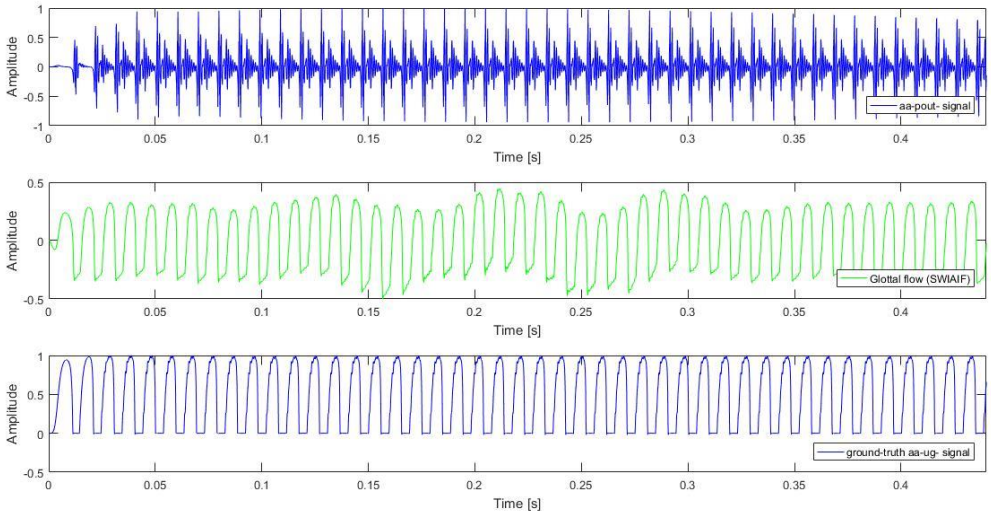
Πανεπιστήμιο Κρήτης
Τμήμα Επιστήμης Υπολογιστών

Διπλωματική Εργασία

Μία Επαναληπτική Προσέγγιση για Γλωττιδική Ανάλυση με Χρήση της Σταθεροποιημένης Γραμμικής Πρόβλεψης με Βάρη

An Iterative Approach for Glottal Analysis using Stabilized Weighted Linear Prediction



Καλοζούμη Αλεξάνδρα

ΑΜ: 3433

Επιβλέπων Καθηγητής: Καφεντζής Γεώργιος

Ηράκλειο Κρήτης 2020

Περιεχόμενα

1 Εισαγωγή

Σε αυτό το κεφάλαιο θα περιγραφούν οι βασικές αρχές της παραγωγής φωνής, καθώς και το μοντέλο πηγής φίλτρου, το οποίο θα μελετηθεί από τη σκοπιά της μηχανικής υπολογιστών. Επίσης, θα γίνει παρουσίαση της τεχνικής του αντίστροφου φιλτραρίσματος (inverse filtering), όπως και το πρόβλημα και η τεχνική που ακολουθήθηκε.

1.1 Το σύστημα παραγωγής φωνής

Όσον αφορά την ανατομία της παραγωγής της φωνής, αυτή αποτελείται κυρίως από τον πνεύμονα, τον λάρυγγα, τη γλωττίδα και τη στοματική και ρινικη κοιλότητα.

Στην Εικόνα 1 μπορεί κανείς να δει τα όργανα που συμμευέχουν στην παραγωγή φωνής στο ανθρώπινο σώμα.

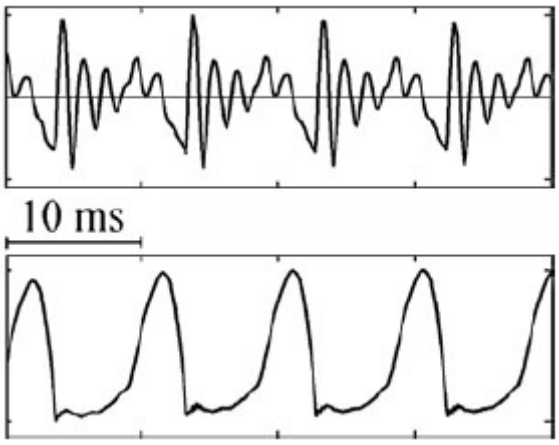
Εικόνα 1 [6]

Αναλυτικότερα, η παραγωγή φωνής ξεκινά από τον αέρα που βρίσκεται στους πνεύμονες. Κατά τη διαδικασία της εκπνοής ο αέρας προχωρά στην τραχεία. Στο σημείο αυτό παρατηρείται η αύξηση της τραχειακής πίεσης. Έπειτα, ο αέρας φτάνει στον λάρυγγα, όπου "ενεργοποιούνται" οι φωνητικές χορδές. Όταν αυτές βρίσκονται σε απόσταση, ο αέρας έχει ελεύθερη δίοδο προς τη στοματική κοιλότητα και η διαδικασία αυτή αποτελεί την αναπνοή. Η τελική μορφή του ήχου καθορίζεται τον φάρυγγα και τη στοματική κοιλότητα, η οποία περιλαμβάνει τη γλώσσα, τα δόντια, τα χείλη και τη ρινική κοιλότητα. [1]

Κατά το άνοιγμα και κλείσιμο των φωνητικών χορδών περιοδικά, διέρχεται ανάμεσα από αυτές ο αέρας. Στην περίπτωση αυή ο ήχος που ακούγεται μοιάζει με βοή. Όταν ο ήχος αυτός εξέλθει στο περιβάλλον θα αποδώσει τα φωνήεντα. Ένας μη περιοδικός ήχος παράγεται όταν η ροή του αέρα περάσει από ένα στενό πέρασμα των φωνητικών χορδών. Τέτοιοι ήχοι αποτελούν τα σύμφωνα.

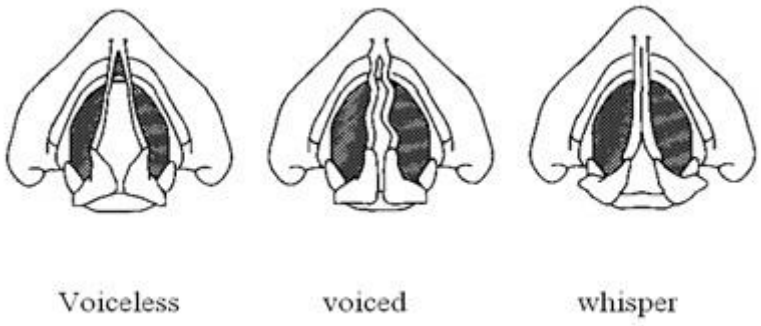
Αξίζει να σημειωθεί πως η πίεση στην τραχεία και η ροή του αέρα είναι μεγαλύτερες κατά τη διάρκεια μεγάλης έντασης στη φωνή και μικρότερες κατά την κανονική ομιλία. Στην περίπτωση του ψίθυρου είναι ακόμη μικρότερες.

Αναφορικά με τη λειτουργία των φωνητικών χορδών, αυτό που επιτυγχάνεται με το "άνοιγμα" και "κλείσιμό" τους είναι η διαμόρφωση της ροής του εκπεμπόμενου αέρα. Η κίνηση αυτή ονομάζεται «φώνηση». Σε κατάσταση ηρεμίας, η γλωττίδα είναι ανοιχτή για να διευκολύνεται η διαδικασία της αναπνοής, αλλά οι φωνητικές χορδές κλείνουν εώς κάποιο βαθμό την αναπνευστική οδό. Όταν η γλωττίδα είναι κλειστή, η γλωττιδική ροή είναι μηδενική. Κατά τη διαδικασία ανοίματος της γλωττίδας, η ροή του αέρα αυξάνεται και φτάνει στο μέγιστο σημείο. Έπειτα, η γλωτττίδα αρχίζει τη διαδικασία κλεισίματος και τότε η ροή του αέρα ελαττώνεται μέχρι να φτάσει στο ελάχιστο. Η διαδικασία αυτή απεικονίζεται στην Εικόνα 2.



Εικόνα 2 [4]

Στην Εικόνα 3 φαίνεται η ανατομία των φωνητικών χορδών κατά τη διάρκεια της παραγωγής άφωνων ήχων, της παραγωγής φωνής και του ψιθύρου.



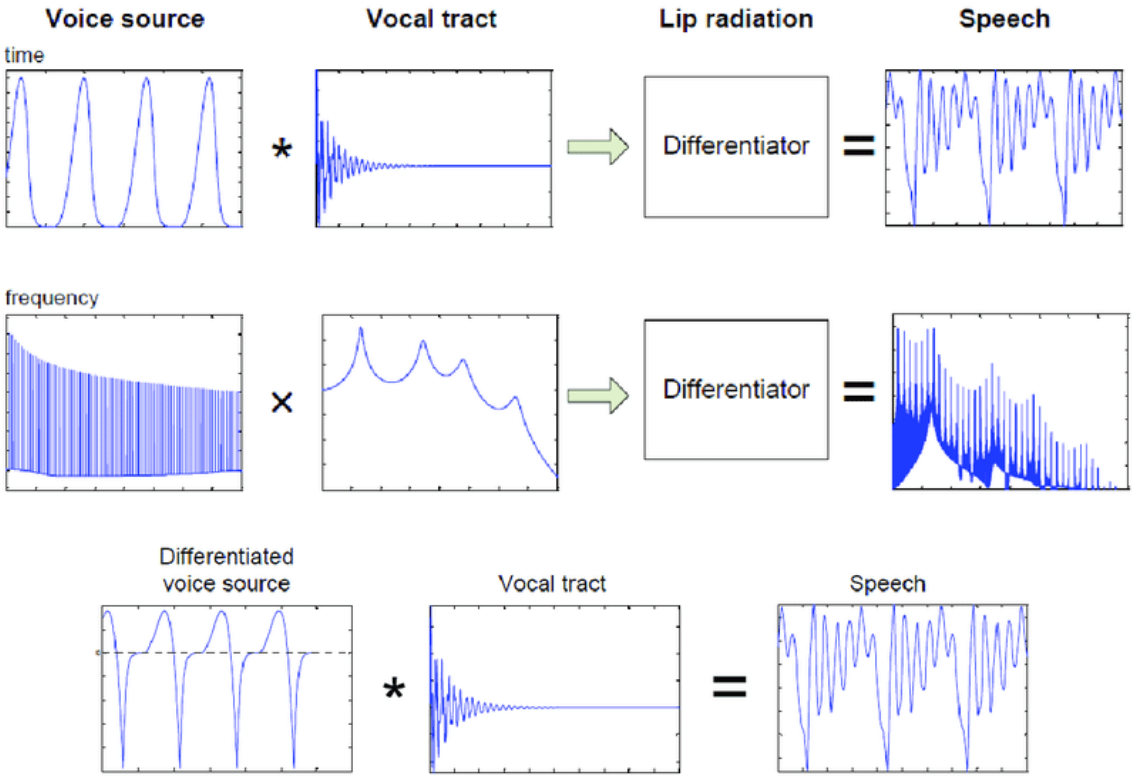
Εικόνα 3 [7]

1.2 Το μοντέλο πηγής-φίλτρου

Ο αέρας που ξεκινά το ταξίδι του από τους πνεύμονες έως να ακουστεί στο περιβάλλον, δεν είναι άλλο από ένα σήμα που μπορεί να θεωρηθεί ότι περνά και διαμορφώνεται από διάφορα συστήματα. Γι' αυτό η παραγωγή φωνής μπορεί να ποσεγγιστεί από ένα μοντέλο πηγής-φίλτρου. Πιο συγκεκριμένα, οι φωνητικές χορδές διαμορφώνουν τη ροή του αέρα από τους πνεύμονες και δημιουργούν το σήμα που ονομάζεται πηγή (source). Αυτό το σήμα αποτελεί ένα σήμα ταχύτητας ροής αέρα (glottal velocity waveform). Αναφορικά με τη φωνητική οδό (φάρυγγας, στοματική και ρινική κοιλότητα), αποτελεί ένα σύστημα, το οποίο δέχεται το σήμα πηγής, το διαμορφώνει συχνοτικά ενισχύοντας ή καταστέλλοντας συγκεκριμένες συχνότητες και εξάγει ένα σήμα ροής που τελικά διαμορφώνεται από τα χείλη, μετατρέποντάς το από σήμα ταχύτητας ροής αέρα σε σήμα μεταβολής πίεσης.

Στο σχήμα από την Εικόνα 4 βλέπει κανείς τη χρονική και συχνοτική αναπαράσταση του μοντέλου πηγής-φίλτρου. Συγκεκριμένα (α) στην πάνω σειρά εικόνων, παρατηρούμε το σήμα ταχύτητας ροής αέρα για έναν έμφωνο ήχο, ο οποίος συνελίσσεται με την κρουστική απόκριση του συστήματος της φωνητικής οδού. Το αποτέλεσμα περνά από ένα διαφοριστή που μοντελοποιεί την επίδραση των χειλιών. Επίσης, (β) στη μεσαία σειρά εικόνων, βλέπουμε το αντίστοιχο συχνοτικό περιεχόμενο: ένας έμφωνος ήχος είναι περιοδικός λόγω του ανοίγματος-κλεισίματος των φωνητικών χορδών, οπότε το φάσμα του εμφανίζει περιοδικότητα. Η συνέλιξη στο χώρο του χρόνου γίνεται γινόμενο στο χώρο της συχνότητας, με το σύστημα να ενισχύει ή καταστέλλει συγκεκριμένα εύρη συχνοτήτων. Οι περιοχές ενίσχυσης συχνοτήτων (οι κορυφές στο φάσμα του σήματος της φωνητικής οδού) ονομάζονται φωνοσυντονισμοί (formants), και καθορίζονται κυρίως από το σχήμα που παίρνει η φωνητική οδός όταν εκφέρουμε ένα φώνημα.

Το αποτέλεσμα περνά ξανά από ένα διαφοριστή για τους ίδιους λόγους, με το διαφοριστή στο χώρο της συχνότητας να εκφράζεται ως ένα υψιπερατό φίλτρο. Τέλος, (γ) στην τελευταία σειρά εικόνων παρουσιάζεται το ίδιο σύστημα, με το διαφοριστή ενσωματωμένο στο σήμα ταχύτητας ροής αέρα, που παράγει έτσι ένα σήμα που ονομάζεται η παράγωγος του σήματος γλωττιδικής ροής (glottal airflow derivative). Αυτό δικαιολογείται από το γεγονός ότι υποέτουμε ότι το μοντέλο πηγής-φίλτρου είναι γραμμικό.



Εικόνα 4 [8]

1.3 Η τεχνική του αντίστροφου φιλτραρίσματος

Η τεχνική του αντίστροφου φιλτραρίσματος (Glottal Inverse Filtering - GIF) είναι μία μέθοδος εκτίμησης του σήματος ταχύτητας ροής αέρα απ' ευθείας από το σήμα της ομιλίας. Αυτό που πρακτικά συμβαίνει κατά το αντίστροφο φιλτράρισμα είναι ότι έχοντας ως δεδομένο ένα ηχογραφημένο σήμα φωνής, μοντελοποιείται το φίλτρο της φωνητικής οδού και "αφαιρείται" από το σήμα φωνής περνώντας το μέσα από το "αντίστροφο φίλτρο" της φωνητικής οδού. Το σήμα

που έχει απομείνει είναι μία αξιόλογη προσέγγιση του "καθαρού" σήματος που παράγουν οι φωνητικές χορδές.

Μία τέτοια τεχνική θα μπορούσε να χρησιμεύσει σε πολλές ιατρικές και μη εφαρμογές, για τη λήψη του σήματος των φωνητικών χορδών προκειμένου να εντοπιστεί η ύπαρξη ή η απουσία παθολογίας. Με τον τρόπο αυτό αποφεύγεται η έως τώρα διαδικασία που επέβαλε αρχικά την επίσκεψη σε γιατρό, καθώς και η επίπονη εξέταση της τοποθέτησης ενός μικροφώνου στο βαθιά έως το φάρυγγα.

Με όρους επεξεργασίας σήματος, το σήμα φωνής $s[n]$ μπορεί να αναλυθεί ως η συνέλιξη τριών σημάτων:

$$s[n] = g[n] * v[n] * l[n]$$

με $g[n]$ το σήμα της ταχύτητας ροής αέρα, $v[n]$ την κρουστική απόκρισητου συστήματος που μοντελοποιεί τη φωνητική οδό και $l[n]$ την κρουστική απόκριση του συστήματος που περιγράφει το σύστημα του διαφοριστή.

Στο χώρο του μετασχηματισμού Z η σχέση αυτή γράφεται ως εξής:

$$S[z] = G[z] \cdot V[z] \cdot L[z]$$

Λύνοντας ως προς $G[z]$ καταλήγουμε στη σχέση:

$$G[z] = S[z] \cdot \frac{1}{V[z]} \cdot \frac{1}{L[z]}$$

Η σχέση αυτή αποδίδει το μετασχηματισμό Z του σήματος ταχύτητας ροής αέρα, γνωρίζοντας τα υποσυστήματα της φωνητικής οδού και του διαφοριστή. Είναι αντιληπτό πως η εκτίμηση των δύο αυτών συστημάτων από το σήμα της φωνής είναι κρίσιμη για την όσο το δυνατόν ακριβέστερη ανάκτηση του σήματος της ταχύτητας ροής αέρα. Αξίζει να σημειωθεί ότι στην πραγματικότητα η παραγωγή ανθρώπινης ομιλίας δεν ακολουθεί αυστηρά γραμμική σχέση μεταξύ των εμπλεκόμενων μελών.

1.4 Το πρόβλημα που μελετάμε και γιατί είναι σημαντικό

Αναλογιζόμενος κανείς την ιδέα του αντίστροφου φιλτραρίσματος, μπορεί να βρει αρκετούς αλγορίθμους και μάλιστα σχετικά αποδοτικούς. Δύο από τους πιο σημαντικούς, οι οποίοι θα αναλυθούν εκτενέστερα σε αυτήν την εργασία, είναι η μέθοδος Iterative Adaptive Inverse Filering (Κεφάλαιο 2.1) και η Stabilized Weighted Linear Prediction (Κεφάλαιο 2.2). Η πρώτη είναι μια αρκετά διαδεδομένη πλήρης μέθοδος αντίστροφου φιλτραρίσματος, ενώ η δεύτερη είναι μια μέθοδος εκτίμησης του συστήματος της φωνητικής οδού. Βάση αυτής της εργασίας είναι η σύμπτηξη και ο συνδυασμός αυτών τω δύο αλγορίθμων με σκοπό την απόδειξη της μεγαλύτερης αποδοτικότητας που θα οδηγήσει σε ακριβέστερη και πιο εύρωστη εκτίμηση του σήματος των φωνητικών χορδών.

Άξιο αναφοράς είναι ένα σημαντικό πρόβλημα της αξιολόγησης τέτοιου είδους τεχνικών αντίστροφου φιλτραρίσματος είναι ότι δεν μας είναι γνωστό το σήμα γλωττιδικής ταχύτητας ροής αέρα, ώστε να γίνει σύγκριση με τις εκτιμήσεις διάφορων αλγορίθμων. Σε αυτή την εργασία οι τεχνικές μας εφαρμόζονται πάνω σε μία βάση δεδομένων (Κεφάλαιο 3.1), η οποία έχει δημιουργηθεί από ένα φυσικό, μηχανικό μοντέλο παραγωγής ομιλίας, στο οποίο μοντέλο είναι διαθέσιμο τόσο το σήμα της γλωττιδικής ροής όσο και το τελικό σήμα φωνής που παράγεται μέσα από το μοντέλο. Με τον τρόπο αυτό είναι δυνατή η χρήση αυτού του τελικού σήματος φωνής και η σύγκρισή του με το σήμα της γλωττιδικής ροής, ώστε να εξαχθούν και να αξιολογηθούν τα συμπεράσματά μας.

1.5 Η δομή της εργασίας στα επόμενα κεφάλαια

Στα κεφάλαια που ακολουθούν η δομή έχει ως εξής:

Στο κεφάλαιο 2 (σελίδα 9) θα αναλυθούν οι δύο παραπάνω αλγόριθμοι.

Στο κεφάλαιο 3 (σελίδα 16) θα αναλυθεί η βάση δεδομένων που χρησιμοποιήθηκε, τα αποτελέσματα και οι μετρικές που οδήγησαν σε αυτά.

Στο κεφάλαιο 4 (σελίδα 24) παραθέτονται τα συμπεράσματα και κάποιες σκέψεις για μελοντικές επεκτάσεις αυτής της εργασίας.

2 Τεχνικές Αντίστροφου Φιλτραρίσματος

2.1 Η IAIF

Η τεχνική IAIF (Iterative Adaptive Inverse Filtering) είναι μία μέθοδος που υπολογίζει το σήμα της γλωττιδικής ροής ασύγχρονα σε πολλές θεμελειώδεις περιόδους. Κατά την εκτίμηση της συνάρτησης μεταφοράς των φωνητικών χορδών, η συνάρτηση μεταφοράς της LPC ανάλυσης παίζει πολύ σημαντικό ρόλο.

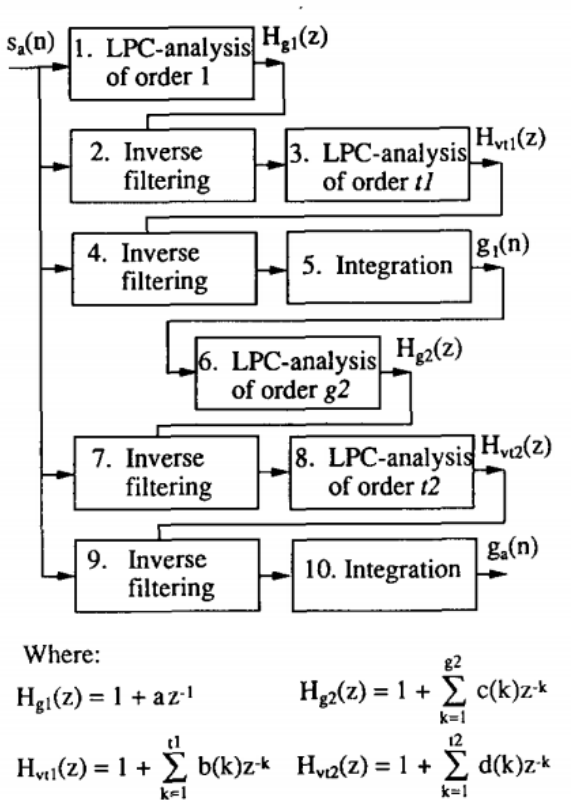
Πιο συγκεριμένα, η εκτίμηση της γλωττιδικής συνεισφοράς και της συνάρτησης μεταφοράς των φωνητικών χορδών υπολογίζονται με τον IAIF αλγόριθμο με επαναληπτική δομή που επαναλαμβάνεται δύο φορές. Στην αρχή της πρώτης επανάληψης, η επίδραση της παλμικής μορφής της γλωττίδας στο φάσμα της ομιλίας υπολογίζεται αρχικά με την LPC ανάλυση πρώτης τάξεως. Έπειτα, η εκτιμώμενη συνεισφορά ακυρώνεται μέσω του αντίστροφου φιλτραρίσματος. Το αρχικό μοντέλο του φωνητικού συστήματος προκύπτει από την εφαρμογή LPC ανάλυσης υψηλότερης τάξεως στο σήμα. Στο σήμα αυτό ακυρώθηκε η επίδραση της πηγής. Μέσω του αντρίστροφου φιλτραρίσματος επιτυγχάνεται η πρώτη εκτίμηση για τη γλωσσική διέγερση. Πιο συγκεκριμένα, εφαρμόζεται ακύρωση της επίδρασης της φωνητικής οδού και των χειλιών από το αρχικό σήμα ομιλίας.

Η εκτίμηση του σήματος της γλωττίδας που προκύπτει από την πρώτη επανάληψη, χρησιμοποιείται στην αρχή της δεύτερης επανάληψης, προκειμένου να υπολογιστεί ένα πιο ακριβές μοντέλο για τη συνολική συνεισφορά. Η επίδραση της γλωττιδικής ροής στο φάσμα της ομιλίας εκτιμάται στη δεύτερη επανάληψη με τη χρήση LPC ανάλυσης, της οποίας η τάξη ισούται με δύο ή τέσσερα. Αξίζει να σημειωθεί ότι η εκτίμηση της συνεισφοράς της γλωττιδικής ροής στην πρώτη επανάληψη, δε θα υπολογιστεί χρησιμοποιώντας μία LPC με τάξη που είναι μεγαλύτερη από ένα. Μία συνέπεια χρήσης LPC ανάλυσης μεγαλύτερης τάξης θα ήταν να μοντελοποιηθούν τα formants με ένα φίλτρο που στοχεύει τη μοντελοποίηση της επίδρασης του σήματος της πηγής.

Μετά την ακύρωση της εκτιμώμενης συνεισφοράς της γλωττίδας στο συνολικό φάσμα, το μοντέλο για τη φωνητική οδό σχηματίζεται άλλη μια φορά χρησιμοποιώντας LPC ανάλυση υψηλότερης τάξης. Το τελικό αποτέλεσμα επιτυγχάνεται με το αντίστροφο φιλτράρισμα της επίδρασης της φωνητικής οδού και των χειλιών από το αρχικό σήμα ομιλίας.

Στο παρακάτω σχήμα στην Εικόνα 5 βλέπει κανείς την IAIF μέθοδο. Το σήμα της φωνής σημειώνεται ως $s_{\mathrm{a}}(n)$ και το αποτέλεσμα που υπολογίστηκε σημειώνεται ως $g_{\mathrm{a}}(n)$ και αποτελεί

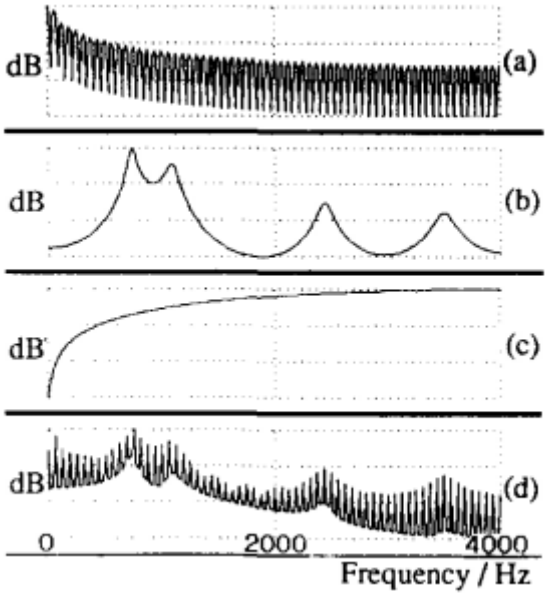
το σήμα που παράγεται ακριβώς πάνω από τις φωνητικές χορδές. Η πρώτη επανάληψη περιλαμβάνει τα βήματα 1 – 5 και η δεύτερη επανάληψη τα βήματα 6 – 10.



Εικόνα 5 [2]

Εικόνα 6 [2]

Στην Εικόνα 7 βλέπουμε: (a) το φάσμα πλάτους της γλωττιδικής ροής, που αποτελεί τη διέγερση στο σύστημα της φωνητικής οδού, (b) το φάσμα πλάτους της φωνητικής οδού, (c) το φάσμα πλάτους του συστήματος που αφορά το φαινόμενο της επίδρασης των χειλιών, (d) το φάσμα της ομιλίας.



Εικόνα 7 [2]

2.2 Η SWLP

Η μέθοδος SWLP (Stabilized Weighted Linear Prediction) είναι μία μέθοδος εκτίμησης της φωνητικής οδού, η οποία έχει πολλές ομοιότητες με τη μέθοδο της αυτοσυσχέτισης της Γραμμικής Πρόβλεψης για την οποία έγινε αναφορά στο προηγούμενο κεφάλαιο. Η κυριότερη διαφορά είναι ότι ο πίνακας αυτοσυσχέτισης λαμβάνει διαφορετικά βάρη στα στοιχεία του, ανάλογα με κάποιο κριτήριο. Η SWLP πιστεύεται ότι μπορεί να αντικαταστήσει την κλασική LPC αυτοσυσχέτισης στις επαναληπτικές διαδικασίες της IAIF, προσφέροντας καλύτερες επιδόσεις, ενώ έχει ήδη δειχθεί [3] ότι υπερτερεί της μεθόδου αυτοσυσχέτισης LPC ως stand-alone μέθοδος αντίστροφου φιλτραρίσματος. Συγκριτικά με τη μέθοδο Autocorrelation εκτιμάται πως η SWLP φαίνεται να δίνει σήμα γλωττιδικής ροής που είναι πιο κοντά στο πραγματικό σήμα, ενώ φαίνεται να είναι συγκρίσιμη με τη μέθοδο Closed Phase Covariance. [4]

Πιο συγκεκριμένα, έχει αποδειχτεί πως η SWLP μπορεί να ξεπεράσει την προσέγγιση της συμβατικής αυτοσυσχέτισης (Autocorrelation), η οποία συνήθως χρησιμοποιείται στο αντίστροφο φιλτράρισμα.

Η SWLP βασίστηκε στην Weighted Linear Prediction (WLP), για την οποία ισχύουν τα εξής:

Για το δείγμα $x[n]$ εκτιμάται με γραμμικό συνδυασμό των προηγούμενων p δειγμάτων:

$$\hat{x}[n] = -\sum_{i=1}^{p} a_i x[n-i]$$

όπου $a_i \in R$.

Το σφάλμα πρόβλεψης $e_n(\mathrm{a})$ ορίζεται ως εξής:

$$e_n(\mathrm{a}) = x[n] - \hat{x}[n] = x[n] + \sum_{i=1}^{p} a_i x[n-i] = \mathrm{a}^T x[n]$$

Όπου $\mathrm{a} = [\mathrm{a}_0 \mathrm{a}_1 \cdots \mathrm{a}_p]^T$ με $\mathrm{a}_0 = 1$ και $\mathrm{x[n]} = [x[n] \cdots x[n-p]]^T$.

Επίσης, η ενέργεια του σφάλματος πρόβλεψης $E(\mathrm{a})$ δίνεται από τον τύπο:

$$E(\mathrm{a}) = \sum_{n=1}^{N+p} (e_n(\mathrm{a}))^2 w_n = \mathrm{a}^T \left(\sum_{n=1}^{N+p} w_n \, x[n] x^T[n]\right) \mathrm{a} = a^T R \mathrm{a}$$

Όπου w_n να είναι το βάρος που επιβάλλεται στο δείγμα n, το N είναι το μήκος του σήματος $x[n]$ και

$$R = \sum_{n=1}^{N+p} w_n \, x[n] x^T[n]$$

Εν τέλη, το μοντέλο WLP ορίζεται ως:

$$H(z) = \frac{1}{A(z)}$$

όπου

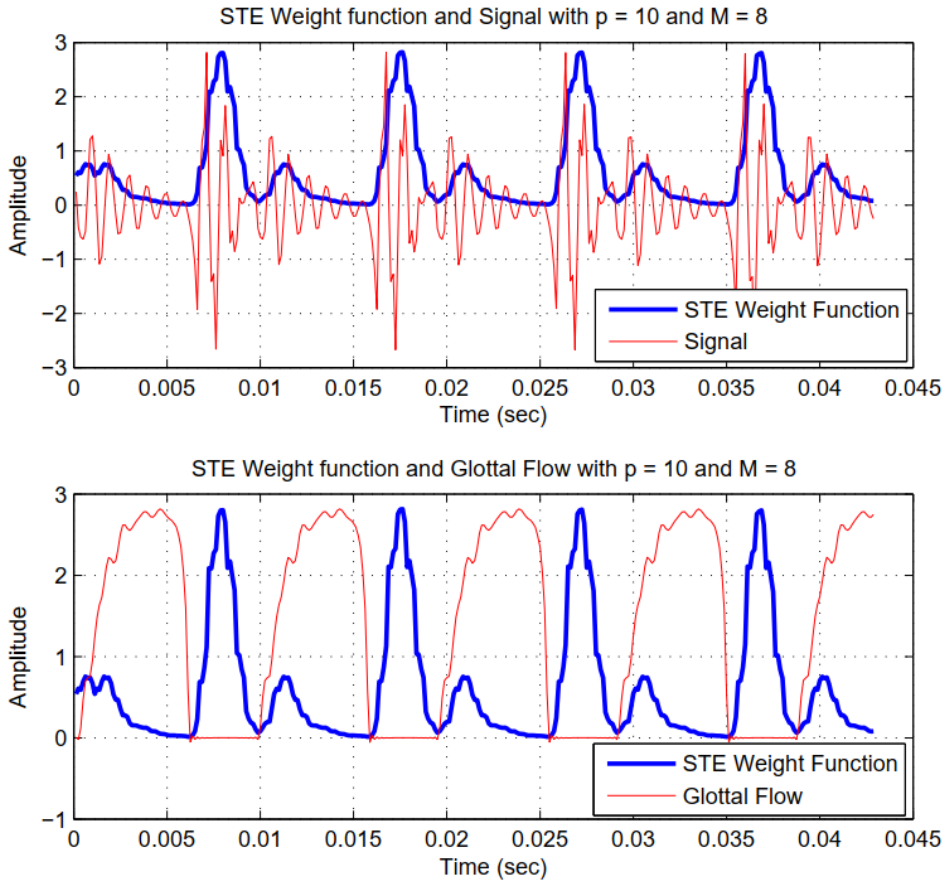
$$A(z) = 1 + \sum_{i=1}^{p} a_i z^{-i}$$

Όσον αφορά τη συνάρτηση στάθμισης (weighting function) στον τομέα του χρόνου, το w_n είναι το βασικό σημείο του WLP:

$$w_n = \sum_{i=0}^{M-1} x[n-i-1]^2$$

όπου M είναι το μήκος του παραθύρου STE (Short-Time-Energy).

Σχετικά με το STE αυτό δίνει έμφαση στα δείγματα ομιλίας μεγάλου πλάτους που εμφανίζονται στο διάστημα κλειστής φάσης της γλωττίδας. Με τον τρόπο αυτό μπορεί να εκτιμηθεί το φίλτρο της φωνητικής οδού με μεγαλύτερη ευκολία. Αυτό συμβαίνει, διότι θεωρούμε πως όταν η γλωττίδα είναι κλειστή, αυτή δεν έχει κάποια αλληλεπίδραση με τη φωνητική οδό. Άρα, αυτά τα δείγματα ομιλίας δεν έχουν επηρρεαστεί από την επίδραση της γλωττίδας, αλλά μόνο από τη φωνητική οδό. Στο σχήμα παρακάτω μπορεί κανείς να διακρίνει τη STE συνάρτηση επάνω στο σήμα φωνής ενός φωνηέντου.



Εικόνα 8 [3]

Στην Εικόνα 8 (επάνω σχήμα) παρατηρείται η συνάρτηση STE, η οποία εκτιμήθηκε επάνω στα δείγματα ενός συνθετικού σήματος ομιλίας, με παράμετρο M = 8. Η ίδια συνάρτηση έχει τοποθετηθεί (κάτω σχήμα) στους ίδιους άξονες με το σήμα γλωττιδικής ροής άερα. Είναι εμφανές ότι οι υψηλές τιμές STE συμπίπτουν με την κλειστή φάση της γλωττίδας, όπου δεν υπάρχει ροή (μηδενική ροή) αέρα ανάμεσα στις φωνητικές χορδές. Αξίζει να σημειωθεί ότι το παράδειγμα είναι πλήρως συνθετικό: η γλωττιδική ροή έχει παράξει το σήμα φωνής μέσα από φυσικό μοντέλο παραγωγής ομιλίας που θα περιγραφεί στο επόμενο κεφάλαιο.

Παρατηρήθηκε για τον αλγόριθμο SWLP ότι συγκρίνοντας τον με με τη συμβατική μέθοδο αυτοσυσχέτισης (LPC) για M=24 βρέθηκε ότι η εκτίμηση με SWLP είναι πιο κοντά στο σήμα της γλωττίδας. [3]

Σε περιπτώσεις φωνηέντων χαμηλού τόνου, όπου η φωνητική ροή έχει μειωμένο κυματισμό στο διάστημα κλειστής φάσης ή για φωνήεντα υψηλότερο τόνου που το διάστημα κλειστής φάσης είναι μικρότερο, τότε μια χαμηλότερη τιμή του M είναι προτιμότερη (π.χ. M=8).

Συμπερασματικά, η μέθοδος SWLP με M=24 έχει καλύτερα αποτελέσματα απ' ότι με M=8 για όλες τις περιπτώσεις φωνημάτων που υπάρχουν στη βάση δεδομένων που χρησιμοποιείται σε αυτή την εργασία. [3]

3 Πειραματικά αποτελέσματα

3.1 Η βάση δεδομένων

Αναφορικά με τη βάση δεδομένων που χρησιμοποιήθηκε σε αυτή την εργασία, αυτή αποτελείται από αρχεία ήχου τύπου ".wav". Κάποια από αυτά τα αρχεία είναι αρχεία συνθετικής φωνής που περιλμβάνουν τα εξής φωνήματα: /aa/, /ae/, /eh/ και /ih/. Στην ουσία, πρόκειται για την παραγώμενη τελική φωνή που εξέρχεται από έναν άνθρωπο, μόνο που σε αυτή την περίπτωση είναι τεχνητή.

Όσον αφορά το φυσικό μοντέλο παραγωγικής φωνής, η δόνηση των φωνητικών χορδών έχει προσομοιωθεί με τρεις μάζες που συνδέονται μεταξύ τους μέσω παραγόντων ακαμψίας και απόσβεσης. Η διάταξη των μαζών αυτών σχεδιάστηκε με τέτοιο τρόπο, ώστε να αποτυπώσει τη δομή των φωνητικών χορδών. [9]

Με τη βοήθεια αεροδυναμικών και ακουστικών εκτιμήσεων, το τεχνητό μοντέλο φωνητικών χορδών συνδυάστηκε με πιέσεις και ροή αέρα στην τραχεία και στη φωνητική οδό. Έπειτα, υπολογίστηκε η διάδοση των ακουστικών κυμάτων στην περιοχή της τραχείας, αλλά και στη φωνητική οδό σε συγχρονισμό με το μοντέλο των φωνητικών χορδών. Ο υπολογισμός αυτός έγινε με την προσέγγιση μιας αντανάκλασης κύματος στο σημείο, όπου η λειτουργία των φωνητικών χορδών και η τραχεία διαχωρίστηκαν σε μικρούς κυλινδρικούς τομείς. Μέσω των συντελεστών ανάκλασης και μετάδοσης (οι οποίοι υπολογίστηκαν στις διασταυρώσεις των μικρών κυλινδρικών τομών) υπολογίστηκαν η πίεση και η ταχύτητα της έντασης, ώστε να διαδοθούν τα ακουστικά κύματα μέσω του συστήματος των φωνητικών χορδών. Παράλληλα, η γλωττιδική ροή προσδιορίστηκε από την αλληλεπίδραση της γλωττίδας με πιέσεις σε διάφορες χρονικές στιγμές. Τέλος, στα χείλη, τα κύματα πίεσης υποβάλλονται σε ένα φορτίο ακτινοβολίας ως αντίσταση παράλληλα με επαγωγή. Μάλιστα, η πίεση εξόδου στο μοντέλο αυτό αντκατοπτρίζει την πίεση που εκπέμπεται στα χείλη. Η υπολογισμένη πίεση εξόδου θα μπορούσε να παρομοιαστεί με την πίεση που προκαλεί ένα μικρόφωνο σε μη ανακλαστικό περιβάλλον. [9]

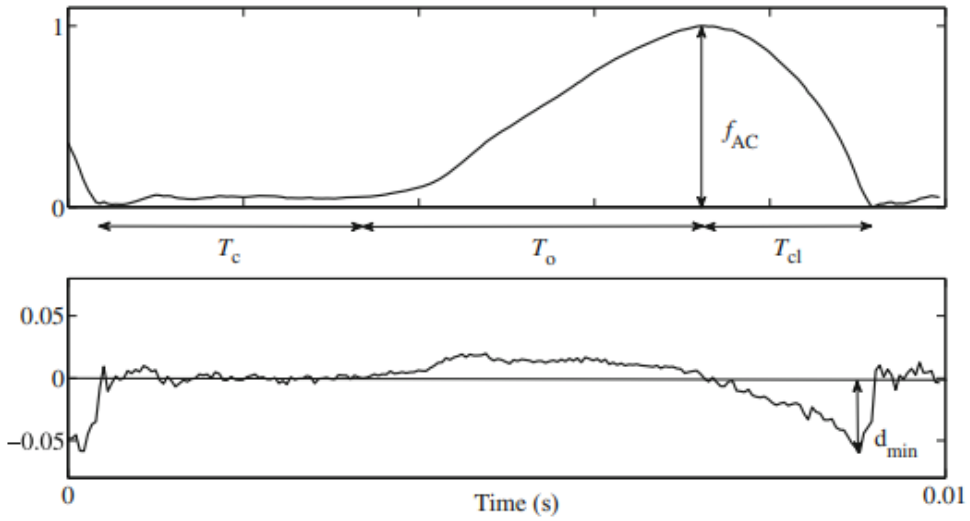
Επίσης, στη βάση αυτή, περιλαμβάνονται και αρχεία ήχου του σήματος της γλωττίδας, τα οποία εκτιμήθηκαν μέσω του αντίστροφου φιλτραρίσματος. Και αυτά τα αρχεία ήχου αφορούν προαναφερθέντα φωνήματα. Τα αρχεία ήχου που χρησιμοποιήθηκαν είχαν συχνότητα 8 kHz.

3.2 Μετρικές

Προκειμένου να εκτιμηθούν τα αποτελέσματα του γλωττιδικού σήματος που παράχθηκε και να αποφανθούμε για το πόσο κοντά είναι το σήμα αυτό με το πραγματικό σήμα, έγινε χρήση των παρακάτω μετρικών.

3.2.1 NAQ - Normalized amplitude quotient

Η πρώτη παράμετρος είναι η **Normalized amplitude quotient (NAQ)**. Πρόκειται για μία μετρική που υπολογίζεται στο πεδίο του χρόνου. Πιο συγκεκριμένα, το γλωττιδικό σήμα μπορεί να χωριστεί σε χρονικές θυρίδες και να οριστούν ως χρονικές στιγμές το άνοιγμα και το κλείσιμο της γλωττίδας, καθώς και η στιγμιαία μέγιστη ροή του αέρα. Η παράμετρος NAQ υπολογίζεται από το πλάτος της ροής του αέρα (AC) και το πλάτος της αρνητιικής κορυφής της παραγώγου της γλωττιδικής ροής. Μάλιστα, έχει αποδειχθεί πως ο λόγος αυτών των δύο τιμών έχουν ως αποτέλεσμα μία ποσότητα στο χρονικό πεδίο. Ο λόγος αυτός χαρακτηρίστηκε απο τον Fant ως η προβολή στον άξονα του χρόνου μιας εφαπτόμενης προς τη συνολική γλωττιδική ροή στο σημείο της διέγερσης, περιοριζόμενη απο τις τιμές 0 και το πλάτος AC της ροής. [4]



Εικόνα 9 [4]

Οπότε, η μετρική NAQ έχει τον εξής τύπο:

$$NAQ = \frac{f_{AC}}{d_{min} \cdot T}$$

Όπου $T = T_c + T_0 + T_{cl}$ και T_c είναι το χρονικό διάστημα κλειστής φάσης, T_0 το χρονικό διάστημα ανοιχτής φάσης και T_{cl} το χρονικό διάστημα φάσης κλεισίματος. [4]

3.2.2 QOQ - Quasi-open quotient

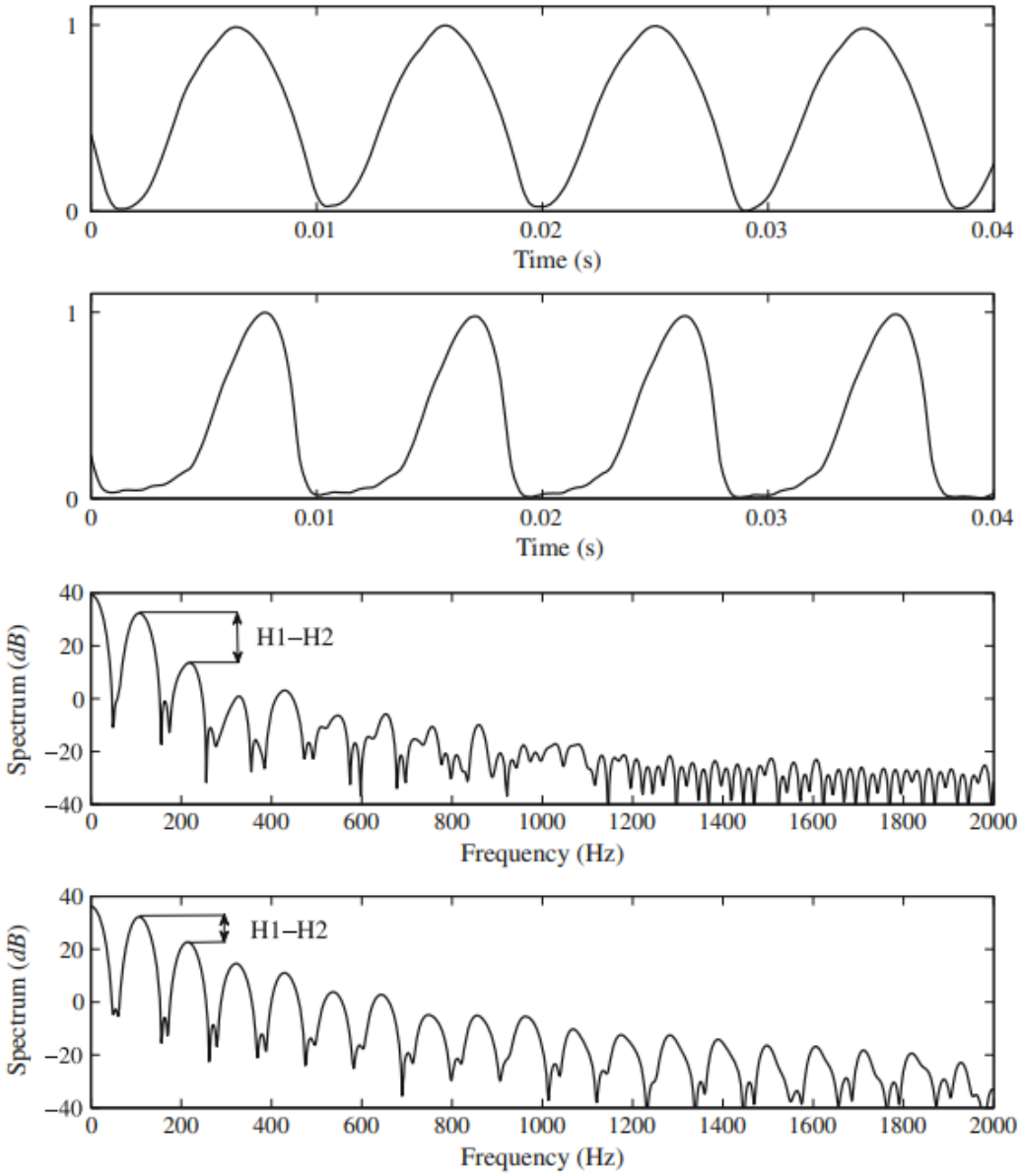
Η δεύτερη μετρική είναι η **Quasi-open quotient (QOQ)**. Αρχικά, η QOQ προέρχεται από την μετρική Open quotient (OQ), η οποία ορίστηκε με την εξαγωγή κρίσιμων χρονικών στιγμών από τις κυματομορφές της γλωττιδικής διέγερσης. Αναφορικά με την μετρική QOQ, αυτή αποτελεί την αναλογία του χρόνου του κύματος στον θετικό άξονα προς την περίοδο του γλωττιδικού σήματος. [5]

3.2.3 HRF - Harmonic richness factor

Μία ακόμη μετρική είναι η **Harmonic richness factor (HRF)**. Η μετρική αυτή αφορά τον τομέα της συχνότητας. Αναλυτικότερα, η μέτρηση της φασματικής φθοράς (spectral decay) της πηγήςυπολογίζεται χρησιμοποιώντας το πλάτος της θεμελειώδους συχνότητας (F_0) και τους πολλαπλάσιους ακέριαους αριθμούς, τις αρμονικές (harmonics). Στην ουσία, η μετρική HRF ορίζεται από το φάσμα της εκτιμώμενης γλωσσικής ροής ως ο λόγος μεταξύ του αθροίσματος των τιμών φάσματος πλάτους των αρμονικών πλην της θεμελειώδους προς τη φασματική τιμή της θεμελειώδους συχνότητας. [4]

3.2.4 H1H2 - Difference in the first two harmonic amplitudes

Η επόμενη μετρική είναι η **Difference in the first two harmonic amplitudes (H1H2)**. Εδώ, η φθορά της της πηγής υπολογίστηκε από τη διαφορά μεταξύ του πλάτους της θεμελειώδους και της δεύτερης αρμονικής. Τα H1 και H2 φαίνονται στο παρακάτω σχήμα (Εικόνα 10), όπου φαίνονται τα φάσματα πλάτους δύο διαφορετικών σημάτων γλωττιδικής ροής. [4]



Εικόνα 10 [4]

Τα διαγράμματα στην Εικόνα 10 αφορούν άνδρα ομιλητή. Στο πρώτο διάγραμμα απεικονίζεται η γλωττιδική ροή (ως προς το χρόνο) κατά την αναπνοή (breathy phonation). Στο δεύτερο απεικονίζει την γλωττιδική ροή (ως προς το χρόνο) κατά τη φάση ανοίγματος μετά από ένα διάστημα μιας μακράς κλειστής φάσης (pressed phonation). Στο τρίτο απεικονίζεται το φάσμα της φωνής κατά την αναπνοή (breathy phonation) και στο τέταρτο απεικονίζεται το φάσμα της φωνής κατά τη φάση ανοίγματος μετά από ένα διάστημα μιας μακράς κλειστής φάσης (pressed phonation). [4]

3.2.5 PSP - Parabolic spectral parameter

Η τελευταία μετρική που χρησιμοποιείται είναι η **Parabolic spectral parameter (PSP)**. Η μετρική αυτή βασίστηκε σε ένα pitch-synchronously υπολογισμένο φάσμα .Η PSP ταιριάζει ένα πολυώνυμο δεύτερης τάξης με το φάσμα της γλωττιδικής ροής που υπολογίζεται σε έναν γλωττιδικό κύκλο. [4]

3.3 Αποτελέσματα

Στα σήματα που μελετήθηκαν για τις δύο διαφορετικές μεθόδους αντίστροφου φιλτραρίσματος (IAIF και SWIAIF) σε αυτή την εργασία, έγινε εφαρμογή των μετρικών που αναλύθηκαν στο κεφάλαιο *3.2 Μετρικές*. Σκοπός της εφαρμογής των μετρικών στα δύο σήματα είναι η σύγκριση της μέσης τιμής του απόλυτου σφάλματος (Mean Absolute Error) που προκύπτει συγκρίνοντας τις τιμές των παραμέτρων για τα δύο μοντέλα (IAIF και SWIAIF) με τις τιμές των παραμέτρων του αρχικού γλωττιδικού σήματος της πηγής (original glottal source signal).

Στους παρακάτω πίνακες βρίσκει κανείς αυτή τη σύγκριση για δέκα διαφορετικές τιμές του M για τη μέθοδο SWIAIF. Αυτό συμβαίνει, διότι πρέπει να αποδειχθεί για ποια τιμή του M (από M=4 έως M=40 με βήμα 4) η μέθοδος SWIAIF δίνει καλύτερη μέση τιμή του απόλυτου σφάλματος από την IAIF.

Τα αποτελέσματα αυτά εξάχθηκαν από την εφαρμογή των μεθόδων IAIF και SWIAIF σε ένα αρχείο συνθετικής φωνής που περιλμβάνει το φώνημα /aa/ σε συχνότητα 8 kHz.

Mean Absolute Errors for all parameters

	IAIF	**SWIAIF (M=4)**
NAQ	0.0061 (± 0.0002)	0.0070 (± 0.0001)
QOQ	0.0159 (± 0.0014)	0.0186 (± 0.0017)
H1H2	0.4010 (± 0.0164)	0.4078 (± 0.0050)
HRF	1.0737 (± 0.0102)	3.1237 (± 0.0113)
PSP	0.0353 (± 0.0018)	0.0404 (± 0.0014)

Mean Absolute Errors for all parameters

	IAIF	**SWIAIF (M=8)**

NAQ	0.0061 (± 0.0002)	0.0063 (± 0.0001)
QOQ	0.0159 (± 0.0014)	0.0250 (± 0.0017)
H1H2	0.4010 (± 0.0164)	0.2779 (± 0.0061)
HRF	1.0737 (± 0.0102)	2.6100 (± 0.0169)
PSP	0.0353 (± 0.0018)	0.0408 (± 0.0015)

Mean Absolute Errors for all parameters

	IAIF	**SWIAIF (M=12)**
NAQ	0.0061 (± 0.0002)	0.0052 (± 0.0002)
QOQ	0.0159 (± 0.0014)	0.0234 (± 0.0016)
H1H2	0.4010 (± 0.0164)	0.2590 (± 0.0079)
HRF	1.0737 (± 0.0102)	1.9060 (± 0.0180)
PSP	0.0353 (± 0.0018)	0.0372 (± 0.0015)

Mean Absolute Errors for all parameters

	IAIF	**SWIAIF (M=16)**
NAQ	0.0061 (± 0.0002)	0.0058 (± 0.0002)
QOQ	0.0159 (± 0.0014)	0.0214 (± 0.0016)
H1H2	0.4010 (± 0.0164)	0.3042 (± 0.0122)
HRF	1.0737 (± 0.0102)	1.5199 (± 0.0180)
PSP	0.0353 (± 0.0018)	0.0371 (± 0.0018)

Mean Absolute Errors for all parameters

	IAIF	**SWIAIF (M=20)**
NAQ	0.0061 (± 0.0002)	0.0065 (± 0.0001)
QOQ	0.0159 (± 0.0014)	0.0189 (± 0.0014)
H1H2	0.4010 (± 0.0164)	0.3919 (± 0.0171)
HRF	1.0737 (± 0.0102)	0.4612 (± 0.0078)
PSP	0.0353 (± 0.0018)	0.0336 (± 0.0016)

Mean Absolute Errors for all parameters

	IAIF	**SWIAIF (M=24)**

NAQ	0.0061 (± 0.0002)	0.0073 (± 0.0001)
QOQ	0.0159 (± 0.0014)	0.0168 (± 0.0012)
H1H2	0.4010 (± 0.0164)	0.4582 (± 0.0206)
HRF	1.0737 (± 0.0102)	0.2434 (± 0.0062)
PSP	0.0353 (± 0.0018)	0.0336 (± 0.0015)

Mean Absolute Errors for all parameters

	IAIF	SWIAIF (M=28)
NAQ	0.0061 (± 0.0002)	0.0076 (± 0.0001)
QOQ	0.0159 (± 0.0014)	0.0183 (± 0.0012)
H1H2	0.4010 (± 0.0164)	0.4874 (± 0.0216)
HRF	1.0737 (± 0.0102)	0.2773 (± 0.0060)
PSP	0.0353 (± 0.0018)	0.0336 (± 0.0015)

Mean Absolute Errors for all parameters

	IAIF	SWIAIF (M=32)
NAQ	0.0061 (± 0.0002)	0.0079 (± 0.0001)
QOQ	0.0159 (± 0.0014)	0.0185 (± 0.0012)
H1H2	0.4010 (± 0.0164)	0.4996 (± 0.0219)
HRF	1.0737 (± 0.0102)	0.3146 (± 0.0059)
PSP	0.0353 (± 0.0018)	0.0337 (± 0.0015)

Mean Absolute Errors for all parameters

	IAIF	SWIAIF (M=36)
NAQ	0.0061 (± 0.0002)	0.0079 (± 0.0002)
QOQ	0.0159 (± 0.0014)	0.0185 (± 0.0012)
H1H2	0.4010 (± 0.0164)	0.4941 (± 0.0213)
HRF	1.0737 (± 0.0102)	0.3263 (± 0.0059)
PSP	0.0353 (± 0.0018)	0.0337 (± 0.0015)

Mean Absolute Errors for all parameters

	IAIF	SWIAIF (M=40)

NAQ	0.0061 (± 0.0002)	0.0080 (± 0.0002)
QOQ	0.0159 (± 0.0014)	0.0188 (± 0.0012)
H1H2	0.4010 (± 0.0164)	0.4864 (± 0.0208)
HRF	1.0737 (± 0.0102)	0.3279 (± 0.0059)
PSP	0.0353 (± 0.0018)	0.0336 (± 0.0015)

4 Συμπεράσματα και μελλοντική εργασία

Αναλύοντας κανείς τα αποτελέσματα στο κεφάλαιο *3.3 Αποτελέσματα* παρατηρεί για κάθε μετρική ξεχωριστά τα εξής.

Για τη μετρική NAQ φαίνεται πως για M=12 η μέθοδος SWIAIF δίνει καλύτερη μέση τιμή του απόλυτου σφάλματος από την IAIF.

Ενώ για τη μετρική QOQ είναι φανερό πως για καμία τιμή του M από 4 έως 40, η μέθοδος SWIAIF δεν δίνει καλύτερη μέση τιμή του απόλυτου σφάλματος από την IAIF. Φαίνεται πως μόνο για M=24 η μέση τιμή του απόλυτου σφάλματος της μεθόδου SWIAIF πλησιάζει τη μέση τιμή του απόλυτου σφάλματος της IAIF, αλλά δεν είναι μικρότερη της.

Για τη μετρική H1H2 και M=12 η μέθοδος SWIAIF δίνει καλύτερη μέση τιμή του απόλυτου σφάλματος από την IAIF.

Για την επόμενη μετρική, την HRF, φαίνεται πως για τις μεγαλύτερες τιμές του M η μέθοδος SWIAIF δίνει καλύτερη μέση τιμή του απόλυτου σφάλματος από την IAIF. Πιο συγκεκριμένα, για M=20, M=24, M=28, M=32, M=36 και M=40 η μέθοδος SWIAIF δίνει καλύτερη μέση τιμή του απόλυτου σφάλματος, με τη βέλτιστη τιμή για M=24.

Όσον αφορά την τελευταία μετρική, την PSP, για M=20, M=24, M=28 και M=40 η SWIAIF δίνει καλύτερη μέση τιμή του απόλυτου σφάλματος από την IAIF.

Αναφορικά με κάποια μελλοντικη εργασία, θα μπορούσε κανείς να χρησιμοποιήσει μία άλλη βάση δεδομένων, η οποία θα περιέχει αρχεία ήχου πραγματικής φωνής.

5 Βιβλιογραφία

[1] HY-578: Digital Speech Signal Processing (https://www.csd.uoc.gr/~hy578/)

[2] Glottal wave analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering, Paavo Alku, Helsinki University of Technology, Acoustics Laboratory, Otakaari5A, SF-02150Espoo, Finland

[3] GLOTTAL INVERSE FILTERING USING STABILISED WEIGHTED LINEAR PREDICTION, George P. Kafentzis1 , Yannis Stylianou1 , and Paavo Alku2, 1Institute of Computer Science, FORTH, and Multimedia Informatics Lab, CSD, UoC, Greece, 2Department of Signal Processing and Acoustics, Aalto University, Helsinki, Finland

[4] Glottal inverse filtering analysis of human voice production — A review of estimation and parameterization methods of the glottal excitation and their applications, PAAVO ALKU, Department of Signal Processing and Acoustics, Aalto University, Otakaari 5 A, P.O. Box 13000, 00076 Aalto, Finland

[5] Hacki, T. Klassifizierung von glottisdysfunktionen mit hilfe der elektroglottographie. Folia Phoniatrica, pages 43–48, 1989.

[6] https://opencourses.uoa.gr/modules/document/file.php/DI37/%CE%94%CE%B9%CE%B4%CE%B1%CE%BA%CF%84%CE%B9%CE%BA%CF%8C%20%CE%A0%CE%B1%CE%BA%CE%AD%CF%84%CE%BF/%CE%A0%CE%B1%CF%81%CE%BF%CF%85%CF%83%CE%B9%CE%AC%CF%83%CE%B5%CE%B9%CF%82/PDF/2.pdf Φεβρουάριος 2021

[7] William O'Grady, John Archibald, Francis Katamba, Michael Dobrovolsky, Ewa Czaykowska-Higgins, Contemporary Linguistics, Longman (2001)

[8] https://www.researchgate.net/profile/Gang-Chen-30/publication/277020864/figure/fig2/AS:669404161392658@1536609797780/The-linear-source-filter-model-of-speech-production-Fan70-Top-panels-the-model-in.png Φεβρουάριος 2021

[9] https://www.csd.uoc.gr/~kafentz/Thesis_Final.pdf