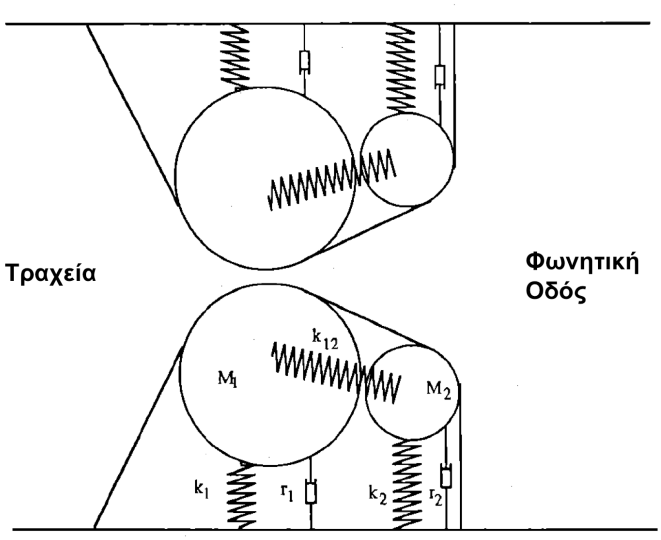
$$-\frac{\partial p}{\partial x} = \rho_0 \frac{\partial}{\partial t}\left(\frac{U}{A}\right) + \rho_0 \frac{U_0}{A}\frac{\partial}{\partial x}\left(\frac{U}{A}\right) + \left[R + \rho_0 \frac{U_0}{A}\frac{\partial}{\partial x}\left(\frac{1}{A}\right)\right]U + \frac{1}{c_0^2}\left[\frac{\partial}{\partial t}\left(\frac{U_0}{A}\right) + \frac{U_0^2}{A}\frac{\partial}{\partial x}\left(\frac{1}{A}\right)\right]p \tag{4.9}$$

Προκειμένου να διερευνηθεί η σημασία των επιπλέον όρων, το σύστημα αρχικά προσομοιώθηκε στο πεδίο των συχνοτήτων. Υπολογίστηκε η απόκριση συχνότητας του συστήματος για διάφορες γεωμετρίες και διάφορες τιμές μέσης παροχής στο εύρος $[0, 1000\ cm^3/sec]$, που μπορεί να εμφανιστεί κατά την παραγωγή φωνής [10]. Για την προσομοίωση ακολουθήθηκε η προσέγγιση που περιγράφτηκε στην Ενότητα 3.3.1.

Ενδεικτικά αποτελέσματα παρουσιάζονται στο Σχήμα 4.3 για μια κατάσταση της φωνητικής οδού, δηλαδή συγκεκριμένη συνάρτηση εμβαδού, που αντιστοιχεί στον τυρβώδη ήχο /χ/ [85]. Το φάσμα που αντιστοιχεί σε μέση παροχή $U_0 = 600cm^3/sec$ έχει μετακινηθεί προς τα κάτω κατά $10dB$ ώστε να είναι περισσότερο ευδιάκριτες οι διαφορές μεταξύ των δύο παρουσιαζόμενων φασμάτων. Η πιο σημαντική επίδραση της μη μηδενικής μέσης παροχής μπορεί να παρατηρηθεί κυρίως στις χαμηλές συχνότητες όπου έχουμε μια σημαντική μείωση του πρώτου συντονισμού της φωνής. Για γεωμετρίες της φωνητικής οδού που αντιστοιχούν σε φωνήεντα η απόκριση συχνότητας δεν επηρεάζεται παρά ελάχιστα. Αυτό πιθανώς οφείλεται στη σχετική σημασία της χωρικής παραγώγου $\partial/\partial x(1/A)$. Η μέση παροχή θεωρείται χρονομεταβλητή και για τον υπολογισμό της κάθε χρονική στιγμή εφαρμόζουμε το μοντέλο που περιγράφτηκε στην Ενότητα 4.2.



Σχήμα 4.8: *Απόκριση συχνότητας για μη μηδενικές μέσες παροχές όγκου. Η συνάρτηση εμβαδού που χρησιμοποιήθηκε αντιστοιχεί στον τυρβώδη ήχο /χ/ και φαίνεται επίσης στο σχήμα. Το φάσμα για μέση παροχή $U_0 = 600cm^3/sec$ έχει μετακινηθεί προς τα κάτω κατά $10dB$ για καλύτερη οπτικοποίηση. Παρατηρούνται επιπτώσεις στις χαμηλότερες συχνότητες. Πιο συγκεκριμένα φαίνεται ότι το σχετικό πλάτος μεταξύ των δύο πρώτων συντονισμών έχει αλλάξει.*



Σχήμα 4.9: *Βελτιωμένο μοντέλο δύο μαζών [125].*

4.4 Εφαρμογή βελτιωμένου αεροδυναμικού και μηχανικού μοντέλου για τη γλωττίδα

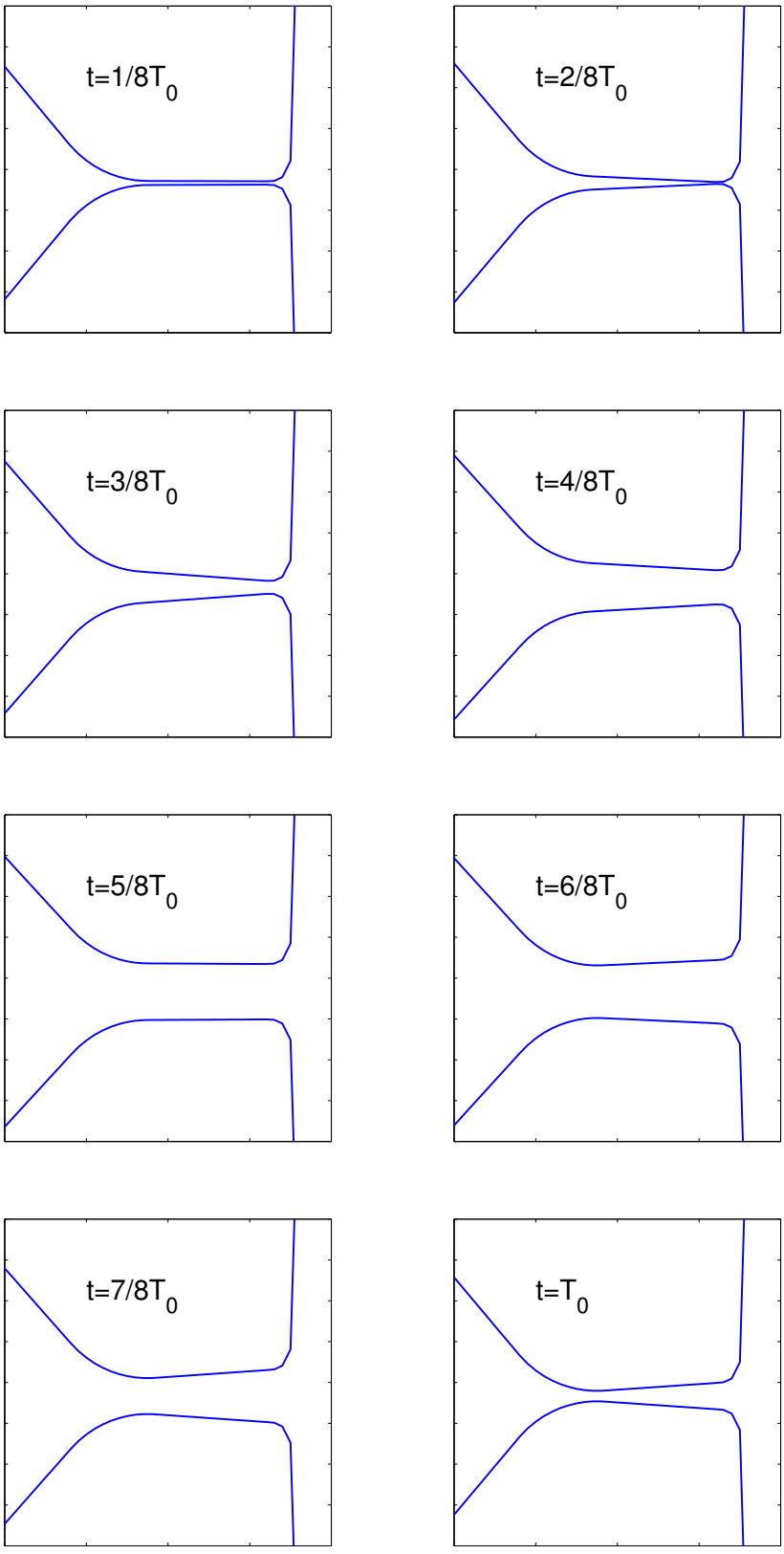
Προκειμένου να ληφθεί υπόψη η βελτιωμένη αεροδυναμική προσέγγιση για τη γλωττίδα που προτάθηκε από τον Pelorson και τους συνεργάτες του υιοθετήθηκαν βελτιώσεις του μοντέλου των δύο μαζών που προτείνονται στα [97, 125]. Το βελτιωμένο μοντέλο των δύο μαζών πέρα από το διαφορετικό σχήμα των μαζών (Σχήμα 4.9), ως βασικό του χαρακτηριστικό έχει ότι λαμβάνει υπόψη αεροδυναμικά φαινόμενα στα συνοριακά στρώματα της ροής στη γλωττίδα και επίσης ότι προβλέπει μετακινούμενο σημείο διαχωρισμού της ροής κατά την είσοδό της στη φωνητική οδό. Για την πρόβλεψη της θέσης του σημείου διαχωρισμού τελικά προτιμήθηκε η προσέγγιση που προτείνεται στο [97] ως λιγότερο πολύπλοκη υπολογιστικά παρά του ότι η μέθοδος που προτείνεται στο [125] και βασίζεται στη θεωρία συνοριακού στρώματος ελεύθερης πλάκας , μπορεί να είναι περισσότερο ακριβής από φυσική άποψη. Για τη μοντελοποίηση του ομαλού μηχανισμού κλεισίματος των φωνητικών χορδών που οδηγεί σε φυσικότερη ακουστική διέγερση, ακολουθήθηκαν οι προτάσεις του [125].

Στην είσοδο της γλωττίδας αμελείται ενδεχόμενο φαινόμενο vena contracta με τη δικαιολογία ότι η είσοδος από την τραχεία είναι ομαλή οπότε δεν υπάρχουν οι απαραίτητες πρϋποθέσεις εμφάνισης τέτοιου φαινομένου. Η βασική διαφοροποίηση πάντως στο αεροδυναμικό μοντέλο σε σχέση με την κλασσική θεώρηση είναι ότι λαμβάνεται μετακινούμενο σημείο αποκόλλησης της ροής. Αυτή η τροποποίηση έχει τελικά ως αποτέλεσμα να περιορίζονται οι απότομες μεταβολές στην κυματομορφή της ογκικής ταχύτητας κατά το κλείσιμο της γλωττίδας. Στο Σχήμα 4.10 φαίνονται διαδοχικά στιγμιότυπα της γλωττίδας για έναν κύκλο μεταβολής της. Με T_0 σημειώνεται η θεμελιώδης περίοδος της ταλάντωσης. Στα Σχήματα 4.11, 4.12 συγκρίνονται οι γλωττιδικές ογκικές ταχύτητες για εναλλακτικά μοντέλα γλωττίδας και τα αντίστοιχα ακουστικά σήματα στην έξοδο της φωνητικής οδού.

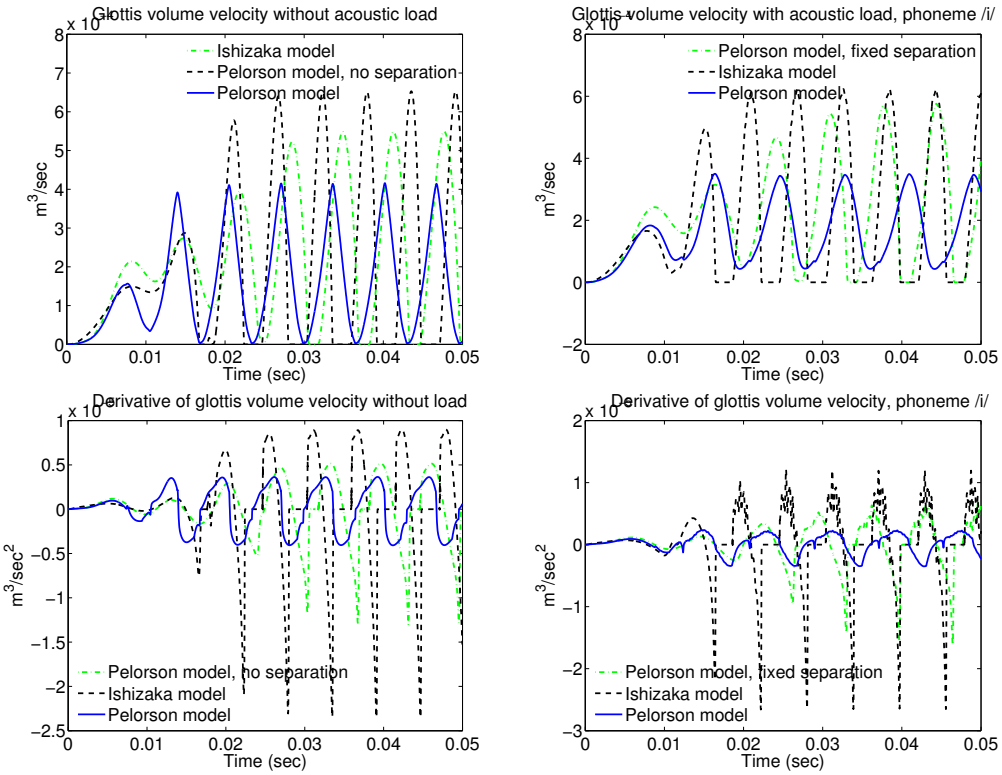
4.5 Αεροακουστική διέγερση στη γλωττίδα και σε στενώσεις

Με την αξιοποίηση του αεροδυναμικού μοντέλου τόσο για τη γλωττίδα όσο και για τη φωνητική οδό είναι δυνατή η εκτίμηση των πηγών ήχου κατά την παραγωγή φωνής με βάση την αεροακουστική θεωρία που παρουσιάστηκε στο Κεφάλαιο 2, [68, 71, 91, 150].

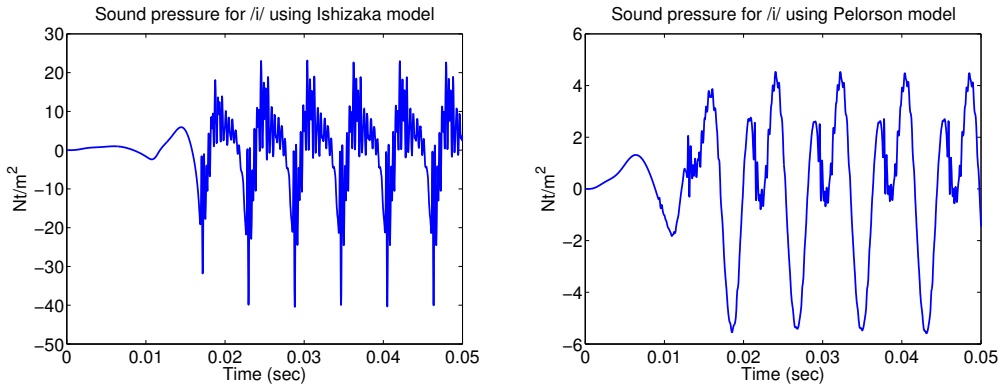
Σύμφωνα με την υπόθεση της προσεγγιστικής στατικότητας η διπολική πηγή ήχου στη



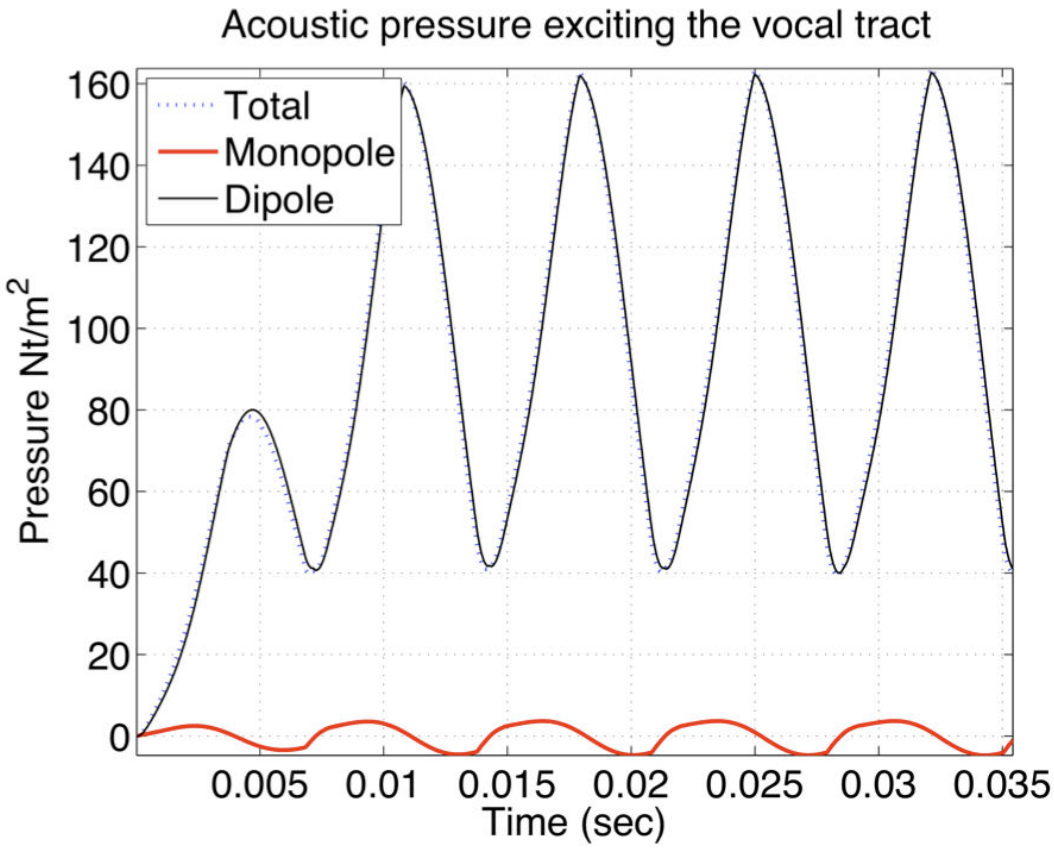
Σχήμα 4.10: *Στιγμιότυπα του σχήματος της γλωττίδας για έναν κύκλο μεταβολής. Με T_0 συμβολίζεται η θεμελιώδης περίοδος ταλάντωσης της γλωττίδας.*



Σχήμα 4.11: *Γλωττιδική ογκική ταχύτητα και η χρονική της παράγωγος, όπως υπολογίζεται με το κλασσικό μοντέλο και με το βελτιωμένο μοντέλο δύο μαζών για τη γλωττίδα. Απεικονίζεται και η ογκική ταχύτητα αν συμπεριληφθεί κινούμενο σημείο αποκόλλησης της ροής. Δίνονται οι περιπτώσεις απουσίας και ύπαρξης ακουστικού φορτίου, στην αριστερή και δεξιά στήλη αντίστοιχα.*



Σχήμα 4.12: *Ακουστικό σήμα στα χείλη για την περίπτωση του φωνήματος /i/ εφαρμόζοντας το κλασσικό ή το βελτιωμένο μοντέλο γλωττίδας.*



Σχήμα 4.13: *Ενδεικτικές κυματομορφές της μονοπολικής και της διπολικής συνιστώσας της ηχητικής πηγής στη γλωττίδα όπως προβλέπονται από την αεροακουστική θεωρία [71]. Η διπολική συνιστώσα είναι σημαντικά ισχυρότερη.*

γλωττίδα δίνεται από τη σχέση [71]:

$$p_d^g(t) = \frac{\rho_0 c_0^2}{2} \frac{\sigma A_{gm}(t)}{A_1(t)} \left(\sqrt{\left(\frac{\sigma A_{gm}(t)}{A_1(t)} \right)^2 + \frac{4P_s(t)}{\rho_0 c_0^2}} - \frac{\sigma A_{gm}(t)}{A_1(t)} \right) \tag{4.10}$$

όπου A_1 είναι η επιφάνεια εγκάρσιας διατομής της φωνητικής οδού αμέσως μετά τη γλωττίδα και A_{gm} είναι η επιφάνεια εγκάρσιας διατομής της γλωττίδας στο σημείο μεγαλύτερης στένωσης. Το μετακινούμενο σημείο αποκόλλησης της ροής της γλωττίδας μεταβάλλει το συντελεστή σ που είναι γνωστός και ως λόγος συστολής της ροής. Λόγω των μικρών περιοδικών αλλαγών του όγκου του αέρα μέσα στη γλωττίδα εμφανίζεται και μια μονοπολική πηγή ήχου που δίνεται ως [71] :

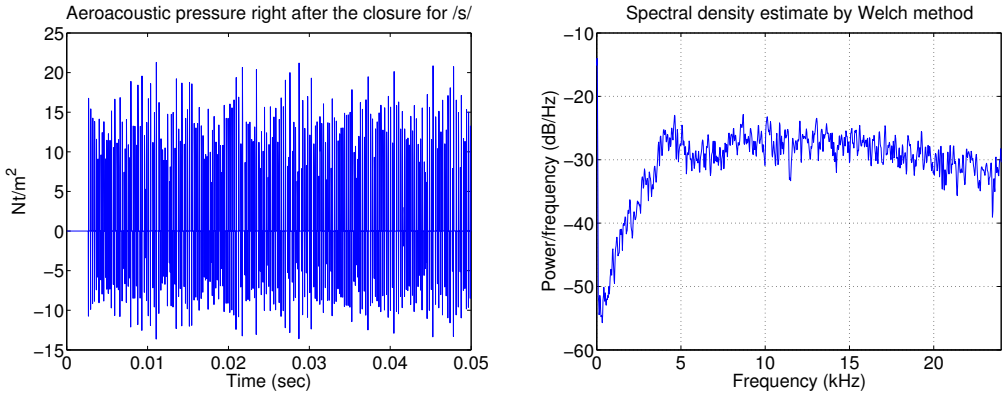
$$p_m^g(t) = -P_s(t) \frac{\rho c_0 l_g}{2A_1} \frac{\partial}{\partial t} \int_{glottis} h(x,t) dx \tag{4.11}$$

όπου $h(x,t)$ είναι το άνοιγμα της γλωττίδας. Στο Σχήμα 4.13 δίνονται ενδεικτικές κυματομορφές της μονοπολικής και της διπολικής συνιστώσας της πηγής ήχου στη γλωττίδα. Η διπολική πηγή είναι σημαντικά πιο ισχυρή.

Στις στενώσεις, η ακουστική πίεση λόγω ενός στροβίλου μπορεί να θεωρηθεί ότι είναι [91] :

$$p = \frac{-\rho_0}{A} \int_0^{2\pi} [\Gamma(\boldsymbol{\epsilon_\theta} \times \mathbf{v}) \cdot \mathbf{U}^*] d\theta \tag{4.12}$$

$$= \frac{-2\pi r_\omega \rho_0}{A} [\Gamma(\boldsymbol{\epsilon_\theta} \times \mathbf{v}) \cdot \mathbf{U}^*] \tag{4.13}$$



Σχήμα 4.14: *Αεροακουστική πίεση και το φάσμα της αμέσως μετά τη στένωση για το φώνημα* /s/.

όπου ϵ_θ είναι το μοναδιαίο διάνυσμα στην κατεύθυνση της στροβιλότητας, A είναι η εγκάρσια επιφάνεια διατομής της φωνητικής οδού στο σημείο που βρίσκεται ο στρόβιλος και r_ω είναι η ακτίνα του στροβιλώδους δακτυλίου. Δεδομένου του ότι η φωνητική οδός έχει πεπερασμένο μήκος πρέπει να ληφθούν υπόψη και ενδεχόμενες ανακλάσεις από τα άκρα, κάτι που γίνεται στην πράξη με την ακουστική προσομοίωση. Καθώς μεταφέρονται οι στρόβιλοι, διεγείρουν διαφορετικά σημεία της φωνητικής οδού με αποτελέσμα η ακουστική πηγή να είναι τελικά χωρικά και χρονικά κατανεμημένη. Για να αποφευχθούν ασυνέχειες στα σημεία όπου ένας στρόβιλος αλλάζει διάστημα διακριτοποίησης, εφαρμόζεται το σχήμα εξομάλυνσης που προτείνεται στο [150] και πρακτικά προβλέπει ότι ένας στρόβιλος μπορεί να διεγείρει και το τμήμα της φωνητικής οδού στο οποίο πλησιάζει χωρίς ακόμα να είναι μέσα σε αυτό. Στο Σχήμα 4.14 δίνονται οι διαταραχές της πίεσης και το φάσμα τους για την περίπτωση του φωνήματος /s/ αμέσως μετά τη στένωση. Έχει γίνει η υπόθεση εργασίας ότι δεν υπάρχουν ανακλάσεις στα άκρα της φωνητικής οδού.

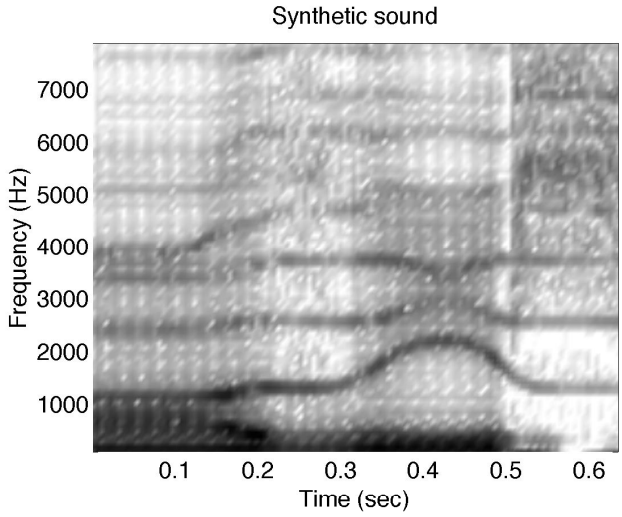
4.6 Πειράματα σύνθεσης ακολουθιών της μορφής Φωνήεν - Σύμφωνο - Φωνήεν

Το προτεινόμενο μοντέλο παραγωγής φωνής εφαρμόζεται για τη σύνθεση ακολουθιών Φωνήεν-Τυρβώδες Σύμφωνο-Φωνήεν $\Phi_1 T \Phi_2$, όπου το τυρβώδες σύμφωνο στη μέση μπορεί να είναι ένα έμφωνο ή άφωνο. Ενδεικτικά δίνονται τα σπεκτρογραφήματα που αντιστοιχούν στα αποτελέσματα της σύνθεσης των ακολουθιών / αξής /, / άσος /, / όφις / στο Σχήμα 4.15. Έχουν χρησιμοποιηθεί πραγματικές συναρτήσεις εμβαδού όπως έχουν μετρηθεί με τη χρήση αξονικής τομογραφίας.

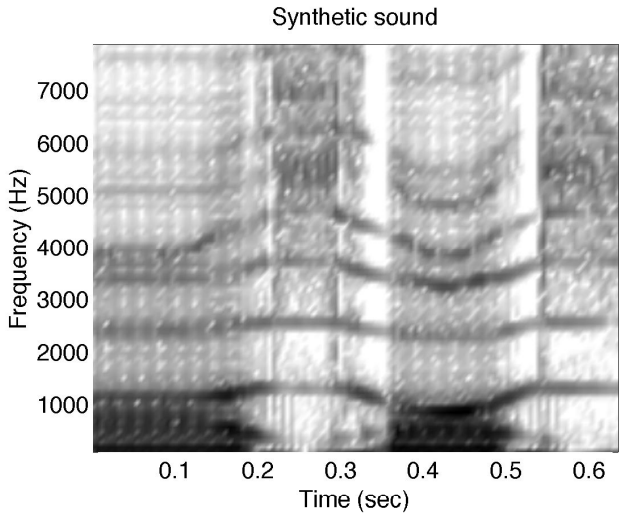
Για τον έλεγχο της άρθρωσης ορίζονται σε όλη τη διάρκεια του χρονικού διαστήματος σημαντικά γεγονότα και η διάρκειά τους για κάθε παράμετρο άρθρωσης από τις παρακάτω : υπογλωττιδική πίεση, αναλογικό μήκος του μη ταλαντούμενου τμήματος της φωνητικής οδού, ρινικότητα, και παράγοντας τάσης των φωνητικών χορδών. Για παράδειγμα, στα 120 ms ορίζεται να πραγματοποιηθεί μεταβολή της υπογλωττιδικής πίεσης στην τιμή 700 $\mathrm{Nt/m^2}$ μέσα σε 40 ms. Ορίζεται επίσης και ο τρόπος της μεταβολής, για παράδειγμα κυβική ή γραμμική ή με χρήση σπλήνας. Με αυτόν τον τρόπο καθορίζονται οι παράμετροι άρθρωσης για όλο το χρονικό διάστημα που μελετάται.

4.7 Συζήτηση

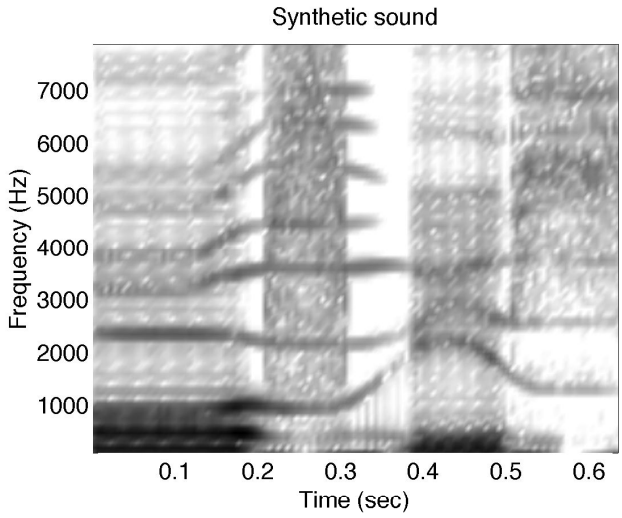
Παρουσιάστηκε η αεροδυναμική-αεροακουστική μοντελοποίηση που αναπτύχθηκε στο πλαίσιο της διδακτορικής διατριβής για τη φωνητική οδό. Έγινε προσπάθεια να είναι δυνατή η περιγραφή σημαντικών φαινομένων για την παραγωγή φωνής χωρίς όμως να αυξάνεται



(α') / αζής /



(β') / άσος /



(γ') / όφις /

Σχήμα 4.15: *Σύνθεση των ακολουθιών φωνημάτων / αζής /, / άσος /, / όφις / με τη χρήση αρθρωτών. Έχει χρησιμοποιηθεί το γενικό προτεινόμενο πλαίσιο που περιλαμβάνει ξεχωριστό αεροδυναμικό μοντέλο για τη φωνητική οδό τόσο για το αστρόβιλο όσο και για το στροβιλώδες πεδίο. Περιλαμβάνει επίσης πρόβλεψη των πηγών ήχου με βάση την αεροακουστική θεωρία και τέλος ένα κατάλληλα τροποποιημένο μοντέλο δύο μαζών ώστε να λαμβάνονται υπόψη σημαντικές λεπτομέρειες για την αποκόλληση της ροής.*

σημαντικά η υπολογιστική πολυπλοκότητα. Για το πεδίο ροής θεωρήθηκε ότι μπορεί να διαχωριστεί σε αστρόβιλο και στροβιλώδες πεδίο τα οποία αναπαρίστανται διαφορετικά. Το μέτρο του αστρόβιλου (δυναμικού πεδίου) προσδιορίζεται με βάση ένα απλοποιημένο ηλεκτρικό ανάλογο που είναι κατάλληλα συζευγμένο με ένα μοντέλο δύο μαζών για τη γλωττίδα. Το τελευταίο συνδυάζει βασικά χαρακτηριστικά του κλασσικού μοντέλου των Ishizaka, Flanagan [73] με βελτιώσεις που προτάθηκαν από τους Pelorson, Lous και τους συνεργάτες τους [97, 125]. Το σημαντικότερο ίσως είναι ότι επιτρέπει μετακινούμενο σημείο διαχωρισμού της ροής, ώστε το κλείσιμο της γλωττίδας να είναι πιο ομαλό. Με βάση τους αεροδυναμικούς υπολογισμούς γίνεται στη συνέχεια δυνατός ο προσδιορισμός των πηγών ήχου στο ακουστικό πεδίο, όπως αυτό προσομοιώνεται με το σύστημα που περιγράφτηκε στο Κεφάλαιο 3. Για την περιγραφή των πηγών χρησιμοποιήθηκαν σύγχρονα συμπεράσματα που έχουν προκύψει από την αεροακουστική θεωρία. Έτσι, για παράδειγμα, στη γλωττίδα θεωρείται μια διπολική και μια μονοπολική πηγή ήχου, με την πρώτη να είναι πιο σημαντική [71, 183] ενώ στις στενώσεις θεωρείται διπολική πηγή ήχου με χαρακτηριστικά που περιγράφονται στο [91]. Με αυτόν τον τρόπο επιτυγχάνεται τελικά η σύνθεση φωνής αξιοποιώντας μια σημαντικά πιο ακριβή αναπαράσταση της φυσικής της φωνητικής οδού, που ήταν και το ζητούμενο.

Κεφάλαιο 5

Οπτικοακουστική Αντιστροφή Φωνής

5.1 Εισαγωγή

Η θεώρηση και αξιοποίηση της πολυμεσικότητας της φωνής έχει οδηγήσει σε ενδιαφέρουσες εξελίξεις των τεχνολογιών φωνής τα τελευταία χρόνια [1]. Για παράδειγμα, με κατάλληλη αξιοποίηση οπτικής πληροφορίας από το πρόσωπο του ομιλητή, τα συστήματα αναγνώρισης φωνής μπορούν να γίνουν περισσότερο ανθεκτικά στο θόρυβο [129]. Η εισαγωγή ομιλούντων προσώπων σε συστήματα σύνθεσης φωνής βελτιώνει τη φυσικότητα και την καταληπτότητά τους [18]. Γενικά, η αξιοποίηση της οπτικής συνιστώσας της φωνής με τρόπους εμπνευσμένους από τους φυσικούς μηχανισμούς παραγωγής [178] και αντίληψης της φωνής [111] μπορεί να είναι ιδιαίτερα ωφέλιμη για την αυτόματη επεξεργασία φωνής και για τις διεπαφές ανθρώπου-μηχανής.

Σε αυτό το πλαίσιο, το ενδιαφέρον έγκειται στη ανάκτηση ιδιοτήτων του συστήματος παραγωγής φωνής, συγκεκριμένα του σχήματος και της δυναμικής της φωνητικής οδού, χρησιμοποιώντας όχι μόνο το ακουστικό σήμα φωνής αλλά και το κινούμενο πρόσωπο του ομιλητή. Το πρόβλημα στη γενική του μορφή θα μπορούσε να αναφερθεί ως οπτικοακουστική αντιστροφή φωνής. Επιπλέον της θεωρητικής του σημασίας, θα μπορούσε να επιτρέψει την αναπαράσταση των ακουστικών και οπτικών συνιστωσών της φωνής μέσω της αντίστοιχης κατάστασης του φωνητικού συστήματος. Μια τέτοια αναπαράσταση μπορεί να είναι χρήσιμη σε σημαντικές εφαρμογές όπως είναι η σύνθεση φωνής [145], η αναγνώριση φωνής [87], η κωδικοποίηση φωνής [144] και η γλωσσική εκπαίδευση [50].

Η αντιστροφή φωνής παραδοσιακά θεωρείται ως ο προσδιορισμός του σχήματος της φωνητικής οδού μόνο από το ακουστικό σήμα φωνής [145]. Σύγχρονες προσεγγίσεις ακουστικής μόνο αντιστροφής φωνής βασίζονται σε εξελιγμένες τεχνικές μηχανικής μάθησης. Για παράδειγμα, στο [121] βελτιστοποιείται η αναζήτηση σε βιβλία κωδικών ώστε να είναι δυνατή η ανάκτηση σχημάτων της φωνητικής οδού από συντονισμούς του σήματος φωνής. Το σύστημα αντιστροφής στο [137] είναι βασισμένο σε νευρωνικά δίκτυα. Στο [170] προτείνεται μια απεικόνιση βασισμένη στην εφαρμογή ενός μείγματος Γκαουσιανών για αντιστροφή από Mel συντελεστές cepstrum, ενώ στο [63] παρουσιάζεται μια ακουστική απεικόνιση σε παραμέτρους του φωνητικού συστήματος βασισμένη σε κρυφά Μαρκοβιανά μοντέλα. Κάθε φώνημα μοντελοποιείται με ένα κρυφό Μαρκοβιανό μοντέλο που εξαρτάται από τα συμφραζόμενα. Σε κάθε κατάσταση αυτού του μοντέλου αντιστοιχεί ένα ξεχωριστό γραμμικό μοντέλο μεταξύ των παρατηρούμενων Mel συντελεστών cepstrum και των αντίστοιχων παραμέτρων του συστήματος. Παρόμοιες μέθοδοι έχουν εφαρμοστεί στο συμπληρωματικό πρόβλημα της αντιστροφής από τη φωνή στα χείλια, δηλαδή του συγχρονισμού των χειλιών με βάση το ακουστικό σήμα

[1] Σημαντικό τμήμα του κεφαλαίου έχει προκύψει από τη μετάφραση του δημοσιευμένου άρθου [84].

φωνής [29,48,177]. Οι παράμετροι των χειλιών και του ακουστικού σήματος μοντελοποιούνται από κοινού χρησιμοποιώντας ανά φώνημα ένα κρυφό Μαρκοβιανό μοντέλο με μείγμα Γκαουσιανών σε κάθε κατάσταση στο [32] ενώ στο [176] χρησιμοποιούνται πιο σύνθετα δυναμικά δίκτυα Bayes που εισάγουν και πληροφορία σχετική με το σύστημα παραγωγής φωνής.

Ένα εγγενές μειονέκτημα των συστημάτων αντιστροφής φωνής μόνο από το ακουστικό σήμα είναι ότι η απεικόνιση από τον ακουστικό χώρο στο χώρο όπου περιγράφεται το σχήμα της φωνητικής οδού δεν είναι ένα προς ένα [121], με την έννοια ότι υπάρχει ένα μεγάλος αριθμός καταστάσεων της φωνητικής οδού που μπορούν να παράξουν το ίδιο ακουστικό σήμα και οπότε το πρόβλημα της αντιστροφής είναι σημαντικό υποορισμένο. Η ενσωμάτωση της οπτικής συνιστώσας της φωνής μπορεί να βελτιώσει σημαντικά την ακρίβεια της αντιστροφής. Σημαντικοί αρθρωτές όπως τα χείλια, το σαγόνι, τα δόντια και η γλώσσα είναι σε κάποιο βαθμό ορατοί. Γι' αυτό, τα οπτικά στοιχεία μπορούν να περιορίσουν δραστικά τον χώρο των λύσεων και να αντιμετωπίσουν κατά κάποιο τρόπο τις επιβαρυντικές ιδιότητες του προβλήματος. Πράγματι, έχουν πραγματοποιηθεί αρκετές μελέτες που δείχνουν ότι υπάρχει σημαντική συσχέτιση μεταξύ του προσώπου ενός ομιλητή και της κίνησης σημαντικών αρθρωτών της φωνητικής οδού, όπως, για παράδειγμα, της γλώσσας [17,49,76,89,178]. Ο Yehia και οι συνεργάτες του στο [178] εξερευνούν απλά καθολικά γραμμικά μοντέλα για να φανερώσουν συσχετίσεις μεταξύ της συμπεριφοράς των δεδομένων του προσώπου και των κινήσεων των αρθρωτών κατά τη διάρκεια της ομιλίας. Δείχνουν ότι η ανάλυση μπορεί να επιτευχθεί πραγματοποιώντας μια διαδικασία μείωσης των διαστάσεων του προβλήματος κατά την οποία προσδιορίζονται οι συνιστώσες που κυρίως επηρεάζουν τη σχέση μεταξύ των χώρων των οπτικών και των αρθρωτικών δεδομένων. Τα πειραματικά τους δεδομένα περιλαμβάνουν μετρήσεις των θέσεων δεικτών πάνω στο πρόσωπο και ηλεκτρομαγνητικών αισθητήρων μέσα στη φωνητική οδό, καθώς επίσης και τα εξαγόμενα ακουστικά δεδομένα από τη φωνή, για δύο ομιλητές. Συμπεραίνουν ότι ένα υψηλό ποσοστό της μεταβλητότητας που παρατηρείται στα δεδομένα της φωνητικής οδού (80%) μπορεί να ανακτηθεί από τα οπτικά δεδομένα του προσώπου. Το συμπέρασμα αυτό επιβεβαιώνεται επίσης στο [76] σε παρόμοια δεδομένα, δηλαδή 20 ανακλαστές που είναι κολλημένοι στο πρόσωπο και ιχνηλατώνται από ανάλογο σύστημα. Και πάλι η αντιστροφή επιτυγχάνεται μέσω καθολικής πολυμεταβλητής γραμμικής μοντελοποίησης. Στην εν λόγω δουλειά, οι συγγραφείς κυρίως επικεντρώνουν το ενδιαφέρον τους στις μεταβολές των οπτικο-αρθρωτικών σχέσεων για διάφορες συλλαβές του τύπου ΣΦ (Σύμφωνο-Φωνήεν) και πώς αυτές επηρεάζουν την καταληπτότητα της φωνής. Πιο πρόσφατα, στα [49,88,89] ανακτώνται αρθρωτικές παράμετροι από οπτικά και από ακουστικά δεδομένα είτε με τη χρήση μηχανών διανυσμάτων συσχέτισης ή με καθολικά γραμμικά μοντέλα.

Παρά τα πολλά υποσχόμενα αποτελέσματα, κάποιος μπορεί να αναγνωρίσει δύο κύρια μειονεκτήματα σε αυτές τις προσεγγίσεις οπτικοακουστικής αντιστροφής φωνής. Πρώτον, η οπτική συνιστώσα καταγράφεται με τη χρήση πολύπλοκων οπτικών συστημάτων που περιορίζουν την εφαρμοσιμότητα των εν λόγω τεχνικών μόνο μέσα στο εργαστήριο. Σε πιο ρεαλιστικές συνθήκες η καταγραφή του προσώπου αναμένεται να γίνεται από μία και μόνο κάμερα και χωρίς τη χρήση οποιουδήποτε σημαδευτή. Δεύτερον, οι παραπάνω μελέτες χρησιμοποιούν μια καθολική απεικόνιση. Ενώ πιο γενικές στατικές μη γραμμικές απεικονίσεις μπορεί να είναι πιο αποτελεσματικές, είναι πιο δύσκολο να εκπαιδευτούν, ιδιαίτερα όταν τα διαθέσιμα δεδομένα είναι περιορισμένα, και δεν επιτρέπουν την εύκολη ενσωμάτωση της δυναμικής της φωνής στη διαδικασία της αντιστροφής.

Στην παρούσα ερευνητική εργασία που πραγματοποιήθηκε στα πλαίσια της διδακτορικής διατριβής [80,82–84], αντιμετωπίζονται αποτελεσματικά και τα δύο παραπάνω ζητήματα. Όσον αφορά στην ανάλυση του προσώπου, προτείνεται μια προσέγγιση με τη χρήση τεχνικών όρασης υπολογιστών για την αυτόματη εξαγωγή οπτικών χαρακτηριστικών από την μπροστινή όψη του προσώπου χωρίς τη χρήση σημαδευτών. Το οπτικό σύστημα επεξεργασίας που χρησιμοποιείται βασίζεται σε ενεργά μοντέλα εμφάνισης [35]. Πρόκειται για αναγεν-

νητικά μοντέλα εικόνας που επιτρέπουν την αποτελεσματική και εύρωστη μοντελοποίηση του προσώπου. Το κύριο πλεονέκτημά τους σε σύγκριση με τις τεχνικές που είναι βασισμένες σε μετασχηματισμούς, όπως το σχήμα ανάλυσης σε ανεξάρτητες συνιστώσες στο [88], είναι ότι λαμβάνουν υπόψη τόσο μεταβολές του σχήματος του προσώπου όσο και της υφής του. Η αρχικοποίηση του μοντέλου γίνεται αυτόματα με τη χρήση ενός ανιχνευτή προσώπου βασισμένου στον αλγόριθμο Adaboost [174]. Το συνολικό σύστημα οπτικής επεξεργασίας επιτρέπει την αξιόπιστη εξαγωγή χαρακτηριστικών σχήματος και υφής του προσώπου τα οποία στη συνέχεια αξιοποιούνται για την αντιστροφή.

Επιπρόσθετα, για να ξεπεραστούν οι δυσκολίες μιας καθολικής απεικόνισης από οπτικοακουστική πληροφορία σε πληροφορία του φωνητικού συστήματος και εμπνευσμένο από τις τεχνικές για ακουστική μόνο αντιστροφή των [63] και [45], προτείνεται ένα προσαρμοστικό πλαίσιο οπτικακουστικής αντιστροφής φωνής το οποίο επιτρέπει την εναλλαγή ανάμεσα σε ξεχωριστές (ανά φώνημα, οπτικό ή ακουστικό) γραμμικές απεικονίσεις. Ο μηχανισμός εναλλαγής προσδιορίζεται από μια κρυφή διαδικασία Markov που επιτρέπει την εφαρμογή περιορισμών στη δυναμική συμπεριφορά των παραμέτρων του συστήματος παραγωγής φωνής. Παρά την απλότητα της κάθε γραμμικής απεικόνισης, η προκύπτουσα κατά τμήματα προσέγγιση μπορεί να περιγράψει τις πολύπλοκες οπτικοακουστικές συσχετίσεις με τις κρυμμένες φωνητικές παραμέτρους. Επιπλέον, οι συνιστώσες απεικονίσεις μπορούν να υπολογιστούν μέσω αποδοτικών πολυμεταβλητών μεθόδων ανάλυσης. Συγκεκριμένα, συζητείται η χρήση της ανάλυσης κανονικής συσχέτισης η οποία αρμόζει για τον υπολογισμό γραμμικών μοντέλων με τα περιορισμένα δεδομένα που αντιστοιχούν σε κάθε συγκεκριμένη κλάση του συνολικού μοντέλου. Το προτεινόμενο σχήμα αντιστροφής απαιτεί τον προσδιορισμό της κρυφής Markov ακολουθίας καταστάσεων για κάθε εκφώνηση. Για αυτό το σκοπό, μελετήθηκαν εναλλακτικές τεχνικές αντιστοίχισης των καταστάσεων με τις παρατηρήσεις. Οι τεχνικές αυτές συνδυάζουν την οπτική και ακουστική πληροφορία σε διαφορετικά επίπεδα συγχρονισμού [44]. Στην περίπτωση της σύγχρονης σύμμειξης, οι δύο συνιστώσες μοιράζονται μια κοινή κατάσταση και αντιστοιχίζονται από κοινού χρησιμοποιώντας συγχρονισμένα ανά κατάσταση πολυκαναλικά κρυφά Μαρκοβιανά μοντέλα, ενώ στην περίπτωση της πλήρως ασύγχρονης εκ των υστέρων σύμμειξης, οι συνιστώσες είναι σε ανεξάρτητες μεταξύ τους καταστάσεις και ευθυγραμμίζονται ξεχωριστά χρησιμοποιώντας ξεχωριστά κρυφά Μαρκοβιανά μοντέλα. Με δεδομένη την κρυφή ακολουθία καταστάσεων, η αντιστροφή πραγματοποιείται δίνοντας τα κατάλληλα βάρη σε κάθε συνιστώσα, λαμβάνοντας υπόψη την αξιοπιστία της καθεμιάς. Η προτεινόμενη μέθοδος αξιολογείται στη βάση MOCHA [175] (MultiCHannel Articulatory) και QSMT (Qualisys-Movetrack) [51] που περιλαμβάνουν ταυτόχρονα καταγεγραμμένα ήχο, βίντεο και ηλεκτρομαγνητικά δεδομένα άρθρωσης. Ο στόχος είναι η αυτόματη πρόβλεψη των τροχιών των πηνίων που είναι κολλημένα πάνω σε σημαντικούς αρθρωτές, όπως η γλώσσα και τα δόντια, και ιχνηλατούνται με ηλεκτρομαγνητικά μέσα.

5.2 Αντιστροφή με γραμμικά μοντέλα

Σε ένα πιθανοτικό πλαίσιο, η λύση στο πρόβλημα της ανάκτησης των κινήσεων των αρθρωτών από τη φωνή και το πρόσωπο του ομιλητή μπορεί να θεωρηθεί ότι είναι η κατάσταση των αρθρωτών που μεγιστοποιεί την ύστερη πιθανότητα των χαρακτηριστικών των αρθρωτών $\mathbf{x}$ δεδομένων των διαθέσιμων οπτικοακουστικων μετρήσεων $\mathbf{y} = \{\mathbf{y}_\mathrm{a}, \mathbf{y}_\mathrm{v}\}$:

$$p(\mathbf{x}|\mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})/p(\mathbf{y}). \tag{5.1}$$

Θα ήταν επωφελές διαισθητικά να θεωρήσουμε πρώτα τη στατική περίπτωση στην οποία τόσο οι αρθρωτές όσο και τα οπτικοακουστικά χαρακτηριστικά δε μεταβάλλονται με το χρόνο. Το διάνυσμα παραμέτρων $\mathbf{x}$ (n στοιχείων) παρέχει μια κατάλληλη αναπαράσταση της φωνητικής οδού. Αυτή η αναπαράσταση μπορεί είτε να είναι άμεση, συμπεριλαμβάνοντας

χωρικές συντεταγμένες πραγματικών αρθρωτών, είτε έμμεση, περιγράφοντας ένα κατάλληλο αρθρωτικό μοντέλο για παράδειγμα. Το οπτικοακουστικό διάνυσμα παραμέτρων $\mathbf{y}$ (m στοιχείων), που περιλαμβάνει τις ακουστικές και οπτικές παραμέτρους $\mathbf{y}_\mathrm{a}$ και $\mathbf{y}_\mathrm{v}$ αντίστοιχα, θα πρέπει ιδανικά να περιέχει όλη την πληροφορία που σχετίζεται με τη φωνητική οδό και μπορεί να εξαχθεί από το ακουστικό σήμα από τη μία και από το πρόσωπο του ομιλητή από την άλλη. Ως μέσο αναπαράστασης της φωνής έχουν χρησιμοποιηθεί οι συντονισμοί της φωνητικής οδού, γραμμικές φασματικές συχνότητες ή συντελεστές cepstrum στη συχνότητα Mel. Για το πρόσωπο, θα μπορούσαν να χρησιμοποιηθούν χωρικές συντεταγμένες σημαντικών σημείων, για παράδειγμα γύρω από το στόμα, ή, εναλλακτικά, παράμετροι ενός πιο σύνθετου μοντέλου, όπως είναι το ενεργό μοντέλο εμφάνισης.

Για τη μεγιστοποίηση, η κατανομή $p(\mathbf{y})$ μπορεί να παραληφθεί δεδομένου του ότι δεν εξαρτάται από το $\mathbf{x}$. Η κατανομή $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \bar{\mathbf{x}}, \Sigma_x)$ υποτίθεται ότι είναι Gaussian με μέση τιμή $\bar{\mathbf{x}}$ και πίνακα συμμεταβλητότητας Σ_x. Η σχέση μεταξύ των οπτικοακουστικών και των αρθρωτικών παραμέτρων αναμένεται γενικά να είναι μη γραμμική αλλά σε πρώτη τάξη θα μπορούσε να προσεγγιστεί στοχαστικά από ένα γραμμικό μοντέλο:

$$\mathbf{y} - \bar{\mathbf{y}} = W(\mathbf{x} - \bar{\mathbf{x}}) + \boldsymbol{\epsilon} \tag{5.2}$$

Το λάθος $\boldsymbol{\epsilon}$ της προσέγγισης θεωρείται μηδενικής μέσης τιμής Γκαουσιανό με συμμεταβλητότητα Q, δίνοντας $p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}; \bar{\mathbf{y}} + W(\mathbf{x} - \bar{\mathbf{x}}), Q)$. Ο στοχαστικός χαρακτήρας αυτής της προσέγγισης δικαιολογείται από το γεγονός ότι οι ακουστικές και οπτικές αναπαραστάσεις μπορεί να μην είναι άμεσα σχετιζόμενες με το σχήμα της φωνητικής οδού. Για παράδειγμα μια φασματική απεικόνιση του ακουστικού σήματος επηρεάζεται επίσης από την πηγή στη γλωττίδα ενώ μια απεικόνιση υφής για το πρόσωπο μπορεί να μεταβάλλεται και με βάση μια συγκεκριμένη έκφραση του προσώπου. Επιπλέον, αβεβαιότητα στη μοντελοποίηση και στις μετρήσεις πρέπει επίσης να ληφθεί υπόψη. Η λύση που προκύπτει από τη μεγιστοποίηση της εκ των υστέρων πιθανότητας είναι:

$$\hat{\mathbf{x}} = (\Sigma_x^{-1} + W^T Q^{-1} W)^{-1} \left(\Sigma_x^{-1}\bar{\mathbf{x}} + W^T Q^{-1}(\mathbf{y} - \bar{\mathbf{y}} + W\bar{\mathbf{x}})\right) \tag{5.3}$$

Η εκτιμώμενη λύση είναι ένας σταθμισμένος μέσος της παρατήρησης και των πρότερων μοντέλων. Τα βάρη είναι ανάλογα με τη σχετική αξιοπιστία των δύο προσθετέων.

5.2.1 Υπολογισμός του γραμμικού μοντέλου

Το γραμμικό μοντέλο μπορεί να προσδιοριστεί με τη χρήση τεχνικών πολυμεταβλητής γραμμικής ανάλυσης. Τέτοιες τεχνικές έχουν μελετηθεί σε βάθος στη στατιστική και σε άλλα πεδία ποσοτικής ανάλυσης δεδομένων· μια ενδιαφέρουσα εισαγωγή γίνεται στο [106]. Μπορούμε εύκολα να δούμε ότι, όταν έχουμε ακριβή γνώση των στατιστικών χαρακτηριστικών δεύτερης τάξης στη μορφή πινάκων συμμεταβλητότητας $R_{xx} = E[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T]$, $R_{yy} = E[(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{y} - \bar{\mathbf{y}})^T]$, και $R_{yx} = E[(\mathbf{y} - \bar{\mathbf{y}})(\mathbf{x} - \bar{\mathbf{x}})^T]$, τότε η βέλτιστη επιλογή υπό την έννοια του ελάχιστου τετραγωνικού λάθους για τον $m \times n$ πίνακα W αντιστοιχεί στο φίλτρο Wiener.

$$W = R_{yx} R_{xx}^{-1}, \tag{5.4}$$

και η συμμεταβλητότητα του λάθους προσέγγισης στην (5.2) είναι:

$$Q = R_{yy} - R_{yx} R_{xx}^{-1} R_{yx}^T. \tag{5.5}$$

Αφού τα δεύτερης τάξης στατιστικά χαρακτηριστικά στην πράξη είναι άγνωστα εκ των προτέρων, θα πρέπει να συμβιβαστούμε με εκτιμήσεις τους από τα διαθέσιμα κάθε φορά δεδομένα. Αν έχουμε N δείγματα $\mathbf{x}_t$ και $\mathbf{y}_t$, όπου $t = 1, \ldots, N$, τότε λογικές εκτιμήσεις της μέσης τιμής και της συμμεταβλητότητας του $\mathbf{x}$ είναι $\bar{\mathbf{x}} \approx \frac{1}{N}\sum_{t=1}^{N} \mathbf{x}_t$ και $R_{xx} \approx \frac{1}{N}\sum_{t=1}^{N}(\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})^T$, αντίστοιχα, και παρόμοια για $\bar{\mathbf{y}}$, R_{yy} και R_{yx}. Αυτές οι εκτιμήσεις μπορεί να

μην είναι αρκετά αξιόπιστες όταν το μέγεθους N του συνόλου δεδομένων εκπαίδευσης είναι μικρό σε σχέση με τις διαστάσεις των δεδομένων n του $\mathbf{x}$, m του $\mathbf{y}$, και, κατά συνέπεια, όταν εισάγονται στην (5.4) για να δώσουν τον πίνακα W, μπορεί να οδηγήσουν σε ιδιαίτερα φτωχές επιδόσεις όταν αργότερα εφαρμόσουμε τον γραμμικό προβλέπτη (5.2) σε άγνωστα δεδομένα. Θα δούμε ότι η ανάλυση κανονικής συσχέτισης, μεταξύ άλλων πλεονεκτημάτων, παρέχει έναν αξιόπιστο μηχανισμό για την επιλογή πολυμεταβλητών γραμμικών μοντέλων περιορισμένης τάξης τα οποία μπορεί να έχουν πολύ καλύτερες επιδόσεις από τα αντίστοιχα μοντέλα πλήρους τάξης σε περιπτώσεις μικρών συνόλων δεδομένων εκπαίδευσης.

5.2.2 Ανάλυση κανονικής συσχέτισης

Η ανάλυση κανονικής συσχέτισης είναι μια πολυμεταβλητή στατιστική τεχνική ανάλυσης για τη μελέτη της συμμεταβολής δύο συνόλων μεταβλητών $\mathbf{x}$ και $\mathbf{y}$ [106, Κεφ. 10]. Όμοια με την περισσότερο γνωστή ανάλυση σε πρωτεύουσες συνιστώσες (Principal Component Analysis, PCA), η ανάλυση κανονικής συσχέτισης περιορίζει τη μεγάλη διάσταση των συνόλων δεδομένων και έτσι δημιουργεί περισσότερο συμπυκνωμένες και οικονομικές αναπαραστάσεις τους. Ανόμοια όμως με την PCA, είναι ειδικά σχεδιασμένη έτσι ώστε οι διατηρούμενοι υποχώροι των $\mathbf{x}$ και $\mathbf{y}$ να είναι όσο το δυνατόν περισσότερο συσχετισμένοι. Γι' αυτό η ανάλυση κανονικής συσχέτισης είναι ιδιαίτερα κατάλληλη για εφαρμογές μοντελοποίησης, όπως στο εξεταζόμενο πρόβλημα. Στην περίπτωση που τα $\mathbf{x}$ και $\mathbf{y}$ είναι Γκαουσιανά, μπορεί να αποδειχτεί ότι οι υποχώροι που δίνονται από την ανάλυση κανονικής συσχέτισης είναι επίσης βέλτιστοι υπό την έννοια ότι διατηρούν όσο το δυνατόν περισσότερη από την από κοινού πληροφορία μεταξύ των $\mathbf{x}$ και $\mathbf{y}$ [142]. Η ανάλυση κανονικής συσχέτισης σχετίζεται επίσης και με τη γραμμική διακριτική ανάλυση (LDA). Όμοια με την LDA, η ανάλυση κανονικής συσχέτισης μειώνει τις διαστάσεις του $\mathbf{x}$ διακριτικά. Η διαφορά είναι ότι η μεταβλητή-στόχος στην ανάλυση κανονικής συσχέτισης είναι διανυσματική και συνεχής ενώ στην LDA είναι βαθμωτή και διακριτή.

Υποθέτοντας κεντραρισμένα δεδομένα (μετά την αφαίρεση της μέσης τιμής τους), στην ανάλυση κανονικής συσχέτισης (ΑΚΣ) αναζητούμε διευθύνσεις, $\mathbf{a}$ (στο $\mathbf{x}$ χώρο) και $\mathbf{b}$ (στον $\mathbf{y}$ χώρο), έτσι ώστε οι προβολές των δεδομένων στις αντίστοιχες διευθύνσεις να είναι κατά το μέγιστο συσχετισμένες. Αυτό επιτυγχάνεται μέσω μεγιστοποίησης ως προς τα $\mathbf{a}$ και $\mathbf{b}$ του συντελεστή συσχέτισης μεταξύ των προβεβλημένων δεδομένων $\mathbf{a}^T\mathbf{x}$ και $\mathbf{b}^T\mathbf{y}$

$$\rho(\mathbf{a},\mathbf{b}) = \frac{\mathbf{a}^T R_{xy}\mathbf{b}}{\sqrt{\mathbf{a}^T R_{xx}\mathbf{a}}\sqrt{\mathbf{b}^T R_{yy}\mathbf{b}}}. \tag{5.6}$$

Έχοντας βρει το πρώτο ζευγάρι από τις *διευθύνσεις κανονικής συσχέτισης* $(\mathbf{a}_1,\mathbf{b}_1)$, μαζί με τον αντίστοιχο *κανονικό συντελεστή συσχέτισης* ρ_1, συνεχίζουμε επαναληπτικά για να βρούμε το επόμενο ζευγάρι $(\mathbf{a}_2,\mathbf{b}_2)$ διανυσμάτων που μεγιστοποιεί το συντελεστή $\rho(\mathbf{a},\mathbf{b})$, με τους περιορισμούς $\mathbf{a}_1^T R_{xx}\mathbf{a}_2 = 0$ και $\mathbf{b}_1^T R_{yy}\mathbf{b}_2 = 0$. Η ανάλυση συνεχίζεται επαναληπτικά και τελικά λαμβάνονται μέχρι και $k = \operatorname{rank}(R_{xy}) \leq \min(m,n)$ ζευγάρια διευθύνσεων $(\mathbf{a}_i,\mathbf{b}_i)$ και συντελεστών ρ_i της ΑΚΣ, με:

$$1 \geq \rho_1 \geq \ldots \geq \rho_k > 0, \tag{5.7}$$

που, με φθίνουσα βαρύτητα, περιγράφουν τις διευθύνσεις της συμμεταβολής των $\mathbf{x}$ και $\mathbf{y}$. Για περαιτέρω πληροφορίες σχετικά με την ΑΚΣ και αλγορίθμους, δείτε το [106].

Είναι ενδιαφέρον το ότι ο πίνακας του φίλτρου Wiener (5.4) του πολυμεταβλητού μοντέλου πρόβλεψης μπορεί να εκφραστεί με ευκολία με τη χρήση ΑΚΣ ως

$$W = R_{yx}R_{xx}^{-1} = R_{yy}BKA^T, \tag{5.8}$$

όπου $A = [\mathbf{a}_1 \ldots \mathbf{a}_k]$ και $B = [\mathbf{b}_1 \ldots \mathbf{b}_k]$ έχουν τις κανονικές διευθύνσεις ως στήλες και $K = \operatorname{diag}(\rho_1,\ldots,\rho_k)$ είναι ένας διαγώνιος πίνακας των διατεταγμένων συντελεστών κανονικής

συσχέτισης. Μπορεί να αποδειχτεί [142] ότι κρατώντας μόνο τις r πρώτες, $1 \leq r \leq k$, κανονικές διευθύνσεις, χρησιμοποιώντας δηλαδή το φίλτρο Wiener *μειωμένης τάξης*:

$$W_r \triangleq R_{yy} B_r K_r A_r^T, \tag{5.9}$$

με $A_r = [\mathbf{a}_1 \dots \mathbf{a}_r]$ και $B_r = [\mathbf{b}_1 \dots \mathbf{b}_r]$, και $K_r = \mathrm{diag}(\rho_1, \dots, \rho_r)$, επιτυγχάνεται βέλτιστο φιλτράρισμα για τα φίλτρα r τάξης με την έννοια του ελάχιστου τετραγωνικού λάθους. Αυτό που είναι πιο σημαντικό για εμάς, όταν το σύνολο εκπαίδευσης είναι πολύ μικρό για να εκτιμηθούν με ακρίβεια οι πίνακες συμμεταβλητότητας που χρειάζονται, αυτοί οι γραμμικοί προβλέπτες μειωμένης τάξης είναι δυνατόν να επιδείξουν βελτιωμένες επιδόσεις πρόβλεψης σε άγνωστα δεδομένα σε σύγκριση με τα μοντέλα πλήρους τάξης [27]. Αυτό είναι ανάλογο με τις βελτιωμένες επιδόσεις των μοντέλων που είναι βασισμένα σε ανάλυση σε πρωτεύουσες συνιστώσες και χρησιμοποιούνται σε καλά μελετημένα προβλήματα αναγνώρισης προτύπων, όπως είναι η αναγνώριση προσώπου, όταν διατηρείται μόνο ένα υποσύνολο από τις κυρίαρχες διευθύνσεις.

5.3 Δυναμική και οπτικοακουστική σύμμειξη

5.3.1 Δυναμικά εναλλασσόμενη απεικόνιση για προσαρμοστική αντιστροφή

Το περιγραφόμενο πλαίσιο μπορεί να επεκταθεί και στην περίπτωση υπολογισμού αρθρωτικών παραμέτρων από χρονομεταβλητές οπτικοακουστικές ακολουθίες δεδομένων. Οι πιθανότητες στην Εξίσωση (5.1) θα αφορούν τώρα σε ακολουθίες διανυσμάτων. Το κύριο πρόβλημα είναι η εύρεση καλών πρότερων μοντέλων και μοντέλων παρατήρησης που θα έκαναν δυνατή την εύρεση λύσης. Αυτό δεν είναι απλό δεδομένης της πολυπλοκότητας της σχέσης μεταξύ του ακουστικού χώρου και του χώρου των αρθρωτικών παραμέτρων, η οποία σε γενικές γραμμές είναι μη γραμμική και όχι ένα προς ένα. Επιπλέον, η οπτική πληροφορία πρέπει να αξιοποιηθεί κατάλληλα έτσι ώστε να μειώσει κάπως τον αριθμό των πιθανών λύσεων.

Διαισθητικά, στην περίπτωση συνεχούς λόγου, αναμένουμε η γραμμική προσέγγιση της Εξίσωσης (5.2) να είναι αποδεκτή μόνο για περιορισμένα χρονικά διαστήματα που αντιστοιχούν σε συγκεκριμένο φώνημα, ή τουλάχιστον σε μέρος του φωνήματος, δηλαδή σε μετάβαση ή στη σταθερή κατάσταση. Το ίδιο ισχύει και για το πρότερο μοντέλο άρθρωσης, δηλαδή την κατανομή πυκνότητας πιθανότητας του $\mathbf{x}$. Γι' αυτό αναμένουμε ότι χρησιμοποιώντας διαφορετικές απεικονίσεις και πρότερα μοντέλα για κάθε φώνημα (ή για μεταβάσεις μεταξύ φωνημάτων όπως στο [45]) θα έχουμε καλύτερα αποτελέσματα από ότι αν χρησιμοποιήσουμε μια καθολική γραμμική προσέγγιση. Αυτό απαιτεί τον προσδιορισμό της διαδικασίας εναλλαγής μεταξύ των επιμέρους μοντέλων, ουσιαστικά οδηγώντας σε μια κατά τμήματα γραμμική προσέγγιση της σχέσεις μεταξύ των παρατηρούμενων παραμέτρων και των παραμέτρων της φωνητικής άρθρωσης.

Για αυτό το σκοπό μπορούν να χρησιμοποιηθούν κρυφά Μαρκοβιανά μοντέλα ανά φώνημα [63]. Κάθε κατάσταση αντιστοιχεί σε διαφορετικό πρότερο μοντέλο $p(\mathbf{x})$ για τις παραμέτρους άρθρωσης και μοντέλο παρατήρησης $p(\mathbf{y}|\mathbf{x})$ για τη γραμμική απεικόνιση μεταξύ των παρατηρήσεων και των χαρακτηριστικών άρθρωσης. Πιο συγκεκριμένα, επεκτείνοντας την ανάλυση στην Ενότητα 5.2, η πρότερη και η υπό συνθήκη κατανομές πιθανότητας στην κατάσταση c θεώρουνται ως

$$p_c(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \bar{\mathbf{x}}_c, \Sigma_{x,c}) \tag{5.10}$$

$$p_c(\mathbf{y}|\mathbf{x}) = \mathcal{N}\left(\mathbf{y}; \bar{\mathbf{y}}_c + W_c(\mathbf{x} - \bar{\mathbf{x}}_c), Q_c\right). \tag{5.11}$$

Τότε (π.χ. [25, Ενότητα. 2.3.3]) η αντίστοιχη κατανομή του $\mathbf{y}$ είναι

$$p_c(\mathbf{y}) = \mathcal{N}(\mathbf{y}; \bar{\mathbf{y}}_c, \Sigma_{y,c}), \tag{5.12}$$

με $\Sigma_{y,c} = W_c \Sigma_{x,c} W_c^T + Q_c$, και η υπό συνθήκη κατανομή του $\mathbf{x}$ δεδομένου του $\mathbf{y}$ είναι

$$p_c(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_c, \Sigma_{\hat{x},c}), \text{ με} \tag{5.13}$$

$$\hat{\mathbf{x}}_c = \Sigma_{\hat{x},c} \left(\Sigma_{x,c}^{-1} \bar{\mathbf{x}}_c + W_c^T Q_c^{-1} (\mathbf{y} - \bar{\mathbf{y}}_c + W_c \bar{\mathbf{x}}_c) \right) \tag{5.14}$$

$$\Sigma_{\hat{x},c}^{-1} = \Sigma_{x,c}^{-1} + W_c^T Q_c^{-1} W_c. \tag{5.15}$$

Σημειώνεται ότι η Εξίσωση (5.14) είναι η γενίκευση του εκτιμητή στην Εξίσωση (5.3) για πολλαπλά μοντέλα. Σε αυτό το πλαίσιο, για να εκτιμηθεί η διαδικασία εναλλαγής μεταξύ των επιμέρους μοντέλων $\mathcal{M}_c = \{W_c, Q_c, \bar{\mathbf{x}}_c, \bar{\mathbf{y}}_c, \Sigma_{x,c}\}$ (ένα για κάθε κατάσταση), η αντιστροφή απαιτεί την εύρεση της βέλτιστης ακολουθίας καταστάσεων $\mathbf{c}^*$ δεδομένων των παρατηρήσεων (ακολουθίες $\mathcal{Y}$ ακουστικών, οπτικών ή οπτικοακουστικών χαρακτηριστικών)

$$\mathbf{c}^* = \arg\max_{\mathbf{c}} P(\mathcal{Y}|\mathbf{c}). \tag{5.16}$$

Δεδομένης της Εξίσωσης (5.12), αυτό μπορεί να επιτευχθεί χρησιμοποιώντας τον αλγόριθμο Viterbi όπως με τα συμβατικά κρυφά Μαρκοβιανά μοντέλα [63]. Στη συνέχεια, για κάθε διάνυσμα παρατήρησης που έχει αντιστοιχηθεί σε κάποια κατάσταση το αντίστοιχα διάνυσμα των παραμέτρων άρθρωσης μπορεί να υπολογιστεί χρησιμοποιώντας τον εκτιμητή στη συγκεκριμένη κατάσταση, Εξίσωση (5.14). Για την επιβολή συνέχειας στις εκτιμώμενες τροχιές των παραμέτρων άρθρωσης είναι δυνατή η εφαρμογή ενός εκ των υστέρων σταδίου επεξεργασίας όπως στο [63] με τη χρησιμοποίηση των παραγώγων των παρατηρήσεων και των παραμέτρων άρθρωσης ή η υιοθέτηση ενός πιο σύνθετου πρότερου μοντέλου για το χώρο καταστάσεων σε μια συνδυασμένη προσέγγιση κρυφών Μαρκοβιανών μοντέλων και φιλτραρίσματος Kalman [80].

Η εκπαίδευση των πρότερων πιθανοτήτων και των πιθανοτήτων μετάβασης των κρυφών Μαρκοβιανών μοντέλων, όπως επίσης και οι μέσες τιμές και μεταβλητότητες ανά κατάσταση $\{\bar{\mathbf{y}}_c, \Sigma_{y,c}\}$ που αντιστοιχούν στις παρατηρήσεις $\mathbf{y}$ είναι δυνατή με το συμβατικό τρόπο μέσω μεγιστοποίησης της πιθανοφάνειας με τον αλγόριθμο μεγιστοποίησης της προσδοκίας [63]. Με δεδομένες τις τελικές πιθανότητες ανάθεσης σε κάθε κατάσταση $\gamma_t(c)$, που καθεμιά τους είναι η πιθανότητα η παρατήρηση $\mathbf{y}_t$ να αντιστοιχεί στην κατάσταση c τη χρονική στιγμή t και υπολογίζεται με χρήση της διαδικασίας forward-backward [133], έχουμε

$$\bar{\mathbf{x}}_c = \frac{\sum_t \gamma_t(c) \mathbf{x}_t}{\sum_t \gamma_t(c)} \tag{5.17}$$

$$\Sigma_{x,c} = \frac{\sum_t \gamma_t(c) (\mathbf{x}_t - \bar{\mathbf{x}}_c)(\mathbf{x}_t - \bar{\mathbf{x}}_c)^T}{\sum_t \gamma_t(c)} \tag{5.18}$$

όπου $\mathbf{x}_t$ είναι το διάνυσμα των χαρακτηριστικών άρθρωσης της στιγμή t. Για να βρούμε τον πίνακα W_c θα πρέπει να λύσουμε τις εξισώσεις [63]

$$\sum_t \gamma_t(c) [(\mathbf{y}_t - \bar{\mathbf{y}}_c) - W_c (\mathbf{x}_t - \bar{\mathbf{x}}_c)] (\mathbf{x}_t - \bar{\mathbf{x}}_c)^T = 0 \tag{5.19}$$

που είναι ίδιες με τις εξισώσεις που προκύπτουν όταν λύνεται το σταθμισμένο πρόβλημα ελαχίστων τετραγώνων όπου τα $\mathbf{x}_t$ και $\mathbf{y}_t$ σταθμίζονται με $\gamma_t(c)^{1/2}$ [40]. Εκτιμούμε τον W_c μέσω ανάλυσης κανονικής συσχέτισης όπως περιγράφεται στην Ενότητα 5.2.2 χρησιμοποιώντας ακριβώς αυτές τις σταθμισμένες εκδόσεις των δεδομένων. Η βέλτιστη τάξη του μοντέλου της ανάλυσης κανονικής συσχέτισης προσδιορίζεται μέσω μιας διαδικασίας πολλαπλής επιβεβαίωσης (cross validation) όπως συζητείται περαιτέρω στην Ενότητα 5.5. Τέλος, για τη συμμεταβλητότητα Q_c του λάθους έχουμε

$$Q_c = \frac{\sum_t \gamma_t(c) \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T}{\sum_t \gamma_t(c)}, \text{ όπου } \boldsymbol{\epsilon}_t = \mathbf{y}_t - \bar{\mathbf{y}}_c - W_c (\mathbf{x}_t - \bar{\mathbf{x}}_c). \tag{5.20}$$

5.3.2 Οπτικοακουστική σύμμειξη για αντιστροφή

Η ταυτοποίηση της κρυφής δυναμικής της φωνής και η ανάκτηση των ιδιοτήτων της άρθρωσης μπορούν να ωφεληθούν σημαντικά από την κατάλληλη εισαγωγή οπτικής πληροφορίας στο προτεινόμενο σχήμα. Οι διαδικασίες εναλλαγής των οπτικών και ακουστικών απεικονίσεων μπορούν να αλληλεπιδρούν σε διάφορα επίπεδα συγχρονισμού. Διερευνώνται διάφορες εναλλακτικές οπτικακουστικής σύμμειξης.

5.3.2.1 Περίπτωση πλήρους συγχρονισμού (πολυκαναλικά κρυφά Μαρκοβιανά μοντέλα)

Το σενάριο του πλήρους συγχρονισμού βασίζεται στην υπόθεση ότι οι μεταβολές της άρθρωσης αντανακλώνται ταυτόχρονα τόσο στην ακουστική όσο και στην οπτική συνιστώσα της φωνής. Η από κοινού δυναμική μπορεί να απεικονιστεί αποτελεσματικά με τη χρήση πολυκαναλικών κρυφών Μαρκοβιανών μοντέλων. Τέτοια μοντέλα έχουν χρησιμοποιηθεί ευρέως και με επιτυχία για οπτικακουστική αναγνώριση φωνής [44, 122, 157]. Η από κοινού αντιστοίχιση (alignment) των καταστάσεων είναι εφικτή μέσω κατάλληλης εφαρμογής του αλγορίθμου Viterbi. Με αυτόν τον τρόπο, η συμμετοχή της κάθε συνιστώσας στην αντιστοίχιση είναι ανεξάρτητα ελεγχόμενη, το οποίο δεν ισχύει για τα απλά κρυφά Μαρκοβιανά μοντέλα. Το τροποποιημένο σκορ σε κάθε κατάσταση είναι

$$b(\mathbf{y}|c) \propto \mathcal{N}(\mathbf{y}_{\mathrm{a}}; \bar{\mathbf{y}}_{\mathrm{a},c}, \Sigma_{\mathrm{a},c})^{w_{\mathrm{a}}} \mathcal{N}(\mathbf{y}_{\mathrm{v}}; \bar{\mathbf{y}}_{\mathrm{v},c}, \Sigma_{\mathrm{v},c})^{w_{\mathrm{v}}} \tag{5.21}$$

όπου c είναι η κοινή κατάσταση και για τις δύο συνιστώσες και τα βάρη w_{a} και w_{v} αθροίζουν στη μονάδα. Παρά του ότι αυτή η προσέγγιση παρέχει έναν ευθύ τρόπο για το συνδυασμό των δύο καναλιών πληροφορίας, μπορεί να είναι ιδιαίτερα περιοριστική όσον αφορά στο συγχρονισμό. Πιο ευέλικτες παραλλαγές κρυφών Μαρκοβιανών μοντέλων όπως τα κρυφά Μαρκοβιανά μοντέλα που προκύπτουν από γινόμενο μοντέλων ξεχωριστών για κάθε κανάλι [99] θα μπορούσαν μερικώς να αντιμετωπίσουν το εν λόγω πρόβλημα.

5.3.2.2 Τελείως ασύγχρονη περίπτωση (Εκ των υστέρων σύμμειξη)

Στο άλλο άκρο, η δυναμική των απεικονίσεων από τις συνιστώσες του ήχου και της εικόνες στις παραμέτρους της άρθρωσης μπορεί να μοντελοποιηθεί ξεχωριστά και χωρίς κανένα περιορισμό συγχρονισμού. Θεωρείται ότι προσδιορίζεται από ξεχωριστές διαδικασίες εναλλαγής και διαφορετικά κρυφά Μαρκοβιανά μοντέλα χρησιμοποιούνται για κάθε κανάλι. Συνδυασμός της συμπληρωματικής πληροφορίας επιτυγχάνεται κατ' επέκταση σε ένα μετέπειτα στάδιο, αφού το κάθε κανάλι παρατήρησης έχει ανεξάρτητα αντιστραφεί σε τροχιές των χαρακτηριστικών άρθρωσης. Αξιοποιώντας την προκύπτουσα ευελιξία, μπορούν να υιοθετηθούν πιο αντιπροσωπευτικά και ακριβή μοντέλα για κάθε κανάλι, π.χ. κρυφά Μαρκοβιανά μοντέλα ανά οπτικό φώνημα για το πρόσωπο και ανά ακουστικό φώνημα για το ακουστικό σήμα. Τα οπτικά φωνήματα αντιστοιχούν σε ομάδες φωνημάτων τα οποία δεν μπορούν να διακριθούν μεταξύ τους οπτικά και αποτελούν περισσότερο φυσικές δομικές μονάδες για τον οπτικό λόγο [129]. Για παράδειγμα, το οπτικό φώνημα *Π* αντιστοιχεί στο σύνολο των ακουστικών φωνημάτων /π/,/μπ/,/μ/. Το προτεινόμενο ασύγχρονο σχήμα είναι μερικώς περιορισμένο με την έννοια ότι δεν αξιοποιεί συσχετίσεις μεταξύ των καναλιών για να προσδιορίσει την ακολουθία των σύνθετων καταστάσεων της άρθρωσης. Προσφέρει όμως ευελιξία στη μοντελοποίηση και δε χρειάζεται κάποια πρότερη γνώση ή υπόθεση σχετικά με το συγχρονισμό των εμπλεκόμενων καναλιών πληροφορίας.

Δεδομένης της σύνθετης κατάστασης $c = \{c_a, c_{\mathrm{v}}\}$ κάθε στιγμή, δηλαδή της εναλλασσόμενης ακολουθίας που καθορίζει την εφαρμοζόμενη κατά τμήματα γραμμική απεικόνιση από την οπτικοακουστική πληροφορία σε πληροφορία άρθρωσης, το οπτικό και το ακουστικό κανάλι συνεισφέρουν στη διαδικασία αντιστροφής σταθμισμένα με τη σχετική τους αξιοπιστία.

Αυτό επιτυγχάνεται τόσο στην σύγχρονη, όπου $c_a = c_\mathrm{v}$, όσο και στην ασύγχρονη περίπτωση. Υποθέτοντας ανεξαρτησία των λαθών παρατήρησης σε κάθε κανάλι, η τελική εκτίμηση της κατάστασης άρθρωσης είναι

$$\hat{\mathbf{x}}_c = \Sigma_{f,c}\big(\Sigma_{x,c}^{-1}\bar{\mathbf{x}}_c + W_{\mathrm{a},c_\mathrm{a}}^T Q_{\mathrm{a},c_\mathrm{a}}^{-1}(\mathbf{y}_\mathrm{a} - \bar{\mathbf{y}}_{\mathrm{a},c} + W_{\mathrm{a},c_\mathrm{a}}\bar{\mathbf{x}}_c) + W_{\mathrm{v},c_\mathrm{v}}^T Q_{\mathrm{v},c_\mathrm{v}}^{-1}(\mathbf{y}_\mathrm{v} - \bar{\mathbf{y}}_{\mathrm{v},c} + W_{\mathrm{v},c_\mathrm{v}}\bar{\mathbf{x}}_c)\big) \quad (5.22)$$

όπου

$$\Sigma_{f,c}^{-1} = \Sigma_{x,c}^{-1} + W_{\mathrm{a},c_\mathrm{a}}^T Q_{\mathrm{a},c_\mathrm{a}}^{-1} W_{\mathrm{a},c_\mathrm{a}} + W_{\mathrm{v},c_\mathrm{v}}^T Q_{\mathrm{v},c_\mathrm{v}}^{-1} W_{\mathrm{v},c_\mathrm{v}}$$

δίνει την αβεβαιότητα $\Sigma_{f,c}$ του σύνθετου αποτελέσματος της αντιστροφής που περιλαμβάνει τόσο την αβεβαιότητα του πρότερου μοντέλου όσο και του μοντέλου παρατήρησης. Γραμμικά μοντέλα $W_{s,c}, Q_{s,c}$ στη μορφή που δίνεται στην Εξίσωση (5.2) υπολογίζονται όπως περιγράφεται στην Ενότητα 5.3.1, το καθένα να αντιστοιχεί σε διαφορετικό κανάλι s. Όσο περισσότερο ακριβές είναι ένα κανάλι πληροφορίας, δηλαδή όσο μικρότερη είναι η συμμεταβλητότητα του λάθους $Q_{s,c}$, τόσο περισσότερο επηρεάζει την τελική εκτίμηση. Χαλαρώνοντας την υπόθεση ανεξαρτησίας, στην περίπτωση της σύγχρονης σύμμειξης, μπορούμε να λάβουμε υπόψη μας συσχετίσεις μεταξύ των εμπλεκόμενων καναλιών χρησιμοποιώντας ένα σύνθετο οπτικακουστικό γραμμικό μοντέλο ανά κατάσταση $W_{\mathrm{av},c}, Q_{\mathrm{av},c}$. Η προβλεπόμενη άρθρωση τότε είναι

$$\hat{\mathbf{x}}_c = \Sigma_{f,c}\left(\Sigma_{x,c}^{-1}\bar{\mathbf{x}}_c + W_{\mathrm{av},c}^T Q_{\mathrm{av},c}^{-1}(\mathbf{y} - \bar{\mathbf{y}}_c + W_{\mathrm{av},c}\bar{\mathbf{x}}_c)\right) \quad (5.23)$$

όπου σε αυτή την περίπτωση η ακρίβεια πρόβλεψης

$$\Sigma_{f,c}^{-1} = \Sigma_{x,c}^{-1} + W_{\mathrm{av},c}^T Q_{\mathrm{av},c}^{-1} W_{\mathrm{av},c}$$

εξάγεται με βάση την αβεβαιότητα του πρότερου μοντέλου και του σύνθετου μοντέλου οπτικοακουστικής παρατήρησης.

5.4 Ανάλυση του προσώπου με ενεργά μοντέλα εμφάνισης

Χρησιμοποιούνται *ενεργά μοντέλα εμφάνισης* [35] προσώπων [2] για να ακολουθηθούν με ακρίβεια οι κινήσεις του προσώπου του ομιλητή και να εξαχθούν οπτικά χαρακτηριστικά τόσο από το σχήμα όσο και από την υφή του. Τα ενεργά μοντέλα εμφάνισης είναι αναγεννητικά μοντέλα της εμφάνισης ενός αντικειμένου και έχουν αποδειχτεί αποτελεσματικά στη μοντελοποίηση ανθρώπινων προσώπων για διάφορες εφαρμογές, όπως αναγνώριση προσώπων ή ιχνηλάτηση. Σε αυτό το πλαίσιο το σχήμα ενός αντικειμένου μοντελοποιείται ως μία μάσκα που ορίζεται από ένα σύνολο L σημαντικών σημείων, των οποίων οι συντεταγμένες σχηματίζουν ένα διάνυσμα σχήματος $\mathbf{s}$ μήκους $2L$. Επιτρέπουμε αποκλίσεις από το μέσο σχήμα $\mathbf{s}_0$ δεχόμενοι ότι το $\mathbf{s}$ βρίσκεται σε έναν N_p-διάστατο υπόχωρο και παίρνουμε

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{N_p} p_i \mathbf{s}_i \quad (5.24)$$

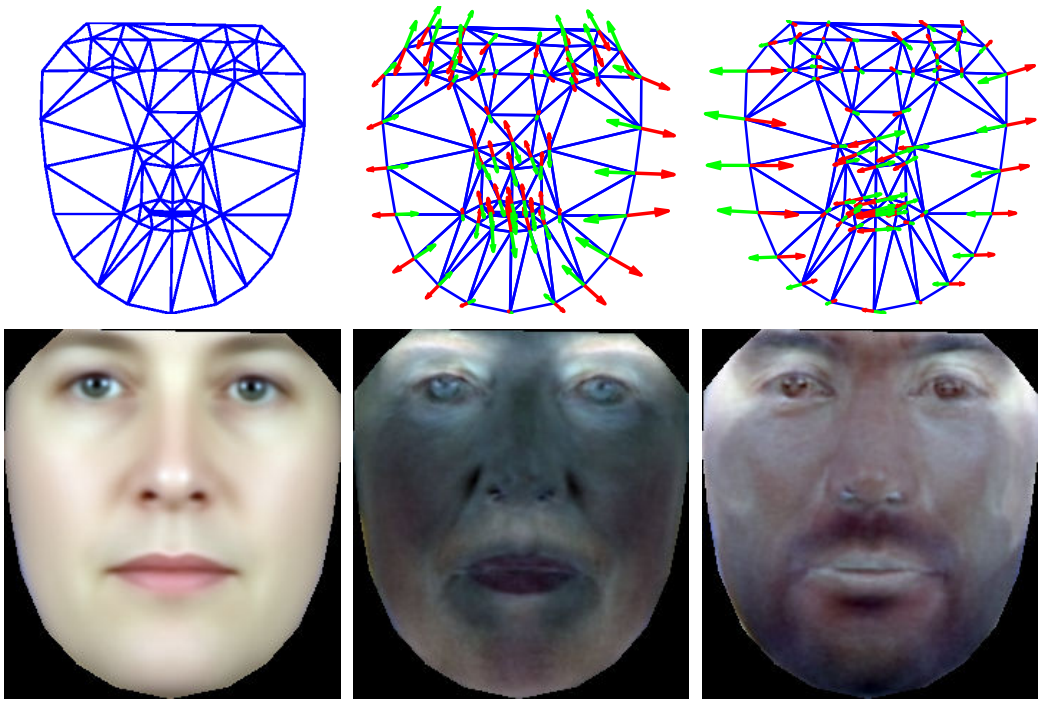
Η διαφορά του σχήματος $\mathbf{s}$ από το μέσο σχήμα $\mathbf{s}_0$ ορίζει ένα μετασχηματισμό $\mathbf{W}(\mathbf{p})$, που εφαρμόζεται για να φέρει το πρόσωπο στην τρέχουσα εικόνα I σε αντιστοιχία με το πρότυπο του μέσου προσώπου. Μετά από αυτή τη διαδικασία η έγχρωμη υφή του προσώπου μπορεί να μοντελοποιηθεί ως ένα σταθμισμένο άθροισμα 'ιδιοπροσώπων' $\{A_i\}$, δηλαδή

$$I(\mathbf{W}(\mathbf{p})) \approx A_0 + \sum_{i=1}^{N_\lambda} \lambda_i A_i, \quad (5.25)$$

[2] Σε συνεργασία με τον Γ. Παπανδρέου

όπου A_0 είναι η μέση υφή των προσώπων. Τόσο οι βάσεις των ιδιοσχημάτων όσο και των ιδιοπροσώπων δημιουργούνται σε μια φάση εκπαίδευσης, χρησιμοποιώντας ένα αντιπροσωπευτικό σύνολο από εικόνες που έχουν σημαδευτεί κατάλληλα με το χέρι, [35].

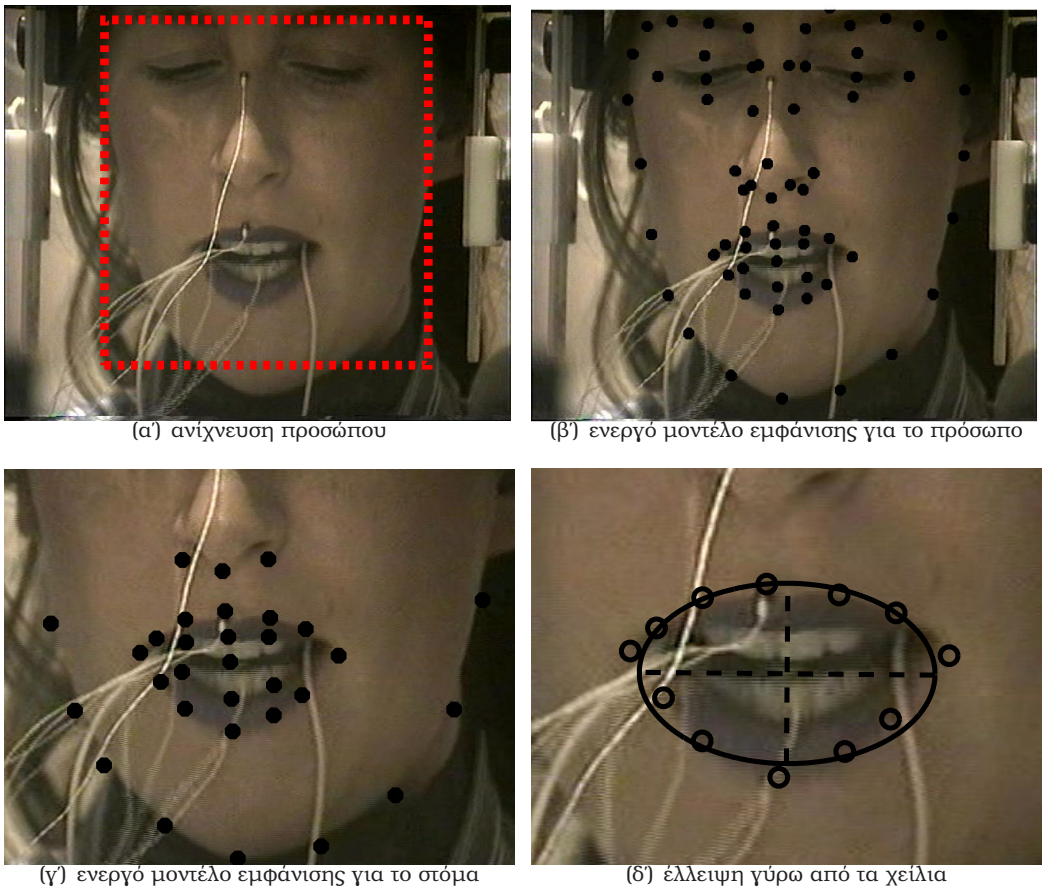
Τα σχήματα του συνόλου εκπαίδευσης πρώτα ευθυγραμμίζονται και στη συνέχεια μια ανάλυση πρωτεουουσών συνιστωσών δίνει τις βασικές κατευθύνσεις μεταβολής του σχήματος, $\{\mathbf{s}_i\}$. Παρόμοια, οι αρχικές πρωτεύουσες συνιστώσες των εκπαιδευτικών διανυσμάτων υφής αποτελούν το σύνολο ιδιοπροσώπων $\{A_i\}$. Τα πρώτα τρία ιδιοσχήματα και ιδιοπρόσωπα που εξάγονται με αυτή τη διαδικασία φαινονται στο Σχήμα 5.1.



Σχήμα 5.1: *Ενεργά μοντέλα εμφάνισης. Πάνω: Μέσο σχήμα* $\mathbf{s}_0$ *και τα δύο πρώτα ιδιοσχήματα* $\mathbf{s}_1$ *και* $\mathbf{s}_2$*. Κάτω: Μέση υφή* A_0 *και τα δύο πρώτα ιδιοπρόσωπα* A_1 *και* A_2*.*

Δοσμένου ενός εκπαιδευμένου ενεργού μοντέλου εμφάνισης, η μοντελοποίηση έγκειται στην εύρεση για κάθε πλαίσιο I_t του βίντεο τις παραμέτρους $\mathbf{q}_t \equiv \{\mathbf{p}_t, \boldsymbol{\lambda}_t\}$ που ελαχιστοποιούν το τετραγωνικό λάθος ανακατασκευής $I_t(\mathbf{W}(\mathbf{p}_t)) - A_0 - \sum_{i=1}^{N_\lambda} \lambda_{t,i} A_i$. Εφαρμόστηκαν αποδοτικοί επαναληπτικοί αλγόριθμοι που περιγράφονται στο [124] για την επίλυση αυτού του μη γραμμικού προβλήματος ελαχίστων τετραγώνων. Λόγω της επαναληπτικής φύσης των αλγορίθωμν μοντελοποίησης με τη χρήση ενεργών μοντέλων εμφάνισης, η μάσκα σχήματος του μοντέλου πρέπει να αρχικοποιηθεί σχετικά κοντά στο πρόσωπο ώστε να υπάρχει επιτυχία. Για την αυτοματοποίηση της αρχικοποίησης της μάσκας του μοντέλου εφαρμόζουμε έναν ανιχνευτή προσώπου βασιμένο στον αλγόριθμό Adaboost [174]. Με αυτόν τον τρόπο παίρνουμε τη θέση του προσώπου στο αρχικό πλαίσιο και αρχικοποιούμε το σχήμα του ενεργού μοντέλου εμφάνισης, όπως φαίνεται στο Σχήμα 5.2(α'). Στη συνέχεια, για κάθε επόμενο πλαίσιο, αρχικοποιούμε τον αλγόριθμο με το σχήμα όπως έχει προκύψει από το μοντέλο για το προηγούμενο πλαίσιο μετά τη σύγκλιση.

Στα πειράματα χρησιμοποιήθηκε μια ιεραρχία δύο ενεργών μοντέλων εμφάνισης. Το πρώτο *μουτέλο προσώπου*, βλέπε Σχήμα 5.2(β'), καλύπτει όλο το πρόσωπο και μπορεί με αξιοπιστία να παρακολουθήσει τον ομιλητή σε μακριές ακολουθίες βίντεο. Το δεύτερο, *μουτέλο της περιοχής ενδιαφέροντος*, βλέπε Σχήμα 5.2(γ'), καλύπτει μόνο την περιοχή ενδιαφέροντος γύρω από το στόμα και για αυτό είναι πιο εστιασμένο στο τμήμα του προσώπου που είναι περισσότερο πληροφοριακό στον οπτικό λόγο. Λόγω του ότι το δεύτερο μοντέλο καλύπτει πολύ μικρή περιοχή του προσώπου ώστε να επιτρέψει αξιόπιστη παρακολούθηση



(α') ανίχνευση προσώπου (β') ενεργό μοντέλο εμφάνισης για το πρόσωπο

(γ') ενεργό μοντέλο εμφάνισης για το στόμα (δ') έλλειψη γύρω από τα χείλια

Σχήμα 5.2: *Ανάλυση του προσώπου της ομιλήτριας της βάσης MOCHA με τη χρήση ενεργών μοντέλων εμφάνισης. (α) Αποτέλεσμα αυτόματης ανίχνευσης του προσώπου για αρχικοποίηση του ενεργού μοντέλου. (β) Τελείες που αντιστοιχούν σε σημαντικά σημεία για το μοντέλο του προσώπου, όπως αυτές εντοπίζονται μέσω αυτόματης μοντελοποίησης. (γ) Σημαντικά σημεία για το μοντέλο της περιοχής ενδιαφέροντος. (δ) Οι μικροί κύκλοι είναι το υποσύνολο των σημαντικών σημείων ενδιαφέροντος του μουτέλου του στόματος που περιγράφουν τα χείλια του ομιλητή. Η έλλειψη που φαίνεται είναι αυτή που ταιριάζει βέλτιστα στα συγκεκριμένα σημεία.*

του προσώπου, χρησιμοποιείται μόνο για την ανάλυση του σχήματος και της υφής της περιοχής του στόματος , όπως έχει ήδη εντοπιστεί από το μοντέλο προσώπου. Ως διάνυσμα οπτικών χαρακτηριστικών χρησιμοποιούμε τις παραμέτρους ανάλυσης $\mathbf{q}_t$ του μοντέλου της περιοχής ενδιαφέροντος.

Έχοντας εντοπίσει σημαντικά σημεία στο πρόσωπο με τη χρήση του ενεργού μοντέλου εμφάνισης, είναι δυνατή η εξαγωγή εναλλακτικών μετρήσεων του προσώπου που είναι απλό να ερμηνευθούν γεωμετρικά. Για του λόγου το αληθές, χρησιμοποιούμε τα σημεία που έχουν εντοπιστεί με τη μοντελοποίση γύρω από τα χείλια για να ταιριάξουμε μια έλλειψη που θα περιγράφει το σχήμα του στόματος ακολουθώντας την τεχνική του [54]. Το αποτέλεσμα φαίνεται στο Σχήμα 5.2(δ'). Οι άξονες της έλλειψης αντιστοιχούν στο οριζόντιο και κατακόρυφο άνοιγμα του στόματος.

5.5 Πειραματικά αποτελέσματα και συζήτηση

Στα πειράματα που ακολουθούν επιδεικνύεται ότι με την προτεινόμενη προσέγγιση είναι δυνατή η αποτελεσματική εξαγωγή και αξιοποίηση οπτικής πληροφορίας από το πρόσωπο του ομιλητή σε συνδυασμό με το ακουστικό σήμα για την ανάκτηση πληροφορίας που αφορά στην άρθρωση. Οι ακολουθίες των οπτικών και ακουστικών χαρακτηριστικών της φωνής συν-

δυάζονται κατάλληλα για να ανακτήσουν τις αντίστοιχες τροχιές των παραμέτρων άρθρωσης. Πρόκειται για τις τροχιές σημείων πάνω σε σημαντικούς αρθρωτές, όπως είναι η γλώσσα, τα δόντια και τα χείλια, και στην ουσία παρέχουν ένα απλό μέσο αναπαράστασης της φωνητικής οδού κατά την ομιλία. Για την εκπαίδευση των μοντέλων χρησιμοποιήθηκαν ταυτόχρονα καταγεγραμμένα ακουστικά, βίντεο δεδομένα καθώς και δεδομένα αρθρωτών καταγεγραμμένα ηλεκτρομαγνητικά. Τα τελευταία στην ουσία είναι οι συνιστώσες μικρών πηνίων που έχουν κολληθεί πάνω στους αρθρωτές και ιχνηλατώνται με τη χρήση ειδικού εξοπλισμού. Μέρος των διαθέσιμων δεδομένων εξαιρέθηκε ώστε να χρησιμοποιηθεί για αξιολόγηση.

5.5.1 Κριτήρια αξιολόγησης

Το σχήμα και η δυναμική των προβλεπόμενων τροχιών των παραμέτρων άρθρωσης συγκρίνονται με τις μετρηθείσες τροχιές μέσω δύο ποσοτικών κριτηρίων, που είναι η ρίζα του μέσου τετραγωνικού (RMS) λάθους e_{RMS} και ο συντελεστής συσχέτισης Pearson $\rho_{x\hat{x}}$. Το λάθος RMS φανερώνει τη συνολική διαφορά μεταξύ των εκτιμώμενων και των τροχιών που έχουν μετρηθεί, $\hat{x}_{1:T}$ και $x_{1:T}$ αντίστοιχα. Για την παράμετρο i της άρθρωσης και για διάρκεια T της αντίστοιχης τροχιάς υπολογίζεται ως

$$e_{RMS}[i] = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{\mathbf{x}}_t[i] - \mathbf{x}_t[i])^2}, \quad i = 1 \dots n \tag{5.26}$$

και παρέχει ένα μέτρο εκτίμησης της απόδοσης στις ίδιες μονάδες με τις μετρηθείσες τροχιές, δηλαδή σε χιλιοστά. Για να ληφθεί όμως μια εκτίμηση που να μπορεί καλύτερα να συνοψίζει την απόδοση της αντιστροφής για όλους τους αρθρωτές, χρησιμοποιούμε το αδιάστατο κανονικοποιημένο μέσο λάθος RMS, $\bar{e}_{NRMS}$. Αυτό ορίζεται ως

$$\bar{e}_{NRMS} = \frac{1}{n}\sum_{i=1}^{n}\frac{e_{RMS}[i]}{\sigma_i} \tag{5.27}$$

και επιτρέπει να ληφθεί επίσης υπόψη το γεγονός ότι οι τυπικές αποκλίσεις ($\{\sigma_i\},\ i = 1 \dots n$) των διαφορετικών παραμέτρων άρθρωσης δεν είναι οι ίδιες.

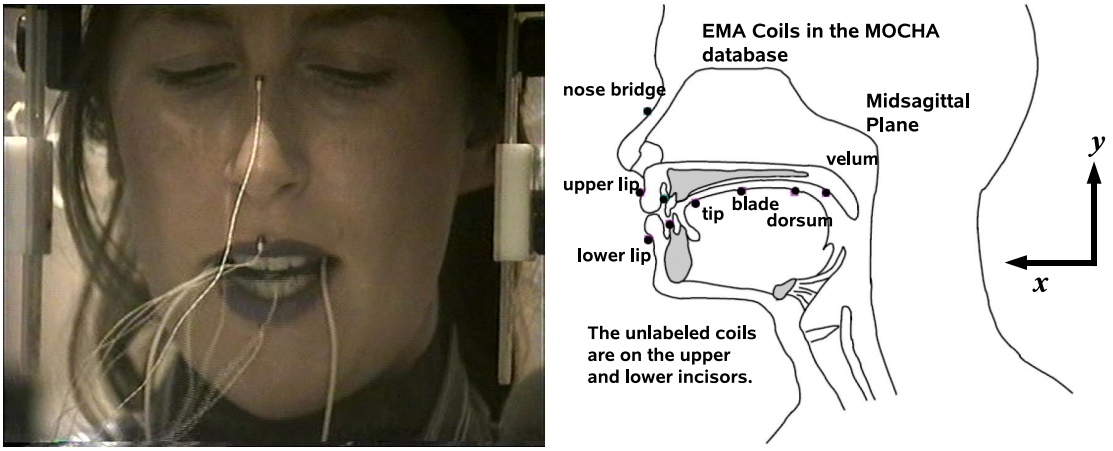
Ο μέσος συντελεστής συσχέτισης μετράει το βαθμό της ομοιότητας του πλάτους και του συγχρονισμού των τροχιών και ορίζεται ως

$$\rho_{x\hat{x}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{t=1}^{T}(\mathbf{x}_t[i] - E[\mathbf{x}[i]])(\hat{\mathbf{x}}_t[i] - E[\hat{\mathbf{x}}[i]])}{\sqrt{\sum_{t=1}^{T}(\mathbf{x}_t[i] - E[\mathbf{x}[i]])^2}\sqrt{\sum_{t=1}^{T}(\hat{\mathbf{x}}_t[i] - E[\hat{\mathbf{x}}[i]])^2}}. \tag{5.28}$$

Αυτά τα κριτήρια είναι εύκολο να εκτιμηθούν και παρέχουν ένα τρόπο ποσοτικοποίησης της ακρίβειας της αντιστροφής φωνής.

5.5.2 Περιγραφή των βάσεων δεδομένων

Τα πειράματα και η αξιολόγηση πραγματοποιήθηκαν στις βάσεις MOCHA και QSMT, οι οποίες περιέχουν δεδομένα άρθρωσης και παράλληλα καταγεγραμμένα οπτικοακουστικά δεδομένα. Η βάση MOCHA [175] είναι μια βάση πλούσια σε δεδομένα που είναι διαθέσιμη στο διαδίκτυο και χρησιμοποιείται ευρέως. Μεταξύ άλλων, περιλαμβάνει ακουστικές καταγραφές και ταυτόχρονες μετρήσεις των κινήσεων των αρθρωτων του φωνητικού συστήματος, δηλαδή της γλώσσας, των χειλιών και του σαγονιού, μέσω ηλεκτρομαγνητικών μεθόδων. Συλλέχθηκε κυρίως με σκοπό την έρευνα για αναγνώριση φωνής με αξιοποίηση γνώσης για το σύστημα παραγωγής και περιλαμβάνει καταγραφές ομιλητών για 460 Βρετανικές προτάσεις

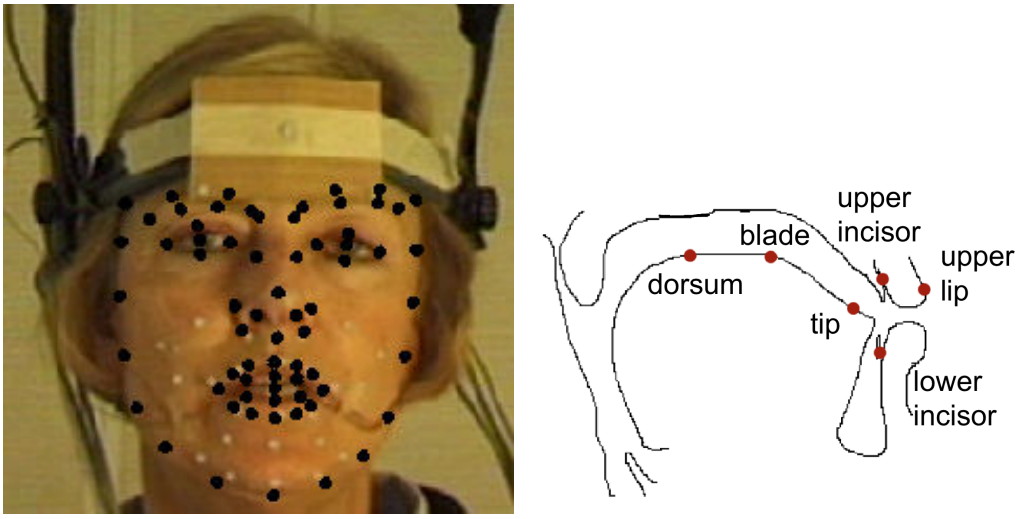


Σχήμα 5.3: *Αριστερά, εικόνα του προσώπου της ομιλήτριας fsew0 από τη βάση δεδομένων MOCHA. Δεξιά φαίνεται η τοποθέτηση των διαφόρων πηνίων οι κινήσεις των οποίων καταγράφονται ηλεκτρομαγνητικά με ειδικό σύστημα. Τα πηνία στη μύτη και στον πάνω κόφτη χρησιμοποιούνται για διόρθωση ενδεχόμενης κίνησης του κεφαλιού.*

από τη βάση TIMIT. Οι ηλεκτρομαγνητικές μετρήσεις είναι στα 500 Hz και έχουν υποδειγματοληπτηθεί στα 60 Hz ώστε να υπάρχει κοινή αναφορά με το QSMT σύνολο δεδομένων. Συνολικά, ιχνηλατούνται οι κινήσεις 7 ηλεκτρομαγνητικών πηνίων. Είναι κολλημένα στα άνω και κάτω χείλη, στον κάτω κόφτη, στην άκρη, μέση και πίσω μέρος της γλώσσας, όπως φαίνεται στο Σχήμα 5.3. Ένα πηνίο στη γλώσσα κι ένα στον πάνω κόφτη χρησιμοποιούνται για να διορθώσουν τις όποιες αναπόφευκτες κινήσεις του κεφαλιού. Για τα πειράματα που περιγράφονται διατέθηκε και το ανεπεξέργαστο βίντεο μιας ομιλήτριας κατά τη συλλόγή των δεδομένων το οποίο δεν είχε χρησιμοποιηθεί προηγουμένως. Είναι η πρώτη φορά που γίνεται προσπάθεια αξιοποίησης των οπτικών δεδομένων της MOCHA. Στην παρούσα φάση, είναι διαθέσιμες οι καταγραφές βίντεο μόνο για την ομιλήτρια 'fsew0', Σχήμα 5.3.

Το σύνολο δεδομένων QSMT διατέθηκε από την ομάδα φωνής του τμήματος Φωνής, Μουσικής και Ακοής του πολυτεχνείου ΚΤΗ της Στοκχόλμης και περιγράφεται με λεπτομέρεια στο [51]. Περιλαμβάνει ταυτόχρονες μετρήσεις του ακουστικού σήματος και των κινήσεων της γλώσσας και του προσώπου. Σε συντομία, εκτός από το ηχητικό σήμα το οποίο δειγματοληπτείται στα $16kHz$ και το βίντεο που είναι στα 30 πλαίσια το δευτερόλεπτο, κάθε πλαίσιο του συνόλου δεδομένων (με το ρυθμό των $60fps$) περιέχει τις 3Δ συντεταγμένες 25 ανακλαστήρων κολλημένων στο πρόσωπο του ομιλητή (διάνυσμα x 75 διαστάσεων, όπως μετράται από ένα σύστημα ιχνηλάτησης κίνησης), όπως επίσης και τις 2Δ συντεταγμένες των 6 πηνίων του συστήματος ηλεκτρομαγνητικής καταγραφής των αρθρωτών (ΗΚΑ) στο εγκάρσιο επίπεδο που τέμνει τη φωνητική οδό σε αριστερό και δεξί τμήμα, κατοπτρικό το ένα του άλλου. Τα πηνία είναι κολλημένα στη γλώσσα του ομιλητή, στα δόντια και στα χείλη του (δωδεκαδιάστατο διάνυσμα μετρήσεων). Συνολικά, υπάρχουν περίπου 60000 πολυμεσικά πλαίσια δεδομένων. Αυτά αντιστοιχούν σε μια εκφώνηση 135 συμμετρικών λέξεων του τύπου ΦΣΦ (Φωνήεν-Σύμφωνο-Φωνήεν), 37 ΣΦΣ (Σύμφωνο-Φωνήεν-Σύμφωνο) και 134 σύντομων καθημερινών σουηδικών προτάσεων. Εκτός από το βίντεο, όλα τα δεδομένα είναι χρονικά στοιχισμένα και συμπεριλαμβάνονται μεταγραφές σε επίπεδο φωνήματος. Μια ενδεικτική εικόνα από τη βάση μαζί με ένα σκίτσο όπου φαίνονται τα σημεία τοποθέτησης των ηλεκτρομαγνητικών πηνίων πάνω στους αρθρωτές φαίνονται στο Σχήμα 5.4. Τα τρία σημεία στο μέτωπο χρησιμοποιούνται για να αντισταθμίσουν την κίνηση του κεφαλιού και τα τα πηνία στο πάνω χείλος και στον πάνω κόφτη χρησιμοποιούνται για να ευθυγραμμίσουν οπτικά και αρθρωτικά δεδομένα.

Το γεγονός ότι η βάση δεδομένων QSMT περιλαμβάνει καταγεγραμμένες τις συντεταγμένες εξωτερικών σημαδευτών στο πρόσωπο την καθιστά ιδιαίτερα ενδιαφέρουσα αφού επιτρέπει την ευκολότερη αξιολόγηση της μοντελοποίησης με τα ενεργά μοντέλα εμφάνισης και



Σχήμα 5.4: *Βάση δεδομένων Qualisys-Movetrack. Αριστερά: Σημαντικά σημεία πάνω στο πρόσωπο του ομιλητή έχουν εντοπιστεί με τη χρήση ενεργών μοντέλων εμφάνισης και φαίνονται ως μαύρες τελείες. Οι λευκές τελείες είναι σημαδευτές κολλημένοι πάνω στο πρόσωπο και ιχνηλατούνται από ειδικό σύστημα κατά την καταγραφή των δεδομένων. Δεξιά: Οι τελείες αντιστοιχούν σε πηνία πάνω στη γλώσσα του ομιλητή (πίσω μέρος, κέντρο, άκρη από αριστερά προς τα δεξιά), στα δόντια και στα χείλη των οποίων οι κινήσεις καταγράφονται μέσω συστήματος ηλεκτρομαγνητικής καταγραφής αρθρωτών. Η βάση περιέχει επίσης και παράλληλες ηχητικές καταγραφές.*

της εξαγωγής οπτικών χαρακτηριστικών.

Ένα πρακτικό θέμα που έπρεπε να αντιμετωπιστεί τόσο για τη βάση QSMT όσο και για τη MOCHA ήταν η έλλειψη κατάλληλης δεικτοδότησης των δεδομένων βίντεο. Το πρόβλημα επιλύθηκε επιτυχώς με την αξιοποίηση των ήδη υπαρχουσών μεταγραφών για τα ακουστικά δεδομένα και την αυτόματη στοίχιση των μεταγεγραμμένων ηχητικών δεδομένων με τα ηχητικά δεδομένα του μη επεξεργασμένου βίντεο. Τα εξαχθέντα οπτικά χαρακτηριστικά υπερδειγματοληπτήθηκαν στα 60 Hz έτσι ώστε να ταιριάζουν με τα δεδομένα της άρθρωσης. Επιπρόσθετα, θέματα συγχρονισμού αντιμετωπίστηκαν μέσω μεγιστοποίησης της συσχέτισης του καθενός καναλιού χαρακτηριστικών με τα δεδομένα άρθρωσης. Η συσχέτιση μετρήθηκε μέσω ανάλυσης κανονικής συσχέτισης όπως προτείνεται στο [141]. Σημαντική καθολική έλλειψη συγχρονισμού, δηλαδή μεγαλύτερη των 120 ms, εντοπίστηκε και διορθώθηκε μόνο μεταξύ των δεδομένων άρθρωσης και του βίντεο για τη βάση QSMT.

5.5.3 Γενικό πειραματικό πλαίσιο

Τα πειράματα πραγματοποιήθηκαν ανεξάρτητα για τα δύο σύνολα δεδομένων. Ξεχωριστά μοντέλα εκπαιδεύτηκαν σε κάθε βάση και η αξιολόγηση πραγματοποιήθηκε παράλληλα. Για τη βάση MOCHA η επιλεγμένη αναπαράσταση της φωνητικής οδού περιλαμβάνει 14 παραμέτρους, δηλαδή τις συντεταγμένες των 7 πηνίων πάνω στους φωνητικούς αρθρωτές, όπως περιγράφτηκαν προηγουμένως. Για τη βάση QSMT χρησιμοποιούνται 8 παράμετροι που αντιστοιχούν στις συντεταγμένες των πηνίων πάνω στη γλώσσα και τον κάτω κόφτη. Για να αποφευχθεί πιθανή μεροληψία στα αποτελέσματά μας λόγω της περιορισμένης ποσότητας δεδομένων, ακολουθείται μια διαδικασία πολλαπλής επιβεβαίωσης με 10 πτυχές (10-fold cross validation). Τα δεδομένα σε κάθε περίπτωση χωρίζονται σε δέκα ξεχωριστά σύνολα, τα εννιά από τα οποία χρησιμοποιούνται για εκπαίδευση και τα υπόλοιπα για αξιολόγηση, με κύλιση.

Για αναφορά, πρώτα μελετάται η απόδοση ενός καθολικού γραμμικού μοντέλου όπως περιγράφεται στην Ενότητα 5.2, για αντιστροφή του ήχου, του βίντεο ή των οπτικοακουστικών παραμέτρων και την πρόβλεψη των τροχιών των παραμέτρων της άρθρωσης. Αυτό επίσης επιτρέπει μια αρχική εκτίμηση των πλεονεκτημάτων των γραμμικών μοντέλων που

υπολογίζονται με τη χρήση της ανάλυσης κανονικής συσχέτισης. Εφαρμόστηκε μια απλή μέθοδος για την επιλογή της τάξης του μοντέλου η οποία περιγράφεται στην Ενότητα 5.5.4. Τα προκύπτοντα αποτελέσματα επιβεβαιώνουν ότι τα μοντέλα μειωμένης τάξης που βασίζονται σε ανάλυση κανονικής συσχέτισης μπορούν πράγματι να έχουν καλύτερη απόδοση από τις παραλλαγές τους πλήρους τάξης ειδικά στην περίπτωση που τα δεδομένα εκπαίδευσης είναι περιορισμένα.

Στη συνέχεια ακολουθεί συστηματική αξιολόγηση της απόδοσης της αντιστροιφής φωνής με τη χρήση μεμονωμένα των ακουστικών ή των οπτικών χαρακτηριστικών της φωνής. Για κάθε ακουστικό φώνημα δημιουργείται ένα μοντέλο για τα σύνολα των Mel συντελεστών cepstrum ή γραμμικών φασματικών συχνοτήτων, ενώ για τις διάφορες παραλλαγές οπτικών χαρακτηριστικών βασισμένων στην ενεργή μοντελοποίηση της εμφάνισης του προσώπου χρησιμοποιούνται μοντέλα για οπτικά φωνήματα. Οι Mel συντελεστές cepstrum μαζί με το συντελεστή μηδενικής τάξης, φαίνεται ότι έχουν την καλύτερη απόδοση σε σύγκριση με τις εναλλακτικές ακουστικές αναπαραστάσεις. Από την πλευρά των οπτικών χαρακτηριστικών, ο συνδυασμός χαρακτηριστικών σχήματος και υφής ήταν που έδωσε τα καλύτερα αποτελέσματα.

Τέλος, διερευνάται η σύμμειξη των μεμονωμένων συνιστωσών ώστε να επιτευχθεί οπτικοακουστική αντιστροφή φωνής με την εφαρμογή των διαφόρων σχημάτων που περιγράφονται στην Ενότητα 5.3.2. Η εκ των υστέρων σύμμειξη βρίσκεται να δίνει τα καλύτερα αποτελέσματα, αποδίδοντας καλύτερα τόσο από τα απλά κρυφά Μαρκοβιανά μοντέλα όσο και από τα πολυκαναλικά, που επιτυγχάνουν πρώιμη και ενδιάμεση σύμμειξη αντίστοιχα. Η ποιοτική ερμηνεία των αποτελεσμάτων σχετικά με το πόσο καλά αντιστρέφονται ορισμένα φωνήματα και πόσο ακριβής είναι η πρόβλεψη της κάθε παραμέτρου άρθρωσης φαίνεται να οδηγεί σε διαισθητικά συμπεράσματα.

5.5.4 Πείραμα με καθολικό μοντέλο ελαττωμένης τάξης με χρήση ανάλυσης κανονικής συσχέτισης

Παρουσιάζουμε πρώτα ένα πείραμα το οποίο επιδεικνύει την προοπτική για βελτιωμένες επιδόσεις του γραμμικού μοντέλου ελαττωμένης τάξης σε σχέση με το συμβατικό πολυμεταβλητό μοντέλο. Στόχος του πειράματος είναι με χρήση ενός καθολικού γραμμικού μοντέλου να προβλεφθεί το 12-διάστατο διάνυσμα $\mathbf{x}$ περιγραφής της φωνητικής οδού για τη βάση QSMT (χρησιμοποιήσαμε όλες τις διαθέσιμες συντεταγμένες για το εν λόγω πείραμα) από το διάνυσμα 75 διαστάσεων $\mathbf{y}_{\mathrm{v}}$ που περιέχει τις τρισδιάστατες συντεταγμένες των σημαδευτών πάνω στο πρόσωπο. Έχουμε χωρίσει το σύνολο των δεδομένων μας σε δεδομένα εκπαίδευσης και δεδομένα αξιολόγησης. Υπολογίζουμε τα στατιστικά χαρακτηριστικά δεύτερης τάξης πάνω στο σύνολο εκπαίδευσης και υπολογίζουμε από αυτά είτε τον πίνακα του γραμμικού μοντέλου W ή τις μειωμένης τάξης εκδοχές του W_r, $r = 1, \ldots, 12$, από τις Εξισώσεις (5.4) και (5.9), αντίστοιχα. Σημειώνεται ότι για το συγκεκριμένο σύνολο δεδομένων $W = W_k$, με $k = 12$.

Το Σχήμα 5.5 δείχνει το λάθος πρόβλεψης του μοντέλου όταν υπολογίζεται η άρθρωση $\mathbf{x}$ από το πρόσωπο $\mathbf{y}_{\mathrm{v}}$ για μεταβλητή τάξη r· κάθε γράφημα στο Σχήμα αντιστοιχεί σε διαφορετικό μέγεθος του συνόλου εκπαίδευσης, $N = 1000, 5000, 50000$ δείγματα. Παρατηρούμε ότι για μικρά μεγέθη του συνόλου εκπαίδευσης, $N = 1000, 5000$, τα μοντέλα W_r ελαττωμένης τάξης με $r = 5$ ή 6 γενικεύουν καλύτερα από ότι το μοντέλα πλήρους τάξης με $W = W_{12}$. Ακόμα και στην περίπτωση του μεγάλου συνόλου εκπαίδευσης με $N = 50000$ δείγματα, παρά του ότι το μοντέλο πλήρους τάξης έχει τις καλύτερες επιδόσεις, τα μοντέλα ελαττωμένης τάξης με $r \geq 7$ έχουν εξίσου καλά αποτελέσματα. Αυτά τα αποτελέσματα είναι ιδιαίτερα σχετικά και ενθαρρυντικά για την ενσωμάτωση της προσέγγισης με την ανάλυση κανονικής συσχέτισης στο σύστημα που βασίζεται σε κρυφά Μαρκοβιανά μοντέλα και περιγράφεται στην Ενότητα 5.3.1. Το σύστημα αυτό ενσωματώνει μεμονωμένους προβλέπτες για κάθε κατάσταση του κάθε κρυφού Μαρκοβιανού μοντέλου και γι' αυτό τα δεδομένα εκπαίδευσης για

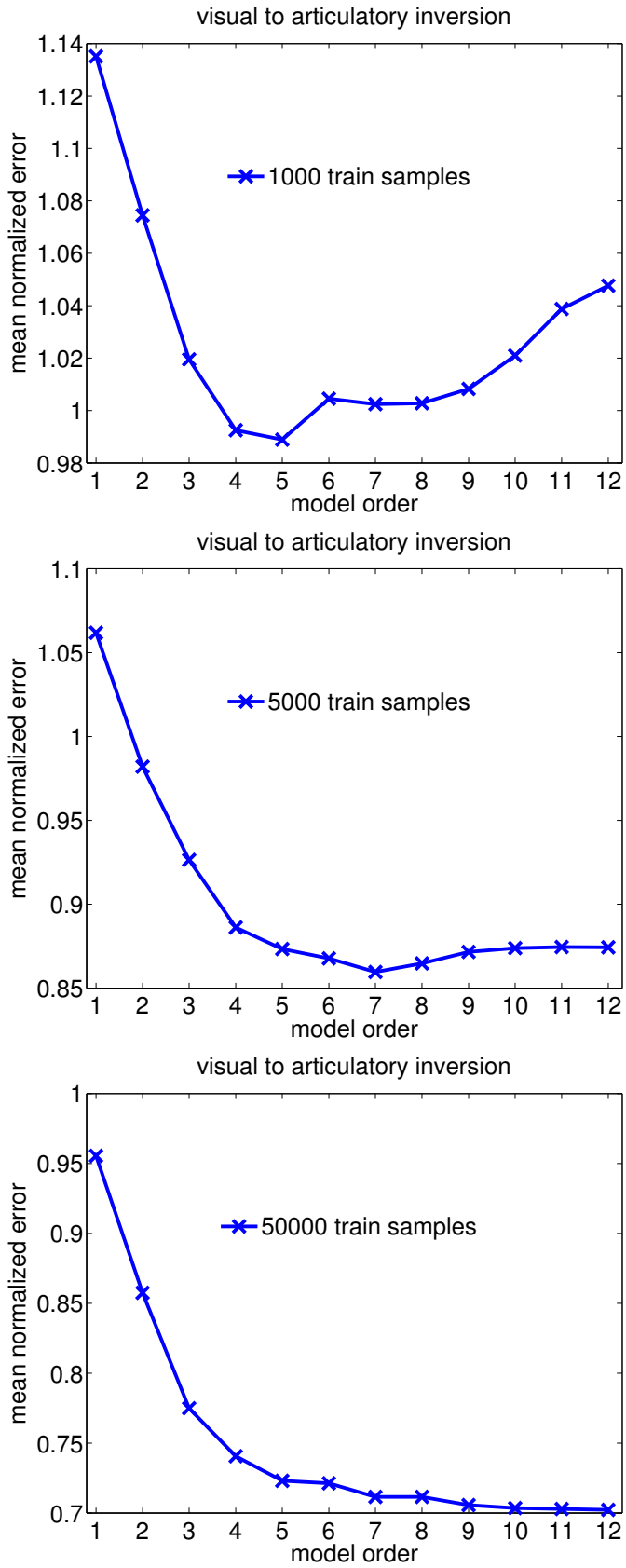
κάθε προβλέπτη μπορεί να είναι πολύ περιορισμένα.

Η αυτόματη επιλογή της τάξης του μοντέλου γίνεται μέσω μιας διαδικασίας πολλαπλής επιβεβαίωσης (cross validation). Για την εύρεση της βέλτιστης τάξης, χωρίζουμε τα δεδομένα εκπαίδευσης του μοντέλου σε δύο σύνολα και προσπαθούμε να προβλέψουμε το μικρότερο σύνολο χρησιμοποιώντας ένα μοντέλο εκπαιδευμένο στο άλλο σύνολο για διάφορες τάξεις. Αυτό επαναλαμβάνεται για κάθε πτυχή της διαδικασίας. Επιλέγεται η τάξη που δίνει το μικρότερο τετραγωνικό λάθος ως η πιο κατάλληλη και το τελικό μοντέλο εκπαιδεύεται με όλα τα δεδομένα.

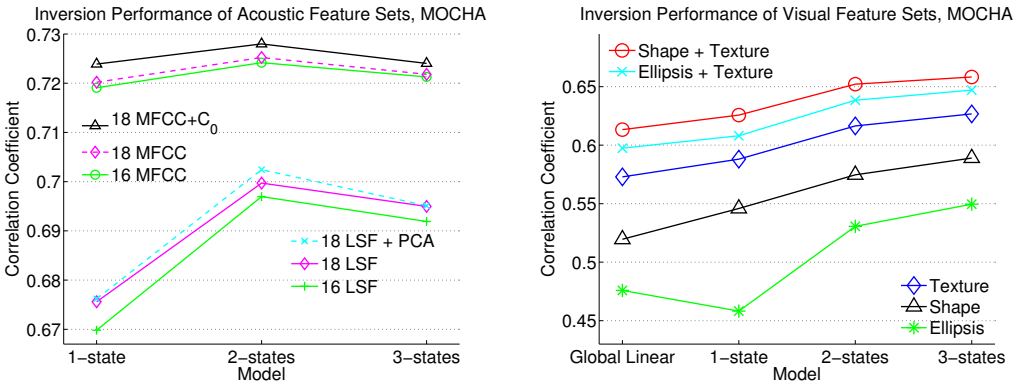
5.5.5 Αντιστροφή φωνής με χρήση κάθε συνιστώσας ξεχωριστά

Στη συνέχεια συζητώνται τα πειράματα ανάκτησης της άρθρωσης είτε από ακουστική είτε από οπτική μόνο πληροφορία. Όσον αφορά στο ακουστικό σήμα, πραγματοποιήθηκαν πειράματα με δύο βασικές ακουστικές αναπαραστάσεις, είτε Mel συντελεστές cepstrum όπως δίνονται στο [179] είτε γραμμικές φασματικές συχνότητες (LSFs) [152]. Και τα δύο σύνολα χαρακτηριστικών βρέθηκαν να αποδίδουν παρόμοια για αντιστροφή φωνής χρησιμοποιώντας νευρωνικά δίκτυα [131]. Στην προκειμένη περίπτωση, εξάγονται στα 60 Hz από πλαίσια του σήματος διάρκειας 30 ms στα οποία έχει εφαρμοστεί προέμφαση (συντελεστής 0.97) και παραθύρωση Hamming. Ο ρυθμός επελέγη ώστε να ταιριάζει με το ρυθμό με τον οποίο είναι δειγματοληπτημένα τα οπτικά δεδομένα και τα δεδομένα άρθρωσης της βάσης QSMT. Για τους Mel συντελεστές cepstrum χρησιμοποιήθηκαν 24 φίλτρα ενώ για τις γραμμικές φασματικές συχνότητες ο αριθμός των συντελεστών που χρησιμοποιούνταν κάθε φορά ήταν ο ίδιος με την τάξη της αντίστοιχης ανάλυσης γραμμικής πρόβλεψης που προηγείται. Εξετάστηκε η σημασία του συνολικού αριθμού εξαχθέντων χαρακτηριστικών (από 12 σε 22) καθώς επίσης και η σημασία της συμπερίληψης του πρώτου (γνωστού και ως μηδενικού) Mel συντελεστή cepstrum. Αφού στα πειράματα χρησιμοποιήθηκαν κρυφά Μαρκοβιανά μοντέλα για ακουστικά φωνήματα με διαγώνιους πίνακες συμμεταβλητότητας $\Sigma_{y,c}$ έγινε επίσης προσπάθεια να εκτιμηθεί η αξία της εφαρμογής ανάλυσης σε πρωτεύουσες συνιστώσες των γραμμικών φασματικών συχνοτήτων. Οι τελευταίες δεν αναμένονται γενικά να έχουν σχεδόν διαγώνιο πίνακα συμμεταβλητότητας εξ ορισμού, όπως συμβαίνει με τους Mel συντελεστές cepstrum. Στο Σχήμα 5.6(α') δίνονται ενδεικτικά αποτελέσματα για τη βάση MOCHA. Και στις δύο βάσεις τα συμπεράσματα είναι παρόμοια· οι Mel συντελεστές cepstrum δίνουν καλύτερα αποτελέσματα από τις γραμμικές φασματικές συχνότητες ακόμα και στην περίπτωση που η απόδοση των τελευταίων βελτιώνεται κάπως με την εφαρμογή της ανάλυσης σε πρωτεύουσες συνιστώσες. Επιπλέον, η συμπερίληψη του μηδενικού Mel συντελεστή cepstrum είναι επωφελής ενώ βρίσκεται ότι αν κρατηθούν 18 από τους συντελεστές επιτυγχάνεται αρκετά ικανοποιητική απόδοση.

Για τα ακουστικά μοντέλα, στο περιγραφόμενο πλαίσιο, τα καλύτερα αποτελέσματα επιτυγχάνονται για δικατάστατα κρυφά Μαρκοβιανά μοντέλα της μορφής από αριστερά προς τα δεξιά για κάθε ακουστικό φώνημα. Μοντέλα για διφωνήματα μπορεί να είχαν ακόμα καλύτερα αποτελέσματα αν ήταν διαθέσιμα αρκετά δεδομένα για εκπαίδευση [63]. Στη MOCHA εκπαιδεύονται 46 μοντέλα συνολικά, 44 για τα ακουστικά φωνήματα και 2 για την αναπνοή και τη σιωπή, ενώ στη βάση QSMT εκπαιδεύονται 52 μοντέλα για τα 51 ακουστικά φωνήματα και τη σιωπή που εμφανίζονται στις φωνηματικές μεταγραφές των δεδομένων. Για την οπτική αντιστροφή φωνής από την άλλη είναι δυνατή η επίτευξη καλύτερων αποτελεσμάτων με τη χρήση μοντέλων βασισμένων σε οπτικά φωνήματα. Παρά του ότι κατά καιρούς έχουν προταθεί διάφορα σύνολα οπτικών φωνημάτων για το σύνολο των ακουστικών φωνημάτων της βάσης TIMIT (MOCHA), θεωρήθηκε ότι θα ήταν καταλληλότερος ο προσδιορισμός των αντίστοιχων οπτικών φωνημάτων από τα διαθέσιμα δεδομένα [61]. Ξεκινώντας από τάξεις που περιείχαν μόνο ένα ακουστικό φώνημα (όπως προσδιορίζονται από τις φωνηματικές μεταγραφές), ακουλουθήθηκε μια προσέγγιση ομαδοποίησης από κάτω προς τα πάνω ώστε τελικά να οριστούν 14 οπτικά φωνήματα στη MOCHA και 34 στην QSMT. Οι τάξεις των οπτι-



Σχήμα 5.5: *Ανάκτηση της άρθρωσης από οπτική πληροφορία του προσώπου του ομιλητή. Λάθος γενίκευσης για το καθολικό γραμμικό μοντέλο για διάφορες τάξεις του μοντέλου και διαφορετικό πλήθος δεδομένων εκπαίδευσης. Τα μοντέλα περιορισμένης τάξης που έχουν εκπαιδευτεί με ανάλυση κανονικής συσχέτισης μπορούν να αντιμετωπίσουν αποτελεσματικά περιπτώσεις περιορισμένων δεδομένων.*



(α′) (β′)

Σχήμα 5.6: *Βάση MOCHA: Απόδοση της αντιστροφής από τις μεμονωμένες συνιστώσες της φωνής στη MOCHA. Εναλλακτικές ακουστικές / οπτικές μόνο αναπαραστάσεις συγκρίνονται με βάση το μέσο συντελεστή συσχέτισης των αποτελεσμάτων της αντιστροφής με τις μετρήσεις. Αριστερά: Αντιστροφή από τον ήχο μόνο με χρήση των συντελεστών cepstrum ή των γραμμικών φασματικών συχνοτήτων. Δεξιά: Αντιστροφή φωνής από οπτική πληροφορία μόνο χρησιμοποιώντας εναλλακτικά σύνολα χαρακτηριστικών βασισμένων στην ενεργή μοντελοποίηση εμφάνισης του προσώπου.*

κών φωνημάτων για τη βάση MOCHA δίνονται στον Πίνακα 5.1. Η αυτόματη ομαδοποίηση φαίνεται να έχει οδηγήσει σε διαισθητικά αποτελέσματα στις περισσότερες περιπτώσεις.

Για τη δημιουργία των γραμμικών απεικονίσεων μεταξύ των παρατηρήσεων και των παραμέτρων άρθρωσης σε κάθε κατάσταση εφαρμοσαμε ανάλυση κανονικής συσχέτισης όπως περιγράφτηκε στην Ενότητα 5.2.2 και με περισσότερη λεπτομέρεια στην Ενότητα 5.5.4. Σε αυτή τη διαδικασία, μη επαρκή διαθέσιμα δεδομένα μπορεί να οδηγήσουν σε ακατάλληλους συντελεστές κανονικής συσχέτισης, πιο συγκεκριμένα ο πρώτος συντελεστής συσχέτισης μπορεί να ισούται με τη μονάδα [171], ή σε εκφυλισμένες εκτιμήσεις της συμμεταβλητότητας του λάθους του μοντέλου. Για να αντιμετωπιστούν τέτοια προβλήματα, ομαδοποιείται το προβληματικό μοντέλο με το κοντινότερό του (με βάση την Ευκλίδεια απόσταση) και γίνεται επανεκτίμηση.

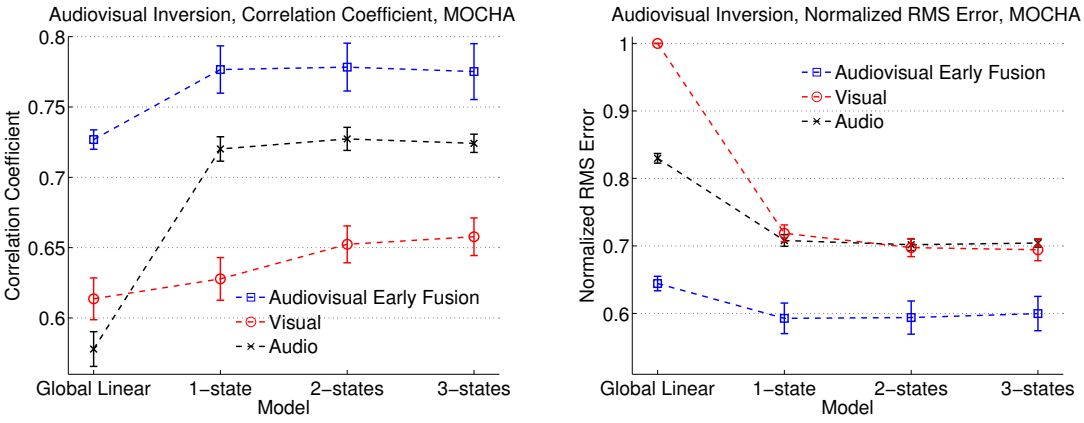
Σε αυτό το πλαίσιο, για να διερευνηθεί η απόδοση διαφορετικών οπτικών αναπαραστάσεων, πραγματοποιήθηκαν πειράματα ώστε να αναδειχτεί ποια θα πρέπει να είναι η φύση των οπτικών χαρακτηριστικών του προσώπου για αντιστροφή. Ο αριθμός των χαρακτηριστικών σχήματος από την ενεργή μοντελοποίηση εμφάνισης, 12 για τη MOCHA, 9 για τη QSMT και χαρακτηριστικών υφής, 27 για τη MOCHA και 24 για τη QSMT, αντιστοιχεί στο 95% της παρατηρούμενης μεταβλητότητας στα δεδομένα του προσώπου σε κάθε βάση. Το σχήμα εναλλακτικά μπορεί να περιγραφτεί με συμπαγή τρόπο με το σύνολο των γεωμετρικών χαρακτηριστικών που βασίζονται στην έλλειψη που περιγράφεται στην Ενότητα 5.4. Είναι ενδιαφέρον ότι και αυτή η αναπαράσταση είναι αποτελεσματική, όχι όμως τόσο όσο το αρχικό διάνυσμα χαρακτηριστικών σχήματος του ενεργού μοντέλου εμφάνισης. Τα αποτελέσματα της αντιστροφής για διάφορα σενάρια συνοψίζονται στο Σχήμα 5.6(β′) για τη MOCHA. Συνολικά συμπεραίνουμε ότι τα καλύτερα αποτελέσματα επιτυγχάνονται με τον συνδυασμό των χαρακτηριστικών σχήματος και υφής.

5.5.6 Πειράματα οπτικοακουστικής αντιστροφής της φωνής

Για οπτικακουστική αντιστροφή φωνής, πρώτα πραγματοποιούνται πειράματα με πρώιμη σύμμειξη των ακουστικών και οπτικών διανυσμάτων χαρακτηριστικών. Τα αντίστοιχα διανύσματα συνενώνονται κάθε χρονική στιγμή ώστε να σχηματίσουν ένα σύνθετο οπτικοακουστικό διάνυσμα χαρακτηριστικών και απλά κρυφά Μαρκοβιανά μοντέλα, ένα για κάθε ακουστικό φώνημα, εκπαιδεύονται ώστε να γίνει δυνατός ο προσδιορισμός της κατά τμήματα γραμμικής προσέγγισης. Τα αποτελέσματα συνοψίζονται στα Σχήματα 5.7 και 5.8 για τη MOCHA και τη

P'inakas 5.1: *Τάξεις οπτικών φωνημάτων όπως προσδιορίζονται στη βάση MOCHA ακολουθώντας μια προσέγγιση αυτόματης ομαδοποίησης από κάτω προς τα πάνω. Τα φωνητικά σύμβολα και τα αντίστοιχα αποτελέσματα είναι όπως στις φωνηματικές μεταγραφές της βάσης MOCHA.*

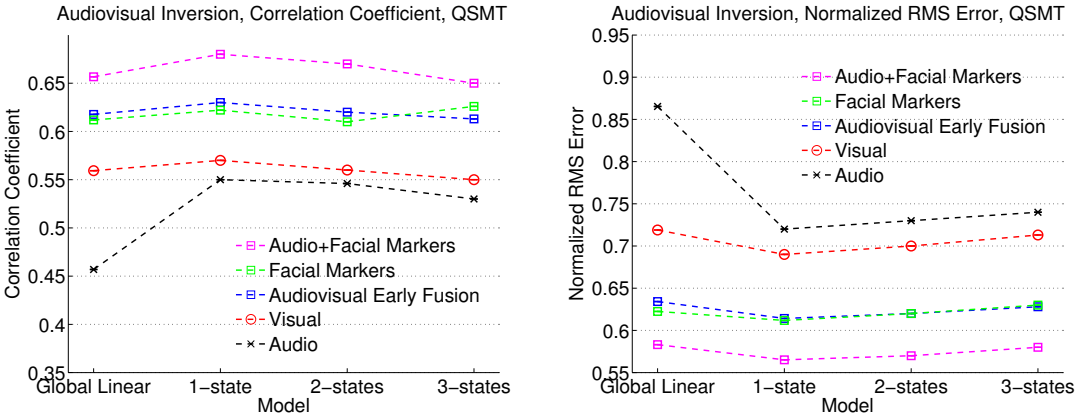
Viseme	Phonemes	Word Examples from MOCHA
v_1	zh, sh, jh, ch	pleasure, educa**ti**on, **g**eological, porch
v_2	uh, uu, ow	lunch, too, found
v_3	i@, ii, iy	year, peel, bare**ly**,
v_4	a, ai, e, ei, eir, h	N**a**ncy, pie, **e**lderly, day, where, her
v_5	b, m, p	obtain, more, public
v_6	dh, th	the, wealth
v_7	f, v	for, live
v_8	k, breath, g, i, n, ng	cream, good, Acropol**i**s, hand, mango
v_9	d, s, sil, t, z	wild, seldom, to, is
v_{10}	@, l, y	w**a**s, lily, why
v_{11}	@@, aa	were, arm
v_{12}	o, oi, ou, u	often, enj**oy**, do, would
v_{13}	oo, w	all, why
v_{14}	r	rabbits



Σχήμα 5.7: *Βάση MOCHA: Συντελεστής συσχέτισης και κανονικοποιημένο RMS λάθος μεταξύ των αρχικών και των προβλεπόμενων τροχιών των παραμέτρων άρθρωσης για αυξανόμενο αριθμό καταστάσεων των κρυφών Μαρκοβιανών μοντέλων χρησιμοποιώντας μόνο οπτική πληροφορία, μόνο ακουστική πληροφορία (μέσω Mel συντελεστών cepstrum), και οπτικοακουστική πληροφορία. Η επίδοση του καθολικού γραμμικού μοντέλου δίνεται επίσης για σύγκριση.*

QSMT αντίστοιχα για ένα καθολικό γραμμικό μοντέλο και αυξανόμενο αριθμό καταστάσεων σε κάθε Μαρκοβιανό μοντέλο. Τα διαστήματα εμπιστοσύνης που σημειώνονται αντιστοιχούν στην τυπική απόκλιση της αντίστοιχης εκτίμησης του συντελεστή συσχέτισης ή του μέσου κανονικοποιημένου λάθους, όπως δίνονται μέσω της διαδικασίας πολλαπλής επιβεβαίωσης. Για σύγκριση, συμπεριλαμβάνονται και τα αποτελέσματα της αντιστροφής με χρήση της κάθε συνιστώσας της φωνής ξεχωριστά. Και στα δύο σύνολα δεδομένων ο συνδυασμός των δύο συνιστωσών είναι σίγουρα επωφελής. Στη βάση QSMT όπου είναι διαθέσιμα οπτικά δεδομένα του προσώπου για αναφορά, η οπτικοακουστική αντιστροφή φωνής είναι σχεδόν τόσο καλή όσο η αντιστροφή που επιτυγχάνεται με συνδυασμό του ήχου και των πραγματικών συντεταγμένων των σημαδευτών πάνω στο πρόσωπο. Επιπλέον, η μέτρηση των χαρακτηριστικών με βάση τη ενεργή μοντελοποίηση εμφάνισης είναι πολύ περισσότερο πρακτική αφού δεν απαιτεί κάποιον ειδικό ή πολύπλοκο εξοπλισμό αλλά μόνο βίντεο της μπροστινής όψης του προσώπου του ομιλητή.

Τα αποτελέσματα βελτιώνονται όταν εφαρμόζεται ενδιάμεση σύμμειξη, δηλαδή όταν χρησιμοποιούνται πολυκαναλικά αντί για απλά κρυφά Μαρκοβιανά μοντέλα. Στο Σχήμα 5.9

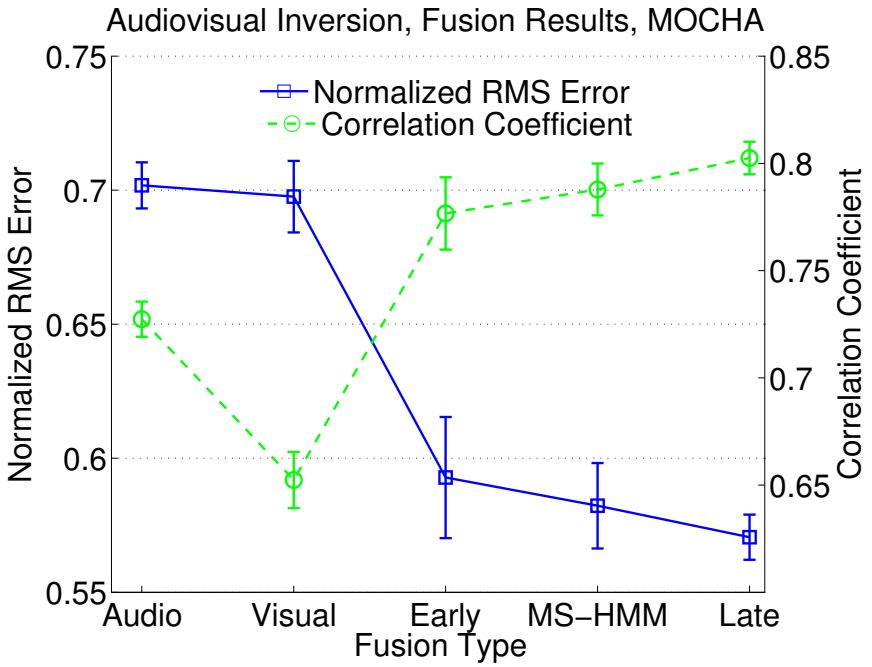


Σχήμα 5.8: *Βάση QSMT: Συντελεστής συσχέτισης και κανονικοποιημένο RMS λάθος μεταξύ των αρχικών και των προβλεπόμενων τροχιών των παραμέτρων άρθρωσης για αυξανόμενο αριθμό καταστάσεων των κρυφών Μαρκοβιανών μοντέλων χρησιμοποιώντας μόνο οπτική πληροφορία (μέσω ενεργών μοντέλων εμφάνισης ή συντεταγμένων των σημαδευτών στο πρόσωπο), μόνο ακουστική πληροφορία (μέσω Mel συντελεστών cepstrum), και οπτικοακουστική πληροφορία. Η επίδοση του καθολικού γραμμικού μοντέλου δίνεται επίσης για σύγκριση.*

φαίνονται τα καλύτερα αποτελέσματα με τη χρήση ενδιάμεσης σύμμειξης για τη MOCHA, που επιτυγχάνονται όταν εφαρμόζονται δικατάστατα πολυκαναλικά μοντέλα. Τα βάρη για τα κανάλια χρησιμοποιούνται για τον προσδιορισμό της βέλτιστης ακολουθίας καταστάσεων μέσω του αλγορίθμου Viterbi, όπως εξηγείται στην Ενότητα 5.3.2. Η διαδικασία αυτή είναι μια διαδικασία στοίχισης και όχι αναγνώρισης, αφού θεωρείται ότι είναι γνωστό το φωνηματικό περιεχόμενο κάθε εκφώνησης. Βρήκαμε ότι η επίδοση είναι βέλτιστη στην περίπτωση που η στοίχιση γίνεται μόνο με τη χρήση των ακουστικών χαρακτηριστικών, δηλαδή θέτοντας μηδενικό βάρος στο κανάλι της οπτικής πληροφορίας. Αυτή η παρατήρηση είναι σε συμφωνία με παρόμοια εμπειρία σε οπτικοακουστική αναγνώριση φωνής για την περίπτωση απουσίας θορύβου στο ηχητικό κανάλι. [44]. Στο προτεινόμενο σχήμα, φαίνεται ότι το ακουστικό κανάλι είναι αξιόπιστο για τη στοίχιση αλλά αφού προσδιοριστεί η βέλτιστη ακολουθία καταστάσεων, η συνεισφορά της οπτικής συνιστώσας είναι επίσης πολύ σημαντική σε κάθε περίπτωση.

Οι επιδόσεις γίνονται ακόμα καλύτερες αν θεωρηθεί ότι τα δύο κανάλια είναι ασύγχρονα και μοντελοποιηθούν ανεξάρτητα, δηλαδή χρησιμοποιώντας δικατάστατα Μαρκοβιανά μοντέλα για το ηχητικό κανάλι και τρικαταστάτα μοντέλα (ένα για κάθε οπτικό φώνημα) για το οπτικό κανάλι. Οι τελικές εκτιμήσεις για τις τροχιές των παραμέτρων άρθρωσης βρίσκονται τότε μετά από εκ των υστέρων σύμμειξη με βάση την Εξίσωση (5.22), όπως περιγράφεται στην Ενότητα 5.3.2. Αυτό είναι ουσιαστικά το καλύτερο δυνατό σενάριο όπως φαίνεται στο Σχήμα 5.9. Η αποτελεσματικότητα της ασύγχρονης θεώρησης θα πρέπει να αποδοθεί στην ευελιξία που παρέχει στην επιλογή της βέλτιστης τοπολογίας αλλά και φύσης (οπτικά/ακουστικά φωνήματα) του κρυφού Μαρκοβιανού μοντέλου για κάθε συνιστώσα της φωνής. Παρόμοια αποτελέσματα, μόνο με μεγαλύτερη αβεβαιότητα, λόγω του μικρού μεγέθους του συνόλου των δεδομένων, μπορούν να εξαχθούν και για τη βάση QSMT επίσης.

Για τη βάση MOCHA στο Σχήμα 5.10 παρουσιάζονται οι αρθρωτές για τους οποίους η αντιστροφή είναι περισσότερο επιτυχής. Φαίνονται τόσο το λάθος RMS, ώστε να δοθεί και αίσθηση της αποτελεσματικότητας σε φυσικές μονάδες (mm), καθώς και η κανονικοποιημένη έκδοσή του. Όπως ήταν ίσως αναμενόμενο, η πρόβλεψη των κινήσεων των χειλιών βελτιώνεται σημαντικά, σε σύγκριση με την περίπτωση όπου χρησιμοποιείται μόνο το ακουστικό σήμα. Γενικά, η σχετική βελτίωση είναι μεγαλύτερη για τις y-συντεταγμένες, το οποίο μπορεί να θεωρηθεί λογικό αφού η διδιάστατη μπροστινή εικόνα του προσώπου είναι αρκετά δύσκολο να δώσει πληροφορία για τις κινήσεις στη x-διάσταση που είναι ορατές μόνο έμμεσα. Αυτή η παρατήρηση θα μπορούσε για παράδειγμα να εξηγήσει γιατί η κίνηση του κάτω κόφτη δεν ανακτάται με τόση ακρίβεια στη x-διάσταση. Είναι ενδιαφέρον το ότι υπάρχουν και



Σχήμα 5.9: *Βάση MOCHA: Δίνονται τα καλύτερα αποτελέσματα για κάθε σενάριο αντιστροφής, δηλαδή δικατάστατα ακουστικά κρυφά Μαρκοβιανά μοντέλα, τρικατάστατα οπτικά κρυφά Μαρκοβιανά μοντέλα, απλό οπτικοακουστικό κρυφό Μαρκοβιανό μοντέλο με μία κατάσταση, πολυκαναλικό οπτικοακουστικό μοντέλο με μιά κατάσταση και το σενάριο με εκ των υστέρων σύμμειξη δικατάστατων ακουστικών και τρικαταστάτων οπτικών μοντέλων*

βελτιώσεις στην πρόβλεψη των κινήσεων της γλώσσας, εκτός των άλλων.

Αυτές οι διαπιστώσεις θα μπορούσαν πιθανόν να δικαιολογήσουν και τις βελτιώσεις στην αντιστροφή όταν αυτές θεωρούνται από την πλευρά των φωνημάτων όπως στο Σχήμα 5.11. Φαίνεται το λάθος RMS για τα 20 καλύτερα οπτικακουστικά αντριστρέψιμα φωνήματα. Δίνεται επίσης η σχετική βελτίωση.

Στο Σχήμα 5.12 δίνεται ένα ποιοτικό αποτέλεσμα· οι προβλεφθείσες και οι μετρηθείσες τροχιές για το πάνω χείλος και την άκρη της γλώσσας (y συντεταγμένες) για μία εκφώνηση στη βάση MOCHA. Η οπτικοακουστική αντιστροφή εμφανίζεται ως περισσότερο ακριβής.

5.6 Μοντελοποιώντας τη δυναμική των παραμέτρων άρθρωσης

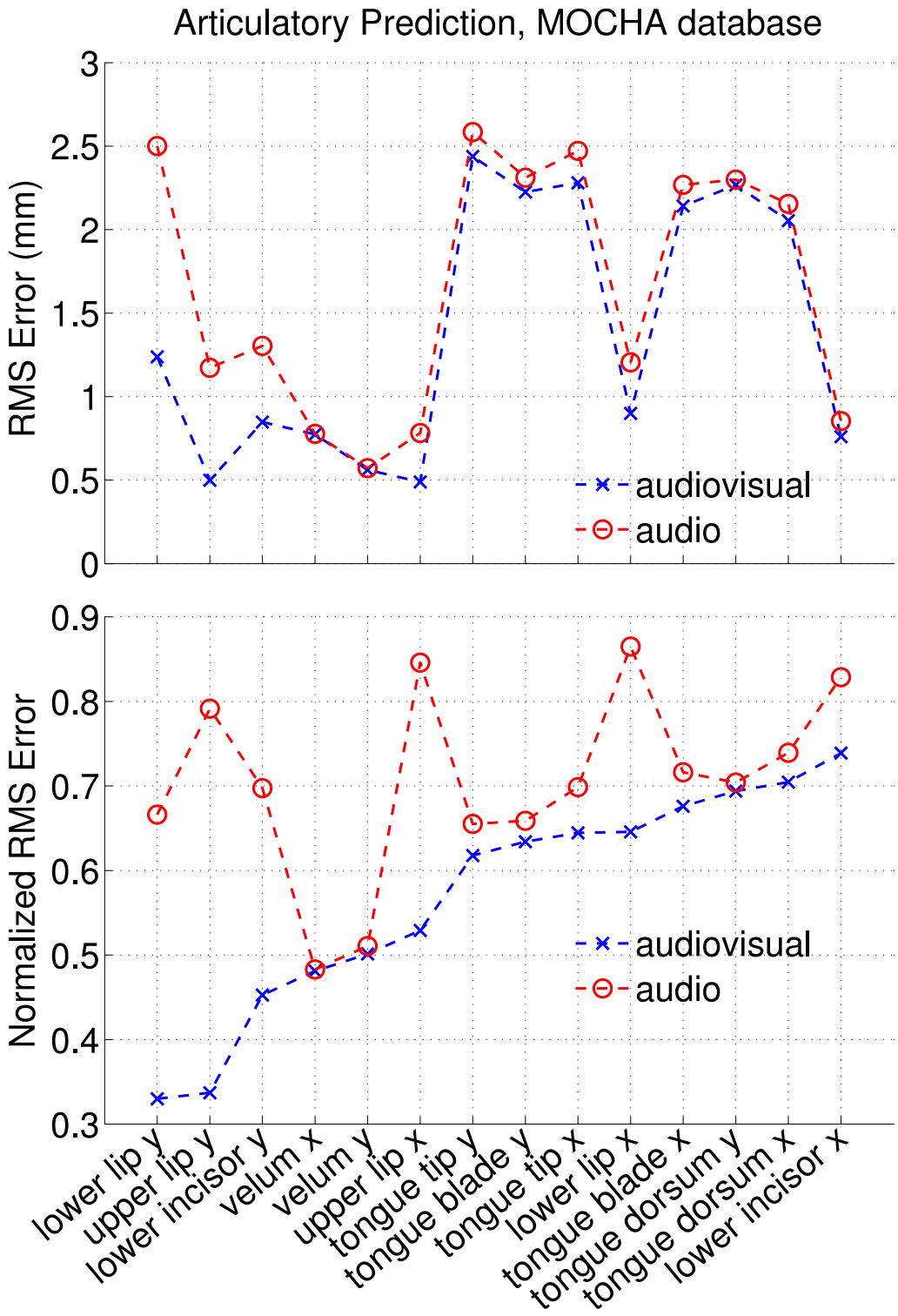
Επεκτείνοντας την παρουσίαση της Ενότητας 5.2, μπορούμε να θεωρήσουμε τη λύση του προβλήματος αντιστροφής σε μια χρονική στιγμή t ως την κατάσταση της φωνητικής οδού $\mathbf{x}_t$ που μεγιστοποιεί την εκ των υστέρων πιθανότητα των παραμέτρων άρθρωσης δεδομένης της διαθέσιμης οπτικοακουστικής πληροφορίας μέχρι τη χρονική στιγμή t, δηλαδή την ακολουθία $\mathbf{Y}_t = \{\mathbf{y}_1, \ldots, \mathbf{y}_t\}$:

$$p(\mathbf{x}_t|\mathbf{Y}_t) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{Y}_{t-1})}{p(\mathbf{y}_t|\mathbf{Y}_{t-1})}. \tag{5.29}$$

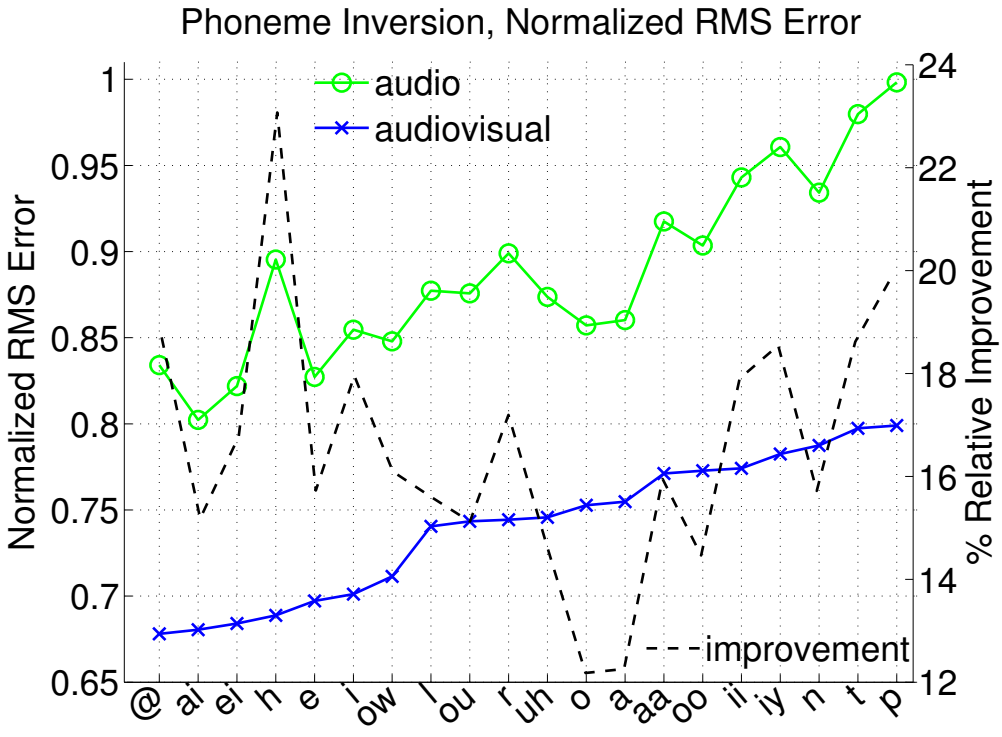
Έχοθμε υποθέσει ότι η παρατήρηση $\mathbf{y}_t$ τη στιγμή t εξαρτάται μόνο από την τρέχουσα κατάσταση $\mathbf{x}_t$. Μπορούμε επιπλέον να έχουμε:

$$p(\mathbf{x}_t|\mathbf{Y}_{t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{Y}_{t-1})d\mathbf{x}_{t-1} \tag{5.30}$$

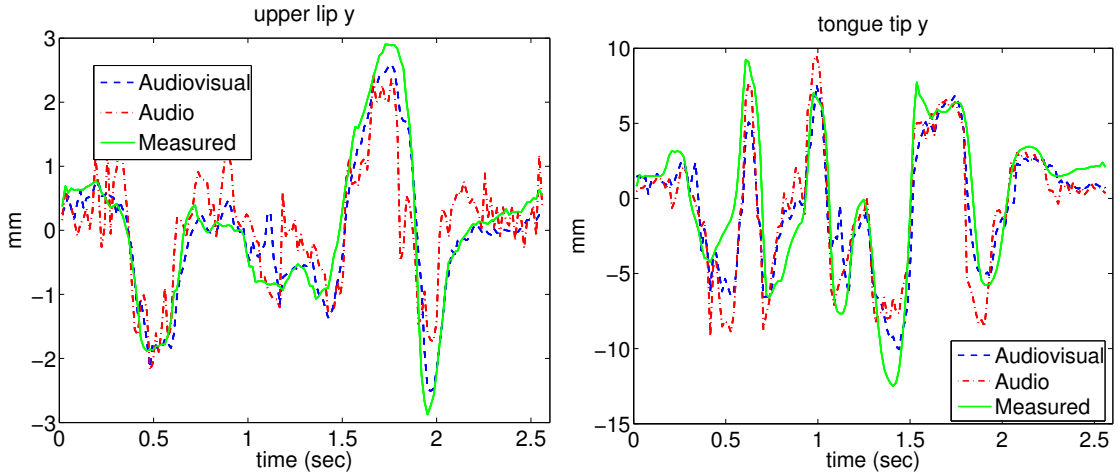
ολοκληρώνοντας ως προς όλες τις προηγούμενες πιθανές καταστάσεις $\mathbf{x}_{t-1}$. Το διάνυσμα παραμέτρων $\mathbf{x}_t$ (n στοιχεία) παρέχει μια κατάλληλη αναπαράσταση της φωνητικής οδού. Αυτή η αναπαράσταση θα μπορούσε να είναι είτε ευθεία, περιλαμβάνοντας τις συντεταγμένες



Σχήμα 5.10: *Βάση MOCHA: Το RMS λάθος πρόβλεψης και η κανονικοποιημένη εκδοχή του για τους αρθρωτές στη φωνητική οδό, χρησιμοποιώντας ακουστική μόνο ή οπτικοακουστική πληροφορία. Τα αποτελέσματα αντιστοιχούν στο καλύτερο σενάριο και για τις δύο περιπτώσεις. Χρησιμοποιούνται δικατάστατα κρυφά Μαρκοβιανά μοντέλα για την ακουστική πληροφορία και συνδυάζονται με τρικατάστατα οπτικά κρυφά Μαρκοβιανά μοντέλα με εκ των υστέρων σύμμειξη.*



Σχήμα 5.11: *Βάση MOCHA: Μέσο κανονικοποιημένο λάθος RMS για τα φωνήματα που αντιστράφηκαν με ελάχιστο λάθος. Παρουσιάζονται τα αποτελέσματα τόσο για την περίπτωση που χρησιμοποιήθηκε μόνο ακουστική πληροφορία όσο και για όταν χρησιμοποιείται και οπτική πληροφορία. Χρησιμοποιούνται δικατάστατα μοντέλα για τον ήχο και συνδυάζονται με τρικατάστατα οπτικά μοντέλα με εκ των υστέρων σύμμειξη.*



Σχήμα 5.12: *y-συντεταγμένες του πάνω χείλους και της άκρης της γλώσσας όπως μετρήθηκαν με το σύστημα ηλεκτρομαγνητικής καταγραφής και όπως προβλέφθηκαν από ακουστικές ή οπτικακουστικές παρατηρήσεις για μια ενδεικτική εκφώνηση της βάσης MOCHA.*

των διάφορων αρθρωτών του συστήματος είτε έμμεση, βασισμένη σε ένα μοντέλο άρθρωσης για παράδειγμα. Το οπτικοακουστικό διάνυσμα παραμέτρων $\mathbf{y}_t$ (m στοιχεία), που περιέχει ακουστικές και οπτικές παραμέτρους $\mathbf{y}_t^{\mathrm{a}}$ και $\mathbf{y}_t^{\mathrm{v}}$, πρέπει ιδανικά να περιέχει όλη την πληροφορία που σχετίζεται με τη φωνητική οδό και μπορεί να εξαχθεί από το ακουστικό σήμα από τη μία και το πρόσωπο του ομιλητή από την άλλη. Διάφορες πιθανές αναπαραστάσεις μπορούν να χρησιμοποιηθούν, βλ. Ενότητα 5.5.5. Αν υποθέσουμε ότι:

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + \mathbf{w}_t \tag{5.31}$$

$$\mathbf{y}_t = C\mathbf{x}_t + \mathbf{v}_t \tag{5.32}$$

όπου $\mathbf{w} \sim N(0, Q)$ και $\mathbf{v} \sim N(0, R)$ ανεξάρτητες διαδικασίες θορύβου και επιπλέον $\mathbf{x}_0 \sim N(\boldsymbol{\mu}_0, V_0)$, τότε η λύση της μέγιστης εκ των υστέρων πιθανότητας σε αυτό το πρόβλημα δίνεται από το φίλτρο Kalman, [5].

Όπως και στην Ενότητα 5.3.1 στην περίπτωση συνεχούς λόγου, περιμένουμε η γραμμική προσέγγιση της Εξίσωσης (5.32) να ισχύει μόνο ως εξίσωση παρατήρησης για περιορισμένα χρονικά διαστήματα, αντίστοιχα ενός φωνήματος ή και μέρους φωνήματος, είτε στη μετάβαση είτε στη σταθερή κατάσταση. Το ίδιο ισχύει για την εξίσωση κατάστασης που ελέγχει τη δυναμική των παραμέτρων άρθρωσης. Είναι λοιπόν φυσικό η χρήση διαφορετικών ανά φώνημα (ή ανά διάστημα μεταξύ φωνημάτων όπως στο [45]) εξισώσεων κατάστασης και παρατήρησης να είναι καταλληλότερη από ένα καθολικό γραμμικό δυναμικό σύστημα. Το προτεινόμενο διακοπτόμενο γραμμικό δυναμικό σύστημα είναι:

$$\mathbf{x}_t = A_{1,c}\mathbf{x}_{t-1} + A_{2,c}\mathbf{x}_{t-2} + B_c\mathbf{u}_c + \mathbf{w}_t \tag{5.33}$$

$$\mathbf{y}_t = C_c\mathbf{x}_t + \mathbf{v}_t \tag{5.34}$$

Ουσιαστικά, υπάρχει ένα ξεχωριστό γραμμικό δυναμικό σύστημα που αντιστοιχεί σε κάθε τάξη c. Για κάθε τέτοια τάξη, η ακολουθία των διαφορετικών καταστάσεων του συστήματος μοντελοποιείται ως δεύτερης τάξης autoregressive διαδικασία. Η επιλογή αυτή βασίζεται σε επιχειρήματα φυσιολογίας [45]. Επιπλέον, θεωρείται $B_c = I - (A_{1,c} + A_{2,c})$ έτσι ώστε η μέση τιμή της κατάστασης άρθρωσης σε κάθε κατάσταση να είναι u_c. Οι συμμεταβλητότητες των θορύβων Q_c, R_c εξαρτώνται επίσης από την κατάσταση c.

Η εκπαίδευση και η πρόβλεψη για ένα διακοπτόμενο γραμμικό δυναμικό μοντέλο μπορούν να επιτευχθούν μέσω μεταβολικών προσεγγίσεων όπως περιγράφεται στο [59]. Για απλότητα στην τρέχουσα εργασία, δεχόμαστε ότι ο διαχωρισμός των οπτικοακουστικών δεδομένων και των δεδομένων άρθρωσης σε ξεχωριστές τάξεις μπορεί να προσδιοριστεί ανεξάρτητα και σχετίζεται με φωνηματικές ιδιότητες. Για την επίτευξη αυτού του διαχωρισμού χρησιμοποιούνται οπτικοακουστικά πολυκαναλικά κρυφά Μαρκοβιανά μοντέλα για τα ακουστικά φωνήματα, όπως περιγράφεται στην Ενότητα 5.3.2.1. Σε κάθε κατάσταση ενός κρυφού Μαρκοβιανού μοντέλου αντιστοιχεί ένα ξεχωριστό γραμμικό δυναμικό μοντέλο, όπως περιγράφεται από τις Εξισώσεις (5.33) και (5.34). Τα κρυφά Μαρκοβιανά μοντέλα είναι εκπαιδεύσιμα κατά το συμβατικό τρόπο, όπως περιγράφεται και στην Ενότητα 5.3.1. Μετά από μία διαδικασία εξαναγκασμένης αντιστοίχισης (forced alignment) των οπτικοακουστικών δεδομένων υπολογίζονται οι πιθανότητες ανάθεσης σε κάθε κατάσταση/τάξη c και έτσι συγκεντρώνονται τα δεδομένα εκπαίδευσης που ανιστοιχούν σε κάθε γραμμικό δυναμικό μοντέλο. Η εξίσωση κατάστασης κάθε δυναμικού μοντέλου ταυτοποιείται μέσω μεγιστοποίησης της πιθανοφάνειας δεδομένων των εκπαιδευτικών διανυσμάτων των παραμέτρων άρθρωσης. Οι παράμετροι των εξισώσεων παρατήρησης προσδιορίζονται μέσω ανάλυσης κανονικής συσχέτισης περιορισμένης τάξης όπως στην Ενότητα 5.2.2.

Σε αυτό το πλαίσιο, η αντιστροφή απαιτεί την εύρεση της βέλτιστης ακολουθίας καταστάσεων δεδομένων των παρατηρήσεων (ακολουθίες ακουστικών, οπτικών ή οπτικοακουστικών χαρακτηριστικών), η οποία στην ουσία καθορίζει την εναλλαγή μεταξύ των ξεχωριστών γραμμικών δυναμικών μοντέλων. Για κάθε διάνυσμα παρατήρησης που έχει αντιστοιχηθεί σε

Πίνακας 5.2: *Λάθη RMS σε mm για τρεις διαφορετικές τεχνικές αντιστροφής, χρησιμοποιώντας ένα καθολικό γραμμικό δυναμικό σύστημα (LDS), χρησιμοποιώντας κρυφά Μαρκοβιανά μοντέλα (HMM) ή το προτεινόμενο διακοπτόμενο γραμμικό δυναμικό σύστημα (SLDS). Δίνονται οι περιπτώσεις χρήσης ακουστικής και οπτικοακουστικής πληροφορίας.*

	Λάθος RMS		
Τύπος αντιστροφής			
	LDS	HMM	SLDS
Ακουστική	2.15	1.76	1.78
Οπτικοακουστική	1.89	1.53	1.43

μια κατάσταση το αντίστοιχο διάνυσμα των παραμέτρων άρθρωσης εκτιμάται με το γραμμικό δυναμικό μοντέλο σε αυτή την κατάσταση με τη χρήση φίλτρου Kalman.

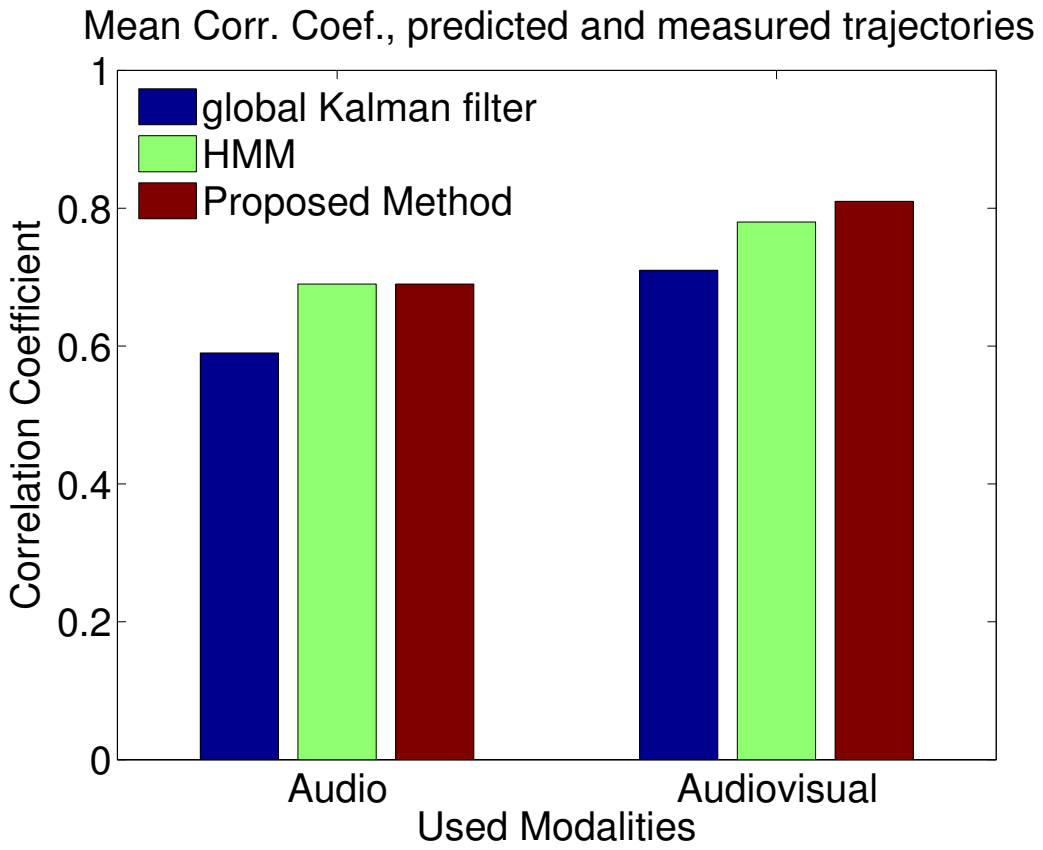
5.7 Πειράματα με περιορισμούς στη δυναμική των αρθρωτών

Ο στόχος είναι η αξιολόγηση της προτεινόμενης επέκτασης σε σύγκριση με την οπτικακουστική αντιστροφή με τη χρήση ενός καθολικού γραμμικού δυναμικού συστήματος ή με το βασικό πλαίσιο που βασίζεται σε κρυφά Μαρκοβιανά μοντέλα και προτείνεται στην Ενότητα 5.3. Τα πειράματα πραγματοποιούνται στη βάση MOCHA, βλ. Ενότητα 5.5.2. Για αναφορά παρουσιάζονται επίσης τα αποτελέσματα χρησιμοποιώντας τις ακουστικές μόνο παρατηρήσεις. Τα κρυφά Μαρκοβιανά μοντέλα που χρησιμοποιήθηκαν είναι βασισμένα σε ακουστικά φωνήματα και έχουν μία μόνο κατάσταση. Μοντέλα με περισσότερες καταστάσεις δεν ήταν δυνατό να εκπαιδευτούν επαρκώς και για αυτό η επίδοσή τους χειροτέρευσε ελαφρώς. Συνολικά, εκπαιδεύονται 46 μοντέλα, ένα για κάθε ακουστικό φώνημα που εμφανίζεται στη βάση MOCHA και δύο ακόμα για την αναπνοή και τη σιωπή. Σε κάθε μοντέλο, ενσωματώνονται και δύο καταστάσεις στις οποίες δεν αντιστοιχούν παρατηρήσεις, μία στην αρχή και μία στο τέλος, έτσι ώστε και οι μεταβάσεις μεταξύ των μοντέλων να μπορούν να ληφθούν υπόψη [179]. Για τις εκφωνήσεις που χρησιμοποιήθηκαν στην αξιολόγηση, το φωνηματικό τους περιεχόμενο θεωρείται γνωστό και οπότε είναι δυνατή η αποφυγή της αναγνώρισης και η πραγματοποίηση απλά μιας διαδικασίας εξαναγκασμένης αντιστοίχισης.

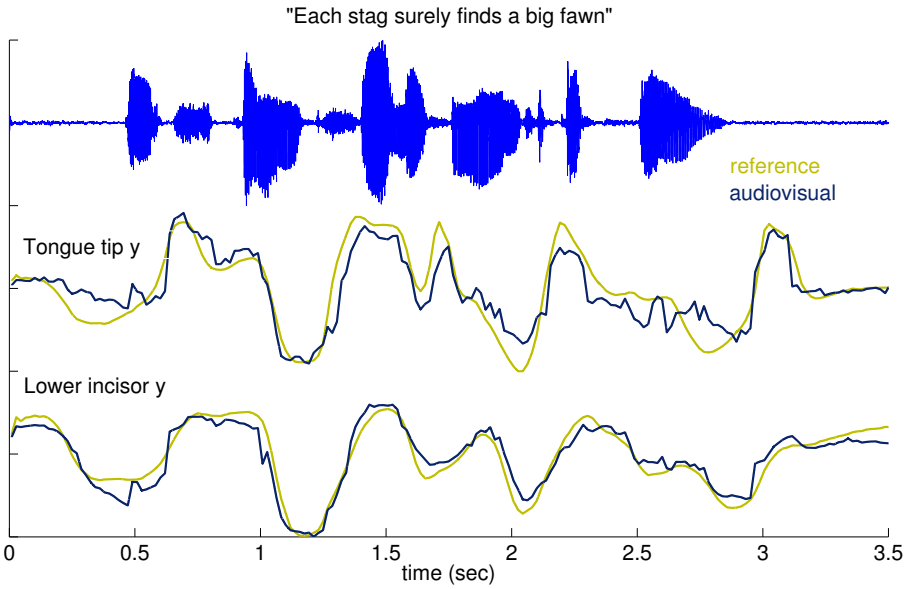
Υπολογίζεται ο μέσος συντελεστής συσχέτισης και RMS διαφορές μεταξύ των εκτιμώμενο και των μετρηθείσων τροχιών των αρθρωτικών παραμέτρων. Όπως φαίνεται στο Σχήμα 5.13, η ενσωμάτωση της δυναμικής μοντελοποίησης στο προτεινόμενο πλαίσιο οδηγεί σε βελτιώσεις. Αυτό φαίνεται επίσης στον Πίνακα 5.2 όπου δίνονται τα αντίστοιχα μέσα λάθη RMS. Για σύγκριση, δίνεται και το λάθος για την περίπτωση ενός καθολικού γραμμικού δυναμικού συστήματος. Στο Σχήμα 5.14 δίνεται ένα παράδειγμα τροχιών των y-συντεταγμένων της άκρης της γλώσσας και του κάτω κόφτη, όπως προβλέπονται και όπως έχουν μετρηθεί για μια εκφώνηση στη βάση MOCHA.

5.8 Μίμηση φωνής

Από τη σκοπιά του μηχανικού, η αντιστροφή της φωνής σε πληροφορία άρθρωσης αναφέρεται στο πρόβλημα της ταυτοποίησης του υποκείμενου συστήματος παραγωγής φωνής δεδομένης της παρατηρούμενης εξόδου. Στο συγκείμενο της παρούσας διατριβής η αντιστροφή εφαρμόζεται για την ταυτοποίηση του προτεινόμενου υπολογιστικού μοντέλου της φωνής. Πιο συγκεκριμένα, ο στόχος είναι η μίμηση φωνής, ο προσδιορισμός δηλαδή της ακολουθίας των κρυφών καταστάσεων του μοντέλου ώστε να είναι στη συνέχεια δυνατή η επανασύνθεση του αρχικού, παρατηρούμενου σήματος φωνής. Η πρακτική αυτή εφαρμογή σε πραγματικά δεδομένα επιτρέπει την αξιολόγηση του μοντέλου φωνής και δίνει ένα πλαίσιο για την περαιτέρω μελέτη και ανάπτυξη τεχνικών σύνθεσης φωνής βασισμένων σε πιστές αναπαραστάσεις των αντίστοιχων φυσικών ανθρώπινων λειτουργιών.



Σχήμα 5.13: *Αξιολόγηση της ακουστικής ή οπτικοακουστικής πληροφορίας με βάση το μέσο συντελεστή συσχέτισης μεταξύ των μετρηθείσων και των εκτιμώμενων τροχιών άρθρωσης. Δίνονται τρεις περιπτώσεις για σύγκριση, με τη χρήση ενός καθολικού γραμμικού δυναμικού συστήματος, με τη χρήση μόνο κρυφών Μαρκοβιανών μοντέλων ή με το προτεινόμενο διακοπτόμενο γραμμικό δυναμικό μοντέλο.*



Σχήμα 5.14: *Οι προβλεπόμενες τροχιές του κάτω κόφτη και της άκρης της γλώσσας (y-συντεταγμένες) όπως βρίσκονται με το προτεινόμενο σχήμα οπτικοακουστικής αντιστροφής. Για αναφορά, δίνονται οι αντίστοιχες μετρήσεις για τις συγκεκριμένες τροχιές με ανοιχτό χρώμα.*

Για την ανάπτυξη ενός πλαισίου αντιστροφής, τρία είναι τα κύρια θέματα σχεδίασης που πρέπει να αντιμετωπιστούν. Είναι σημαντική η επιλογή του τρόπου αναπαράστασης της φωνής, η περιγραφή του συστήματος παραγωγής φωνής και τέλος η υιοθέτηση κατάλληλου υπολογιστικού σχήματος. Οι σχετικές αποφάσεις συνήθως βασίζονται σε κάποια θεώρηση της θεωρίας παραγωγής φωνής, στη φύση και το πλήθος των διαθέσιμων δεδομένων και στους συγκεκριμένους στόχους της εφαρμογής που θα πρέπει να εξυπηρετούνται με την αντιστροφή. Στο πλαίσιο αντιστροφής που παρουσιάζεται ως επέκταση του συστήματος που έχει παρουσιαστεί ως τώρα στο Κεφάλαιο 5, η φωνή αναπαρίσταται τόσο με ακουστικά φασματικά δεδομένα όσο και με οπτική πληροφορία από το πρόσωπο του ομιλητή. Το σύστημα παραγωγής περιγράφεται μέσω ένος κατάλληλου μοντέλου άρθρωσης και η απεικόνιση της φωνής στην άρθρωση προσεγγίζεται με κατά τμήματα γραμμικό τρόπο.

Όπως σημειώθηκε και πρωτύτερα, παραδοσιακά η αντιστροφή φωνής θεωρείται ως ο προσδιορισμός του σχήματος της φωνητικής οδού μόνο από το ακουστικό σήμα φωνής [63, 113, 121]. Συχνότητες συντονισμών, γραμμικές φασματικές συχνότητες ή Mel συντελεστές cepstrum είναι αναπαραστάσεις που έχουν χρησιμοποιηθεί. Όπως φάνηκε όμως, βλ. Ενότητα 5.5.6, η εισαγωγή της οπτικής συνιστώσας στη διαδικασία της αντιστροφής μπορεί να βελτιώσει σημαντικά την ακρίβεια των αποτελεσμάτων. Η ανάλυση ανεξάρτητων συνιστωσών της περιοχής γύρω από τα χείλια [88] ή η ενεργή μοντελοποίηση εμφάνισης του προσώπου, Ενότητα 5.4, παρέχουν τα πρακτικά μέσα για να επιτευχθεί η κατάλληλη οπτική αναπαράσταση. Εναλλακτικά, είναι δυνατή η χρησιμοποίηση των συντεταγμένων κατάλληλων σημαδευτών στο πρόσωπο, για παράδειγμα πάνω στα χείλια ή στο σαγόνι, μετά από ιχνηλάτησή τους με βάση στερεοσκοπικά δεδομένα (βλ. Ενότητα 5.8.3).

Όσον αφορά στην περιγραφή της φωνητικής οδού, έχουν προταθεί διάφορες εναλλακτικές, η καθεμιά από τις οποίες ικανοποιεί συγκεκριμένες απαιτήσεις. Για παράδειγμα, το μοντέλο με τους κυλίνδρους στο [143] επιτρέπει αντιστροφή από τους συντονισμούς της φωνής με βάση τη γραμμική θεωρία παραγωγής φωνής. Είναι όμως αρκετά περιοριστικό και δεν επιτρέπει άμεσα την ανάλυση ήχων διαφορετικών των φωνηέντων. Η αναπαράσταση μέσω των συντεταγμένων διάφορων σημείων πάνω σε σημαντικούς αρθρωτές, όπως χρησιμοποιείται στα [45, 63, 137, 170] και στο πλαίσιο που παρουσιάστηκε στην Ενότητα 5.5.2, είναι περισσότερο ρεαλιστική αλλά δεν είναι τόσο λεπτομερής και τόσο πληροφοριακή για την κατάσταση ολόκληρης της φωνητικής οδού. Το καλό είναι όμως ότι τέτοια δεδομένα μπορούν να συλλεχθούν σχετικά εύκολα με τη χρήση συστήματος ηλεκτρομαγνητικής καταγραφής των αρθρωτών και έχουν επιτρέψει την εφαρμογή τεχνικών μηχανικής μάθησης για την επίτευξη της αντιστροφής. Μια αρκετά περισσότερο πληροφοριακή αναπαράσταση του φωνητικού συστήματος είναι αυτή που επιτυγχάνεται μέσω ενός μοντέλου άρθρωσης [113, 121] το οποίο μπορεί να περιγράψει τη γεωμετρία ή τις επιφάνειες των εγκάρσιων τομών της φωνητικής οδού κι ελέγχεται από έναν περιορισμένο αριθμό παραμέτρων. Τέτοια μοντέλα έχουν φτιαχτεί από πραγματικά δεδομένα άρθρωσης, που έχουν συγκεντρωθεί από εικόνες είτε ακτίνων-Χ, όπως το στατιστικό μοντέλο στο [102] ή το γεωμετρικό μοντέλο στο [112] που έχει επίσης επεκταθεί στις τρεις διαστάσεις [23], είτε μαγνητικής τομογραφίας [11, 14, 113] της φωνητικής οδού. Το πλήθος των αντίστοιχων δεδομένων είναι περιορισμένο όμως και για αυτό τέτοια δεδομένα δεν είναι εύκολα χρησιμοποιήσιμα σε σενάρια αντιστροφής με τεχνικές μηχανικής μάθησης.

Όσον αφορά στο υπολογιστικό σχήμα αντιστροφής, έχουν αναφερθεί τόσο μέθοδοι βασισμένες σε μοντέλο [12, 121, 143] όσο και μέθοδοι βασισμένες σε μηχανική μάθηση [63, 181]. Για παράδειγμα, στο [121] περιγράφονται αποδοτικοί τρόποι για χρήση βιβλίων κωδικών που να συσχετίζουν συντονισμούς φωνής και τις παραμέτρους ενός μοντέλου άρθρωσης. Για να αξιοποιηθεί επιπλέον και δυναμική πληροφορία, μια κατά τμήματα γραμμική προσέγγιση της σχέσης ήχου-συστήματος παρουσιάζεται στο [63]. Το κάθε φώνημα μοντελοποιείται, όπως παρουσιάστηκε και στο Κεφάλαιο 5, με ένα κρυφό Μαρκοβιανό μοντέλο και ένα ξεχωριστό γραμμικό μοντέλο εκπαιδεύεται σε κάθε κατάσταση μεταξύ των παρατηρούμενων ακουστικών παραμέτρων και των παραμέτρων άρθρωσης.

Σε αυτό το πλαίσιο, προτείνεται ένα σχήμα οπτικοακουστικής αντιστροφής της φωνής που αναπτύσσεται με τη χρήση πολυμεσικών δεδομένων άρθρωσης και επιτρέπει την επανασύνθεση του παρατηρούμενης ακουστικής συνιστώσας της φωνής. Με αυτόν τον τρόπο γίνεται δυνατή η μίμηση του ανθρώπινου φωνητικού συστήματος. Η οπτική πληροφορία αξιοποιείται με τη χρήση των τρισδιάστατων συντεταγμένων σημαδευτών που είναι ζωγραφισμένοι πάνω στο πρόσωπο του ομιλητή και ιχνηλατούνται μέσω στέρεο-οπτικής, βλ. Ενότητα 5.8.3. Το σχήμα της φωνητικής οδού περιγράφεται από ένα μοντέλο άρθρωσης το οποίο κατασκευάζεται από δεδομένα ακτίνων-Χ στα οποία έχει προηγηθεί ένα στάδιο ανθρώπινης επεξεργασίας και εξαγωγής καμπυλών, βλ. Ενότητα 5.8.2. Η αντιστροφή επιτυγχάνεται μέσω ενός πλαισίου παρόμοιου με αυτό της ενότητας 5.3 βασισμένο σε κρυφά Μαρκοβιανά μοντέλα. Λεπτομέρειες δίνονται στην Ενότητα 5.8.4. Τα πειράματα πραγματοποιούνται σε ένα σύνολο δεδομένων άρθρωσης που συλλέχθηκε πρόσφατα και περιλαμβάνει ταυτόχρονες καταγραφές ήχου, στέρεο-οπτικής του προσώπου του ομιλητή, δεδομένα ηλεκτρομαγνητικής καταγραφής των αρθρωτών και βίντεο της γλώσσας με τη χρήση υπερήχων, βλ. Ενότητα 5.8.1. Για την εξαγωγή κατάλληλων παραμέτρων άρθρωσης από αυτό το σύνολο αντιστοιχίσαμε το μοντέλο άρθρωσης στο ορατό τμήμα της καμπύλης της γλώσσας σε κάθε πλαίσιο των δεδομένων που είναι καταγεγραμμένα με υπερήχους, βλ. Ενότητα 5.8.3.4. Η αντιστοίχιση του συστήματος αναφοράς των εικόνων ακτίνων-Χ στις εικόνες των υπερήχων επιτυγχάνεται με κατάλληλη αξιοποίηση των διαθέσιμων δεδομένων μαγνητικής τομογραφίας του κεφαλιού του ομιλητή και τα στέρεο-οπτικά δεδομένα, βλ. Ενότητα 5.8.3. Τα εκτιμώμενα σχήματα της φωνητικής οδού μετά την αντιστροφή είναι αρκετά κοντά στα αρχικά και επιδεικνύουν τις προοπτικές της προσέγγισης.

5.8.1 Πολυτροπικά δεδομένα άρθρωσης

Το σύνολο των δεδομένων άρθρωσης με το οποίο έχει αναπτυχθεί το περιγραφόμενο πλαίσιο περιγράφεται με λεπτομέρεια στο [8]. Περιλαμβάνει φωνητικά μεταγεγραμμένο ήχο (44kHz) στέρεο-βίντεο (120Hz) του προσώπου, εικόνες της γλώσσας με υπερήχους (65Hz) και καταγραφές ηλεκτρομαγνητικών αισθητήρων (40Hz) που είναι τοποθετημένοι πάνω στον καταγραφέα υπερήχων, τη γλώσσα και το κεφάλι του ομιλητή. Στα πειράματα που πραγματοποιηθηκαν, αξιοποιήθηκαν περίπου 6 λεπτά από τις καταγραφές αυτές. Ο ομιλητής είναι Γάλλος και το σώμα κειμένων που εκφωνείται περιλαμβάνει μια μεγάλη ποικιλία από μεμονωμένες ακολουθίες φωνημάτων (Φωνήεν-Σύμφωνο-Φωνήεν ή Φωνήεν-Φωνήεν) και ένα σύνολο από φωνηματικά εξισορροπημένες Γαλλικές προτάσεις. Επιπρόσθετα, είναι διαθέσιμα τρισδιάστατα δεδομένα μαγνητικής τομογραφίας του κεφαλιού του ομιλητή (για τρία παρατεταμένα φωνήεντα) ενώ υπάρχουν και περίπου 700 σχήματα της φωνητικής οδού, που έχουν εξαχθεί με ημιαυτόματη επεξεργασία από 30 δευτερόλεπτα βίντεο (25Hz) της φωνητικής οδού του ομιλητή που έχει καταγραφεί με ακτίνες - Χ. Ο ουρανίσκος, η μαλακή υπερώα και το φαρυγγικό τοίχωμα που θεωρούνται ότι δεν αλλάζουν από εικόνα σε εικόνα, προσδιορίστηκαν ημιαυτόματα με τη χρήση κατάλληλης γραφικής διεπαφής. Η ημιαυτόματη εξαγωγή της γλώσσας, των χειλιών και του λάρυγγα έγινε από την ερευνητική ομάδα Parole του LORIA[3] στο Νανσύ και το IPS[4] στο Στρασβούργο με βάση τεχνικές εμπνευσμένες από το [58]. Οι καμπύλες που περιγράφουν τη φωνητική οδό χρησιμοποιήθηκαν για την εκπαίδευση κατάλληλου μοντέλου άρθρωσης.

5.8.2 Ανάπτυξη μοντέλου άρθρωσης

Ο ρόλος του μοντέλου άρθρωσης είναι πολύ σημαντικός στο προτεινόμενο πλαίσιο. Περιγράφει το σχήμα της φωνητικής οδού στο μεσοκάθετο επίπεδο που περνάει από τη μύτη και

[3]Lorraine Laboratory of IT Research and its Applications
[4]Institute Phonetique de Strasbourg

χωρίζει τη φωνητική οδό σε δύο κατοπτρικά τμήματα (μέσο οβελιαίο επίπεδο). Δημιουργείται κατά βάση ακολουθώντας τη μέθοδο που περιγράφεται στα [100, 102]. Ένα ημιπολικό πλέγμα τοποθετείται κατάλληλα στο μεσοκάθετο επίπεδο, Σχήμα 5.15(α'), και βρίσκονται οι συντεταγμένες των σημείων τομής των γραμμών του πλέγματος με τα όρια της φωνητικής οδού, όπως φαίνεται στο Σχήμα 5.15(β') και περιγράφεται στην Ενότητα 3.8.1.

Έχοντας τοποθετήσει το πλέγμα πάνω στις εικόνες της φωνητικής οδού, ακολουθεί αυτόματη εξαγωγή των σημείων τομής των γραμμών του πλέγματος με τις καμπύλες που περιγράφουν το εσωτερικό και το εξωτερικό τοίχωμα της φωνητικής οδού. Το εσωτερικό τοίχωμα θεωρείται ότι σχηματίζεται κυρίως από τη γλώσσα και το υπεργλωττιδικό τοίχωμα του λάρυγγα ενώ καταλήγει στον κάτω κόφτη. Αμελείται η επιγλωττίδα στην παρούσα ανάλυση. Το εξωτερικό τοίχωμα περιλαμβάνει τον πάνω κόφτη, τον ουρανίσκο, τη μαλακή υπερώα (στη θέση όπου το ρινοφαρυγγικό άνοιγμα είναι κλειστό) το φαρυγγικό τοίχωμα και το υποφαρυγγικό τοίχωμα του λάρυγγα. Τα χείλη εξαιρούνται αφού θεωρείται ότι λόγω της διαμήκους μεταβλητότητάς τους δεν περιγράφονται κατάλληλα από το συγκεκριμένο ημιπολικό πλέγμα. Τα σημεία τομής που εξάγονται αρχικά στο καρτεσιανό σύστημα συντεταγμένων [5] μετατρέπονται τελικά στο σύστημα συντεταγμένων του ημιπολικού πλέγματος. Το κάθε σημείο τομής δηλαδή αναπαρίσταται από τον αύξοντα αριθμό της αντίστοιχης γραμμής πλέγματος και την απόστασή του από την αρχή της γραμμής αυτής. Έτσι τα δύο τοιχώματα της φωνητικής οδού σε κάθε εικόνα ακτίνων - Χ περιγράφονται το καθένα από ένα N-διάστατο διάνυσμα αποστάσεων όπου N είναι το πλήθος των γραμμών του ημιπολικού πλέγματος.

Το εξωτερικό τοίχωμα θεωρείται ότι παρουσιάζει ελάχιστη μεταβλητότητα στις διαθέσιμες εικόνες και στο αρθρωτικό μοντέλο περιλαμβάνεται μόνο το μέσο διάνυσμα περιγραφής του. Για το εσωτερικό τοίχωμα και με στόχο τον προσδιορισμό μιας συμπαγούς περιγραφής του εφαρμόζεται πιθανοτική ανάλυση σε πρωτεύουσες συνιστώσες [25, σελ. 570-577] των διανυσμάτων αποστάσεων. Το διάνυσμα συντελεστών στο χώρο των πρωτευουσών συνιστωσών έχει διάσταση M και θεωρείται ότι ακολουθεί την πρότερη κατανομή

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \mathbf{0}, I)$$

ενώ το παρατηρούμενο διάνυσμα διάστασης N μοντελοποιείται ως :

$$\mathbf{y} = A\mathbf{x} + \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

όπου $\boldsymbol{\mu} = E[\mathbf{y}]$ και το $\boldsymbol{\epsilon}$ είναι κανονικά κατανεμημένο με μέση τιμή μηδέν και συμμεταβλητότητα $\sigma^2 I$. Αυτό σημαίνει ότι :

$$p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x} = N(\mathbf{y}|\boldsymbol{\mu}, C)$$

με $C = AA^T + \sigma^2 I$. Ο πίνακας A μπορεί πρακτικά να υπολογιστεί μέσω μεγιστοποίησης της πιθανοφάνειας του μοντέλου με βάση τις παρατηρήσεις [25]. Τελικά, σχηματίζεται από τα M πρωτεύοντα ιδιοδιανύσματα όπως αυτά προκύπτουν από την κλασική ανάλυση πρωτευουσών συνιστωσών. Αντίστοιχα, η παράμετρος σ^2 εκτιμάται ως :

$$\sigma^2 = \frac{1}{N-M} \sum_{M+1}^{N} \lambda_i$$

όπου $\lambda_i, i = 1, \ldots, N$ είναι οι ιδιοτιμές όπως προκύπτουν από την κλασσική ανάλυση. Στην ουσία δηλαδή, με την πιθανοτική ανάλυση επιτυγχάνεται η σωστή αναπαράσταση της μεταβλητότητας των δεδομένων κατά μήκος των πρωτευουσών κατευθύνσεων και η προσέγγιση της μεταβλητότητας στις υπόλοιπες κατευθύνσεις με μία μόνο μέση τιμή σ^2.

[5] Χρησιμοποιείται η συνάρτηση polyxpoly του MATLAB που γενικότερα εντοπίζει σημεία τομής μεταξύ πολυγωνικών καμπυλών.

Η ανάλυση αυτή καθορίζει τις συνιστώσες ενός γραμμικού μοντέλου το οποίο μπορεί να περιγράψει περίπου 96% της μεταβλητότητας του σχήματος με τη χρήση μόνο 6 παραμέτρων. Οι αντίστοιχες έξι πρωτεύουσες συνιστώσες δίνονται στα Σχήματα 5.16(α')-5.16(β') για τις μέγιστες και τις ελάχιστες τιμές των ιδιο-παραμέτρων στα δεδομένα εκπαίδευσης. Αναπαρίσταται επίσης και το εξωτερικό τοίχωμα. Στην παρούσα φάση δεν έχει γίνει κάποια προσπάθεια διαισθητικής ερμηνείας των επιμέρους συνιστωσών ή κάποια προσπάθεια καθοδήγησης της διαδικασίας εξαγωγής του μοντέλου ώστε να επιτευχθεί η αντιστοίχιση των στατιστικών παραμέτρων με φυσικές [102]. Θα μπορούσε πάντως να πει κάποιος ότι η πρώτη συνιστώσα αντιστοιχεί σε συνδυασμένη κίνηση του σαγονιού και κίνηση της γλώσσας μπρος - πίσω. Από την άλλη, η δεύτερη συνιστώσα σχετίζεται περισσότερο με το ύψος της γλώσσας, όπως είναι γνωστή στη φωνολογιά η θέση της γλώσσας που αφορά την κίνησή της πάνω-κάτω.

Ο στόχος είναι στη συνέχεια να ταιριάξει αυτό το μοντέλο στα δεδομένα της γλώσσας που έχουν καταγραφεί με υπερήχους ώστε να είναι δυνατή η εξαγωγή μιας αποδοτικής αναπαράστασης του σχήματος της φωνητικής οδού για ολόκληρο το σύνολο δεδομένων. Για αυτό το σκοπό, είναι αναγκαία η αντιστοίχιση του ημιπολικού συστήματος συντεταγμένων στις εικόνες των υπερήχων.

5.8.3 Η αντιστοίχιση των δεδομένων άρθρωσης

Για την αντιστοίχιση του μοντέλου άρθρωσης στα δεδομένα που έχουν καταγραφεί με τη βοήθεια υπερήχων [6], εφαρμόζεται ένα στάδιο προεπεξεργασίας ξεχωριστά στους διάφορους τύπους δεδομένων. Το εξωτερικό τοίχωμα της φωνητικής οδού του ομιλητή και η επιφάνεια του μετώπου ανακατασκευάζονται από τα τρισδιάστα δεδομένα μαγνητικής τομογραφίας του κεφαλιού του, Σχήμα 5.17. Οι τρισδιάστατες θέσεις των ζωγραφισμένων σημαδευτών στο πρόσωπο κάθε χρονική στιγμή ιχνηλατούνται αυτόματα χρησιμοποιώντας τις ακολουθίες εικόνων από το ζευγάρι των στερεο-καμερών, Σχήμα 5.18. Οι ακολουθίες εικόνων που έχουν καταγραφεί με τη χρήση υπερήχων φιλτράρονται όπως περιγράφεται στο [8], οπότε η καμπύλη της γλώσσας ξεχωρίζει περισσότερο από το υπόβαθρο, Σχήμα 5.19.

Τα βασικά βήματα για το συνδυασμό και την αξιοποίηση των διαφορετικών τύπων αρθρωτικών δεδομένων έχουν ως εξής :

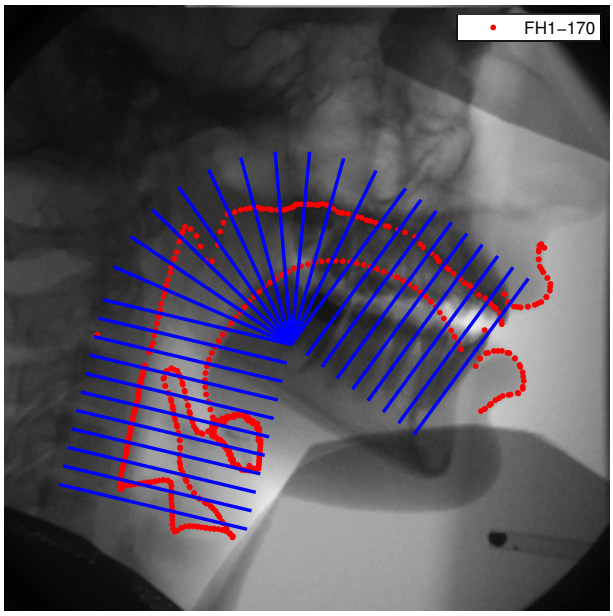
5.8.3.1 Μετατροπή στο σύστημα συντεταγμένων του ηλεκτρομαγνητικού συστήματος καταγραφής

Οι τρισδιάστατες θέσεις των τριών ηλεκτρομαγνητικών αισθητήρων (πίσω από τα αυτιά και πάνω στον καταγραφέα υπερήχων) προσεγγίζονται στο στέρεο-οπτικό σύστημα συντεταγμένων με χρήση των στέρεο-εικόνων σε κάποιες χρονικές στιγμές. Με αυτόν τον τρόπο, είναι δυνατό να έχουμε τις ίδιες θέσεις τόσο στο στέρεο-οπτικό σύστημα όσο και στο σύστημα συντεταγμένων των ηλεκτρομαγνητικών αισθητήρων. Οπότε, μέσω αντιστοίχισης, μπορεί να εκτιμηθεί ο μετασχηματισμός από το ένα σύστημα στο άλλο.

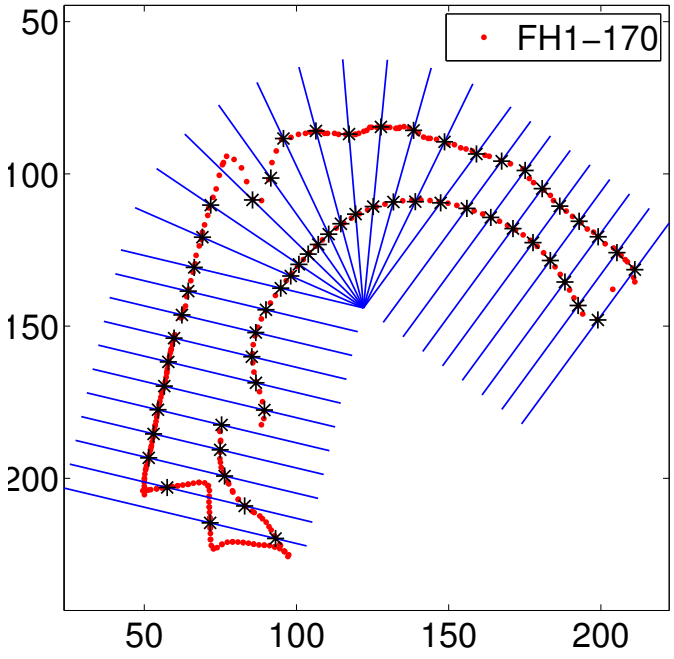
5.8.3.2 Αντιστάθμιση της κίνησης του κεφαλιού

Χρησιμοποιείται ένα σύστημα συντεταγμένων το οποίο έχει σταθερή θέση και προσανατολισμό στο χρόνο με αναφορά το κεφάλι του ομιλητή. Το σύστημα αυτό θα αναφέρεται ως σύστημα αναφοράς του κεφαλιού. Ο μετασχηματισμός συντεταγμένων από το σύστημα των ηλεκτρομαγνητικών αισθητήρων στο σύστημα του κεφαλιού υπολογίζεται μέσω αντιστοίχισης των θέσεων των αισθητήρων στο πάνω μέρος του κεφαλιού μια συγκεκριμένη χρονική στιγμή αναφοράς. Οι τρισδιάστατες τροχιές των ηλεκτρομαγνητικών αισθητήρων και των υπόλοιπων σημαδευτών του προσώπου εκφράζονται στο σύστημα αναφοράς του κεφαλιού, Σχήμα 5.17.

[6] Σε συνεργασία με τον Α. Ρούσσο

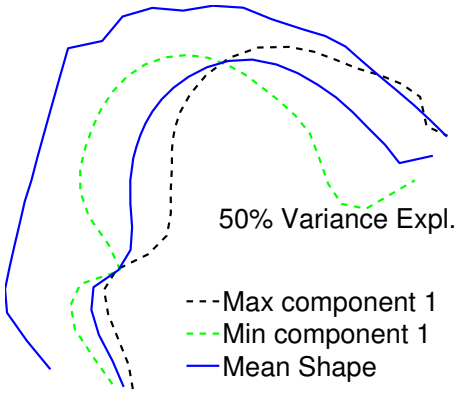


(α') Ακτίνες - X, καμπύλες και πλέγμα

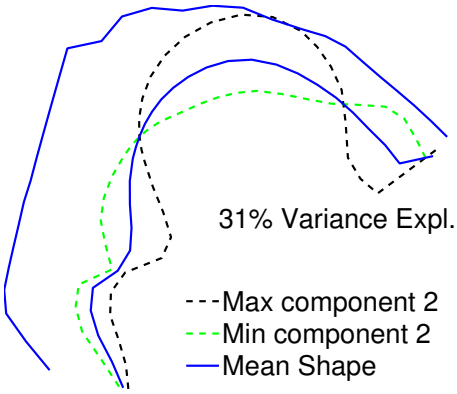


(β') Σημεία τομής

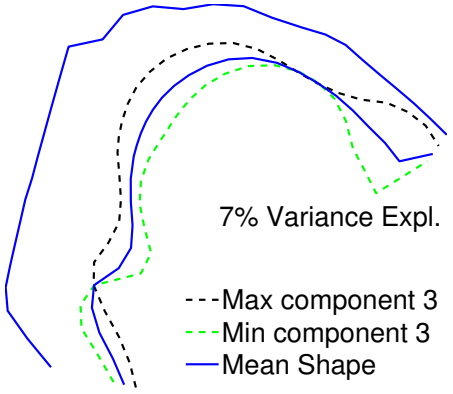
Σχήμα 5.15: *Δημιουργία του μοντέλου άρθρωσης από τα δεδομένα ακτίνων-X· τοποθέτηση του πλέγματος - συστήματος αναφοράς και εύρεση των σημείων τομής με τις ακμές της φωνητικής οδού.*



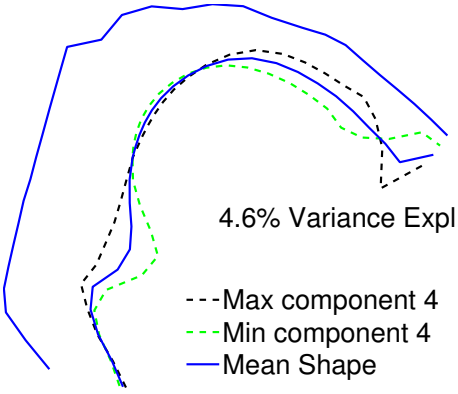
(α') 1η συνιστώσα γραμμικού μοντέλου



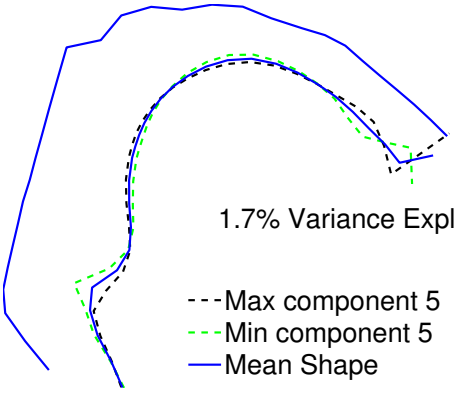
(β') 2η συνιστώσα γραμμικού μοντέλου



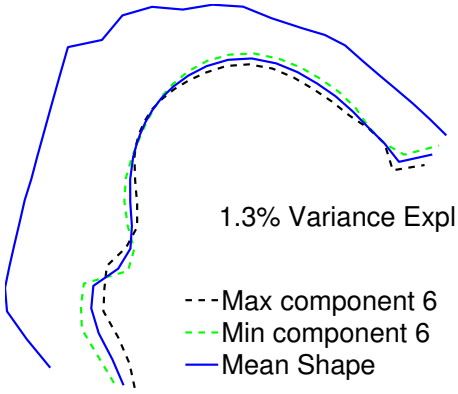
(γ') 3η συνιστώσα γραμμικού μοντέλου



(δ') 4η συνιστώσα γραμμικού μοντέλου

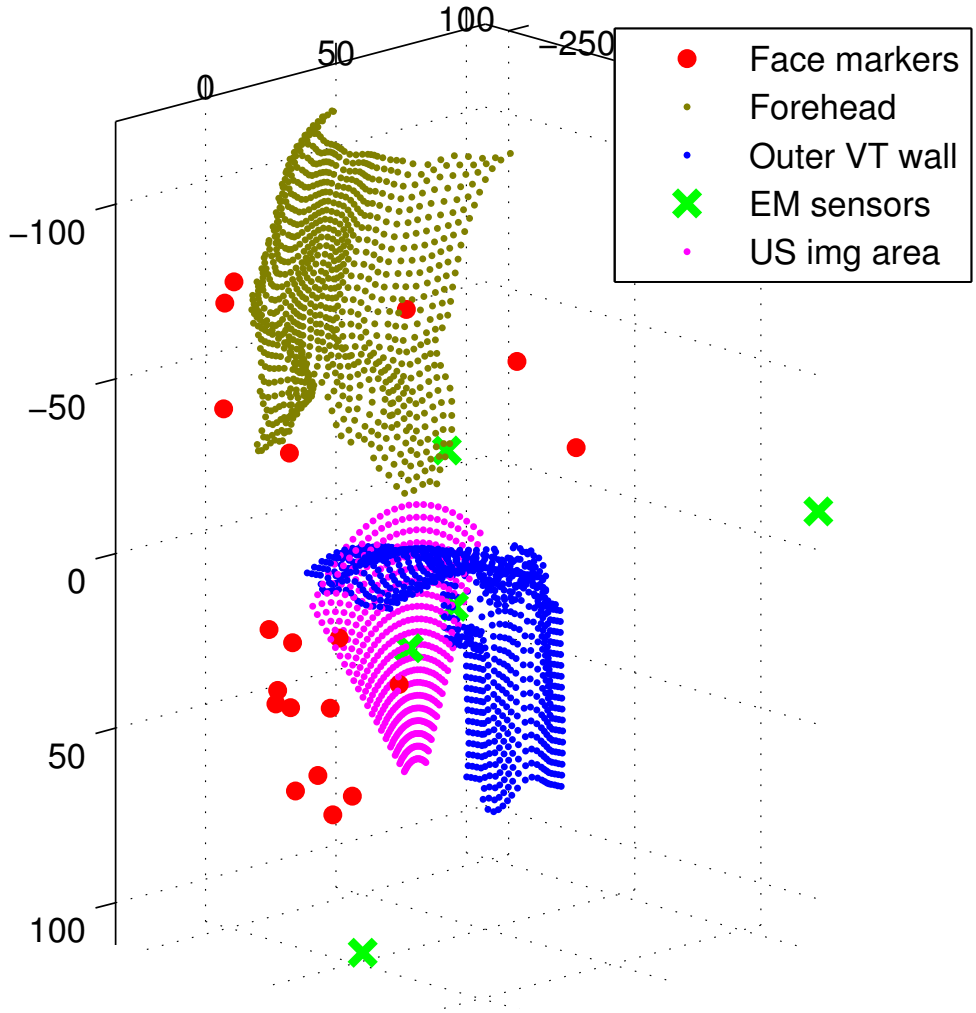


(ε') 5η συνιστώσα γραμμικού μοντέλου

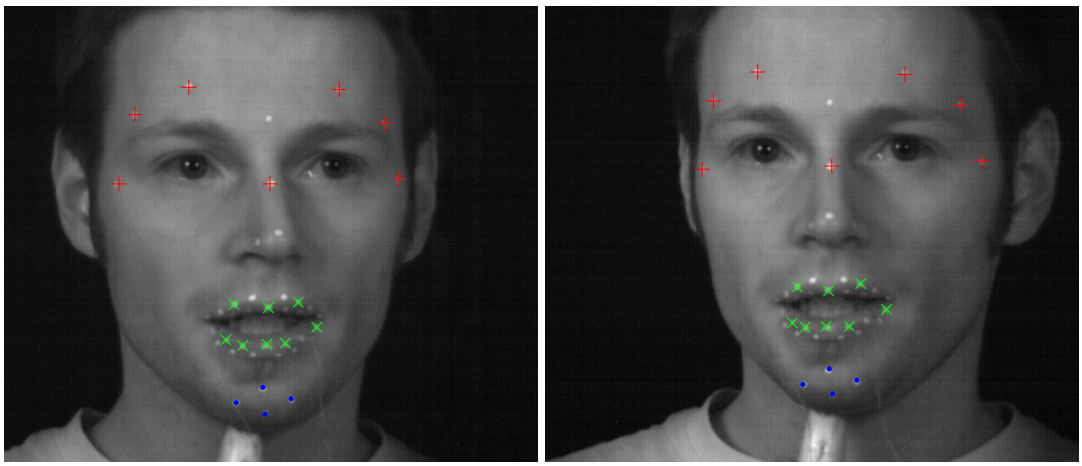


(ϛ') 6η συνιστώσα γραμμικού μοντέλου

Σχήμα 5.16: *Δίνονται οι πρώτες έξι συνιστώσες του μοντέλου μετά την πιθανοτική ανάλυση σε πρωτεύουσες συνιστώσες.*



Σχήμα 5.17: *Η αντιστοίχιση των πολυμεσικών δεδομένων άρθρωσης για μια συγκεκριμένη χρονική στιγμή. Έχει χρησιμοποιηθεί το σύστημα αναφοράς που είναι πακτωμένο στο κεφάλι.*



Σχήμα 5.18: *Ένα ζευγάρι εικόνων από τις στέρεο-κάμερες μαζί με τους σημαδευτές στο πρόσωπο που έχουν σημαδευτεί. Οι σημαδευτές στο πάνω μέρος του κεφαλιού ('+') χρησιμοποιούνται για τη διαδικασία αντιστοίχισης ενώ οι σημαδευτές στα χείλια ('×') χρησιμοποιούνται κατά την αντιστροφή.*

Μεταξύ αυτών, οι τροχιές των σημαδευτών πάνω στα χείλια χρησιμοποιούνται στη συνέχεια για την οπτικοακουστική αντιστροφή, όπως περιγράφεται στην Ενότητα 5.8.4.

5.8.3.3 Η αντιστοίχιση του πλέγματος της φωνητικής οδού στις εικόνες των υπερήχων

Το ημιπολικό πλέγμα/σύστημα αναφοράς αρχικά προσδιορίζεται στις εικόνες ακτίνων-Χ και στη συνέχεια περιγράφεται στο μεσοκάθετο επίπεδο των δεδομένων μαγνητικής τομογραφίας μέσω αντιστοίχισης του εξωτερικού τοιχώματος της φωνητικής οδού στους δύο τύπους δεδομένων. Το πλέγμα επεκτείνεται στην τρίτη διάσταση θεωρώντας ότι είναι σταθερό για όλα τα επίπεδα δεδομένων μαγνητικής τομογραφίες. Στη συνέχεια, εκφράζεται στο σύστημα αναφοράς του κεφαλιού μέσω αντιστοίχισης της επιφάνειας του μετώπου όπως εξάγεται από τις εικόνες μαγνητικής τομογραφίας με τους αντίστοιχους σημαδευτές στο πρόσωπο. Η τρισδιάστατη θέση και ο προσανατολισμός (στο σύστημα αναφοράς του κεφαλιού) του κινούμενου επιπέδου των εικόνων των υπερήχων ανακτώνται χρησιμοποιώντας τον ηλεκτρομαγνητικό αισθητήρα με τους έξι βαθμούς ελευθερίας πάνω στον καταγραφέα υπερήχων, Σχήμα 5.17. Στο τέλος, υπολογίζεται η τομή του επιπέδου της εικόνας των υπερήχων με το τρισδιάστατο πλέγμα της φωνητικής οδού για κάθε χρονική στιγμή.

Μια εναλλακτική και πιθανά ακριβέστερη μέθοδος καταγραφής περιγράφεται στο [9]. Η βασική διαφορά της είναι ότι για την καταγραφή των εικόνων αξονικής τομογραφίας στο ηλεκτρομαγνητικό σύστημα αναφοράς χρησιμοποιεί μια μάσκα του προσώπου όπως αυτή δειγματοληπτείται με τη βοήθεια ενός ηλεκτρομαγνητικού αισθητήρα. Για τα τελικά πειράματα που παρουσιάζουμε χρησιμοποιήσαμε τα αποτελέσματα αυτής της ακριβέστερης μεθόδου, τα οποία όμως επίσης εμφανίζουν αφύσικες αποκλίσεις σε κάποιες περιπτώσεις. Για παράδειγμα, υπάρχουν εικόνες όπου φαίνεται η γλώσσα να διαπερνά τον ουρανίσκο ή να βγαίνει έξω από το στόμα. Για τη διόρθωση αυτών των λαθών προσπαθήσαμε να τροποποιήσουμε ελαφρώς το ημιπολικό πλέγμα που καταγράφεται στις εικόνες των υπερήχων ώστε να μεγιστοποιήσουμε την πιθανοφάνεια του αρθρωτικού μοντέλου που περιγράφει το σχήμα της γλώσσας και που τελικά θέλουμε να ταιριάξουμε στην καμπύλη της γλώσσας όπως αυτή φαίνεται στις εικόνες αυτές. Με τον τρόπο αυτό έγινε δυνατή η διόρθωση μεγάλου μέρους των παρατηρούμενων αποκλίσεων.

5.8.3.4 Ταίριασμα του μοντέλου στα σημεία της γλώσσας όπως φαίνονται με τους υπέρηχους

Το επόμενο βήμα περιλαμβάνει την εξαγωγή της καμπύλης της γλώσσας από τις εικόνες των υπερήχων και την εύρεση των σημείων τομής της με το ημιπολικό πλέγμα. Το πρόβλημα δεν είναι εύκολο αφού η καμπύλη της γλώσσας είναι μόνο μερικώς και ασθενώς ορατή. Η πιο γνωστή μέθοδος ιχνηλάτησης της γλώσσας σε τέτοιες εικόνες έχει δημοσιευτεί στο [93] και βασίζεται σε ένα μοντέλο τύπου Snake. Είναι ημιαυτόματη μέθοδος που δουλεύει αρκετά καλά για ακολουθίες εικόνων όπου φαίνεται το ίδιο τμήμα της γλώσσας αλλά δεν μπορεί εύκολα να ανανήψει από ενδεχόμενα λάθη με αποτέλεσμα να χρειάζεται συχνά επαναρχικοποίηση. Μια βελτίωσή της προτάθηκε στο [8] όπου επιπρόσθετα οι εικόνες φιλτράρονται, όπως στη δική μας περίπτωση, χρησιμοποιείται πληροφορία οπτικής ροής μεταξύ δύο διαδοχικών πλαισίων αλλά και η πληροφορία της θέσης δύο αισθητήρων που είναι προσαρτημένοι πάνω στη γλώσσα.

Η αρχική προσέγγιση που ακολουθήθηκε βασίστηκε σε μερική απλοποίηση του προβλήματος. Αντί για την ανίχνευση ολόκληρης της καμπύλης της γλώσσας, αναζητήθηκαν κατευθείαν τα σημεία τομής της με το ημιπολικό πλέγμα. Η υπόθεση είναι ότι στις φιλτραρισμένες εικόνες ηπερήχων, κάθε γραμμή του πλέγματος της φωνητικής οδού τέμνει το ορατό μέρος της καμπύλης της γλώσσας μόνο αν η μέγιστη ένταση της εικόνας πάνω σε αυτή τη γραμμή είναι μεγαλύτερη από ένα καθολικό κατώφλι. Σε αυτή την περίπτωση, κρατώνται τα σημεία πάνω στη γραμμή των οποίων η ένταση ξεπερνάει ένα κατώφλι συγκεκριμένο για τη γραμμή και τελικά το σημείο που είναι κοντύτερα στο άνω σύνορο της φωνητικής οδού θεωρείται ότι είναι το ζητούμενο, βλ. Σχήμα 5.19 για ένα παράδειγμα. Η μέθοδος είναι αρκετά εύρωστη αλλά δεν είναι ακριβής στον επιθυμητό βαθμό. Για το λόγο αυτό, στη συνέχεια, θέλοντας να αποφύγουμε μια επιπλέον ενδεχόμενη αιτία ανακριβειών, χρησιμοποιήθηκαν οι καμπύλες της γλώσσας όπως αυτές εξάχθηκαν από τις εικόνες των υπερήχων με ημιαυτόματο τρόπο [1]. Τα σημεία τομής των καμπυλών με το ημιπολικό πλέγμα προσδιορίστηκαν όπως και στην περίπτωση των εικόνων ακτίνων-Χ. Εκφράζονται στο σύστημα συντεταγμένων του ημιπολικού πλέγματος.

Στο τελευταίο στάδιο εξαγωγής των παραμέτρων περιγραφής της φωνητικής οδού, το μοντέλο άρθρωσης προσαρμόζεται ώστε να ταιριάζει καλύτερα στις συντεταγμένες των σημείων τομής της γλώσσας με το ημιπολικό πλέγμα πάνω στο επίπεδο των εικόνων υπερήχων. Με βάση τους συμβολισμούς που εισήχθησαν στην Ενότητα 5.8.2, το ζητούμενο είναι να προσδιορίσουμε ένα διάνυσμα παραμέτρων $\mathbf{x}$ του μοντέλου για κάθε διάνυσμα παρατηρήσεων $\mathbf{y}$. Αυτό λαμβάνεται ως το σημείο όπου μεγιστοποιείται η εκ των υστέρων κατανομή $p(\mathbf{x}|\mathbf{y})$:

$$p(\mathbf{x}|\mathbf{y}) = N(\mathbf{x}|M^{-1}A^T(\mathbf{y}-\boldsymbol{\mu}), \sigma^{-2}M)$$

με

$$M = A^T A + \sigma^2 I.$$

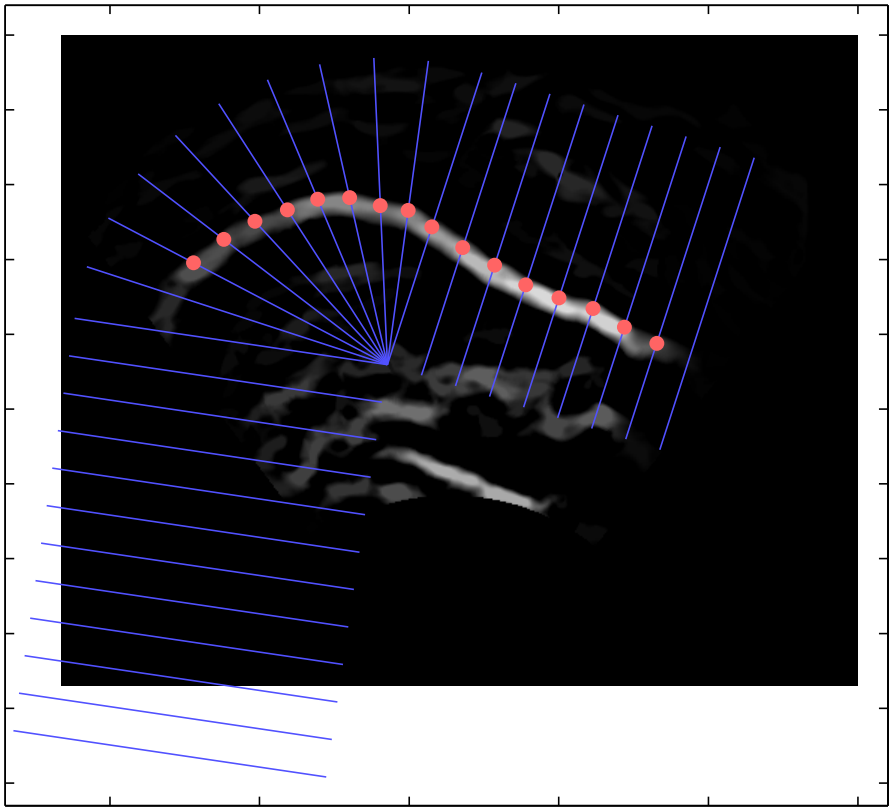
Δηλαδή το εκτιμώμενο διάνυσμα παραμέτρων είναι

$$\hat{\mathbf{x}} = M^{-1}A^T(\mathbf{y}-\boldsymbol{\mu}).$$

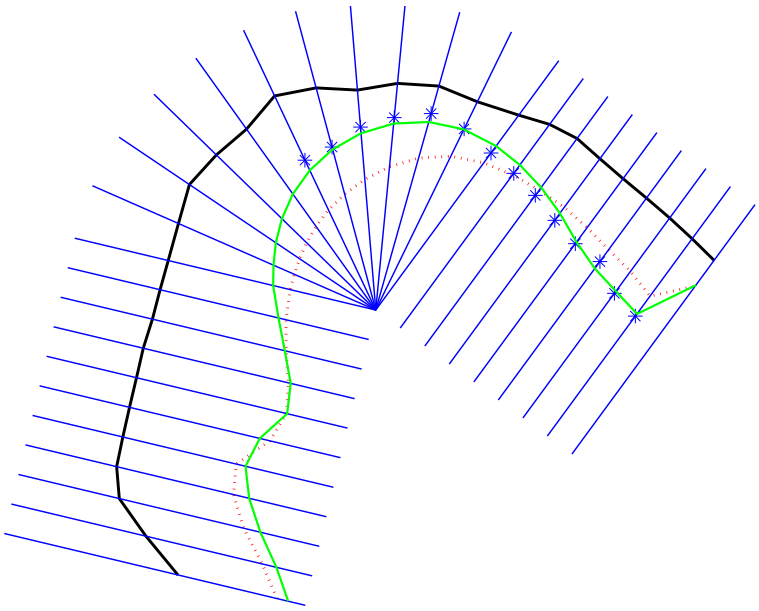
Ένα παράδειγμα εφαρμογής της διαδικασίας προσαρμογής φαίνεται στο Σχήμα 5.20.

Εναλλακτικά, τα πρόβληματα ανίχνευσης της γλώσσας στις εικόνες των υπερήχων και ταιριάσματος του αρθρωτικού μοντέλου στα σημεία τομής με το πλέγμα που παρουσιάστηκαν μπορούν να θεωρηθούν ενιαία στο πλαίσιο των ενεργών μοντέλων εμφάνισης για τη γλώσσα. Τα πρώτα σχετικά αποτελέσματα είναι αρκετά ενθαρρυντικά [139].

Στο Σχήμα 5.21 παρουσιάζονται τα αποτελέσματα της προσαρμογής του μοντέλου για τα δεδομένα υπερήχων που αντιστοιχούν στην εκφώνηση της ακολουθίας /Aku/. Με τους αστερίσκους σημειώνονται τα σημεία τομής του πλέγματος με τις καμπύλες της γλώσσας όπως έχουν επισημειωθεί ημιαυτόματα. Στα σημεία αυτά προσαρμόζεται το αρθρωτικό μοντέλο και



Σχήμα 5.19: *Εξαγωγή των σημείων της γλώσσας (κόκκινες τελείες) πάνω στο πλέγμα της φωνητικής οδού (μπλε γραμμές), για την ίδια χρονική στιγμή όπως στο Σχήμα 5.17. Χρησιμοποιείται το αντίστοιχο προεπεξεργασμένο πλαίσιο των δεδομένων υπερήχων.*



Σχήμα 5.20: *Προσαρμογή του μοντέλου άρθρωσης στα σημεία της γλώσσας '∗' που έχουν προσδιοριστεί από μια εικόνα υπερήχων. Η συνεχής πράσινη γραμμή αντιστοιχεί στο προσαρμοσμένο μοντέλο ενώ η διακεκομμένη κόκκινη γραμμή είναι το μέσο σχήμα.*

στη συνέχεια ανακατασκευάζεται το σχήμα της φωνητικής οδού που στο Σχήμα 5.21 δίνεται με τη συνεχή (μπλε) γραμμή μεγάλου πάχους. Για επαλήθευση υπερτίθεται το σχήμα της φωνητικής οδού όπως έχει σημειωθεί πάνω σε δεδομένα ακτίνων-Χ για την ίδια ακολουθία. Με διακεκομμένη γραμμή δίνεται το ανακατασκευασμένο σχήμα όπως προκύπτει από το προσαρμοσμένο μοντέλο στα δεδομένα αυτά. Δεδομένου του ότι οι αρχικές εκφωνήσεις στις δύο σειρές δεδομένων δεν είχαν την ίδια διάρκεια, ο συγχρονισμός έχει γίνει ευριστικά.

5.8.4 Αντιστροφή φωνής

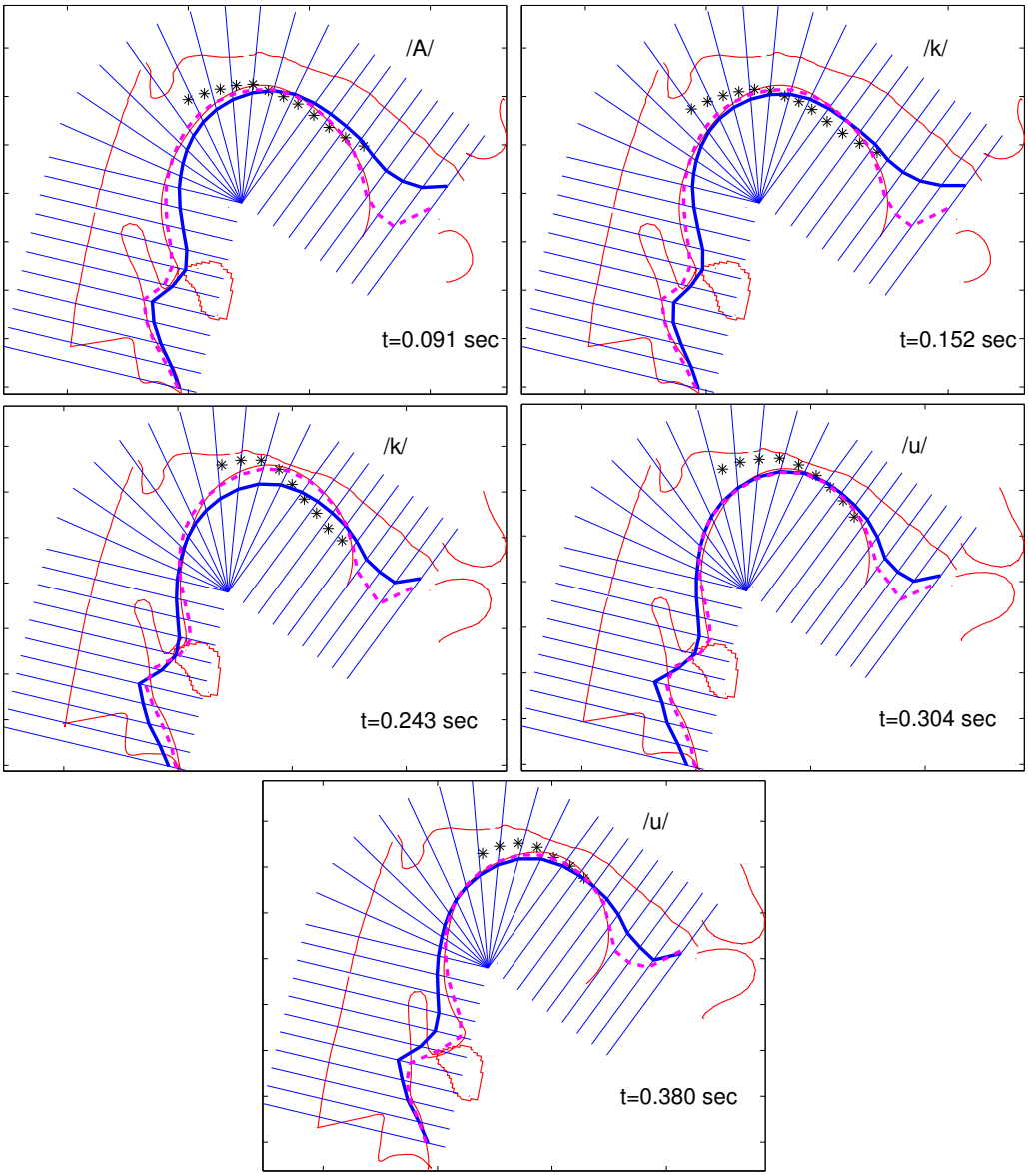
Μετά την εξαγωγή των παραμέτρων άρθρωσης για όλο το σύνολο των δεδομένων, ως αποτέλεσμα της διαδικασίας προσαρμογής του μοντέλου, εκπαιδεύεται μια απεικόνιση από τις οπτικοακουστικές παρατηρήσεις στις παραμέτρους άρθρωσης, όπως περιγράφεται στην Ενότητα 5.3. Η ακουστική πληροφορία αναπαρίσταται μέσω δεκαέξι Mel συντελεστών cepstrum ενώ η οπτική πληροφορία δίνεται με τη μορφή των τρισδιάστατων συντεταγμένων των οχτώ σημαδευτών πάνω στα χείλια του ομιλητή. Εκπαιδεύονται πολυκαναλικά κρυφά Μαρκοβιανά μοντέλα, ένα για κάθε ακουστικό φώνημα. Σε κάθε κατάσταση προσδιορίζεται μια γραμμική απεικόνιση μεταξύ των οπτικοακουστικών και των παραμέτρων άρθρωσης. Για την αντιστροφή, με βάση την οπτικοακουστική πληροφορία βρίσκεται πρώτα η βέλτιστη ακολουθία κρυφών καταστάσεων μέσω του αλγορίθμου Viterbi. Η υποκείμενη ακολουθία των παραμέτρων άρθρωσης υπολογίζεται ως αποτέλεσμα μεγιστοποίησης της εκ των υστέρων πιθανότητας, βλ. Ενότητα 5.2.

Στο Σχήμα 5.22 δίνονται τα σχήματα της φωνητικής οδού για τα φωνήματα /ι/, /ου/, /α/, /ρ/ και /ο/ όπως έχουν προκύψει από την αντιστροφή μαζί με τα αντίστοιχα σχήματα αναφοράς, όπως έχουν εκτιμηθεί με τη διαδικασία προσαρμογής που περιγράφεται στην Ενότητα 5.8.3.

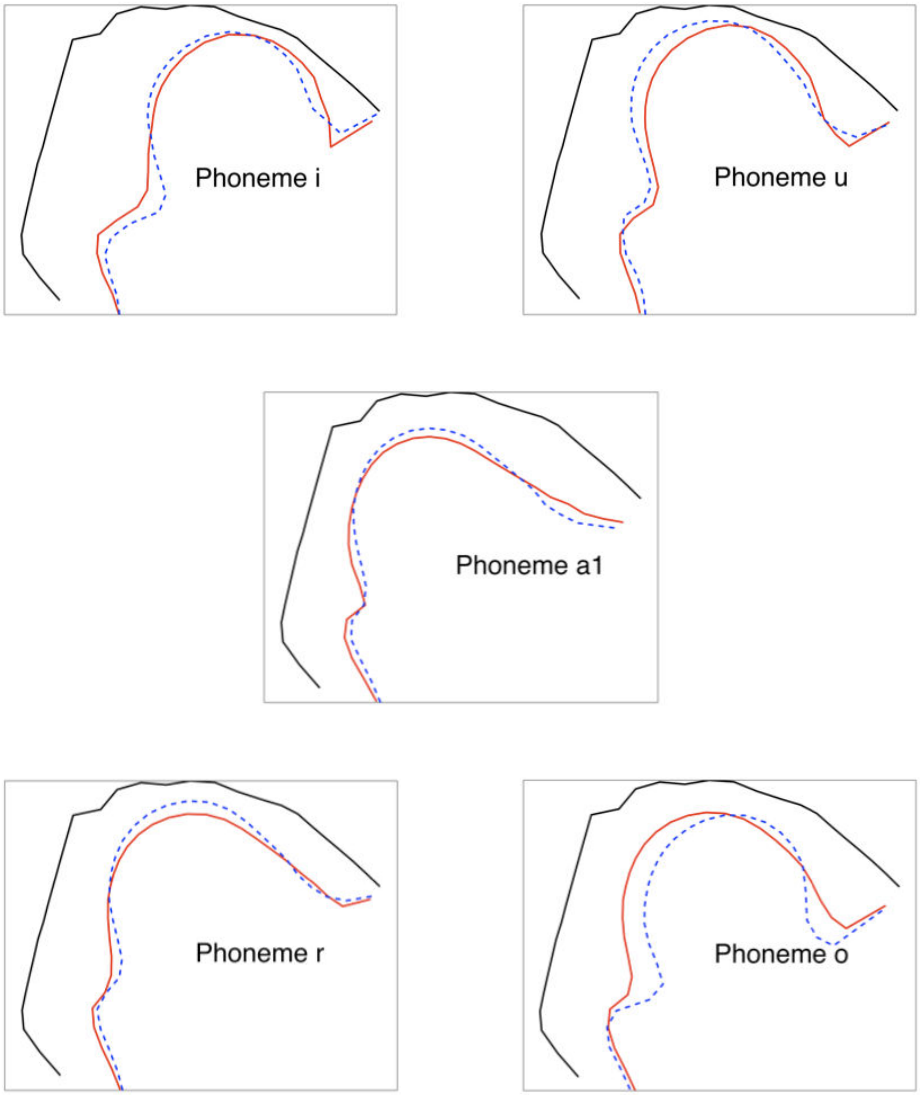
5.9 Συζήτηση

Παρουσιάστηκε ένα στατιστικό πλαίσιο βασισμένο σε κρυφά Μαρκοβιανά μοντέλα και ανάλυση κανονικής συσχέτισης για να προσδιορίστουν χαρακτηριστικά του συστήματος παραγωγής φωνής με βάση πληροφορία προερχόμενη από το ακουστικό σήμα φωνής και από το πρόσωπο του ομιλητή. Τα πειράματα πραγματοποιήθηκαν στις βάσεις MOCHA και QSMT. Η ανάλυση του προσώπου του ομιλητή πραγματοποείται μέσω ενεργής μοντελοποίησης εμφάνισης. Με αυτόν τον τρόπο είναι δυνατόν να χρησιμοποιηθεί οπτική πληροφορία χωρίς ιδιαίτερο εξοπλισμό για ανίχνευση κίνησης που θα απαιτούσε για παράδειγμα την επικόλληση ειδικών σημαδευτών πάνω στο πρόσωπο του ομιλητή. Τα πειράματα που αφορούν σε εναλλακτικές μοντελοποίησης και σύμμειξης δείχνουν ότι μοντελοποιώντας το οπτικό κανάλι σε επίπεδο οπτικού φωνήματος μπορεί να βελτιώσει την απόδοση και ότι τα σχήματα μέσης και εκ των υστέρων σύμμειξης είναι καταλληλότερα για την οπτικοακουστική αντιστροφή φωνής από την πρώιμη σύμμειξη.

Μέσω της προτεινόμενης οπτικοακουστικής μοντελοποίησης της φωνής γίνεται τελικά δυνατή η ταυτοποίηση του υποκείμενου φωνητικού συστήματος μόνο με βάση το παρατηρούμενο σήμα. Το εν λόγω πλαίσιο επεκτάθηκε κατάλληλα ώστε να αξιοποιήσει πολυτροπικά δεδομένα άρθρωσης. Η γεωμετρία της φωνητικής οδού περιγράφεται με μοντέλο άρθρωσης που εκπαιδεύεται σε δεδομένα ακτίνων-Χ της φωνητικής οδού. Για την ενσωμάτωση δυναμικής πληροφορίας γίνεται στη συνέχεια κατάλληλο ταίριασμα του μοντέλου σε δεδομένα υπερήχων της στοματικής κοιλότητας. Τελικά από τα παρατηρούμενα στερεο-οπτικά και ακουστικά δεδομένα γίνεται δυνατός ο προσδιορισμός ολόκληρου του μεσοβελιαίου σχήματος της φωνητικής οδού οπότε με κατάλληλο μοντέλο για τον προσδιορισμό της συνάρτησης επιφανείας μπορεί να πραγματοποιηθεί η επανασύνθεση του παρατηρούμενου σήματος φωνής, βλ. Ενότητα 3.8.



Σχήμα 5.21: *Υπέρθεση ανακατασκευασμένων σχημάτων της φωνητικής οδού με βάση προσαρμοσμένα μοντέλα σε δεδομένα ακτίνων - X και υπερήχων για την ακολουθία φωνημάτων /Aku/. Με τη διακεκομμένη γραμμή είναι το σχήμα όπως προκύπτει από τις ακτίνες - X ενώ με τη συνεχή γραμμή είναι όπως προκύπτει από τους υπερήχους (μετά από τη διαδικασία προσαρμογής του μοντέλου άρθρωσης και ανακατασκευής). Οι αστερίσκοι αντιστοιχούν στα σημεία τομής της καμπύλης της γλώσσας στους υπερήχους με το πλέγμα. Με τη λεπτή συνεχή γραμμή δίνεται το σχήμα της φωνητικής οδού όπως είναι σημειωμένο πάνω στα δεδομένα ακτίνων - X.*



Σχήμα 5.22: *Σχήματα φωνητικής οδού όπως προκύπτουν μετά την αντιστροφή των φωνημάτων /ι/, /ου/, /α/, /ρ/ και /ο/. Τα αποτελέσματα δίνονται με συνεχή κόκκινη γραμμή. Τα σχήματα αναφοράς δίνονται με διακεκομμένη μπλε γραμμή. Με συνεχή μαύρη γραμμή αναπαρίσταται το σταθερό εξωτερικό τοίχωμα της φωνητικής οδού.*

Κεφάλαιο 6

Συμπεράσματα και Κατευθύνσεις για Μελλοντική Έρευνα

Το αντικείμενο της παρούσας διδακτορικής διατριβής θα μπορούσε να συνοψιστεί ως η ανάπτυξη ενός υπολογιστικού μοντέλου φωνής με τη δυνατότητα προσομοίωσης του ανθρώπινου φωνητικού συστήματος και αξιοποίησης σημαντικών σχετικών αεροδυναμικών φαινομένων και η εφαρμογή του για οπτικοακουστική αντιστροφή φωνής σε ένα πολυτροπικό στοχαστικό πλαίσιο. Το προτεινόμενο μοντέλο προσομοιώνει το ανθρώπινο φωνητικό σύστημα τόσο στο φυσικό επίπεδο όσο και στο επίπεδο άρθρωσης (βλ. Κεφάλαιο 1).

6.1 Συνεισφορές - Συμπεράσματα

Οι επιμέρους συνεισφορές της διδακτορικής διατριβής είναι σε δύο βασικούς άξονες κατ' αναλογία με τα επίπεδα που προαναφέρθηκαν. Συνοψίζονται και τα σχετικά συμπεράσματα.

6.1.1 Φυσική μοντελοποίηση της φωνητικής οδού

Η ακριβέστερη φυσική μοντελοποίηση της φωνητικής οδού χωρίς σημαντική υπολογιστική επιβάρυνση ήταν ένας από τους βασικούς στόχους της διδακτορικής διατριβής. Τα επιμέρους επιτεύγματα σε αυτή την κατεύθυνση περιλαμβάνουν :

Βελτιωμένη ακουστική μοντελοποίηση Με αποδοτική αριθμητική προσομοίωση των ακουστικών εξισώσεων μέσα στο φωνητικό σωλήνα υλοποιήθηκε συνθέτης φωνής με δυνατότητα σύνθεσης ακολουθιών φωνηέντων της μορφής Φωνήεν-Σύμφωνο-Φωνήεν. Προσομοιώνεται η δόνηση των τοιχωμάτων, η σύζευξη με επιμέρους ακουστικές κοιλότητες όπως είναι η ρινική και οι piriform fossae καθώς και η ζεύξη του μοντέλου των δύο μαζών για τη γλωττίδα ώστε να επιτρέπεται η αλληλεπίδραση πηγής - φωνητικής οδού. Υπήρξε μέριμνα για τις επιπτώσεις στην ακουστική διάδοση της ύπαρξης μέσου πεδίου αερορροής μέσα στη φωνητική οδό. Για τα πειράματα σύνθεσης χρησιμοποιήθηκαν πραγματικά γεωμετρικά δεδομένα που έχουν καταγραφεί είτε μέσω αξονικής τομογραφίας είτε μέσω ακτίνων-Χ. Η σύγκριση με αντίστοιχα πραγματικά σήματα φωνής φανερώνει την αποτελεσματικότητα του προτεινόμενου μοντέλου.

Αεροδυναμική-αεροακουστική μοντελοποίηση για σύνθεση φωνής Για πρώτη φορά ενσωματώθηκαν σε έναν συνθέτη φωνής εξελιγμένα μοντέλα αεροδυναμικών-αεροακουστικών φαινομένων που εμφανίζονται κατά την παραγωγή τόσο άφωνων όσο και έμφωνων ήχων. Το αεροδυναμικό μοντέλο περιγράφει βασικές ιδιότητες τόσο της στροβιλώδους όσο και της αστρόβιλης συνιστώσας του πραγματικού πεδίου ροής. Για τη γλωττίδα εισάγεται ένα μηχανικό μοντέλο δύο μαζών και ανάλογη αεροδυναμική περιγραφή που συνδυάζει χαρακτηριστικά προηγούμενων μοντέλων με το σημαντικότερο

ίσως να είναι το ότι επιτρέπει τη μετακίνηση του σημείου αποκόλλησης της αερορροής. Επιτυγχάνεται κατάλληλη σύζευξη του μοντέλου της γλωττίδας με το αεροδυναμικό μοντέλο και γίνεται με αυτόν τον τρόπο τελικά δυνατός ο προσδιορισμός των πηγών ήχου τόσο στη γλωττίδα όσο και σε στενώσεις της φωνητικής οδού. Αξιοποιούνται σύγχρονα συμπεράσματα της αεροακουστικής θεωρίας. Οι δυνατότητες του προτεινόμενου συστήματος επιδεικνύονται μέσω επιτυχούς σύνθεσης διαφόρων ακολουθιών φωνημάτων.

6.1.2 Οπτικοακουστική αντιστροφή φωνής με πολυτροπικά δεδομένα

Ο κατάλληλος έλεγχος του φυσικού επιπέδου του μοντέλου ώστε να είναι δυνατή η μίμηση του ανθρώπινου φωνητικού συστήματος ήταν ο δεύτερος στόχος της διδακτορικής διατριβής. Τα επιμέρους επιτεύγματα σε αυτή την κατεύθυνση περιλαμβάνουν :

Οπτικοακουστική αντιστροφή φωνής Με κατάλληλη κατά τμήματα γραμμική προσέγγιση μοντελοποιήθηκε η σχέση μεταξύ οπτικοακουστικής πληροφορίας και πληροφορίας άρθρωσης. Η οπτική πληροφορία αναπαρίσταται μέσω ενεργών μοντέλων εμφάνισης. Η προσέγγιση υλοποιήθηκε με κατάλληλη εφαρμογή ανάλυσης κανονικής συσχέτισης, πλαισίου κρυφών Μαρκοβιανών μοντέλων και φίλτρων Kalman. Η αξιοποίηση της οπτικής πληροφορίας αποδείχτηκε ιδιαίτερα επωφελής για την αντιστροφή. Διερευνήθηκαν εναλλακτικές δυνατότητες σύμμειξης της πληροφορίας σε διαφορετικά επίπεδα συγχρονισμού. Η εκ των υστέρων σύμμειξη ήταν η πλέον αποτελεσματική επιτρέποντας πιο ευέλικτη μοντελοποίηση των δύο ροών πληροφορίας.

Αντιστροφή φωνής με αξιοποίηση πολυτροπικών δεδομένων Με την αξιοποίηση πολυτροπικών δεδομένων άρθρωσης (εικόνες της φωνητικής οδού με ακτίνες Χ, μαγνητική τομογραφία και υπερήχους) αναπτύχθηκε αρθρωτικό μοντέλο της φωνητικής οδού με βάση το οποίο ήταν δυνατή στη συνέχεια η εφαρμογή ένος πλαισίου μηχανικής μάθησης για την αντιστροφή της φωνής από οπτικοακουστικά δεδομένα. Το εν λόγω πλαίσιο διευρύνει το πεδίο εφαρμογής των τεχνικών αντιστροφής και επιτρέπει την αξιοποίηση πληθώρας πραγματικών αρθρωτικών δεδομένων.

Επίσης, με το προτεινόμενο πλαίσιο γίνεται δυνατή η αποτελεσματικότερη αξιολόγηση τόσο της διαδικασίας αντιστροφής της φωνής όσο και της αεροδυναμικής - αεροακουστικής - ακουστικής μοντελοποίησης.

6.2 Μελλοντικές ερευνητικές κατευθύνσεις

Ήδη βρίσκονται σε εξέλιξη διάφορες ερευνητικές προσπάθειες για πιθάνες βελτιώσεις του προτεινόμενου μοντέλου που αφορούν τόσο στο επίπεδο της φυσικής μοντελοποίησης όσο και στο επίπεδο της μοντελοποίησης της άρθρωσης.

Πιο συγκεκριμένα, η γενική αεροδυναμική μοντελοποίηση που θα καλύπτει καταστάσεις της φωνητικής οδού όπως στην παραγωγή γέλιου ή τραγουδιού ή και διαφόρων άλλων ήχων είναι ακόμα ανοιχτό πρόβλημα. Ακόμα και στην περίπτωση των ήχων που μελετήθηκαν βέβαια πρέπει να αξιολογηθεί πιο συστηματικά η απλουστευμένη αεροδυναμική μοντελοποίηση σε σχέση με την πλήρη περιγραφή του τρισδιάστατου πεδίου ροής. Όσον αφορά στην αεροακουστική μοντελοποίηση, είναι σημαντικό να εκτιμηθούν οι ενδεχόμενες επιδράσεις του ακουστικού πεδίου στην πηγή που κατά περιπτώσεις μπορεί να είναι σημαντικές [71,91]. Σε σχέση με τη σύνθεση φωνής από πραγματικά δεδομένα στην παρούσα φάση γίνεται προσπάθεια να εφαρμοστεί ο συνθέτης φωνής σε δεδομένα από εικόνες υπερήχων ακολουθώντας τις τεχνικές που περιγράφτηκαν στη διατριβή. Σε αυτή την κατεύθυνση, θα πρέπει να βελτιωθούν οι τεχνικές αντιστοίχισης του αρθρωτικού μοντέλου στα δεδομένα υπερήχων. Αυτό εκ των πραγμάτων απαιτεί και πιο αποτελεσματική ιχνηλάτηση της γλώσσας [139].

Επίσης, επιδιώκεται ο ρόλος του συνθέτη φωνής να γίνει αναδραστικός στο προτεινόμενο μοντέλο. Το συνθετικό σήμα θα συγκρίνεται με το σήμα εισόδου και το αποτέλεσμα της αντιστροφής θα βελτιώνεται επαναληπτικά. Ως προηγούμενες σχετικές δουλειές σε αυτό το σημείο θα μπορούσαν ενδεχόμενα να αναφερθούν τα [60], [110], [67], ενώ θα ήταν ενδεχόμενα σχετική και η αναφορά αναπαραστάσεων για τη φωνή που επιτρέπουν καλύτερο διαχωρισμό των διαφόρων επιδράσεων στο ακουστικό φάσμα όπως είναι αυτή που παρουσιάζεται στο [86]. Σημαντική επίσης όσον αφορά στην αντιστροφή φωνής είναι και η προσπάθεια ώστε να απλοποιηθεί και να αξιολογηθεί η διαδικασία αντιστοίχισης (registration) του αρθρωτικού μοντέλου στα δεδομένα υπερήχων.

Ένα επίσης σημαντικό θέμα που αναδεικνύεται είναι η προσαρμογή του προτεινόμενου μοντέλου σε έναν διαφορετικό ομιλητή ώστε να ενισχυθεί η χρησιμότητα του και να γίνει πιθανόν δυνατή η χρησιμοποίησή του σε πρακτικές εφαρμογές, όπως είναι η εκμάθηση ξένης γλώσσας. Τεχνικές που ήδη εφαρμόζονται στην αναγνώριση φωνής όπως είναι η προσαρμογή σε ομιλητή μέγιστης πιθανοφάνειας θα μπορούσαν να είναι χρήσιμες γι' αυτό το σκοπό [64]. Τέλος, πολλά υποσχόμενη εμφανίζεται και η προοπτική εφαρμογής του προτεινόμενου υπολογιστικού μοντέλου που συνδυάζει οπτικοακουστική και αρθρωτική πληροφορία για οπτικοακουστική αναγνώριση φωνής [87, 123].

Κατάλογος δημοσιεύσεων του συγγραφέα

Δημοσιεύσεις σε περιοδικά

1. A. Katsamanis, G. Papandreou and P. Maragos, Face Active Appearance Modeling and Speech Acoustic Information to Recover Articulation, IEEE Transactions on Audio, Speech and Language Processing, vol. 17, no. 3, Mar. 2009.

2. G. Papandreou, A. Katsamanis, V. Pitsikalis, and P. Maragos, Adaptive Multimodal Fusion by Uncertainty Compensation with Application to Audio-Visual Speech Recognition, IEEE Transactions on Audio, Speech and Language Processing, vol. 17, no. 3, Mar. 2009.

Δημοσιεύσεις σε διεθνή συνέδρια με κριτή

1. A. Roussos, A. Katsamanis, and P. Maragos, Tongue Tracking in Ultrasound Images with Active Appearance Models, Proc. IEEE Int'l Conf. on Image Processing (ICIP-09), Cairo, Egypt, Nov. 7-11, 2009.

2. S. Theodorakis, A. Katsamanis, and P. Maragos, Product-HMMs for automatic sign language recognition, Proc. IEEE Int'l Conference on Acoustics, Speech, and Signal Processing (ICASSP-2009), Taipei, Taiwan, April 2009.

3. A. Katsamanis, T. Roussos, P. Maragos, M. Aron, and M.-O. Berger, Inversion from Audiovisual Speech to Articulatory Information by Exploiting Multimodal Data, Proc. International Seminar on Speech Production (ISSP 2008), Strasbourg, France, Dec. 2008.

4. A. Katsamanis, G. Ananthakrishnan, G. Papandreou, P. Maragos, and O. Engwall, Audiovisual Speech Inversion by Switching Dynamical Modeling Governed by a Hidden Markov Process, Proc. European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, Aug. 2008.

5. A. Katsamanis, G. Papandreou, and P. Maragos, Audiovisual-to-Articulatory Speech Inversion Using Active Appearance Models for the Face and Hidden Markov Models for the Dynamics, Proc. IEEE Int'l Conference on Acoustics, Speech, and Signal Processing (ICASSP-2008), Las Vegas, NV, U.S.A., Mar.-Apr. 2008.

6. S. Lefkimmiatis, P. Maragos, and A. Katsamanis, Multisensor Multiband Cross-Energy Tracking for Feature Extraction and Recognition, Proc. IEEE Int'l Conference on Acoustics, Speech, and Signal Processing (ICASSP-2008), Las Vegas, NV, U.S.A., Mar.-Apr. 2008.

7. G. Papandreou, A. Katsamanis, V. Pitsikalis, and P. Maragos, Multimodal Fusion and Learning with Uncertain Features Applied to Audiovisual Speech Recognition, Proc. IEEE Workshop on Multimedia Signal Processing (MMSP-2007), pp. 264-267, Chania, Greece, October 1-3, 2007.

8. A. Katsamanis, G. Papandreou, and P. Maragos, Audiovisual-to-Articulatory Inversion Using Hidden Markov Models, Proc. IEEE Workshop on Multimedia Signal Processing (MMSP-2007), pp. 457-460, Chania, Greece, October 1-3, 2007.

9. A. Katsamanis, P. Tsiakoulis, P. Maragos, and A. Potamianos, Investigations in Articulatory Synthesis, Proc. 16th International Congress of Phonetic Sciences (ICPhS-2007), pp. 877-880, Saarbruecken, Germany, August 6-10, 2007.

10. V. Pitsikalis, A. Katsamanis, G. Papandreou, and P. Maragos, Adaptive Multimodal Fusion by Uncertainty Compensation, Proc. Int'l Conference on Spoken Language Processing (ICSLP-2006), pp. 2458-2461, Pittsburgh PA, USA, Sep. 17-21, 2006.

11. A. Katsamanis, G. Papandreou, V. Pitsikalis, and P. Maragos, Multimodal Fusion by Adaptive Compensation for Feature Uncertainty with Application to Audiovisual Speech Recognition, Proc. 14th European Signal Processing Conference (EUSIPCO-2006), Florence, Italy, Sept. 4-8 2006.

12. A. Katsamanis and P. Maragos, Advances in Statistical Estimation and Tracking of AM-FM Speech Components, Proc. Interspeech 2005 - Eurospeech – 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, September 2005.

Κεφάλαια σε βιβλία

1. G. Papandreou, A. Katsamanis, V. Pitsikalis and P. Maragos, Adaptive Multimodal Fusion by Uncertainty Compensation with Application to Audio-Visual Speech Recognition, in Multimodal Processing and Interaction: Audio, Video, Text, edited by P. Maragos, A. Potamianos, and P. Gros, Springer-Verlag, New York, 2008.

2. P. Maragos, P. Gros, A. Katsamanis and G. Papandreou, Cross-Modal Integration for Performance Improving in Multimedia: A Review, in Multimodal Processing and Interaction: Audio, Video, Text, edited by P. Maragos, A. Potamianos, and P. Gros, Springer-Verlag, New York, 2008.

Βιβλιογραφία

[1] Alipour, F., C. Fan, and R.C. Scherer: *A numerical simulation of laryngeal flow in a forced-oscillation glottal model*. Computer Speech and Language, 10:75–93, 1996.

[2] Alipour, F. and R.C. Scherer: *Pulsatile airflow during phonation: An excised larynx model*. J. of the Acous. Soc. Am., 97:1241–1248, 1995.

[3] Alipour, F. and R.C. Scherer: *Flow separation in a computational oscillating vocal fold model*. J. of the Acous. Soc. Am., 116:1710–1719, 2004.

[4] Ananthapadmanabha, T. and G. Fant: *Calculation of true glottal flow and its components*. Speech Communication, 1:167–184, 1982.

[5] Anderson, B.D.O. and J.B. Moore: *Optimal Filtering*. Prentice-Hall, 1979.

[6] Anderson, J.D.: *Fundamentals of Aerodynamics*. McGraw-Hill, 1984.

[7] Arfken, G.B. and H.J. Weber: *Mathematical Methods for Physicists*. Acad. Press, 2000.

[8] Aron, M., A. Roussos, M.O. Berger, E. Kerrien, and P. Maragos: *Multimodality acquisition of articulatory data and processing*. In *Proc. European Signal Processing Conference*, 2008.

[9] Aron, M., A. Toutios, M.O. Berger, E. Kerrien, B. Wrobel-Dautcourt, and Y. Laprie: *Registration of multimodal data for estimating the parameters of an articulatory model*. In *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 2009.

[10] Badin, P.: *Acoustics of voiceless fricatives: Production theory and data*. QPSR Speech Transmission Laboratory Quarterly Progress and Status Report, 3:033–055, 1989.

[11] Badin, P., G. Bailly, L. Revéret, M. Baciu, C. Segebarth, and C. Savariaux: *Three-dimensional linear articulatory modeling of tongue, lips and face based on MR and video images*. Journal of Phonetics, 30:533–553, 2002.

[12] Badin, P., D. Beautemps, R. Laboissiére, and J.L. Schwartz: *Recovery of vocal tract geometry from formants for vowels and fricative consonants using a midsagittal-to-area function conversion model*. Journal of Phonetics, 23:221–229, 1995.

[13] Badin, P. and G. Fant: *Notes on vocal tract computation*. QPSR Speech Transmission Laboratory Quarterly Progress and Status Report, pp. 53–108, 1984.

[14] Badin, P. and A. Serrurier: *Three-dimensional linear modeling of tongue: Articulatory data and models*. In *Proc. Int'l Seminar on Speech Production*, 2006.

[15] Bae, Y. and Y.J. Moon: *Computation of phonation aeroacoustics by an INS/PCE splitting method*. Computers & Fluids, 37:1332–1343, 2007.

[16] Baer, T., J.C. Gore, L.C. Gracco, and P.W. Nye: *Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels*. J. of the Acous. Soc. Am., 90:799–828, 1991.

[17] Bailly, G. and P. Badin: *Seeing tongue movements from outside*. In *Proc. Int'l Conf. on Spoken Language Processing*, 2002.

[18] Bailly, G., M. Bérar, F. Elisei, and M. Odisio: *Audiovisual speech synthesis*. International Journal of Speech Technology, 6(4):331–346, October 2003.

[19] Barney, A., C.H. Shadle, and P. Davies: *Fluid flow in a dynamic mechanical model of the vocal folds and tract. I. measurements and theory*. J. of the Acous. Soc. Am., 105:444–455, 1999.

[20] Batchelor, G.K.: *An Introduction to Fluid Dynamics*. Cambridge Mathematical Library, 2002.

[21] Beautemps, D., P. Badin, and R. Laboissiére: *Deriving vocal-tract area functions from midsagittal profiles and formant frequencies: A new model for vowels and fricative consonants based on experimental data*. Speech Communication, 16:27–47, 1995.

[22] Birkholz, P. and D. Jackèl: *Simulation of flow and acoustics in the vocal tract*. In *Proceedings CFA/DAGA, Strasbourg*, pp. 895–896, 2004.

[23] Birkholz, P., D. Jackèl, and B.J. Kröger: *Construction and control of a three-dimensional vocal tract model*. In *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 2006.

[24] Birkholz, P., D. Jackèl und B.J. Kröger: *Simulation of Losses due to Turbulence in the time-varying vocal tract system*. IEEE Trans. Audio, Speech, and Language Processing, 15:1218–1226, 2007.

[25] Bishop, C.: *Pattern Recognition and Machine Learning*. Springer, 2006.

[26] Boersma, P.: *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. PhD thesis, University of Amsterdam, 1998.

[27] Breiman, L. and J.H. Friedman: *Predicting multivariate responses in multiple linear regression*. J. of Royal Stat. Soc. (Series B), 59(1):3–54, 1997.

[28] Bresch, E., J. Nielsen, K. Nayak, and S. Narayanan: *Synchronized and noise-robust audio recordings during real-time magnetic resonance imaging scans*. J. of the Acous. Soc. Am., 120:1791–1794, 2006.

[29] Chen, T.: *Audiovisual speech processing*. IEEE Signal Process. Mag., 18:9–21, 2001.

[30] Chi, X. and M. Sonderegger: *Subglottal coupling and its influence on vowel formants*. J. of the Acous. Soc. Am., 122:1735–1745, 2007.

[31] Chiba, T. and M. Kajiyama: *The vowel - its nature and structure*. Kaseikan Pub. Co., 1941.

[32] Choi, K., Y. Luo, and J.N. Hwang: *Hidden Markov model inversion for audio-to-visual conversion in an MPEG-4 facial animation systems*. Journal of VLSI Signal Processing, 29:51–61, 2001.

[33] Cohen, L.: *Time-frequency analysis: theory and applications*. Prentice Hall, 1995.

[34] Conte, S. and C.d. Boor: *Elementary Numerical Analysis*. McGraw-Hill, 1980.

[35] Cootes, T.F., G.J. Edwards, and C.J. Taylor: *Active appearance models*. IEEE Trans. Pattern Anal. Mach. Intell., 23(6):681–685, 2001.

[36] Dang, J. and K. Honda: *Morphological and acoustical analysis of the nasal and paranasal cavities*. J. of the Acous. Soc. Am., 96:2088–2100, 1994.

[37] Dang, J. and K. Honda: *Acoustic characteristics of the piriform fossa in models and humans*. J. of the Acous. Soc. Am., 101:456–465, 1997.

[38] Davies, P.: *Practical flow duct acoustics*. Journal of Sound and Vibration, 124:91–115, 1988.

[39] Davies, P., R.S. McGowan, and C.H. Shadle: *Practical flow duct acoustics applied to the vocal tract*. In *Vocal Fold Physiology: Frontiers in Basic Science*. Singular, San Diego, 1993.

[40] DeSarbo, W. and W. Cron: *A maximum likelihood methodology for clusterwise linear regression*. Journal of Classification, 5:249–282, 1988.

[41] Deverge, M., X. Pelorson, C. Vilain, P.Y. Lagrée, F. Chentourf, J. Willems, and A. Hirschberg: *Influence of collision on the flow through in-vitro rigid models of the vocal folds*. J. of the Acous. Soc. Am., 114:3354–3362, 2003.

[42] Dimitriadis, D., P. Maragos, V. Pitsikalis, and A. Potamianos: *Modulation and chaotic acoustic features for speech recognition*. Journal on Intelligent Control Systems, 30:19–26, 2002.

[43] Doel, K. van den and U.M. Ascher: *Real-time numerical solution of Webster's equation on a non-uniform grid*. IEEE Trans. Speech and Audio Process., 16:1163–1172, 2008.

[44] Dupont, S. and J. Luettin: *Audio-visual speech modeling for continuous speech recognition*. IEEE Trans. Multimedia, 2(3):141–151, 2000.

[45] Dusan, S. and L. Deng: *Acoustic-to-articulatory inversion using dynamical and phonological constraints*. In *Proc. Int'l Seminar on Speech Production*, pp. 237–240, 2000.

[46] Einstein, A.: *The foundation of the general theory of relativity*. Annalen der Physik, 1916.

[47] El-Masri, S., X. Pelorson, P. Saguet, and P. Badin: *Development of the transmission line matrix method in acoustics applications to higher modes in the vocal tract and other complex ducts*. International Journal of Numerical Modelling: Electronic Networks, Devices and Fields, 11:133–151, 1998.

[48] Englebienne, G., T. Cootes, and M. Rattray: *A probabilistic model for generating realistic speech movements from speech*. In *Proc. Advances in Neural Information Processing Systems*, 2007.

[49] Engwall, O.: *Introducing visual cues in acoustic-to-articulatory inversion*. In *Proc. Int'l Conf. on Speech Communication and Technology*, pp. 3205–3208, 2005.

[50] Engwall, O., O. Bälter, A.M. Öster, and H. Sidenbladh-Kjellström: *Designing the user interface of the computer-based speech training system ARTUR based on early user tests*. Journal of Behaviour and Information Technology, 25(4):353–365, 2006.

[51] Engwall, O. and J. Beskow: *Resynthesis of 3D tongue movements from facial data*. In *Proc. European Conf. on Speech Communication and Technology*, 2003.

[52] Fant, G.: *Acoustic Theory of Speech Production*. Mouton, Gravenhage, 1960.

[53] Fant, G.: *Vocal tract wall effects, losses, and resonance bandwidths*. QPSR Speech Transmission Laboratory Quarterly Progress and Status Report, 13:28–52, 1972.

[54] Fitzgibbon, A., M. Pilu, and R. Fisher: *Direct least square fitting of ellipses*. IEEE Trans. Pattern Anal. Mach. Intell., 21(5):476–480, May 1999.

[55] Flanagan, J.L.: *Speech Analysis Synthesis and Perception*. Springer-Verlag, Berlin, 1972.

[56] Flanagan, J.L. and L. Cherry: *Excitation of vocal-tract synthesizers*. J. of the Acous. Soc. Am., 45:764–769, 1969.

[57] Flanagan, J.L. and K. Ishizaka: *Automatic generation of voiceless excitation in a vocal cord-vocal tract speech synthesizer*. IEEE Trans. Acoust., Speech, Signal Process., 24:163–170, 1976.

[58] Fontecave, J. and F. Berthommier: *A semi-automatic method for extracting vocal tract movements from x-ray films*. Speech Communication, 51:97–115, 2008.

[59] Ghahramani, Z. and G.E. Hinton: *Variational learning for switching state-space models*. Neural Computation, 12(4):831–864, 2000, ISSN 0899-7667.

[60] Gupta, S.K. and J. Schroeter: *Pitch-synchronous frame-by-frame and segment-based articulatory analysis by synthesis*. J. of the Acous. Soc. Am., 94(5):2517–2530, November 1993.

[61] Hazen, T.J.: *Visual model structures and synchrony constrants for audio-visual speech recognition*. IEEE Trans. Speech and Audio Process., 14:1082–1089, 2006.

[62] Heinz, J.M. and K.N. Stevens: *On the derivation of area functions and acoustic spectra from cinéradiographic films of speech*. J. of the Acous. Soc. Am., 36:1037–1038, 1964.

[63] Hiroya, S. and M. Honda: *Estimation of articulatory movements from speech acoustics using an HMM-based speech production models*. IEEE Trans. Speech and Audio Process., 12(2):175–185, March 2004.

[64] Hiroya, S. and T. Mochida: *Multi-speaker articulatory trajectory formation based on speaker-independent articulatory HMMs*. Speech Communication, 48:1677–1690, 2006.

[65] Hirschberg, A.: *Some fluid dynamic aspects of speech*. Bulletin de la Communication Parlée, 2:7–30, 1992.

[66] Honda, K. and S. Maeda: *Glottal-opening and airflow pattern during production of voiceless fricatives: a new non-invasive instrumentation*. In *Proc. Acoustics*, 2008.

[67] Hooke, R. and T. Jeeves: *"Direct search" solution of numerical and statistical problems*. Journal of the ACM, 8:212–229, 1961.

[68] Howe, M.: *Acoustics of Fluid-Structure Interactions*. Cambridge University Press, 1998.

[69] Howe, M.: *Theory of Vortex Sound*. Cambridge University Press, 2003.

[70] Howe, M. and R. McGowan: *Aeroacoustics of [s]*. Proceedings of the Royal Society A, 461:1005–1028, 2005.

[71] Howe, M. and R. McGowan: *Sound generated by aerodynamic sources near a deformable body, with application to voiced speech*. Journal of Fluid Mechanics, 592:367–392, 2007.

[72] Huang, J., S. Levinson, D. Davis, and S. Slimon: *Articulatory speech synthesis based upon fluid dynamic principles*. In *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 2002.

[73] Ishizaka, K. and J. Flanagan: *Synthesis of voiced sounds from a two-mass model of the vocal cords*. The Bell System Technical Journal, 51(6):1233–1268, July-August 1972.

[74] Ishizaka, K., J. French, and J.L. Flanagan: *Direct determination of vocal tract wall impedance*. IEEE Trans. Acoust., Speech, Signal Process., 23:370–373, 1975.

[75] Jackson, M.T., C. Espy-Wilson, and S. Boyce: *Verifying a vocal tract model with a closed side-branch*. J. of the Acous. Soc. Am., 109:2983–2987, 2001.

[76] Jiang, J., A. Alwan, P.A. Keating, E.T. Auer Jr., and L.E. Bernstein: *On the relationship between face movements, tongue movements, and speech acoustics*. EURASIP Journal on Applied Signal Processing, 11:1174–1188, 2002.

[77] Kaburagi, T.: *On the viscous-inviscid interaction of the flow passing through the glottis*. Acoustical Science and Technology, 29:167–175, 2009.

[78] Kaburagi, T. and Y. Tanabe: *Low-dimensional models of the glottal flow incorporating viscous-inviscid interaction*. J. of the Acous. Soc. Am., 125:391–404, 2009.

[79] Kaiser, J.F.: *Some observations on vocal tract operation from a fluid flow point of view*. In *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, 1983.

[80] Katsamanis, A., G. Ananthakrishnan, G. Papandreou, P. Maragos, and O. Engwall: *Audiovisual speech inversion by switching dynamical modeling governed by a hidden Markov process*. In *Proc. European Signal Processing Conference*, 2008.

[81] Katsamanis, A. and P. Maragos: *Advances in statistical estimation and tracking of AM-FM speech components,*. In *Proc. Int'l Conf. on Speech Communication and Technology*, 2005.

[82] Katsamanis, A., G. Papandreou, and P. Maragos: *Audiovisual-to-articulatory speech inversion using HMMs*. In *Proc. of IEEE Int'l Workshop on Multimedia Signal Processing*, 2007.

[83] Katsamanis, A., G. Papandreou, and P. Maragos: *Audiovisual-to-articulatory speech inversion using active appearance models for the face and hidden Markov models for the dynamics*. In *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 2008.

[84] Katsamanis, A., G. Papandreou, and P. Maragos: *Face active appearance modeling and speech acoustic information to recover articulation*. IEEE Trans. Audio, Speech, and Language Processing, 17:411–422, 2009.

[85] Katsamanis, A., P. Tsiakoulis, P. Maragos, and A. Potamianos: *Investigations in articulatory synthesis*. In *Proc. Int'l Congress on Phonetic Sciences*, 2007.

[86] Kawahara, H., I. Masuda-Katsuse, and A. de Cheveigné: *Restructuring speech representations using a pitch adaptive time-frequency smoothing and instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds*. Speech Communication, 27:187–207, 1999.

[87] King, S., J. Frankel, K. Livescu, E. McDermott, K. Richmond, and M. Wester: *Speech production knowledge in automatic speech recognition*. J. of the Acous. Soc. Am., 121(2):723–742, February 2007.

[88] Kjellstrom, H. and O. Engwall: *Audiovisual-to-articulatory inversion*. Speech Communication, 51:195–209, 2008.

[89] Kjellstrom, H., O. Engwall, and O. Bälter: *Reconstructing tongue movements from audio and video*. In *Proc. Int'l Conf. on Speech Communication and Technology*, pp. 2238–2241, 2006.

[90] Krane, M., M. Barry, and T. Wei: *Unsteady behavior of flow in a scaled-up vocal folds model*. J. of the Acous. Soc. Am., 122:3659–3670, 2007.

[91] Krane, M.H.: *Aeroacoustic production of low-frequency unvoiced speech sounds*. J. of the Acous. Soc. Am., 118(1):410–427, 2005.

[92] Krane, M.H. and T. Wei: *Theoretical assesment of unsteady aerodynamic effects in phonation*. J. of the Acous. Soc. Am., 120:1578–1588, 2006.

[93] Li, M., X. Khambhamettu, and M. Stone: *Automatic contour tracking in ultrasound images*. Clinical Linguistics and Phonetics, 6:545–554, 2005.

[94] Lighthill, M.: *On sound generated aerodynamically*. Proc. of the Royal Society of London, 1952.

[95] Liljencrants, J.: *Speech Synthesis with a Reflection-Type Line Analog*. PhD thesis, KTH, August 1985.

[96] Löfqvist, A., L.L. Koenig, and R.S. McGowan: *Vocal tract aerodynamics in /aca/ utterances: Measurements*. Speech Communication, 16:49–66, 1995.

[97] Lous, N., G. Hofmans, R. Veldhuis, and A. Hirschberg: *A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design*. ACUSTICA, 84:1135–1150, 1998.

[98] Lucero, J. and L.L. Koenig: *Simulations of temporal patterns of oral airflow in men and women using a two-mass model of the vocal folds under dynamic control*. J. of the Acous. Soc. Am., 117:1362–1372, 2004.

[99] Luettin, J., G. Potamianos, and C. Neti: *Asynchronous stream modeling for large vocabulary audio-visual speech recognition*. In *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, 2001.

[100] Maeda, S.: *Une approche statistique d' elaboration d' un modele articulatoire du conduit vocal*. Techn. rep., CNRS, 1978.

[101] Maeda, S.: *A digital simulation method of the vocal-tract system*. Speech Communication, 1:199–229, 1982.

[102] Maeda, S.: *Compensatory articulation during speech: evidence from the analysis and synthesis of vocal tract shapes using an articulatory model.* In Hardacastle, W. and A. Marchal (eds.): *Speech Production and Speech Modeling.* Kluwer Academic Publishers, 1990.

[103] Maeda, S.: *Phonemes as concatenable units: VCV synthesis using a vocal tract synthesizers.* In *Sound Patterns of Connected Speech Description, Models and Explanation, Proceedings of the symposium held at Kiel University*, 1996.

[104] Maragos, P., J.F. Kaiser, and T.F. Quatieri: *Energy separation in signal modulations with application to speech analysis.* IEEE Trans. Signal Process., 41(10):3024–3051, Oct. 1993.

[105] Maragos, P., J.F. Kaiser, and T.F. Quatieri: *On amplitude and frequency demodulation using energy operators.* IEEE Trans. Signal Process., 41(4):1532–1550, Apr. 1993.

[106] Mardia, K.V., J.T. Kent, and J.M. Bibby: *Multivariate Analysis.* Acad. Press, 1979.

[107] McGowan, R. and M. Howe: *Compact Green's functions extend the acoustic theory of speech production.* Journal of Phonetics, 35:259–270, 2006.

[108] McGowan, R., L.L. Koenig, and A. Loefqvist: *Vocal tract aerodynamics in /aca/ utterances: Simulations.* Speech Communication, 16:67–88, 1995.

[109] McGowan, R.S.: *An aeroacoustic approach to phonation.* J. of the Acous. Soc. Am., 83(2):696–704, February 1988.

[110] McGowan, R.S.: *Recovering articulatory movement from formant frequency trajectories using task dynamics and a genetic algorithm: preliminary model tests.* Speech Communication, 14(1):19 – 48, February 1994.

[111] McGurk, H. and J. MacDonald: *Hearing lips and seeing voices.* Nature, 264:746–748, 1976.

[112] Mermelstein, P.: *Articulatory model for the study of speech production.* J. of the Acous. Soc. Am., 53(53):1070–1082, 1973.

[113] Mokhtari, P., T. Kitamura, H. Takemoto, and K. Honda: *Principal components of vocal-tract area functions and inversion of vowels by linear regression of cepstrum coefficients.* Journal of Phonetics, 35(1):20–39, January 2007.

[114] Mokhtari, P., H. Takemoto, and T. Kitamura: *Single-matrix formulation of a time domain acoustic model of the vocal tract with side branches.* Speech Communication, 50:179–190, 2008.

[115] Morse, P.M. and U.K. Ingard: *Theoretical Acoustics.* Princeton University Press, 1986.

[116] Motoki, K.: *Three-dimensional acoustic field in vocal-tract.* Acoustic Science and Technology, 23(23):207–212, 2002.

[117] Motoki, K., X. Pelorson, P. Badin, and H. Matsuzaki: *Computation of 3-D vocal tract acoustics based on mode-matching techniques.* In *Proc. Int'l Conf. on Spoken Language Processing*, 2000.

[118] Narayanan, S. and A. Alwan: *Noise source models for fricative consonants.* IEEE Trans. Speech and Audio Process., 8:328–344, 2000.

[119] Narayanan, S.S., A.A. Alwan, and K. Haker: *An articulatory study of fricative consonants using magnetic resonance imaging*. J. of the Acous. Soc. Am., 98:1325–1347, 1995.

[120] Ohala, J.J.: *A mathematical model of speech aerodynamics*. In *Speech Communication Seminar*, 1974.

[121] Ouni, S. and Y. Laprie: *Modeling the articulatory space using a hypercube codebook for acoustic-to-articulatory inversion*. J. of the Acous. Soc. Am., 118(1):444–460, 2005.

[122] Papandreou, G., A. Katsamanis, V. Pitsikalis, and P. Maragos: *Multimodal fusion and learning with uncertain features applied to audiovisual speech recognition*. In *Proc. of IEEE Int'l Workshop on Multimedia Signal Processing*, 2007.

[123] Papandreou, G., A. Katsamanis, V. Pitsikalis, and P. Maragos: *Adaptive multimodal fusion by uncertainty compensation with application to audio-visual speech recognition*. IEEE Trans. Audio, Speech, and Language Processing, 17:423–435, 2009.

[124] Papandreou, G. and P. Maragos: *Adaptive and constrained algorithms for inverse compositional active appearance model fitting*. In *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2008.

[125] Pelorson, X., A. Hirschberg, R.R. van Hassel, A.P.J. Wijnands, and Y. Auregan: *Theoretical and experimentatl study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model*. J. of the Acous. Soc. Am., 96(6):3416–3431, December 1994.

[126] Pierce, A.D.: *Acoustics*. Acoustical Society of America, 1989.

[127] Portnoff, M.R.: *A quasi-one dimensional digital simulation for the time-varying vocal tract*. Master's thesis, MIT, 1973.

[128] Potamianos, A. and P. Maragos: *Speech analysis and synthesis using an AM-FM modulation models*. Speech Communication, 28:195–209, 1999.

[129] Potamianos, G., C. Neti, G. Gravier, A. Garg, and A. Senior: *Recent advances in the automatic recognition of audiovisual speech*. Proc. IEEE, 91(9):1306–1326, 2003.

[130] Powell, A.: *Theory of vortex sound*. J. of the Acous. Soc. Am., 36:177–195, 1964.

[131] Qin, C. and M. Carreira-Perpinan: *A comparison of acoustic features for articulatory inversion*. In *Proc. Int'l Conf. on Speech Communication and Technology*, 2007.

[132] Quatieri, T.: *Discrete-time speech signal processing*. Prentice Hall, 2002.

[133] Rabiner, L. and B. Juang: *Fundamentals of Speech Recognition*. Prentice Hall, 1993.

[134] Rabiner, L.R. and R.W. Schafer: *Digital Processing of Speech Signals*. Prentice-Hall, 1978.

[135] Ramsay, G. and C. Shadle: *The influence of geometry on the initiation of turbulence in the vocal tract during the production of fricatives*. In *Proc. Int'l Seminar on Speech Production*, 2006.

[136] Remoissenet, M.: *Waves Called Solitons*. Springer, 1996.

[137] Richmond, K., S. King, and P. Taylor: *Modelling the uncertainty in recovering articulation from acoustics*. Computer Speech and Language, 17:153–172, 2003.

[138] Rothenberg, M.: *The breath-stream dynamics of simple-released plosive production*. Bibliotheca Phonetica, 6, 1968.

[139] Roussos, A., A. Katsamanis, and P. Maragos: *Tongue tracking in ultrasound images with active appearance models*. In *Proc. IEEE Int'l Conf. on Image Processing*, 2009.

[140] Ruty, N., X. Pelorson, A. Van Hirtum, I. Lopez-Arteaga, and A. Hirschberg: *An in vitro setup to test the relevance and the accuracy of low-order vocal folds models*. J. of the Acous. Soc. Am., 121:479–490, 2007.

[141] Sargin, M.E., Y. Yemez, E. Erzin, and M. Tekalp: *Audiovisual synchronization and fusion using canonical correlation analysis*. IEEE Trans. Multimedia, 9:1396–1403, 2007.

[142] Scharf, L.L. and J.K. Thomas: *Wiener filters in canonical coordinates for transform coding, filtering, and quantizing*. IEEE Trans. Speech and Audio Process., 46(3):647–654, 1998.

[143] Schoentgen, J. and S. Ciocea: *Kinematic formant-to-area mapping*. Speech Communication, 21(4):227–244, 1997.

[144] Schroeter, J. and M.M. Sondhi: *Speech coding based on physiological models of speech production*. In Furui, S. and M.M. Sondhi (eds.): *Advances in Speech Signal Processing*, ch. 8. NewYork: Marcel Dekker Inc, 1992.

[145] Schroeter, J. and M.M. Sondhi: *Techniques for estimating vocal-tract shapes from the speech signal*. IEEE Trans. Speech and Audio Process., 2:133–150, 1994.

[146] Scully, C.: *Articulatory synthesis*. In *Speech Production and Speech Modeling*. Kluwer Academic Publishers, 1990.

[147] Shadle, C.: *The acoustics of fricative consonants*. PhD thesis, MIT, 1985.

[148] Shadle, C.: *The aerodynamics of speech*. In Hardcastle, W.J. and J. Laver (eds.): *The Handbook of Phonetic Sciences*. Blackwell Publishing, 1999.

[149] Shadle, C.H., A. Barney, and P. Davies: *Fluid flow in a dynamic mechanical model of the vocal folds and tract. II. implications for speech production studies*. J. of the Acous. Soc. Am., 105:456–466, 1999.

[150] Sinder, D.J.: *Speech Synthesis Using An Aeroacoustic Fricative Model*. PhD thesis, Rutgers, The State University of New Jersey, 1999.

[151] Sondhi, M.M. and J. Schroeter: *A hybrid time-frequency domain articulatory speech synthesizer*. IEEE Trans. Acoust., Speech, Signal Process., 35(7):955–967, July 1987.

[152] Soong, F.K. and B.H. Juang: *Line spectrum pair and speech data compression*. In *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, vol. 9, pp. 37–40, 1984.

[153] Soquet, A. amd Lecuit, V., T. Metens, and D. Demolin: *Mid-sagittal cut to area function transformations: Direct measurements of mid-sagittal distance and area with MRI*. Speech Communication, 36:169–180, 2002.

[154] Standring (ed.): *Henry Gray's Anatomy of the Human Body*. Elsevier, 2008.

[155] Stevens, K.N.: *Airflow and turbulence noise for fricative and stop consonants: Static considerations*. J. of the Acous. Soc. Am., 50:1180–1192, 1971.

[156] Stevens, K.N.: *Acoustic Phonetics*. The MIT Press, 1998.

[157] Stork, D. and M. Hennecke (eds.): *Speechreading by Humans and Machines*. Springer, Berlin, Germany, 1996.

[158] Story, B.H.: *A parametric model of the vocal tract area function for vowel and consonant simulation*. J. of the Acous. Soc. Am., 5:3231–3254, 2005.

[159] Story, B.H. and I.R. Titze: *Vocal tract area functions from magnetic resonance imaging*. J. of the Acous. Soc. Am., 100:537–554, 1996.

[160] Story, B.H., I.R. Titze, and E.A. Hoffman: *Vocal tract area functions for an adult female speaker based on volumetric imaging*. J. of the Acous. Soc. Am., 104:471–487, 1998.

[161] Suh, J. and S.H. Frankel: *Numerical simulation of turbulence transition and sound radiation for flow through a rigid glottal model*. J. of the Acous. Soc. Am., 121:3728–3739, 2007.

[162] Suh, J. and S.H. Frankel: *Comparing turbulence models for flow through a rigid glottal model*. J. of the Acous. Soc. Am., 123:1237–1240, 2008.

[163] Takemoto, H., P. Mokhtari, and T. Kitamura: *Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method*. In *Proceedings of Acoustics'08*, 2008.

[164] Teager, H.M.: *Some observations on oral air flow during phonation*. 28:599–601, 1980.

[165] Teager, H.M. and S.M. Teager: *Active fluid dynamic voice production models or there is a unicorn in the garden*. Appears in book Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control.

[166] Teager, H.M. and S.M. Teager: *The effects of separated air flow on vocalization*. In *Vocal Fold Physiology: Contemporary research and clinical issues*. College-Hill Press, 1983.

[167] Teager, H.M. and S.M. Teager: *A phenomenological model for vowel production in the vocal tract*. In *Speech Science: Recent Advances*. College-Hill Press, 1985.

[168] Teager, H.M. and S.M. Teager: *Evidence for nonlinear production mechanisms in the vocal tract*. In *Proceedings of the NATO Advanced Study Institute on Speech Production and Modeling, Chateau Bonas, France, July 17-29*. Kluwer Academic Publishers, 1990.

[169] Thomas, T.: *A finite element model of fluid flow in the vocal tract*. Computer Speech and Language, 1:131–151, 1986.

[170] Toda, T., A.W. Black, and K. Tokuda: *Statistical mapping between articulatory movements and acoustic spectrum using a Gaussian mixture models*. Speech Communication, 50:215–227, 2008.

[171] Tso, M.S.: *Reduced-rank regression and canonical analysis*. J. of Royal Stat. Soc. (Series B), 43:183–189, 1981.

[172] Van Den Berg, J., J.T. Zantema, and P. Doornenbal Jr.: *On the air resistance and the Bernoulli effect of the human larynx*. J. of the Acous. Soc. Am., 29:626–631, 1957.

[173] Vilain, C.E., X. Pelorson, C. Fraysse, M. Deverge, A. Hirschberg, and J. Willems: *Experimental validation of a quasi-steady theory for the flow through the glottis*. J. of the Acous. Soc. Am., 276:475–490, 2004.

[174] Viola, P. and M. Jones: *Rapid object detection using a boosted cascade of simple features*. In *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, vol. I, pp. 511–518, 2001.

[175] Wrench, A. and W. Hardcastle: *A multichannel articulatory speech database and its application for automatic speech recognition*. In *Proc. 5th Seminar on Speech Production, Kloster Seeon, Bavaria*, pp. 305–308, 2000. `http://www.cstr.ed.ac.uk/artic`.

[176] Xie, L. and Z.Q. Liu: *Realistic mouth-synching for speech-driven talking face using articulatory modeling*. IEEE Trans. Multimedia, 9:500–510, 2007.

[177] Yamamoto, E., S. Nakamura, and K. Shikano: *Lip movement synthesis from speech based on hidden Markov models*. Speech Communication, 26:105–115, 1998.

[178] Yehia, H., P. Rubin, and E. Vatikiotis-Bateson: *Quantitative association of vocal-tract and facial behavior*. Speech Communication, 26:23–43, 1998.

[179] Young, S., G. Evermann, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland: *The HTK book (for HTK version 3.2)*. Techn. rep., Cambridge University Engineering Department, Dec. 2002.

[180] Zhang, C., W. Zhao, H.S. Frankel, and L. Mongeau: *Computational aeroacoustics of phonation, part II: Effects of flow parameters and ventricular folds*. J. of the Acous. Soc. Am., 112(5):2147–2154, November 2002.

[181] Zhang, L. and S. Renals: *Acoustic-articulatory modeling with the trajectory HMMs*. IEEE Signal Process. Lett., 15:245–248, 2008.

[182] Zhang, Z., L. Mongeau, S.H. Frankel, S. Thomson, and J.B. Park: *Sound generation by steady flow through glottis-shaped orifices*. J. of the Acous. Soc. Am., 116:1720–1728, 2004.

[183] Zhao, W., C. Zhang, H.S. Frankel, and L. Mongeau: *Computational aeroacoustics of phonation, part I: Computational methods and sound generation mechanisms*. J. of the Acous. Soc. Am., 112(5):2134–2143, November 2002.

Βιογραφικό Σημείωμα Αθανασίου Κατσαμάνη

22 Οκτωβρίου 2009

Ο Νάσος Κατσαμάνης πήρε το διδακτορικό του δίπλωμα και το δίπλωμα του ηλεκτρολόγου μηχανικού και μηχανικού υπολογιστών (με βαθμό άριστα) από το Εθνικό Μετσόβιο Πολυτεχνείο το 2009 και το 2003 αντίστοιχα. Αυτή την περίοδο είναι μεταδιδακτορικός ερευνητής στο Πανεπιστήμιο της Νότιας Καλιφόρνιας, μέλος του Εργαστηρίου Ανάλυσης και Ερμηνείας Σήματος στη Σχολή Ηλεκτρολόγων Μηχανικών. Τα ερευνητικά του ενδιαφέροντα εντοπίζονται στη γενικότερη περιοχή της ανάλυσης φωνής και πολυτροπικών σημάτων, και ειδικότερα στα πεδία της (οπτικοακουστικής) παραγωγής, σύνθεσης, αντιστροφής, αναγνώρισης και επεξεργασίας φωνής. Από το 2004 έως το 2009 εργάστηκε ως διπλωματούχος βοηθός ερευνητής στο Εθνικό Μετσόβιο Πολυτεχνείο, μέλος της ερευνητικής ομάδας Όρασης Υπολογιστών, Επικοινωνίας Λόγους και Επεξεργασίας Σημάτων. Στα πλαίσια της διδακτορικής του διατριβής και Ευρωπαϊκών ερευνητικών προγραμμάτων, από το 2003 έχει ασχοληθεί με πολυτροπική αντιστροφή φωνής, αεροακουστική για σύνθεση φωνής με αρθρωτές, προσαρμογή συστήματος αναγνώρισης φωνής σε αλλοδαπό ομιλητή και πολυτροπική σύμμειξη για αναγνώριση φωνής. Από το 2000 ως το 2002, ήταν προπτυχιακός βοηθός έρευνας στο Ινστιτούτο Επεξεργασίας του Λόγου (Ι.Ε.Λ.) στην Αθήνα, με συμμετοχή σε ερευνητικά προγράμματα σε σύνθεση φωνής, εκπαίδευση επεξεργασίας σημάτων και μηχανική μετάφραση. Στη διάρκεια του καλοκαιριού του 2002, ασχολήθηκε με την ανάπτυξη συστήματος αναγνώρισης φωνής στα Καντονέζικα στο Πολυτεχνικό Πανεπιστήμιο του Χονγκ Κονγκ. Το καλοκαίρι του 2007 επισκέφτηκε το Πανεπιστήμιο Télécom Paris (ENST) όπου συνεργάστηκε με τον καθηγητή Shinji Maeda σε θέματα μοντελοποίησης του συστήματος παραγωγής φωνής.