



Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής  
Πρόγραμμα Μεταπτυχιακών Σπουδών  
«Πληροφορική»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	<b>Μελέτη στατιστικών χαρακτηριστικών (pdf, power density spectrum, autocorrelation) σημάτων φωνής με χρήση του πακέτου λογισμικού Matlab</b> <b>Study of statistical features (pdf, power density spectrum, autocorrelation) of voice signals using the software package Matlab</b>
Όνοματεπώνυμο Φοιτητή	<b>Σοφία Χρύσα Μητράκα</b>
Πατρώνυμο	<b>Ιωάννης</b>
Αριθμός Μητρώου	<b>ΜΠΠΑ/12035</b>
Επιβλέπων	<b>Καθηγητής Γεωργιακώδης Φώτιος</b>

Πανεπιστήμιο Πειραιώς

Πανεπιστήμιο Πειραιώς

**Τριμελής Εξεταστική Επιτροπή**

Καθηγητής  
Γεωργιακώδης Φώτιος

Καθηγητής  
Σαπουνάκης Αριστείδης

Καθηγητής  
Τσικούρας Παναγιώτης

Πανεπιστήμιο Πειραιώς

---



## **ΕΥΧΑΡΙΣΤΙΕΣ**

Για την εργασία αυτή θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή Κύριο **Φώτιο Γεωργιακώδη**, για την εμπιστοσύνη που έδειξε στο πρόσωπό μου με την ανάθεση ενός ιδιαίτερα ενδιαφέροντος θέματος, ως πνευματικό επιστέγασμα της διετούς θητείας μου στο Μεταπτυχιακό Πρόγραμμα της «Πληροφορικής» του Πανεπιστημίου Πειραιώς.

Ο καθηγητής Κύριος Γεωργιακώδης στήριξε την προσπάθειά μου, με καθοδήγησε, μου αφιέρωσε πολύ από τον χρόνο του, προσφέροντάς μου τις πολύτιμες γνώσεις και συμβουλές του, για να φτάσουμε ως εδώ.

Θερμές ευχαριστίες απευθύνω και σε όλους τους καθηγητές που είχα τα δύο χρόνια του Μεταπτυχιακού αυτού Προγράμματος. Με βοήθησαν καθ' όλη τη διάρκεια της παρακολούθησης των μαθημάτων μου και αποτέλεσαν πρότυπα για μένα.

Πανεπιστήμιο Πειραιώς

Αφιερώνεται  
στη μνήμη του «μπάκη» μου

Πανεπιστήμιο Πειραιώς

## Πίνακας Περιεχομένων

Κατάλογος Εικόνων.....	11
Περίληψη .....	13
Abstract .....	15
<b>1</b> Εισαγωγή .....	17
1.1 Γενικά στοιχεία .....	17
1.2 Στόχος εργασίας .....	17
1.3 Δομή εργασίας .....	18
<b>2</b> Σύντομη περιγραφή στατιστικών χαρακτηριστικών .....	19
2.1 Τυχαίες μεταβλητές και συναρτήσεις κατανομής .....	19
2.2 Συνάρτηση πυκνότητας πιθανότητας.....	20
2.3 Φάσμα πυκνότητας ισχύος και Αυτοσυσχέτιση .....	23
2.3.1 Σειρές Fourier .....	23
2.3.2 Σήματα με πεπερασμένη ενέργεια (ενεργειακά σήματα) .....	25
2.3.3 Αιτιοκρατικά και Τυχαία-Στοχαστικά Σήματα .....	25
2.3.4 Τυχαίες και εργοδικές διαδικασίες.....	26
2.3.5 Το φάσμα πυκνότητας ισχύος .....	26
2.3.6 Το θεώρημα Parseval .....	28
2.3.7 Ορισμός Αυτοσυσχέτισης.....	29
2.3.8 Υπολογισμός μετασχηματισμού Fourier με βάση το θεώρημα του Parseval .....	31
2.3.9 Χρήση του DFT για τον υπολογισμό του φάσματος ισχύος (power spectrum) .....	34
<b>3</b> Βασικά χαρακτηριστικά σημάτων φωνής.....	35
3.1 Εισαγωγή .....	35
3.2 Χρόνο - Συχνοτική Ανάλυση των Σημάτων Ομιλίας.....	35
3.3 Το παράθυρο ανάλυσης.....	37
3.4 Επίδραση του Μήκους Παραθύρου .....	38
3.5 Ανάλυση Σημάτων Ομιλίας .....	38
3.5.1 Γραμμική Πρόβλεψη και Ανάλυση Σημάτων Ομιλίας .....	39
<b>4</b> Περιγραφή μεθόδου μελέτης των στατιστικών χαρακτηριστικών με χρήση του πακέτου λογισμικού Matlab.....	41
4.1 Περιγραφή εισόδων και εξόδων του κώδικα .....	41
4.2 Σύντομη παρουσίαση του κώδικα.....	42
<b>5</b> Αποτελέσματα εφαρμογής μεθόδου.....	43
5.1 Αποτελέσματα για 1 <sup>ο</sup> αρχείο ήχου – χαμηλό sampling rate.....	43
5.2 Αποτελέσματα για 1 <sup>ο</sup> αρχείο ήχου – υψηλό sampling rate .....	48
5.3 Αποτελέσματα για 2 <sup>ο</sup> αρχείο ήχου – χαμηλό sampling rate.....	53
5.4 Αποτελέσματα για 2 <sup>ο</sup> αρχείο ήχου – υψηλό sampling rate .....	58

5.5	Αποτελέσματα για 3 <sup>ο</sup> αρχείο ήχου – χαμηλό sampling rate.....	63
5.6	Αποτελέσματα για 3 <sup>ο</sup> αρχείο ήχου – υψηλό sampling rate .....	68
<b>6</b>	Συμπεράσματα .....	73
	Βιβλιογραφία .....	75
	Παράρτημα: Ο Κώδικας.....	77

## Κατάλογος Εικόνων

Εικόνα 1: Παράδειγμα (α) νομοτελειακού σήματος και (β) στοχαστικού σήματος.....	25
Εικόνα 2: 3 παραδείγματα περιοδικών σημάτων.....	27
Εικόνα 3: Σειρές Fourier των 3 <sup>ων</sup> περιοδικών σημάτων.....	27
Εικόνα 4: Σύστημα επεξεργασίας σημάτων ομιλίας.....	38
Εικόνα 5: Γραφικό περιβάλλον επεξεργασίας σημάτων ήχου.....	41
Εικόνα 6: 1 <sup>ο</sup> Αποτελέσματα για το 1 <sup>ο</sup> αρχείο ήχου: Εισαγωγή παραμέτρων.....	43
Εικόνα 7: Αποτελέσματα για το 1 <sup>ο</sup> αρχείο ήχου: Ιστόγραμμα.....	44
Εικόνα 8: Αποτελέσματα για το 1 <sup>ο</sup> αρχείο ήχου: Φάσμα ισχύος.....	45
Εικόνα 9: Αποτελέσματα για το 1 <sup>ο</sup> αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε second).....	46
Εικόνα 10: Αποτελέσματα για το 1 <sup>ο</sup> αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε samples).....	47
Εικόνα 11: 1 <sup>ο</sup> Αποτελέσματα για το 1 <sup>ο</sup> αρχείο ήχου: Εισαγωγή παραμέτρων.....	48
Εικόνα 12: Αποτελέσματα για το 1 <sup>ο</sup> αρχείο ήχου: Ιστόγραμμα.....	49
Εικόνα 13: Αποτελέσματα για το 1 <sup>ο</sup> αρχείο ήχου: Φάσμα ισχύος.....	50
Εικόνα 14: Αποτελέσματα για το 1 <sup>ο</sup> αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε second).....	51
Εικόνα 15: Αποτελέσματα για το 1 <sup>ο</sup> αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε samples).....	52
Εικόνα 16: 2 <sup>ο</sup> Αποτελέσματα για το 1 <sup>ο</sup> αρχείο ήχου: Εισαγωγή παραμέτρων.....	53
Εικόνα 17: Αποτελέσματα για το 2 <sup>ο</sup> αρχείο ήχου: Ιστόγραμμα.....	54
Εικόνα 18: Αποτελέσματα για το 2 <sup>ο</sup> αρχείο ήχου: Φάσμα ισχύος.....	55
Εικόνα 19: Αποτελέσματα για το 2 <sup>ο</sup> αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε second).....	56
Εικόνα 20: Αποτελέσματα για το 2 <sup>ο</sup> αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε samples).....	57
Εικόνα 21: 1 <sup>ο</sup> Αποτελέσματα για το 2 <sup>ο</sup> αρχείο ήχου: Εισαγωγή παραμέτρων (υψηλός ρυθμός δειγματοληψίας).....	58
Εικόνα 22: Αποτελέσματα για το 2 <sup>ο</sup> αρχείο ήχου: Ιστόγραμμα (υψηλός ρυθμός δειγματοληψίας).....	59
Εικόνα 23: Αποτελέσματα για το 2 <sup>ο</sup> αρχείο ήχου: Φάσμα ισχύος (υψηλός ρυθμός δειγματοληψίας).....	60

Εικόνα 24: Αποτελέσματα για το 2 <sup>ο</sup> αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε second) (υψηλός ρυθμός δειγματοληψίας) .....	61
Εικόνα 25: Αποτελέσματα για το 2 <sup>ο</sup> αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε samples) (υψηλός ρυθμός δειγματοληψίας).....	62
Εικόνα 26: 2 <sup>ο</sup> Αποτελέσματα για το 3 <sup>ο</sup> αρχείο ήχου: Εισαγωγή παραμέτρων.....	63
Εικόνα 27: Αποτελέσματα για το 3 <sup>ο</sup> αρχείο ήχου: Ιστόγραμμα .....	64
Εικόνα 28: Αποτελέσματα για το 3 <sup>ο</sup> αρχείο ήχου: Φάσμα ισχύος .....	65
Εικόνα 29: Αποτελέσματα για το 3 <sup>ο</sup> αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε second) .....	66
Εικόνα 30: Αποτελέσματα για το 3 <sup>ο</sup> αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε samples) .....	67
Εικόνα 31: Αποτελέσματα για το 3 <sup>ο</sup> αρχείο ήχου: Εισαγωγή παραμέτρων (υψηλός ρυθμός δειγματοληψίας) .....	68
Εικόνα 32: Αποτελέσματα για το 3 <sup>ο</sup> αρχείο ήχου: Ιστόγραμμα (υψηλός ρυθμός δειγματοληψίας) .....	69
Εικόνα 33: Αποτελέσματα για το 3 <sup>ο</sup> αρχείο ήχου: Φάσμα ισχύος (υψηλός ρυθμός δειγματοληψίας) .....	70
Εικόνα 34: Αποτελέσματα για το 3 <sup>ο</sup> αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε second) (υψηλός ρυθμός δειγματοληψίας) .....	71
Εικόνα 35: Αποτελέσματα για το 3 <sup>ο</sup> αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε samples) (υψηλός ρυθμός δειγματοληψίας).....	72



## Περίληψη

Μέσω της εργασίας αυτής έγινε μια προσπάθεια να μελετηθούν κάποια βασικά στατιστικά χαρακτηριστικά σημάτων φωνής, τα οποία είναι απαραίτητα για την μετέπειτα επεξεργασία τους. Τέτοια στατιστικά χαρακτηριστικά είναι η probability density function (pdf, συνάρτηση πυκνότητας πιθανότητας), το power density spectrum (φάσμα πυκνότητας ισχύος) και το autocorrelation (αυτοσυσχέτιση).

Στην εργασία έγινε πρώτα μια περιγραφή των βασικών χαρακτηριστικών της θεωρίας, σχετικά με έννοιες όπως οι τυχαίες μεταβλητές και η συνάρτηση κατανομής, αλλά και η συνάρτηση πυκνότητας πιθανότητας. Επίσης, παρουσιάστηκαν τα βασικά χαρακτηριστικά του φάσματος πυκνότητας ισχύος και της αυτοσυσχέτισης, με αναφορά σε βασικούς ορισμούς εννοιών όπως οι σειρές Fourier, τα αιτιοκρατικά και τα τυχαία / στοχαστικά σήματα, το θεώρημα Parseval και άλλα. Επιπλέον, έγινε παρουσίαση των βασικών χαρακτηριστικών των σημάτων φωνής, εστιάζοντας στην χρόνο – συχνοτική ανάλυση των σημάτων ομιλίας, στην επίδραση του μήκους παραθύρου αλλά και σε βασικά χαρακτηριστικά της ανάλυσης τέτοιων σημάτων.

Στο πειραματικό μέρος με τη βοήθεια του πακέτου λογισμικού Matlab μελετήθηκαν τα τρία επιμέρους στατιστικά χαρακτηριστικά σημάτων φωνής, δηλαδή η συνάρτηση πυκνότητας πιθανότητας, το φάσμα πυκνότητας ισχύος και η αυτοσυσχέτιση. Παρουσιάστηκαν τα αποτελέσματα της εφαρμογής του κώδικα για τρία διαφορετικά αρχεία ήχου με διαφορετικά χαρακτηριστικά, και για δύο διαφορετικές τιμές της συχνότητας δειγματοληψίας. Τα συμπεράσματα που προέκυψαν είναι ιδιαίτερα ενδιαφέροντα, και η ανάλυση αυτή μπορεί να γίνει για μια πληθώρα σημάτων ήχου με διαφορετικά χαρακτηριστικά, ώστε να αποκτήσει ο ερευνητής πλήρη γνώση των βασικών χαρακτηριστικών του εκάστοτε σήματος προτού προχωρήσει σε περαιτέρω επεξεργασία και ανάλυσή του.

Πανεπιστήμιο Πειραιώς

## Abstract

Through this work an attempt was made to study some basic statistical characteristics of voice signals, which is necessary for subsequent processing. Such statistical characteristics is the probability density function (pdf), the power density spectrum and the autocorrelation function.

In this thesis, first a description of the main features of the theory took place, on concepts such as random variables and the distribution function, and the probability density function. Also the main features of the spectrum power density and autocorrelation are presented, with reference to the basic definitions of concepts such series Fourier, deterministic and random / stochastic signals, Parseval's theorem and more. In addition, a presentation of the key characteristics of voice signals was made, focusing on time - frequency analysis of speech signals, the effect of the window length and key features in the analysis of such signals.

In the experimental part, and using the Matlab software package, we studied three different statistical characteristics of voice signals, i.e. the probability density function, the power density spectrum and the autocorrelation function. The results of the implementation of code were presented for three different audio files with different attributes, and for two different values of the sampling frequency. The conclusions are particularly interesting, and the analysis can be done for a variety of audio signals with different characteristics, in order to acquire the researcher full knowledge of the basic characteristics of each signal before proceeding with further processing and analysis.

Πανεπιστήμιο Πειραιώς

## 1 Εισαγωγή

### 1.1 Γενικά στοιχεία

Η μελέτη και η επεξεργασία σημάτων ήχου αποτελεί μία από τις πιο σημαντικές περιοχές στις οποίες εφαρμόζονται τεχνικές ψηφιακής επεξεργασίας σήματος. Υπάρχουν κάποια ιδιαίτερα σημαντικά προβλήματα επεξεργασίας τέτοιων αρχείων, μεταξύ των οποίων είναι η κατάτμηση του σήματος, η μελέτη του φάσματος ισχύος του, η κωδικοποίηση του, αλλά και άλλες διαδικασίες που αποτελούν ενδιαφέρουσες περιπτώσεις στις οποίες μπορεί να εφαρμοστεί πληθώρα διαφορετικών αλγορίθμων ([2], [6]).

Σε κάθε περίπτωση και ανεξάρτητα από το είδος της επεξεργασίας, είναι πολύ σημαντικό να είναι σαφείς οι ιδιότητες των σημάτων ήχου. Σε αυτό το πλαίσιο, η παρούσα εργασία καλείται να αποτυπώσει τα βασικά χαρακτηριστικά και τις ιδιότητες των σημάτων αυτών, και να καταδείξει τις προκλήσεις που εμφανίζονται στην επεξεργασία αυτών των σημάτων, έχοντας ως βασικό στόχο να αποτυπώσει μερικές από τις ιδιότητές τους και να αναδείξει με αυτόν τον τρόπο την ανάγκη εφαρμογής βασικών τεχνικών της επεξεργασίας σημάτων.

Γενικά, τα σήματα ήχου (αλλά και ειδικότερα τα σήματα ομιλίας) είναι σήματα με ιδιαίτερη πολυπλοκότητα, καθώς περιέχουν όχι μόνο την βασική πληροφορία του μηνύματος, όσο και πληροφορίες που σχετίζονται με τον ομιλητή (παραδείγματος χάριν το φύλο, κάποια ένδειξη για την ηλικία του και άλλα). Η φωνή ουσιαστικά παράγεται μέσα σε έναν αγωγό μεταβλητής διατομής που αρχίζει από τις φωνητικές χορδές και τελειώνει στα χείλη. Ένας πρόσθετος αγωγός, η ρινική κοιλότητα, συμμετέχει στην παραγωγή των ένρινων φωνημάτων, κάνοντας την όλη διαδικασία ιδιαίτερα σύνθετη.

### 1.2 Στόχος εργασίας

Έχοντας υπόψη τα παραπάνω, η παρούσα εργασία αποσκοπεί στο να μελετήσει κάποια βασικά στατιστικά χαρακτηριστικά τέτοιων σημάτων, τα οποία είναι απαραίτητα για την μετέπειτα επεξεργασία τους. Τέτοια στατιστικά χαρακτηριστικά είναι η συνάρτηση πυκνότητας πιθανότητας (pdf, probability density function), το φάσμα πυκνότητας ισχύος (power density spectrum) και το αυτοσυσχέτιση (autocorrelation).

### 1.3 Δομή εργασίας

Σε αυτό το πλαίσιο, γίνεται πρώτα μια περιγραφή των βασικών χαρακτηριστικών της θεωρίας, σχετικά με έννοιες όπως οι τυχαίες μεταβλητές και η συνάρτηση κατανομής, αλλά και η συνάρτηση πυκνότητας πιθανότητας. Εν συνεχεία, παρουσιάζονται τα βασικά χαρακτηριστικά του φάσματος πυκνότητας ισχύος και της αυτοσυσχέτισης, με αναφορά σε βασικούς ορισμούς εννοιών όπως οι σειρές Fourier, τα αιτιοκρατικά και τα τυχαία/ στοχαστικά σήματα, το θεώρημα Parseval και άλλα.

Στο τρίτο κεφάλαιο της εργασίας παρουσιάζονται τα βασικά χαρακτηριστικά των σημάτων φωνής, εστιάζοντας στην χρόνο – συχνотική ανάλυση των σημάτων ομιλίας, στην επίδραση του μήκους παραθύρου αλλά και σε βασικά χαρακτηριστικά της ανάλυσης τέτοιων σημάτων. Εν συνεχεία, στο τέταρτο κεφάλαιο γίνεται περιγραφή της μεθόδου μελέτης των στατιστικών χαρακτηριστικών με χρήση του πακέτου λογισμικού Matlab, και περιγράφονται οι είσοδοι και οι έξοδοι του κώδικα που χρησιμοποιήθηκε καθώς και τα βασικά χαρακτηριστικά του.

Στο 5<sup>ο</sup> κεφάλαιο της εργασίας παρουσιάζονται τα αποτελέσματα της εφαρμογής του κώδικα για τρία διαφορετικά αρχεία ήχου με διαφορετικά χαρακτηριστικά, και για δύο διαφορετικές τιμές της συχνότητας δειγματοληψίας. Τέλος, η εργασία ολοκληρώνεται με την παρουσίαση των βασικότερων συμπερασμάτων.

## 2 Σύντομη περιγραφή στατιστικών χαρακτηριστικών

### 2.1 Τυχαίες μεταβλητές και συναρτήσεις κατανομής

Τα δειγματικά σημεία (στοιχειώδη ενδεχόμενα) ενός δειγματικού χώρου στοχαστικού (τυχαίου) πειράματος (ή φαινομένου) δύνανται να είναι αριθμοί, όπως για παράδειγμα στην περίπτωση που εκφράζουν ποσοτικό χαρακτηριστικό του στοχαστικού πειράματος, ή συμβολικές εκφράσεις με γράμματα της αλφαβήτου, όπως για παράδειγμα στην περίπτωση που περιγράφουν ποιοτικό χαρακτηριστικό του στοχαστικού πειράματος. Οι περιπτώσεις αυτές αντιμετωπίζονται ενιαία με την αντιστοίχιση σε κάθε δειγματικό σημείο ενός πραγματικού αριθμού. Επιπλέον, σε ένα στοχαστικό (τυχαίο) πείραμα (ή φαινόμενο) το ενδιαφέρον και από πρακτική άποψη εστιάζεται στην πραγματοποίηση ή μη αριθμητικών μεγεθών τα οποία αντιστοιχούν σε δειγματικά σημεία.

Έστω λοιπόν  $\Omega$  είναι ο δειγματικός χώρος ενός στοχαστικού (τυχαίου) πειράματος. Μια πραγματική συνάρτηση  $X$  που ορίζεται στο δειγματικό χώρο  $\Omega$  καλείται τυχαία μεταβλητή (τ.μ.). Η συνάρτηση αυτή αντιστοιχεί σε κάθε δειγματικό σημείο  $\omega \in \Omega$  έναν πραγματικό αριθμό  $x = X(\omega)$ .

Σημειώνουμε ότι οι τυχαίες μεταβλητές συμβολίζονται με τα κεφαλαία γράμματα χωρίς δείκτες  $X, Y, Z, W$  ή με δείκτες  $X_1, X_2, \dots, X_k$  και οι τιμές τους με τα αντίστοιχα μικρά γράμματα  $x, y, z, w$  ή  $x_1, x_2, \dots, x_k$ .

Το σύνολο  $R_X \subseteq \mathbb{R}$  των τιμών της τυχαίας μεταβλητής  $X$  αποτελεί το νέο δειγματικό χώρο του στοχαστικού (τυχαίου) πειράματος (ή φαινομένου). Το διάστημα  $(-\infty, x]$  είναι βασικό ενδεχόμενο στο νέο αυτόν δειγματικό χώρο. Οποιοδήποτε άλλο ενδεχόμενο  $B \subseteq R_X$  δύναται να εκφρασθεί (ή να προσεγγιστεί) συναρτήσει τέτοιων διαστημάτων. Είναι επομένως χρήσιμη η εισαγωγή της ακόλουθης συνάρτησης [1].

Αντίστοιχα, η συνάρτηση  $F$  η οποία ορίζεται από τη σχέση

$$F(x) = P(X \leq x) = P(\{\omega \in \Omega: X(\omega) \leq x\}), \quad -\infty < x < \infty$$

καλείται συνάρτηση κατανομής (σ.κ.) ή αθροιστική συνάρτηση κατανομής (α.σ.κ.) της τ.μ.  $X$ .

Στις περιπτώσεις που υπάρχει κίνδυνος σύγχυσης η συνάρτηση κατανομής της τ.μ.  $X$  συμβολίζεται με  $F_X$  και η τιμή της στο  $x$  με  $F_X(x)$ . Σημειώνουμε ότι η συνάρτηση κατανομής, ως πιθανότητα, λαμβάνει τιμές στο διάστημα  $[0, 1]$ :

$$0 \leq F(x) \leq 1, \quad -\infty < x < \infty.$$

Επίσης είναι αύξουσα συνάρτηση,

$$F(x_1) \leq F(x_2), \quad -\infty < x_1 \leq x_2 < \infty,$$

επειδή  $\{\omega \in \Omega: X(\omega) \leq x_1\} \subseteq \{\omega \in \Omega: X(\omega) \leq x_2\}$  και ισχύει

$$F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0, \quad F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1$$

Τέλος σημειώνουμε ότι οποιαδήποτε συνάρτηση κατανομής είναι δεξιά συνεχής. Η πιθανότητα όπως μια τυχαία μεταβλητή βρίσκεται σε συγκεκριμένο διάστημα των πραγματικών αριθμών δύναται να εκφρασθεί συναρτήσει της συνάρτησης κατανομής της. Σε αυτό το πλαίσιο, αν  $F$  η συνάρτηση κατανομής μιας τυχαίας μεταβλητής  $X$ , τότε

$$P(\alpha < X \leq \beta) = F(\beta) - F(\alpha)$$

για κάθε πραγματικούς αριθμούς  $\alpha$  και  $\beta$  με  $\alpha < \beta$  [2].

## 2.2 Συνάρτηση πυκνότητας πιθανότητας

Η μελέτη πολλών σημαντικών εννοιών που συνδέονται με τις τυχαίες μεταβλητές διευκολύνεται με το διαχωρισμό των δύο βασικών κατηγοριών: των διακριτών και των συνεχών τυχαίων μεταβλητών.

Μία τυχαία μεταβλητή  $X$  καλείται διακριτή (ή απαριθμητή) αν παίρνει, με πιθανότητα 1, αριθμήσιμο (πεπερασμένο ή αριθμησίμως άπειρο) σύνολο τιμών  $R_X = \{x_0, x_1, \dots, x_\nu, \dots\}$ . Η συνάρτηση  $f$  η οποία σε κάθε σημείο  $(x_k)$ ,  $k = 0, 1, 2, \dots$ , εκχωρεί την πιθανότητά του

$$f(x_k) = P(X = x_k) = P(\{\omega \in \Omega : X(\omega) = x_k\}), k = 0, 1, 2, \dots$$

καλείται συνάρτηση πιθανότητας της τυχαίας μεταβλητής  $X$  [4].

Στις περιπτώσεις που υπάρχει κίνδυνος σύγχυσης η συνάρτηση πιθανότητας της τ.μ.  $X$  συμβολίζεται με  $f_X$  και η τιμή της στο  $x_k$  με  $f_X(x_k)$ .

Σημειώνουμε ότι, χρησιμοποιώντας την παράσταση  $R_X = \{x_0\} \cup \{x_1\} \cup \dots \cup \{x_\nu\} \cup \dots$ , η συνθήκη  $P(X \in R_X)$  δύναται να γραφεί στη μορφή

$$\sum_{k=0}^{\infty} P(X = x_k) = 1$$

Επίσης η συνάρτηση πιθανότητας, όπως προκύπτει άμεσα από τον ορισμό της, είναι μη αρνητική

$$f(x_k) \geq 0, k = 0, 1, 2, \dots \text{ και } f(x) = 0, x \notin R_X$$

και

$$\sum_{k=0}^{\infty} f(x_k) = 1$$

Στην περίπτωση που το σύνολο των τιμών της τυχαίας μεταβλητής  $X$  είναι πεπερασμένο,  $R_X = \{x_0, x_1, \dots, x_\nu\}$ , η παραπάνω σειρά γίνεται ένα πεπερασμένο άθροισμα [3]:

$$\sum_{k=0}^{\nu} f(x_k) = 1$$

Η συνάρτηση πιθανότητας  $f(x_k) = P(X = x_k)$ ,  $k = 0, 1, 2, \dots$  μιας διακριτής τυχαίας μεταβλητής συνδέεται με τη συνάρτηση κατανομής αυτής  $F(x) = P(X \leq x)$ ,  $-\infty < x < +\infty$ . Συγκεκριμένα, στη μερική περίπτωση που  $x_0 < x_1 < x_2 < L$ , ισχύουν οι σχέσεις

$$F(x) = \sum_{k=0}^r f(x_k), x_r \leq x < x_{r+1}, r = 0, 1, 2, \dots$$

με  $F(x) = 0$  για  $-\infty < x < x_0$  και

$$f(x_k) = F(x_k) - F(x_{k-1}), k = 0, 1, 2, \dots$$

με  $f(x_0) = F(x_0)$ . Γενικότερα ισχύει η σχέση

$$F(x) = \sum_{x_k < x} f(x_k), -\infty < x < \infty$$

όπου η άθροιση εκτείνεται σε όλα τα  $x_k$  τα οποία είναι μικρότερα ή ίσα του  $x$  [5].

Σημειώνουμε ότι η συνάρτηση κατανομής  $F$  μιας διακριτής τ.μ.  $X$  είναι σταθερή κατά διαστήματα και αυξάνει μόνο με άλματα στα σημεία  $x_k \in R_X$ .

Μία τυχαία μεταβλητή  $X$  καλείται συνεχής αν υπάρχει μη αρνητική συνάρτηση

$$f(x) \geq 0, -\infty < x < \infty, \text{ με}$$

$$\int_{-\infty}^{\infty} f(x) dx = 1$$



τέτοια ώστε για κάθε πραγματικούς αριθμούς  $\alpha$  και  $\beta$  με  $\alpha < \beta$  να ισχύει

$$P(\alpha < X \leq \beta) = \int_{\alpha}^{\beta} f(x)dx$$

Η  $f(x)$  καλείται πυκνότητα πιθανότητας ή απλώς πυκνότητα της τυχαίας μεταβλητής  $X$  [7].

Άμεση συνέπεια των ορισμών της συνάρτησης κατανομής  $F(x)$  και της συνάρτησης πυκνότητας  $f(x)$  μιας συνεχούς τυχαίας μεταβλητής  $X$  είναι η σχέση

$$F(x) = \int_{-\infty}^x f(t)dt, -\infty < x < \infty$$

που δείχνει ότι η συνάρτηση κατανομής  $F$  μιας συνεχούς τ.μ.  $X$  είναι συνεχής συνάρτηση. Συνεπώς, αν η  $X$  είναι συνεχής τ.μ., τότε για κάθε

$$x \in \mathbb{R}, P(X < x) = F(x) = P(X \leq x).$$

Αν η συνάρτηση  $f(x)$  είναι συνεχής στο σημείο  $x$  τότε παραγωγίζοντας παίρνουμε την

$$F'(x) = \frac{dF(x)}{dx} = f(x)$$

Οι παραπάνω σχέσεις είναι οι αντίστοιχες των προηγούμενων για συνεχείς τυχαίες μεταβλητές. Η πυκνότητα  $f(x)$ , σε αντίθεση με τη συνάρτηση πιθανότητας, δεν παριστάνει την πιθανότητα κάποιου ενδεχομένου. Η πιθανότητα  $P(X = x_0) = 0$  και επομένως η  $f(x_0)$  δεν παριστάνει βέβαια αυτή την πιθανότητα. Μόνον όταν η συνάρτηση αυτή ολοκληρώνεται μεταξύ δύο σημείων, δίνει κάποια πιθανότητα. Κατά προσέγγιση για μικρό  $\Delta x > 0$  έχουμε  $P(x < X \leq x + \Delta x) \cong f(x)\Delta x$ .

Εστιάζοντας στην κατανομή συνάρτησης μιας τυχαίας μεταβλητής, είναι γεγονός πως στην πιθανοθεωρητική μελέτη ενός στοχαστικού (τυχαίου) πειράματος (ή φαινομένου), όπως επίσης και στη στατιστική συμπερασματολογία, αναφέρεται συχνά η ανάγκη προσδιορισμού της κατανομής μιας τυχαίας μεταβλητής  $Y = g(X)$ , η οποία είναι συνάρτηση μιας άλλης τυχαίας μεταβλητής  $X$  με γνωστή κατανομή [6].

Συνήθως το ενδιαφέρον αφορά την περίπτωση που τόσο η τυχαία μεταβλητή  $X$  όσο και η τυχαία μεταβλητή  $Y$  είναι συνεχείς. Στην περίπτωση αυτή ο προσδιορισμός της κατανομής της  $Y$  επιτυγχάνεται ευκολότερα με την εύρεση, αρχικά, της συνάρτησης κατανομής. Η πυκνότητα της  $Y$  προσδιορίζεται με παραγωγή της συνάρτησης κατανομής. Η έκφραση της συνάρτησης κατανομής της τυχαίας μεταβλητής  $Y = g(X)$ ,  $F_Y(y) = P(Y \leq y) = P[g(X) \leq y]$  συναρτήσει της συνάρτησης κατανομής της τυχαίας μεταβλητής  $X$  απαιτεί τον προσδιορισμό του συνόλου  $\{x : g(x) \leq y\}$ . Τούτο επιτυγχάνεται εύκολα αν ο μετασχηματισμός  $y = g(x)$  είναι ένα προς ένα από το σύνολο  $R_X$ . Η συνάρτηση  $f$  η οποία σε κάθε σημείο  $(x_k)$ ,  $k = 0, 1, 2, \dots$ , εκχωρεί των τιμών της  $X$  επί του συνόλου  $\mathbb{R}_y$ . Η συνάρτηση  $f$  η οποία σε κάθε σημείο  $(x_k)$ ,  $k = 0, 1, 2, \dots$ , εκχωρεί των τιμών της  $Y$  και γνησίως μονότονος. Τότε υπάρχει ο αντίστροφος μετασχηματισμός  $x = g^{-1}(y)$  και είναι γνησίως μονότονος. Στην περίπτωση αυτή η σχέση  $g(x) \leq y$  είναι ισοδύναμη με τη σχέση  $x \leq g^{-1}(y)$ , αν η  $y = g(x)$  είναι γνησίως αύξουσα και με τη σχέση  $x \geq g^{-1}(y)$ , αν η  $y = g(x)$  είναι γνησίως φθίνουσα και επομένως [8]:

$$F_Y(y) = P[X \leq g^{-1}(y)] = F_X(g^{-1}(y))$$

αν η  $y = g(x)$  είναι γνησίως αύξουσα και

$$F_Y(y) = P[X \geq g^{-1}(y)] = 1 - P[X < g^{-1}(y)] = 1 - F_X(g^{-1}(y))$$

αν η  $y = g(x)$  είναι φθίνουσα.

Αν η αντίστροφη συνάρτηση  $x = g^{-1}(y)$  παραγωγίζεται και η παράγωγος  $\frac{dg^{-1}(y)}{dy}$

είναι συνεχής για κάθε  $y$  στο  $R_Y$ , τότε παραγωγίζοντας την ανωτέρω έκφραση της συνάρτησης κατανομής  $F_Y(y)$ , σύμφωνα με τον κανόνα παραγωγίσης σύνθετης συνάρτησης, συνάγουμε τη σχέση [9]

$$f_Y(y) = f_x(g^{-1}(y)) \frac{dg^{-1}(y)}{dy}$$

αν η  $y = g(x)$  είναι γνησίως αύξουσα και τη σχέση

$$f_Y(y) = -f_x(g^{-1}(y)) \frac{dg^{-1}(y)}{dy}$$

αν η  $y = g(x)$  είναι γνησίως φθίνουσα. Επομένως

$$f_Y(y) = f_x(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|$$

Τα αποτελέσματα αυτά συνοψίζονται στην παρακάτω διατύπωση:

Έστω ότι η  $X$  είναι μια συνεχής τυχαιά μεταβλητή με πυκνότητα  $f_X(x)$ ,  $x \in R_X$ . Αν ο μετασχηματισμός  $Y = g(X)$  είναι γνησίως μονότονος από το σύνολο  $R_X$  επί του συνόλου  $R_Y = g(R_X)$  και υπάρχει η παράγωγος  $\frac{dg^{-1}(y)}{dy}$  και είναι συνεχής για κάθε  $y$  στο  $R_Y$ , τότε η

τυχαιά μεταβλητή  $Y = g(X)$  είναι συνεχής με πυκνότητα  $f_Y(y) = f_x(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|$ .

[16]

## 2.3 Φάσμα πυκνότητας ισχύος και Αυτοσυσχέτιση

### 2.3.1 Σειρές Fourier

Η προσέγγιση συναρτήσεων μέσω πολυωνύμων, παρά την αποτελεσματικότητα και την, σχετική, απλότητά της, αποδεικνύεται ανεπαρκής για την περιγραφή/προσέγγιση περιοδικών συναρτήσεων. Το πρόβλημα αυτό αντιμετωπίζεται με την διαδικασία ανάπτυξης συναρτήσεων σε *σειρές Fourier*<sup>1</sup> κατά την οποία οι υπό μελέτη συναρτήσεις προσεγγίζονται από απεικονίσεις της μορφής

$$\phi_k(x) = \sum_{n=0}^k (a_n \cos(nx) + \beta_n \sin(nx))$$

όπου οι συντελεστές  $a_n, \beta_n$  είναι πραγματικοί αριθμοί ενώ ο φυσικός  $k$  (**βαθμός** της  $\phi_k(x)$ ) καθορίζει την περίοδο των προσεγγιστικών συναρτήσεων  $\phi_k(x)$ . Οι τελευταίες ονομάζονται **“τριγωνομετρικά πολυώνυμα”**. Στη συνέχεια δίνουμε μια συνοπτική περιγραφή της μεθόδου βάσει της οποίας γίνεται η προσέγγιση αυτή [12].

Αρχικά παρατηρούμε ότι μέσω κατάλληλου μετασχηματισμού οποιαδήποτε συνάρτηση  $f(t)$  ορισμένη στο διάστημα  $[\alpha, \beta]$  μπορεί να θεωρηθεί ότι έχει ως πεδίο ορισμού το  $[0, 2\pi]$ . Συγκεκριμένα ο μετασχηματισμός αυτός δίνεται από την σχέση

$$x = \frac{2\pi \cdot (t - \alpha)}{\beta - \alpha}$$

Έτσι, χωρίς βλάβη της γενικότητας, θα περιορίσουμε την μελέτη μας σε συναρτήσεις που είναι ορισμένες στο διάστημα  $[0, 2\pi]$ . Από την άλλη μεριά, χρειαζόμαστε ένα κριτήριο προσέγγισης “ολικού” χαρακτήρα, ένα κριτήριο δηλαδή το οποίο θα λαμβάνει υπόψη του την συμπεριφορά της συνάρτησης σε όλο το πεδίο ορισμού της και όχι μόνο στην περιοχή ενός σημείου του, όπως συμβαίνει με το κριτήριο των παραγώγων στην προσέγγιση Taylor. Υιοθετήθηκε λοιπόν η χρήση των ορισμένων ολοκληρωμάτων σε όλο το πεδίο ορισμού  $[0, 2\pi]$  των υπό μελέτη συναρτήσεων. Επομένως, θα έχουμε:

$$\int_0^{2\pi} \phi_n(x) dx = \int_0^{2\pi} f(x) dx$$

$$\int_0^{2\pi} \phi_n(x) \cos(kx) dx = \int_0^{2\pi} f(x) \cos(kx) dx, k = 1, 2, \dots, n$$

$$\int_0^{2\pi} \phi_n(x) \sin(kx) dx = \int_0^{2\pi} f(x) \sin(kx) dx, k = 1, 2, \dots, n$$

<sup>1</sup> Ο Γάλλος μαθηματικός Jean-Baptiste Joseph Fourier (1766 – 1830) συνέβαλε στη μελέτη και τον υπολογισμό της θερμότητας διάχυσης, μελετώντας την αγωγή θερμότητας κατά μήκος ενός θερμικά μονωμένου σωλήνα μεγάλου μήκους, με την εργασία του “Théorie analytique de la chaleur” (1822). Σε αυτή γίνεται εκτενής χρήση των σειρών, οι οποίες μετέπειτα πήραν το όνομά του. [13]

Οι προηγούμενες σχέσεις, αν θεωρήσουμε ότι ο βαθμός  $n$  του προσεγγιστικού τριγωνομετρικού πολυωνύμου τείνει στο άπειρο, οδηγούν στο **ανάπτυγμα Fourier** της συνάρτησης  $f(x)$  [12]:

$$f(x) = \sum_{n=0}^{\infty} (a_n \cos(nx) + \beta_n \sin(nx))$$

Όπου

$$a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx$$

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(kx) dx$$

$$\beta_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(kx) dx$$

Χρησιμοποιώντας ανάλογο μετασχηματισμό, μπορούμε να υποθέσουμε ότι η συνάρτηση που μελετάμε έχει πεδίο ορισμού το διάστημα  $[-\pi, \pi]$ . Προκύπτουν τότε ακριβώς οι παραπάνω τύποι **με όρια ολοκλήρωσης από  $-\pi$  έως  $\pi$** .

Οι τύποι αυτοί απλοποιούνται σημαντικά στην περίπτωση άρτιων ή περιττών συναρτήσεων.

Συγκεκριμένα:

Αν η συνάρτηση  $f(x)$ ,  $x \in [-\pi, \pi]$ , είναι άρτια (ισχύει δηλαδή ότι  $f(-x) = f(x)$ , για κάθε  $x \in [-\pi, \pi]$ ) τότε το ανάπτυγμά της σε σειρά Fourier διαμορφώνεται ως εξής:

$$f(x) = \sum_{n=0}^{\infty} (a_n \cos(nx))$$

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(kx) dx$$

Αν η συνάρτηση  $f(x)$ ,  $x \in [-\pi, \pi]$ , είναι περιττή (ισχύει δηλαδή ότι  $f(-x) = -f(x)$ , για κάθε  $x \in [-\pi, \pi]$ ) τότε το ανάπτυγμά της σε σειρά Fourier γίνεται [12]:

$$f(x) = \sum_{n=1}^{\infty} (\beta_n \sin(nx))$$

$$\beta_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(kx) dx$$

Ας σημειωθεί ότι ο Μετασχηματισμός Fourier αperiοδικών σημάτων είναι συνεχής, δηλαδή το συχνотικό φάσμα των σημάτων περιέχει μη μετρήσιμα άπειρες διαφορετικές συχνότητες. Αντιθέτως ο Μετασχηματισμός Φουριέ περιοδικών σημάτων που μελετάται μέσω των σειρών Fourier είναι διακριτός, δηλαδή το φάσμα των σημάτων περιέχει μετρήσιμα άπειρες διαφορετικές συνιστώσες: ένα ημίτονο της θεμελιώδους συχνότητας (η οποία είναι η συχνότητα του αρχικού, ολικού περιοδικού σήματος) και άπειρα ημίτονα που οι συχνότητες

τους είναι ακέραια πολλαπλάσια της θεμελιώδους. Τα ημίτονα αυτά, που οι συχνότητες τους είναι ακέραια πολλαπλάσια της θεμελιώδους, ονομάζονται **αρμονικές σήματος**.

Τέλος, η κολοβωμένη σειρά Fourier προκύπτει αν στην παραπάνω σχέση το άπειρο αντικατασταθεί με έναν πεπερασμένο αριθμό.

Παραδείγματος χάριν, η κολοβωμένη σειρά Fourier για  $N = 5$  δίνεται από τη σχέση:

$$f(x) = \sum_{n=1}^5 (\beta_n \sin(nx))$$

### 2.3.2 Σήματα με πεπερασμένη ενέργεια (ενεργειακά σήματα)

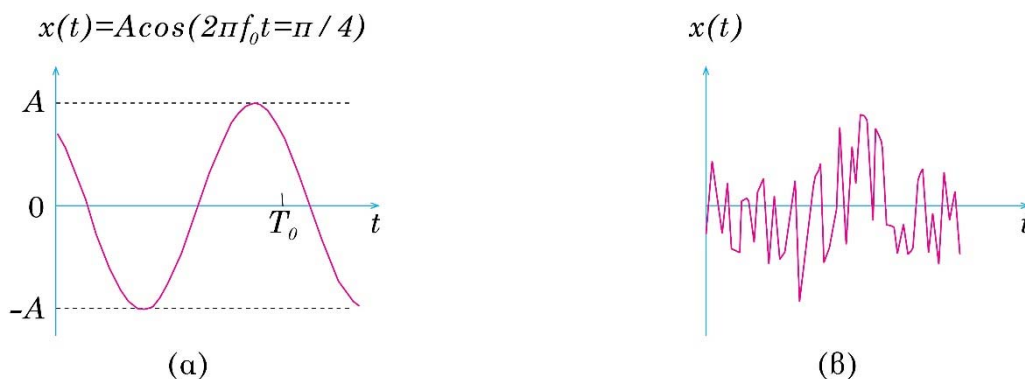
Αν με  $x(t)$  συμβολίσουμε την στιγμιαία ισχύ ενός σήματος, τότε η ενέργεια ενός σήματος ορίζεται από την παρακάτω σχέση:

$$E_{ολ} = \int_{-\infty}^{+\infty} x^2(t) dt$$

Τα σήματα για τα οποία το παραπάνω ολοκλήρωμα έχει πεπερασμένη τιμή, ονομάζονται ενεργειακά. Σε μερικά όμως σήματα το ολοκλήρωμα αυτό απειρίζεται, οπότε το σήμα ονομάζεται μη ενεργειακό. Παράδειγμα ενεργειακού σήματος είναι ο ορθογώνιος παλμός, δηλαδή το σήμα που έχει τιμή  $A$  για ένα συγκεκριμένο χρονικό διάστημα ( $T_1, T_2$ ). Γενικά, τα σήματα που έχουν πεπερασμένη διάρκεια και οι τιμές που παίρνουν δεν απειρίζονται, είναι ενεργειακά σήματα. Αντίθετα, τα περιοδικά σήματα παραδείγματος χάριν είναι μη ενεργειακά σήματα.

### 2.3.3 Αιτιοκρατικά και Τυχαία-Στοχαστικά Σήματα

Όταν οι τιμές που παίρνει ένα σήμα σε κάθε χρονική στιγμή ορίζονται χωρίς αβεβαιότητα το σήμα χαρακτηρίζεται ως *αιτιοκρατικό σήμα* ή *νομοτελειακό σήμα*. Ένα τέτοιο σήμα, για παράδειγμα, είναι το συνημίτονο (όπως φαίνεται στην Εικόνα1α). Στην πράξη, όμως, συναντάμε πολλά σήματα, όπως ο θερμικός θόρυβος, στα οποία η τιμή σε οποιαδήποτε χρονική στιγμή δεν μπορεί να καθορισθεί με βεβαιότητα πριν εμφανιστούν. Τα σήματα αυτά ονομάζονται *τυχαία* ή *στοχαστικά σήματα* (Εικόνα1β).



Εικόνα 1: Παράδειγμα (α) νομοτελειακού σήματος και (β) στοχαστικού σήματος.

### 2.3.4 Τυχαίες και εργοδικές διαδικασίες

Μια τυχαία μεταβλητή αποτελεί μια αντιστοίχιση από τον δειγματικό χώρο ενός πειράματος σε ένα σύνολο πραγματικών αριθμών. Αντίστοιχα, μια διακριτού χρόνου τυχαία διαδικασία είναι μια αντιστοιχία από τον δειγματικό χώρο  $\Omega$  σε μια συλλογή σημάτων διακριτού χρόνου ( $x(n)$ ).

Με άλλα λόγια, μια τυχαία διαδικασία διακριτού χρόνου είναι στην πραγματικότητα μια συλλογή σημάτων διακριτού χρόνου. Ένα απλό παράδειγμα τυχαίας διαδικασίας είναι το εξής. Έστω το πείραμα με το ρίξιμο του ζαριού και έστω το αποτέλεσμα του πειράματος αντιστοιχεί σε μια τυχαία μεταβλητή  $A$ . Δηλαδή η μεταβλητή  $A$  παίρνει τιμές από το 1 ως το 6 με ίση πιθανότητα. Αν θεωρήσουμε το σήμα  $x(n) = A \cos(n\omega)$ , τότε μια τυχαία διαδικασία έχει προκύψει, η οποία αποτελείται από τη συλλογή έξι διαφορετικών, ισοπίθανων σημάτων [11].

Για μια στοχαστική διαδικασία, μπορούμε να ορίσουμε δυο τύπους μέσης τιμής παραμέτρων:

- A. Παρατηρούμε πολλές συναρτήσεις - δείγματα της διαδικασίας και παίρνουμε την μέση τιμή σε δεδομένη χρονική στιγμή  $t_0$  (στατιστικά μέση τιμή).
- B. Βρίσκουμε την μέση τιμή μιας συνάρτησης - δείγματος (χρονικά μέση τιμή)

Αν οι στατιστικά και χρονικά μέσες τιμές συμπίπτουν η διαδικασία ονομάζεται εργοδική.

### 2.3.5 Το φάσμα πυκνότητας ισχύος

Το φάσμα πυκνότητας ισχύος (power density spectrum), ή απλά φάσμα ισχύος, είναι η γραφική παράσταση της πυκνότητας ισχύος (π.χ. σε μονάδες Watt/Hz) ως προς τη συχνότητα (σε μονάδες Hz) και αποδίδει τη διασπορά της ενέργειας ενός σήματος σε διάφορες συχνότητες. Επειδή η ισχύς είναι ανάλογη του τετραγώνου του πλάτους του σήματος, τα φάσματα ισχύος δεν περιέχουν αρνητικές κορυφές, όπως τα φάσματα πλάτους, και στην ουσία αποτελούν τη γραφική παράσταση  $[H(f)]^2$  συναρτήσει της συχνότητας  $f$ .

Από το φάσμα συχνοτήτων κάθε σήματος εντοπίζονται η περιοχή ή οι περιοχές συχνοτήτων που φέρουν το μεγαλύτερο ποσοστό πληροφοριών [10]. Ουσιαστικά αυτό είναι εφικτό, παρατηρώντας στο φάσμα συχνοτήτων τις περιοχές συχνοτήτων όπου η ισχύς λαμβάνει τις μεγαλύτερες τιμές – αυτές, είναι και οι περιοχές με το μεγαλύτερο ποσοστό πληροφοριών.

Έστω για παράδειγμα τα περιοδικά σήματα που φαίνονται στην Εικόνα 1. Στην Εικόνα 2, παρουσιάζονται οι γραφικές παραστάσεις των αντίστοιχων “κολοβωμένων” σειρών Fourier με  $N$  ( $N = 1, 2, 3, 5$  και  $8$ ) ημιτονικούς όρους.



**Γραφική παράσταση σήματος h(t)      Σειρά Fourier (x=2πft)**



1. Τετραγωνικό σήμα

$$h(t) = 4 \left( \frac{\sin x}{1} + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \dots \right)$$

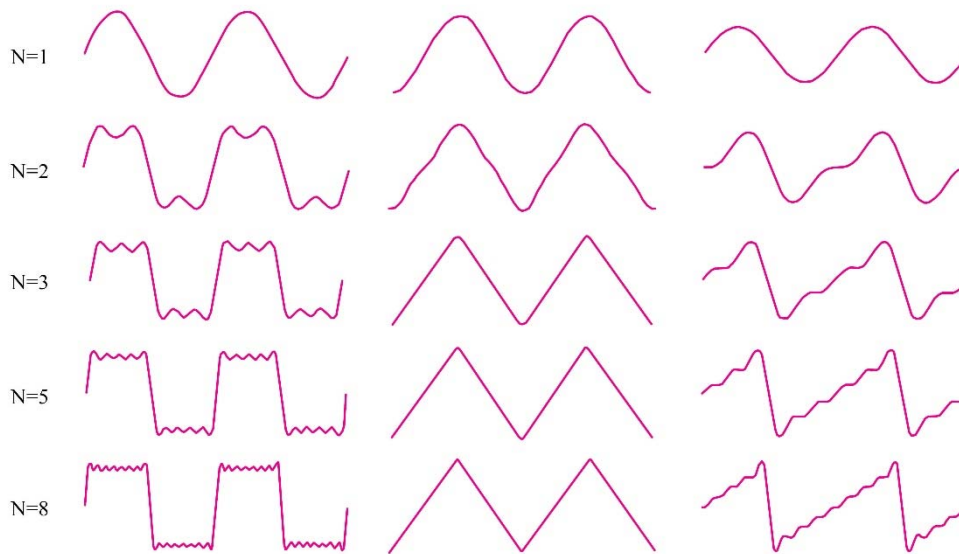
2. Τριγωνικό σήμα

$$h(t) = \frac{\pi}{2} - \frac{4}{\pi} \left( \frac{\cos x}{1^2} + \frac{\cos 3x}{3^2} + \frac{\cos 5x}{5^2} + \dots \right)$$

3. Πριονωτό σήμα

$$h(t) = 2 \left( \frac{\sin x}{1} - \frac{\sin 2x}{2} + \frac{\sin 3x}{3} - \dots \right)$$

**Εικόνα 2: 3 παραδείγματα περιοδικών σημάτων**



**Εικόνα 3: Σειρές Fourier των 3ων περιοδικών σημάτων**

Είναι χαρακτηριστικό το ότι η σχεδόν πλήρης ανασύνθεση του τριγωνικού σήματος πραγματοποιείται με πολύ λιγότερους ημιτονικούς όρους, απ' ό,τι συμβαίνει με το τετραγωνικό και το πριονωτό σήμα. Η διαφορά αυτή μπορεί να ερμηνευθεί με προσεκτική παρατήρηση των αντίστοιχων φασμάτων συχνοτήτων. Οι υψηλές αρμονικές του τετραγωνικού και του πριονωτού σήματος αποτελούν σημαντικό ποσοστό του φάσματος συχνοτήτων, σε αντίθεση με το τριγωνικό, όπου οι υψηλές αρμονικές είναι ουσιαστικά ανύπαρκτες, έχοντας πολύ μικρό πλάτος.

Η περιεκτικότητα του τετραγωνικού και του πριονωτού σήματος σε υψηλές αρμονικές, οφείλεται στις απότομες μεταπτώσεις τους (κάθετες πλευρές). Χαρακτηριστικό αποτέλεσμα της αποκοπής των υψηλών συχνοτήτων (ή ισοδύναμα της μη συμμετοχής τους στην ανασύνθεση) είναι ο κωδωνισμός (ringing), που παρατηρείται στα σημεία απότομων μεταπτώσεων [10].

Στη συνέχεια θα μελετήσουμε τον αναλυτικό υπολογισμό του φάσματος πυκνότητας ισχύος καθώς και την αυτοσυσχέτιση μεταξύ των σημάτων.

### 2.3.6 Το θεώρημα Parseval

Έστω δύο περιοδικά πραγματικά σήματα  $x(t)$ ,  $y(t)$  τα οποία αναπτύσσονται σε σειρά Fourier:

$$x(t) = A_0 + \sum_{k=1}^{\infty} A_k \cos(k\omega_0 t + \phi_k)$$

$$y(t) = B_0 + \sum_{m=1}^{\infty} B_m \cos(m\omega_0 t + \psi_m)$$

Πολλαπλασιάζοντας τα δύο σήματα, και ολοκληρώνοντας το αποτέλεσμα σε μια περίοδο  $T_0$ , έχουμε:

$$\begin{aligned} \int_0^{T_0} x(t)y(t)dt &= A_0 B_0 \underbrace{\int_0^{T_0} dt}_{T_0} + B_0 \sum_{k=1}^{\infty} A_k \underbrace{\int_0^{T_0} \cos(k\omega_0 t + \phi_k) dt}_0 + \\ &+ A_0 \sum_{m=1}^{\infty} B_m \underbrace{\int_0^{T_0} \cos(m\omega_0 t + \psi_m) dt}_0 + \\ &+ \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} A_k B_m \int_0^{T_0} \cos(k\omega_0 t + \phi_k) \cos(m\omega_0 t + \psi_m) dt \\ &= A_0 B_0 T_0 + \sum_{k=1}^{\infty} A_k B_k \int_0^{T_0} \left[ \frac{1}{2} \cos(2k\omega_0 t + \phi_k + \psi_k) + \frac{1}{2} \cos(\phi_k - \psi_k) \right] dt \\ &= A_0 B_0 T_0 + \sum_{k=1}^{\infty} A_k B_k \frac{1}{2} T_0 \cos(\phi_k - \psi_k) \\ \frac{1}{T_0} \int_0^{T_0} x(t)y(t)dt &= A_0 B_0 + \sum_{k=1}^{\infty} \frac{A_k B_k}{2} \cos(\phi_k - \psi_k) \end{aligned}$$



Όπου έχει γίνει χρήση των εξής ταυτοτήτων:

$$\int_0^{T_0} \cos(k\omega_0 t + \phi_k) \cos(m\omega_0 t + \psi_m) dt = \begin{cases} 0, m \neq k \\ \frac{T_0}{2}, m = k \end{cases}$$

$$\int_0^{T_0} \cos(k\omega_0 t + \phi_k) dt = 0, \forall k$$

$$\cos(a)\cos(b) = \frac{1}{2}\cos(a+b) + \frac{1}{2}\cos(a-b)$$

Όταν  $x(t) = y(t)$  τότε:

$$A_0 = B_0$$

$$A_k = B_k, \text{ για κάθε } k.$$

$$\phi_k = \psi_k$$

Έτσι η παραπάνω σχέση παίρνει τη μορφή:

$$\frac{1}{T_0} \int_0^{T_0} x^2(t) dt = A_0^2 + \sum_{k=1}^{\infty} \frac{A_k^2}{2}$$

Που είναι γνωστή σαν το θεώρημα του Parseval. Ερμηνεύοντας το θεώρημα του Parseval, μπορούμε να πούμε πως αυτό αναφέρεται στη διατήρηση της συνολικής ενέργειας του σήματος κατά τη μετάβαση από το πεδίο του χρόνου στο πεδίο της συχνότητας, και δείχνει το πώς κατανέμεται η ενέργεια του σήματος στις διάφορες συχνότητες.

### 2.3.7 Ορισμός Αυτοσυσχέτισης

Έστω ότι έχουμε μια ακολουθία  $x(n)$  η οποία είναι συνήθως το αποτέλεσμα της δειγματοληψίας ενός συνεχούς σήματος  $x_\alpha(t)$  με μια συχνότητα δειγματοληψίας  $F_S$ . Εάν το  $x(t)$  είναι ένα πεπερασμένο ενεργειακά σήμα, τότε η ενέργειά του θα ισούται με:

$$E = \int_{-\infty}^{\infty} |x_\alpha(t)|^2 dt < \infty$$

Τότε υπάρχει ο μετασχηματισμός Fourier και δίνει:

$$X_\alpha(F) = \int_{-\infty}^{\infty} x_\alpha(t) e^{-i2\pi Ft} dt$$

Από το θεώρημα Parseval ισχύει:

$$E = \int_{-\infty}^{\infty} |x_\alpha(t)|^2 dt = \int_{-\infty}^{\infty} |x_\alpha(F)|^2 df$$

Όπως και προηγουμένως, έτσι και εδώ ερμηνεύοντας την εφαρμογή του θεωρήματος του Parseval, μπορούμε να πούμε πως αυτό αναφέρεται στη διατήρηση της συνολικής ενέργειας του σήματος κατά τη μετάβαση από το πεδίο του χρόνου στο πεδίο της συχνότητας.

Η ποσότητα  $|X_a(F)|^2$  αποτελεί την κατανομή της ενέργειας σε συνάρτηση της συχνότητας και ονομάζεται φάσμα της πυκνότητας της ενέργειας (energy density spectrum) του σήματος και συμβολίζεται με  $S_{xx}(F)$

Έτσι η συνολική ενέργεια του σήματος είναι απλά το ολοκλήρωμα της  $S_{xx}(F)$  σε όλες τις συχνότητες  $F$  (δηλαδή η συνολική περιοχή κάτω από την  $S_{xx}(F)$ ). Μπορούμε επίσης να σημειώσουμε ότι το  $S_{xx}(F)$  μπορεί να “προσεγγιστεί” και από μια άλλη σκοπιά. Δηλαδή σαν τον μετασχηματισμό Fourier μιας άλλης συνάρτησης,  $R_{xx}(\tau)$ , η οποία ονομάζεται συνάρτηση αυτοσυσχέτισης του σήματος  $x_a(t)$  (που έχει πεπερασμένη ενέργεια) και ορίζεται ως:

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} x_a^*(t)x(t+\tau)dt$$

Υλοποιώντας τον μετασχηματισμό Fourier της παραπάνω έκφρασης, βάσει του ορισμού του μετασχηματισμού, καταλήγουμε στην εξής έκφραση:

$$\int_{-\infty}^{\infty} R_{xx}(\tau)e^{-i2\pi F\tau}d\tau = S_{xx}(F) = |X_a(F)|^2$$

Έτσι το  $R_{xx}(\tau)$  και το  $S_{xx}(F)$  είναι ένα ζευγάρι μετασχηματισμού Fourier. Η δειγματοληπτημένη έκδοχή του  $x_a(t)$  είναι μια ακολουθία  $x(n)$ ,  $-\infty < n < \infty$ , η οποία έχει μετασχηματισμό Fourier:

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n}$$

ή ισοδύναμα με  $f = \frac{F}{Fs}$ :

$$X(f) = \sum_{n=-\infty}^{\infty} x(n)e^{-j2\pi f n}$$

Για ένα δειγματοληπτημένο σήμα η αυτοσυσχέτιση ορίζεται ως εξής:

$$r_{xx}(k) = \sum_{n=-\infty}^{\infty} x^*(n)x(n+k)$$

της οποίας ο μετασχηματισμός Fourier (Wiener-Khintchine theorem)<sup>2</sup> είναι:

$$S_{xx}(f) = \sum_{k=-\infty}^{\infty} r_{xx}(k)e^{-j2\pi k f}$$

<sup>2</sup> Το Θεώρημα Wiener-Khinchin (ή Θεώρημα Wiener-Khintchine, ή Θεώρημα Khinchin-Kolmogorov) δηλώνει ότι η φασματική πυκνότητα μιας στατικής με την ευρεία έννοια τυχαίας διαδικασίας είναι ο μετασχηματισμός Φουριέ της αντίστοιχης συναρτήσεως αυτοσυσχετισμού.

### 2.3.8 Υπολογισμός μετασχηματισμού Fourier με βάση το θεώρημα του Parseval

Έτσι μπορούμε να υπολογίσουμε το φάσμα της πυκνότητας της ενέργειας με την εφαρμογή του μετασχηματισμού Fourier στην αυτοσυσχέτιση της ακολουθίας  $\{x(n)\}$ . Οπότε μπορούμε να διακρίνουμε δύο μεθόδους για να υπολογίσουμε το φάσμα της πυκνότητας της ενέργειας του σήματος  $x_\alpha(t)$  από τα δείγματα του  $x(n)$ . Η πρώτη είναι η ευθεία μέθοδος, η οποία περιλαμβάνει τον υπολογισμό του μετασχηματισμού Fourier του  $\{x(n)\}$  και μετά:

$$S_{xx}(f) = |X(F)|^2 = \left| \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n} \right|^2$$

Η δεύτερη προσέγγιση ονομάζεται “έμμεση” μέθοδος διότι απαιτεί δύο βήματα. Αρχικά, υπολογίζουμε την αυτοσυσχέτιση  $r_{xx}(k)$  από το  $x(n)$ . Στην συνέχεια βρίσκουμε το μετασχηματισμό Fourier της αυτοσυσχέτισης. Οπότε, με αυτό τον τρόπο υπολογίζουμε το φάσμα της πυκνότητας της ενέργειας. Πρακτικά, μόνο η πεπερασμένης διάρκειας ακολουθία  $x(n)$ , είναι διαθέσιμη για να υπολογίσουμε το φάσμα του σήματος. Άρα περιορίζοντας τη διάρκεια της ακολουθίας  $x(n)$  σε  $N$  σημεία είναι ισοδύναμο με το να πολλαπλασιάσουμε την  $x(n)$  με ένα ορθογώνιο παράθυρο.

Έτσι έχουμε:

$$\tilde{x}(n) = x(n)w(n) = \begin{cases} x(n), 0 \leq n \leq N-1 \\ 0, \text{αλλιώς} \end{cases}$$

Στο πεδίο της συχνότητας, ο πολλαπλασιασμός τους ισοδυναμεί με συνέλιξη, δηλαδή:

$$\tilde{X}(f) = X(f) * W(f) = \int_{-\frac{1}{2}}^{\frac{1}{2}} X(a)W(f-a)da$$

Είναι ξεκάθαρο από τα παραπάνω ότι η πυκνότητα ενέργειας του φάσματος της ακολουθίας (παραθύρου)  $\{x(n)\}$  είναι μια προσέγγιση του επιθυμητού φάσματος της ακολουθίας  $\{x(n)\}$ . Η φασματική πυκνότητα που λαμβάνουμε από το  $\{\tilde{x}(n)\}$  είναι:

$$S_{\tilde{x}\tilde{x}}(f) = \left| \tilde{X}(F) \right|^2 = \left| \sum_{n=0}^{N-1} \tilde{x}(n)e^{-j\omega n} \right|^2$$

Τα φάσμα μπορεί να υπολογιστεί αριθμητικά σε ένα σύνολο από  $N$  σημεία συχνότητας από τον DFT. Έτσι:

$$\tilde{X}(k) = \left| \sum_{n=0}^{N-1} \tilde{x}(n)e^{-j\pi k \frac{n}{N}} \right|^2$$

$$\text{Τότε: } \left| X(F) \right|^2 = S_{xx}(f) \Big|_{f=\frac{k}{N}} = S_{xx}\left(\frac{k}{N}\right)$$

και έτσι

$$S_{xx}\left(\frac{k}{N}\right) = \left| \sum_{n=0}^{N-1} x(n) e^{-j\pi k \frac{n}{N}} \right|^2$$

η οποία αποτελεί μια “παραμορφωμένη” εκδοχή του πραγματικού φάσματος  $S_{xx}(k/N)$ . Τα σήματα πεπερασμένης ενέργειας, όπως είδαμε παραπάνω, έχουν μετασχηματισμό Fourier και χαρακτηρίζονται στο πεδίο της συχνότητας από το φάσμα της πυκνότητας της ενέργειας. Από την άλλη τα σήματα, τα οποία χαρακτηρίζονται ως “στάσιμα”, δεν έχουν πεπερασμένη ενέργεια και έτσι δεν έχουν μετασχηματισμό Fourier. Όμως τέτοια σήματα έχουν πεπερασμένη μέση ισχύ και έτσι χαρακτηρίζονται από το φάσμα της πυκνότητας ισχύος (power density spectrum). Εάν  $x(t)$  είναι ένα στάσιμο σήμα, τότε η συνάρτηση αυτοσυσχέτισης του είναι:

$$\gamma_{xx}(t) = E[x^*(t)x(t+\tau)]$$

όπου το  $E[\cdot]$  αντιστοιχεί στο μέσο όρο της στατιστικής. Τότε διαμέσου του θεωρήματος Wiener-Khintchine, το φάσμα της πυκνότητας ισχύος του στάσιμου σήματος είναι ο μετασχηματισμός Fourier της συνάρτησης αυτοσυσχέτισης:

$$\Gamma_{xx}(F) = \int_{-\infty}^{\infty} \gamma_{xx}(\tau) e^{-j2\pi F t} dt$$

Πρακτικά δεν γνωρίζουμε την πραγματική συνάρτηση αυτοσυσχέτισης  $\gamma_{xx}(t)$  άρα δεν μπορούμε να υπολογίσουμε το μετασχηματισμό Fourier.

Όμως μπορούμε να υπολογίσουμε την συνάρτηση αυτοσυσχέτισης του μέσου χρόνου (time-average autocorrelation function) που είναι ίση με:

$$R_{xx}(\tau) = \frac{1}{2T_0} \int_{-T_0}^{T_0} x_a^*(t)x(t+\tau) dt$$

όπου  $2T_0$  είναι το χρονικό διάστημα παρατήρησης. Εάν η διαδικασία στην περίπτωση του στάσιμου είναι “εργοδική” (ergodic) στην πρώτη και στην δεύτερη χρονική στιγμή (μέση τιμή και συνάρτηση αυτοσυσχέτισης), τότε

$$\gamma_{xx}(\tau) = \lim_{T_0 \rightarrow \infty} R_{xx}(\tau) = \lim_{T_0 \rightarrow \infty} \frac{1}{2T_0} \int_{-T_0}^{T_0} x_a^*(t)x(t+\tau) dt$$

Αυτή η σχέση αποτελεί έναν υπολογισμό της στατιστικής συνάρτησης αυτοσυσχέτισης  $\gamma_{xx}(\tau)$ . Επιπλέον, ο μετασχηματισμός Fourier του  $R_{xx}(\tau)$  αποτελεί έναν υπολογισμό του φάσματος της πυκνότητας ισχύος  $P_{xx}(F)$  και ισούται με:

$$P_{xx}(F) = \int_{-T_0}^{T_0} R_{xx}(\tau) e^{-j2\pi F t} dt = \frac{1}{2T_0} \int_{-T_0}^{T_0} x_a^*(t)x(t+\tau) e^{-j2\pi F t} dt = \frac{1}{2T_0} \left| \int_{-T_0}^{T_0} x(t) e^{-j2\pi F t} dt \right|^2$$

Η πραγματική τιμή του φάσματος της πυκνότητας ισχύος  $P_{xx}(F)$  υπολογίζεται όταν  $T \rightarrow +\infty$ , δηλαδή

$$\Gamma_{xx}(F) = \lim_{T_0 \rightarrow \infty} E[P_{xx}(F)] = \lim_{T_0 \rightarrow \infty} E \left[ \frac{1}{2T_0} \left| \int_{-T_0}^{T_0} x(t) e^{-j2\pi F t} dt \right|^2 \right]$$

Οπότε μπορούμε με δύο τρόπους να υπολογίσουμε την  $P_{xx}(F)$ . Αρχικά, είτε απευθείας είτε έμμεσα υπολογίζοντας αρχικά την  $R_{xx}(\tau)$  και μετά υπολογίζοντας το μετασχηματισμό Fourier της. Στην συνέχεια θεωρούμε ότι το  $x_\alpha(t)$  δειγματοληπτείται με έναν ρυθμό  $F_s > 2B$ , όπου  $B$  είναι η υψηλότερη συχνότητα η οποία περικλείεται στο φάσμα της πυκνότητας ισχύος. Έτσι πετυχαίνουμε μια πεπερασμένης διάρκειας ακολουθία  $x(n)$ . Από αυτά τα δείγματα μπορούμε να υπολογίσουμε την ακολουθία αυτοσυσχέτισης του μέσου χρόνου:

$$r'_{xx}(m) = \frac{1}{N-m} \sum_{n=0}^{N-m-1} x^*(n)x(n+m), m = 0, 1, \dots, N-1$$

$$r'_{xx}(m) = \frac{1}{N-|m|} \sum_{n=|m|}^{N-1} x^*(n)x(n+m), m = (-N+1), \dots, -2, -1$$

$$P'_{xx}(f) = \sum_{m=-N+1}^{N-1} r'_{xx}(m)e^{-j2\pi fm}$$

Για μεγάλες τιμές της παραμέτρου  $m$  (lag), ο υπολογισμός της έχει μεγάλη διακύμανση, ειδικότερα όταν το  $m$  προσεγγίζει το  $N$ . Αυτό συμβαίνει διότι λίγα σημεία δεδομένων εισάγονται στον υπολογισμό για μεγάλες καθυστερήσεις (lags). Ένας εναλλακτικός τρόπος υπολογισμού είναι:

$$r_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-m-1} x^*(n)x(n+m), m = 0, 1, \dots, N-1$$

$$r_{xx}(m) = \frac{1}{N} \sum_{n=|m|}^{N-1} x^*(n)x(n+m), m = (-N+1), \dots, -2, -1$$

Οπότε για να μπορέσουμε να υπολογίσουμε το φάσμα της ισχύος θα χρησιμοποιήσουμε τον παρακάτω τύπο:

$$P_{xx}(f) = \sum_{m=-N+1}^{N-1} r_{xx}(m)e^{-j2\pi fm}$$

Και αν αντικαταστήσουμε την  $r$  στο  $P_{xx}(F)$  τότε θα έχουμε:

$$P_{xx}(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x(n)e^{-j2\pi fm} \right|^2 = \frac{1}{N} |X(f)|^2$$

όπου  $X(f)$  είναι ο μετασχηματισμός Fourier της ακολουθίας  $x(n)$ . Αυτή η μορφή είναι πολύ γνωστή σαν “περιοδόγραμμα” (periodogram).

### 2.3.9 Χρήση του DFT για τον υπολογισμό του φάσματος ισχύος (power spectrum)

Από τις σχέσεις που περιγράψαμε παραπάνω, τα μεγέθη  $S_{xx}(f)$  και  $P_{xx}(f)$  μπορούν να υπολογιστούν χρησιμοποιώντας τον DFT, ο οποίος υπολογίζεται εύκολα από έναν αλγόριθμο FFT. Εάν έχουμε  $N$  σημεία σαν δεδομένα, υπολογίζουμε το DFT για  $N$  σημεία (το ελάχιστο). Αφού θα έχουμε  $N$  σημεία άρα θα υπολογίσουμε τα παραπάνω μεγέθη για τις συχνότητες  $f_k = k/N$  με  $k = 0, 1, \dots, N - 1$ . Δηλαδή θα ισχύει απλά:

$$P_{xx}\left(e^{j\frac{2\pi k}{N}}\right) = \frac{1}{N} |FFT[x(n)]|^2$$

Πρακτικά όμως μια τέτοια αραιή δειγματοληψία του φάσματος δεν μας δίνει μια καλή αναπαράσταση ή μια καλή εικόνα του υπολογισμού του συνεχούς φάσματος  $P_{xx}(f)$ . Αυτό μπορεί να βελτιωθεί εάν υπολογίσουμε το  $P_{xx}(f)$  σε επιπρόσθετες συχνότητες. Ισοδύναμα μπορούμε να αυξήσουμε το μήκος της ακολουθίας συμπληρώνοντας μηδενικά (zero padding). Οπότε στην συνέχεια θα υπολογίσουμε το  $P_{xx}(f)$  σε ένα περισσότερο πυκνό διάστημα συχνοτήτων.

Η τεχνική του zero padding δεν βελτιώνει την ανάλυση στο πεδίο των συχνοτήτων. Απλά μας δίνει μια μέθοδο με την οποία παρεμβάλουμε τις τιμές του υπολογισμού  $P_{xx}(f)$  σε περισσότερες συχνότητες. Η ανάλυση στο πεδίο των συχνοτήτων για τον υπολογισμό του  $P_{xx}(f)$  εξαρτάται από το μήκος  $N$  των δεδομένων.

### 3 Βασικά χαρακτηριστικά σημάτων φωνής

#### 3.1 Εισαγωγή

Η επεξεργασία σημάτων ομιλίας είναι μία από τις πιο γόνιμες περιοχές εφαρμογών τεχνικών της ψηφιακής επεξεργασίας σήματος. Βασικά προβλήματα επεξεργασίας ομιλίας όπως η κατάτμηση, η προέμφαση του σήματος ομιλίας, η ψηφιακή κωδικοποίηση, η αναγνώριση ομιλίας και ομιλητή αποτελούν ενδιαφέρουσες περιπτώσεις στις οποίες μπορούν να εφαρμοστούν αλγόριθμοι ψηφιακής επεξεργασίας σήματος ([2], [6]).

Στις εφαρμογές αυτές είναι πολύ σημαντικό να κατανοήσουμε τις ιδιότητες των σημάτων ομιλίας και να αναζητήσουμε «έξυπνους» τρόπους και σωστές χρήσεις των τεχνικών της Ψηφιακής Επεξεργασίας Σημάτων.

Τα σήματα ομιλίας είναι ιδιαίτερα πολύπλοκα και σύνθετα, αν σκεφτούμε ότι περιέχουν τόσο την «ουσιαστική πληροφορία» που συνδέεται με κάποιο συγκεκριμένο μήνυμα όσο και πληροφορίες που σχετίζονται με τον ομιλητή όπως για παράδειγμα το φύλο, την ηλικία, την συναισθηματική κατάσταση που βρίσκεται ο ομιλών, κ.ά. Η φωνή παράγεται μέσα σε έναν αγωγό μεταβλητής διατομής που αρχίζει από τις φωνητικές χορδές και τελειώνει στα χείλη. Ένας πρόσθετος αγωγός, η ρινική κοιλότητα, συμμετέχει στην παραγωγή των ένρινων φωνημάτων. Η απόκριση του συστήματος παραγωγής ομιλίας είναι στενά συνδεδεμένη με την μορφή της διέγερσης που εφαρμόζεται στην είσοδό του η οποία με την σειρά της καθορίζεται από την θέση των φωνητικών χορδών.

#### 3.2 Χρόνο - Συχνοτική Ανάλυση των Σημάτων Ομιλίας

Ένα βασικό πρόβλημα του Μετασχηματισμού Fourier (όπως και των μετασχηματισμών που βασίζονται σ' αυτόν) είναι ότι δεν παρέχει καμία πληροφορία σχετικά με τον χρόνο αλλαγής του συχνοτικού περιεχομένου ενός σήματος. Ο Μετασχηματισμός Fourier Βραχέους Χρόνου (Short Time Fourier Transform) είναι ένας μετασχηματισμός που μας επιτρέπει να μελετήσουμε την χρονική μεταβολή (αν υπάρχει) του συχνοτικού περιεχομένου ενός σήματος για την διακριτική ικανότητα του μετασχηματισμού και την σχετική αρχή της αβεβαιότητας που την διέπει [7] και ορίζεται από την σχέση ([1], [5]):

$$X_n(e^{j\omega}) = e^{-j\omega n} \sum_{m=-\infty}^{\infty} w[-m]x[n+m]e^{-j\omega m} = e^{-j\omega n} \tilde{X}_n(e^{j\omega})$$

Όπου

$$-\infty < n < +\infty \text{ και } 0 \leq \omega < 2\pi$$



Αν υποθέσουμε ότι επιθυμούμε να υπολογίσουμε τις τιμές του μετασχηματισμού στις συχνότητες

$$\omega_k = \frac{2\pi k}{N}, k = 0, 1, \dots, N-1$$

και ότι το παράθυρο είναι τέτοιο ώστε

$$w[-m] = 0, m < 0$$

και  $m > L - 1$ , η παραπάνω σχέση μπορεί να γραφεί ως ακολούθως:

$$X_n[k] = X_n \left( e^{j\frac{2\pi}{N}k} \right) = \sum_{m=n}^{n+L-1} w[n-m]x[m]e^{-j\frac{2\pi}{N}km}$$

ή

$$X_n[k] = e^{-j\frac{2\pi}{N}kn} \sum_{m=0}^{L-1} w[-m]x[n+m]e^{-j\frac{2\pi}{N}km} = e^{-j\frac{2\pi}{N}kn} \tilde{X}_n[k]$$

Όπου

$$\tilde{w}[m] = w[-m]$$

$$\tilde{X}_n[k] = \sum_{m=0}^{L-1} \tilde{w}[m]x[n+m]e^{-j\frac{2\pi}{N}km}, k = 0, 1, \dots, N-1$$

Τα  $X_n[k], \tilde{X}_n[k]$  διαφέρουν μόνο στον εκθετικό συντελεστή φάσης

$$e^{-j\frac{2\pi}{N}kn} \text{ και άρα:}$$

$$|X_n[k]| = |\tilde{X}_n[k]|$$

Παρατηρήστε επίσης ότι μπορούμε να υπολογίσουμε τον STFT στο σύνολο των  $N$  συχνοτήτων

$$\omega_k = \frac{2\pi k}{N}, k = 0, 1, \dots, N-1$$

με χρήση του DFT (και επομένως με συμπλήρωση μηδενικών αν απαιτείται και τον FFT) ακολουθώντας τα παρακάτω βήματα ([1]):

- Επέλεξε  $L$  δείγματα του σήματος αρχίζοντας από τη χρονική στιγμή  $n$ . Δηλαδή  $\{x[n], x[n+1], \dots, x[n+L-1]\}$

(Για συμμετρικά παράθυρα, είναι πιο βολικό να υποθέσουμε ότι το  $n$  είναι το κέντρο του διαστήματος του παραθύρου).

- Πολλαπλασίασε τα δείγματα του τμήματος ομιλίας με τα δείγματα του παραθύρου.

• Υπολόγισε τον DFT  $N$ -σημείων του πλαισίου ομιλίας μετά την «παραθύρωση» (προσθέτοντας τον αναγκαίο αριθμό μηδενικών στο τέλος του πλαισίου αν  $N > L$ ).

• Πολλαπλασίασε με  $e^{-j\frac{2\pi}{N}kn}$  (αυτό το βήμα μπορεί να παραληφθεί αν υπολογίζεται μόνο το πλάτος του STFT).

- Επανάλαβε τα τέσσερα παραπάνω βήματα για κάθε τιμή του  $n$ .



### 3.3 Το παράθυρο ανάλυσης

Το πρώτο και σημαντικότερο βήμα για τη διαδικασία μετατροπής ενός σήματος από το πεδίο του χρόνου στο πεδίο της συχνότητας είναι η παραθυροποίηση του. Η τελευταία αποτελεί τη διαδικασία εκλογής ενός αριθμού δειγμάτων από το σήμα και τον πολλαπλασιασμό τους με μια συνάρτηση παραθύρου  $w(n)$ .

Ο αριθμός δειγμάτων που προκύπτει σε κάθε βήμα της διαδικασίας καθορίζεται από το μέγεθος του παραθύρου. Το τελευταίο αποτελεί μια σημαντική παράμετρο κυρίως αν ληφθεί υπ' όψιν ότι ο αριθμός των συχνοτικών δειγμάτων που θα δώσει ο DFT στην έξοδό του αντιστοιχεί στο μισό του αριθμού των δειγμάτων στην είσοδό του, διεσπαρμένα στη μισή συχνότητα δειγματοληψίας. Πιο συγκεκριμένα το μέγεθος της συνάρτησης του παραθύρου σχετίζεται με τον τρόπο με τον οποίο το σήμα αναπαρίσταται-καθορίζει αν θα υπάρχει καλή ανάλυση συχνότητας ή καλή χρονική ανάλυση. Έτσι όσο πιο μεγάλο είναι το παράθυρο στο χρόνο τόσο καλύτερη ανάλυση συχνότητας θα έχουμε. Ταυτόχρονα όμως θα έχουμε κακή χρονική ανάλυση. Αυτό το φαινόμενο είναι γνωστό ως εναλλαγή ανάλυσης χρόνου εναντίον συχνότητας (time vs. frequency resolution trade off)

Θα μπορούσε κανείς να ισχυριστεί ότι ένας τρόπος για να πάρει καλύτερη συχνοτική ανάλυση θα ήταν να προστεθούν μηδενικά στο παραθυροποιημένο σήμα έτσι ώστε να προκύπτει μεγαλύτερος FFT (Fast Fourier Transform). Αυτή η διαδικασία είναι γνωστή ως zero-padding και εκφράζει μια παρεμβολή (interpolation) στο πεδίο συχνότητας. Εν τούτοις όταν προστεθούν μηδενικά πριν από την εφαρμογή του DFT δεν αυξάνεται η ικανότητα διάκρισης δύο γειτονικών ημίτονων, αυξάνεται όμως η ανάλυση συχνότητας προσθέτοντας ενδιάμεσα παρεμβεβλημένα συχνοτικά bins. Το είδος του παραθύρου που θα χρησιμοποιηθεί έχει πολύ σημαντική επίδραση στη συχνοτική αναπαράσταση που θα προκύψει.

Δυο χαρακτηριστικά του μετασχηματισμού του παραθύρου στο πεδίο της συχνότητας σχετίζονται με το αν η συγκεκριμένη συνάρτηση παραθύρου είναι χρήσιμη ή όχι: α) το εύρος του κυρίως λοβού και β) η σχέση πλάτους μεταξύ του κυρίου και του δευτερεύοντος λοβού. Το εύρος ζώνης του κυρίως λοβού εκφράζεται σε bins (συχνοτικά δείγματα-spectral samples) και σε σύνδεση με το μέγεθος του παραθύρου καθορίζει την ικανότητα διάκρισης δύο ημιτονοειδών κορυφών. Η ακόλουθη σχέση εκφράζει τη σχέση που θα έπρεπε να έχουν το μέγεθος του παραθύρου σε samples  $M$ , το εύρος ζώνης του κυρίως λοβού  $B$  και η συχνότητα δειγματοληψίας  $f_s$  ώστε να είναι δυνατή η διάκριση δύο ημιτονοειδών με συχνότητες  $f_k$  και  $f_{k+1}$ :

$$M \geq B \frac{f_s}{|f_{k+1} - f_k|}$$

Η σχέση πλάτους μεταξύ του κυρίου και του δευτερεύοντος λοβού εκφράζει το ποσοστό παραμόρφωσης που θα λάβει η ημιτονοειδής κορυφή από γειτονικούς λοβούς. Τώρα, αφού το φάσμα ενός block ή frame έχει υπολογιστεί, το παράθυρο ανάλυσης πρέπει να κινηθεί στην επόμενη θέση πάνω στην κυματομορφή ώστε να χρησιμοποιήσει το επόμενο σύνολο δειγμάτων. Η απόσταση μεταξύ των κέντρων δύο συνεχόμενων παραθύρων είναι γνωστή ως μέγεθος ολίσθησης παραθύρου (window slide ή hop size). Αν το μέγεθος αυτό είναι μικρότερο από το μέγεθος του παραθύρου, τότε θα υπάρξει κάποιο ποσοστό επικάλυψης (overlap), που σημαίνει ότι κάποια δείγματα θα χρησιμοποιηθούν παραπάνω από μια φορά στη διαδικασία ανάλυσης. Γενικότερα, όσο μεγαλύτερο είναι το ποσοστό επικάλυψης τόσο πιο ομαλές θα είναι οι μεταβολές του φάσματος από frame σε frame, αυτό όμως είναι μια υπολογιστικά χρονοβόρα διαδικασία. Ο τύπος του παραθύρου και το μέγεθος ολίσθησης πρέπει να είναι τέτοια ώστε η περιβάλλουσα που προκύπτει να αθροίζεται σε μια σταθερά, σύμφωνα με την εξίσωση:

$$A_w(m) = \sum_{n=-\infty}^{\infty} w(n-m) \approx const$$

Μια μέτρηση της διαφοράς του  $A_w$  από τη σταθερά είναι η σχέση:

$$d_w = 100 \times \frac{\max_m [A_w(m)] - \min_m [A_w(m)]}{\max_m [A_w(m)]}$$

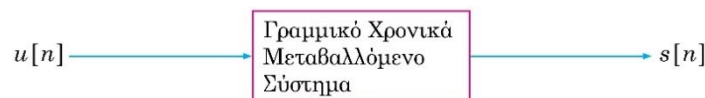
Αυτή η σχέση αναφέρεται ως η διαφόριση πλάτους του παράγοντα επικάλυψης (overlap factor) και η απαίτηση είναι να είναι ίση ή μικρότερη από το ένα τοις εκατό (1%). [15]

### 3.4 Επίδραση του Μήκους Παραθύρου

Το μήκος του παραθύρου είναι μία παράμετρος κλειδί για τον STFT. Αν το παράθυρο είναι μικρό σε σχέση με τα χαρακτηριστικά της κυματομορφής στο χρόνο, τότε ο STFT θα παρακολουθήσει τις αλλαγές αυτών των χαρακτηριστικών. Αν το παράθυρο είναι σχετικά μεγάλο, αλλαγές στο χρόνο θα αλλοιωθούν, αλλά ο STFT θα έχει καλή ανάλυση στη διάσταση της συχνότητας

### 3.5 Ανάλυση Σημάτων Ομιλίας

Είναι φανερό ότι η παραγωγή φωνής είναι ένα σύνθετο φαινόμενο. Ιδανικά θα επιθυμούσαμε να έχουμε μοντέλα αναπαράστασης που να είναι γραμμικά και χρονικά αμετάβλητα. Δυστυχώς ή ευτυχώς, ο μηχανισμός παραγωγής της ανθρώπινης φωνής, δεν ικανοποιεί με ακρίβεια καμία από τις παραπάνω ιδιότητες. Σε πολλές εφαρμογές της επεξεργασίας φωνής είναι θεμελιώδους σημασίας η δυνατότητα να αναπαρίσταται το σήμα ομιλίας από ένα μικρό αριθμό γραμμικών χρονικά μεταβαλλόμενων συστημάτων, όπως αυτό του σχήματος που ακολουθεί.



Εικόνα 4: Σύστημα επεξεργασίας σημάτων ομιλίας

όπου η είσοδος  $u[n]$  είναι είτε λευκός θόρυβος (στην περίπτωση των άηχων φωνημάτων) είτε μια σχεδόν περιοδική ακολουθία παλμών (στην περίπτωση των ηχηρών φωνημάτων). Το γραμμικό σύστημα θεωρούμε πως είναι χρονικά μεταβαλλόμενο με την μεταβολή όμως να συμβαίνει αργά ώστε σε μικρά χρονικά διαστήματα (που καθορίζουν το μέγεθος του πλαισίου στα 20ms - 30ms για σήματα ομιλίας που έχουν δειγματοληπτηθεί με συχνότητα 8KHz) να μπορεί να θεωρηθεί ένα χρονικά αμετάβλητο μοντέλο του πλαισίου ομιλίας που περιγράφει τα σημαντικότερα χαρακτηριστικά του σήματος. Η εξαγωγή αυτών των σημαντικότερων χαρακτηριστικών ενός σήματος ομιλίας είναι ο σκοπός των τεχνικών ανάλυσης ([3], [6]).

### 3.5.1 Γραμμική Πρόβλεψη και Ανάλυση Σημάτων Ομιλίας

Η ανάλυση των σημάτων ομιλίας με γραμμική πρόβλεψη είναι μια από τις πιο γνωστές τεχνικές που χρησιμοποιούνται για το σκοπό αυτό. Σύμφωνα με την τεχνική αυτή το σήμα ομιλίας  $s[n]$  είναι μία διαδικασία αυτοπαλινδρόμησης τάξης  $p$  που προκύπτει από την έξοδο ενός IIR συστήματος με συνάρτηση μεταφοράς:

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

Όπου:

$$A(z) = 1 + \sum_{k=1}^p a_k z^{-k}$$

το φίλτρο σφαλμάτων πρόβλεψης και  $G$  μια σταθερά κέρδους. Μπορεί να διαπιστωθεί εύκολα ότι η είσοδος και η έξοδος ενός τέτοιου συστήματος συνδέονται μέσω της ακόλουθης εξίσωσης διαφορών:

$$s[n] = -\sum_{k=1}^p a_k s[n-k] + Gu[n]$$

όπου  $u[n]$  η διέγερση που εφαρμόζεται στην είσοδο του συστήματος.

Η τεχνική της γραμμικής πρόβλεψης αποσκοπεί στην ανεύρεση των συντελεστών:

$$a_k, k = 1, \dots, p$$

ώστε να ελαχιστοποιείται μία συνάρτηση κόστους η οποία βασίζεται στην ακολουθία σφαλμάτων πρόβλεψης. Γνωστές μέθοδοι οι οποίες βασίζονται στην γραμμική πρόβλεψη και οι οποίες διαφοροποιούνται ως προς τις διαφορετικές υποθέσεις που κάνουμε για την μορφή του σήματος εκτός των χρονικών ορίων του σήματος που έχουμε στην διάθεσή μας και επομένως στο πεδίο ορισμού της συνάρτησης κόστους, είναι αυτή της αυτοσυσχέτισης και αυτή της συνδιασποράς ([2]).

Η μέθοδος που χρησιμοποιείται κατά κανόνα είναι αυτή της αυτοσυσχέτισης κατά την οποία ελαχιστοποιείται το τετραγωνικό σφάλμα πρόβλεψης

$$E = \sum_{n=1}^N f[n]^2,$$

όπου  $f[n]$  το προς τα εμπρός σφάλμα εκτίμησης μεταξύ του σήματος ομιλίας  $s[n]$  και του προβλεπόμενου σήματος  $\hat{s}[n]$  για όλα τα διαθέσιμα δείγματα ( $N$ ) του τμήματος του σήματος ομιλίας.

Στην περίπτωση αυτή η ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος ως προς τους συντελεστές του συστήματος, οδηγεί στις γνωστές εξισώσεις των Yule Walker από την λύση των οποίων προσδιορίζονται οι παράμετροι του μοντέλου αυτοπαλινδρόμησης.

Η συνάρτηση `lpc()` της MATLAB μπορεί να χρησιμοποιηθεί για την λύση του παραπάνω προβλήματος ελαχιστοποίησης και τον προσδιορισμό των βέλτιστων συντελεστών του μοντέλου της γραμμικής πρόβλεψης.

Μια βασική υπόθεση της τεχνικής της γραμμικής πρόβλεψης είναι πως οι συντελεστές του προγνώστη ταυτίζονται με εκείνους του μοντέλου ομιλίας. Συνεπώς, από τον ορισμό του μοντέλου, η έξοδος του γραμμικού φίλτρου πρόβλεψης θα δίνεται από την ακόλουθη σχέση:

$$f[n] = s[n] - \hat{s}[n] = s[n] + \sum_{k=1}^p a_k s[n-k] - Gu[n]$$

Η παραπάνω σχέση μπορεί να χρησιμοποιηθεί για τον προσδιορισμό της παραμέτρου κέρδους  $G$ .

Πράγματι αν  $E_{min}$  η ελάχιστη τιμή του τετραγωνικού σφάλματος πρόβλεψης, τότε η τιμή της σταθεράς κέρδους μπορεί να οριστεί από την ακόλουθη σχέση:

$$G = \sqrt{\frac{E_{min}}{\sum_{n=1}^N u[n]^2}}$$

Πανεπιστήμιο Πειραιώς

## 4 Περιγραφή μεθόδου μελέτης των στατιστικών χαρακτηριστικών με χρήση του πακέτου λογισμικού Matlab

### 4.1 Περιγραφή εισόδων και εξόδων του κώδικα

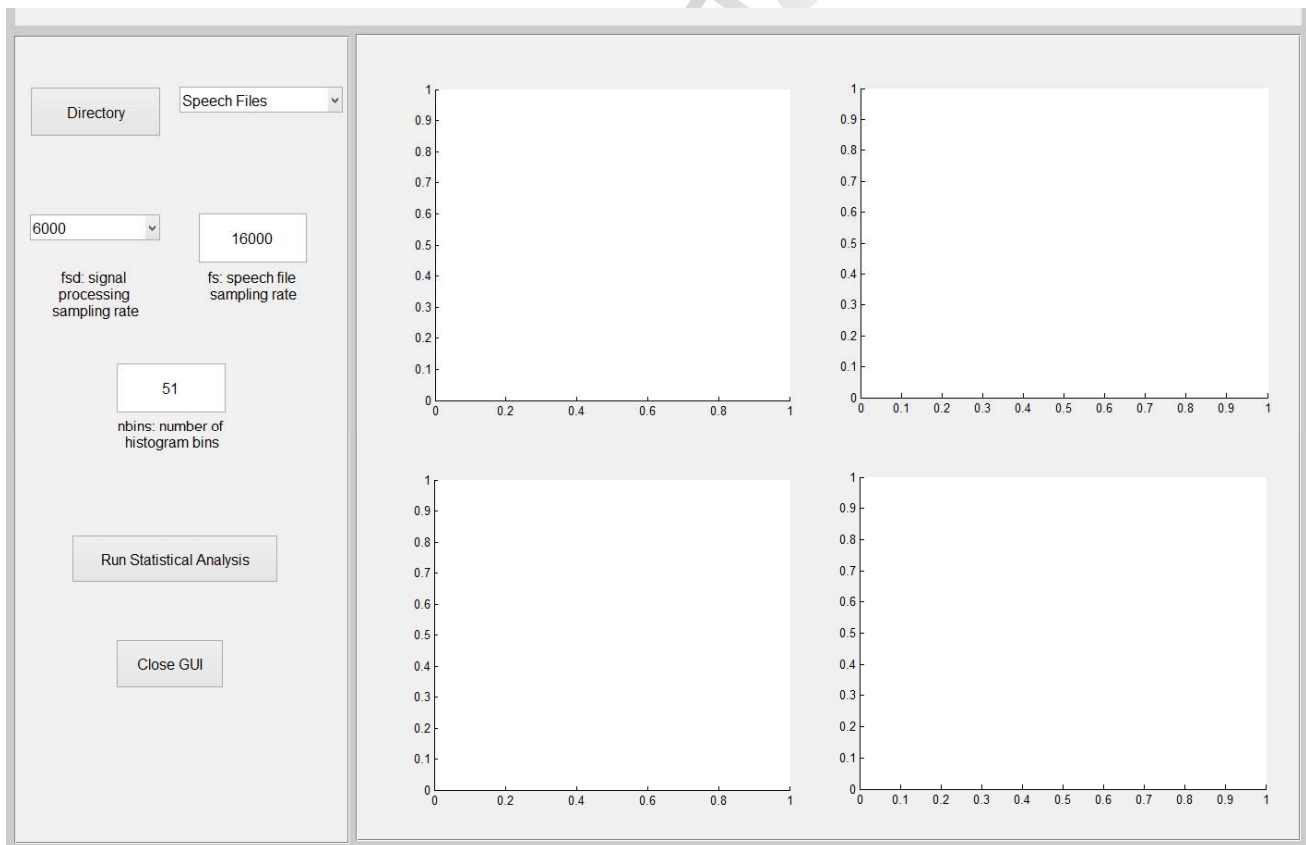
Στα πλαίσια της παρούσης εργασίας χρησιμοποιήθηκε και παραμετροποιήθηκε λογισμικό για την στατιστική επεξεργασία σημάτων ήχου, σε συνδυασμό με ένα γραφικό περιβάλλον χρήστη (Graphical User Interface, GUI), για την εισαγωγή παραμέτρων από τον χρήστη και την επιλογή αρχείων ήχου προς επεξεργασία.

Πιο συγκεκριμένα, ο χρήστης έχει τη δυνατότητα μέσω του περιβάλλοντος αυτού να επιλέγει τιμές για τις εξής παραμέτρους:

- Ρυθμό δειγματοληψίας σήματος
- Ρυθμό δειγματοληψίας επεξεργασίας
- Αριθμό στηλών στο ιστόγραμμα που θα παραχθεί

Επίσης ο χρήστης μπορεί να επιλέξει τον φάκελο και το αρχείο που θα υποστεί την επεξεργασία, και να τρέξει τελικά την προσομοίωση.

Το περιβάλλον φαίνεται στην επόμενη εικόνα.



Εικόνα 5: Γραφικό περιβάλλον επεξεργασίας σημάτων ήχου

Τα αποτελέσματα της ανάλυσης παρουσιάζονται σε τέσσερα επιμέρους διαγράμματα. Το πρώτο εξ αυτών είναι ένα ιστόγραμμα που παρουσιάζει το ηχητικό σήμα που επεξεργαστήκαμε. Το δεύτερο διάγραμμα παρουσιάζει το φάσμα ισχύος του σήματος για τρεις διαφορετικές τιμές παραθύρου: 32, 128 και 512. Το γράφημα παρουσιάζει ισχύ σε dB συναρτήσει της συχνότητας.

Το τρίτο γράφημα των αποτελεσμάτων παρουσιάζει την τιμή της αυτοσυσχέτισης του επιλεγμένου ηχητικού σήματος συναρτήσει του χρόνου. Όπως και στην προηγούμενη περίπτωση, έτσι και εδώ το διάγραμμα παρουσιάζει το φάσμα ισχύος του σήματος για τρεις διαφορετικές τιμές παραθύρου: 32, 128 και 512.

Τέλος, το τελευταίο διάγραμμα παρουσιάζει επίσης την τιμή της αυτοσυσχέτισης του επιλεγμένου ηχητικού σήματος συναρτήσει του χρόνου, αλλά σε μονάδες δειγμάτων (samples) και όχι χρόνου σε δευτερόλεπτα.

## 4.2 Σύντομη παρουσίαση του κώδικα

Ο κώδικας αποτελείται από δύο κύρια αρχεία, το ένα εκ των οποίων αποτελεί το βασικό αρχείο δημιουργίας των απαραίτητων αρχικοποιήσεων και κλήσης της συνάρτησης του γραφικού περιβάλλοντος, και το δεύτερο αποτελεί το αρχείο της συνάρτησης που κατασκευάζει το γραφικό περιβάλλον.

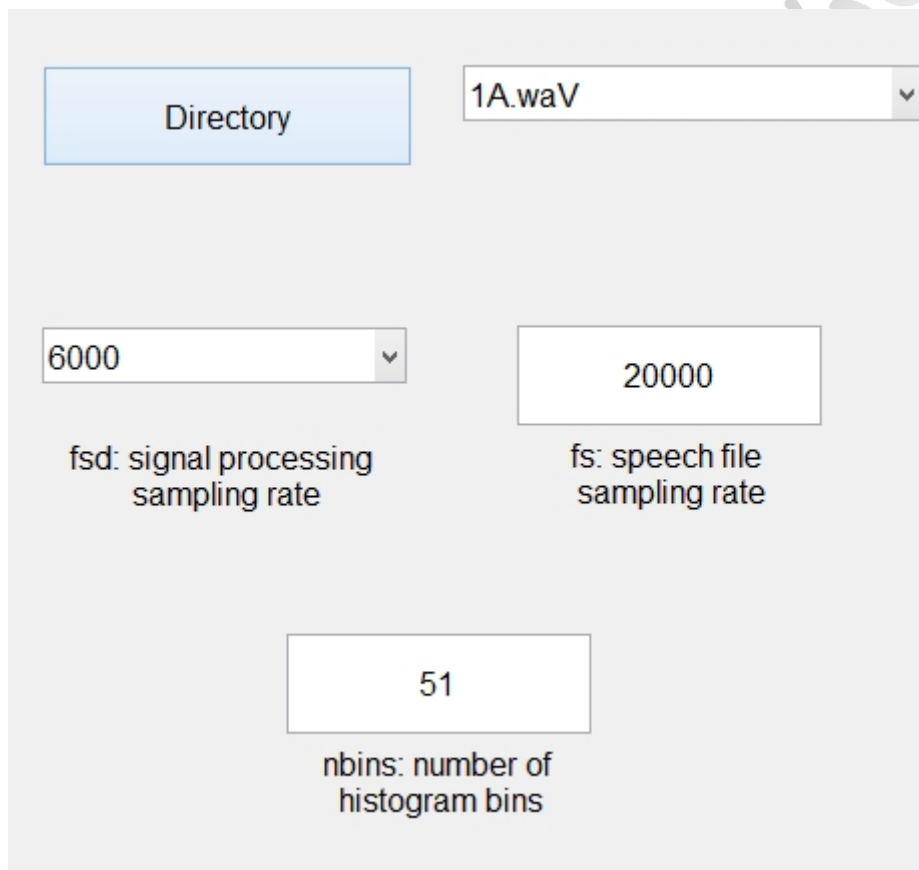
Στο 1<sup>ο</sup> κατά σειρά αρχείο, δημιουργούνται καταρχήν οι απαραίτητες διαδρομές (paths) και αρχικοποιείται το κεντρικό παράθυρο / εικόνα, και στη συνέχεια καλείται η συνάρτηση μέσω της οποίας δημιουργείται το GUI και πραγματοποιούνται οι διάφορες στατιστικές αναλύσεις των αρχείων που έχει επιλέξει ο χρήστης.

Το ιστόγραμμα δημιουργείται απευθείας μέσα στο ίδιο αυτό αρχείο / συνάρτηση, ενώ στον κώδικα περιλαμβάνεται και μια ακόμα συνάρτηση για την παραγωγή του φάσματος ισχύος του ηχητικού σήματος καθώς και των τιμών αυτοσυσχέτισης του ηχητικού σήματος.

## 5 Αποτελέσματα εφαρμογής μεθόδου

### 5.1 Αποτελέσματα για 1<sup>ο</sup> αρχείο ήχου – χαμηλό sampling rate

Σε αυτή την ενότητα επιλέγεται το 1<sup>ο</sup> από μια σειρά αρχείων ήχου, τα οποία θα χρησιμοποιηθούν στην παρούσα εργασία για να μελετηθεί ο αλγόριθμος ανάλυσης που παρουσιάστηκε στο προηγούμενο κεφάλαιο. Αρχικά, από την οθόνη διεπαφής με τον χρήστη, επιλέγονται τα χαρακτηριστικά της ανάλυσης, όπως φαίνεται στην εικόνα που ακολουθεί.

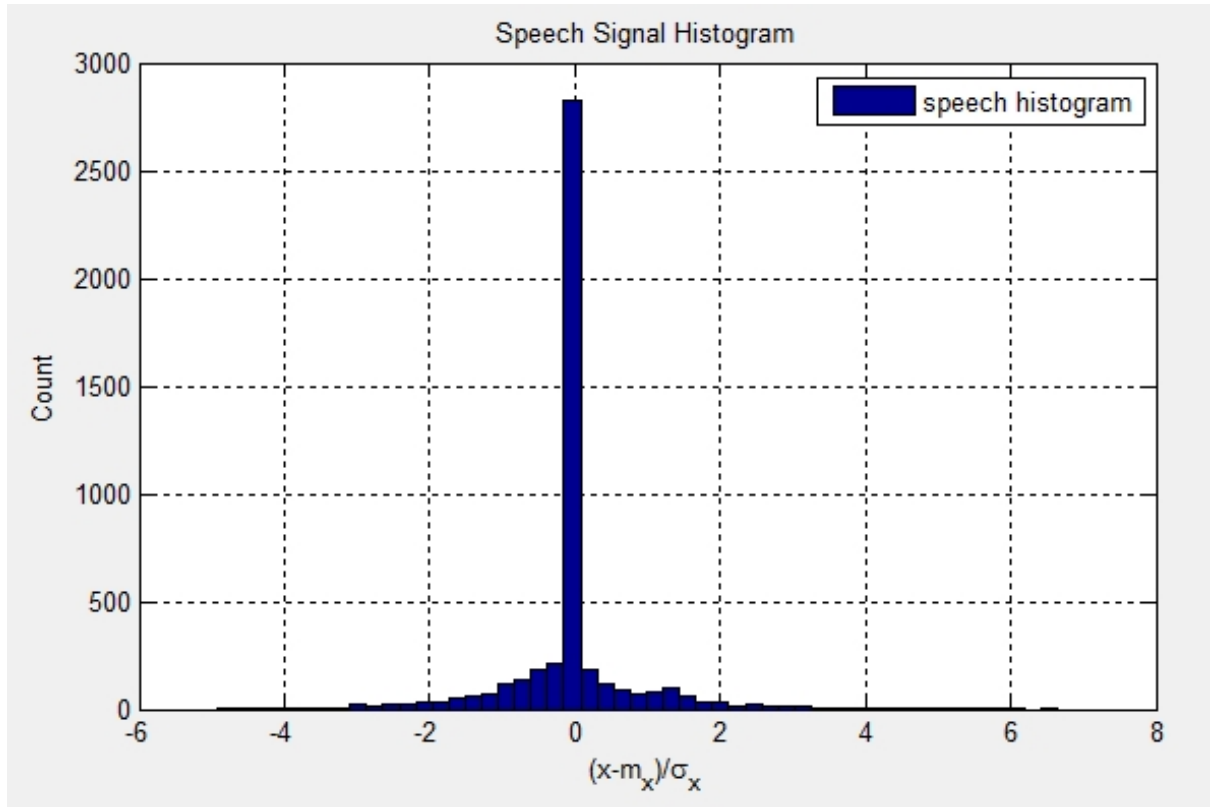


The screenshot shows a software interface for configuring audio analysis parameters. It includes a 'Directory' button, a file name dropdown set to '1A.wav', a 'fsd' dropdown set to '6000', a 'fs' input field set to '20000', and an 'nbins' input field set to '51'. Labels for each parameter are provided below the input fields.

Parameter	Value
Directory	1A.wav
fsd: signal processing sampling rate	6000
fs: speech file sampling rate	20000
nbins: number of histogram bins	51

Εικόνα 6: 1<sup>ο</sup> Αποτελέσματα για το 1<sup>ο</sup> αρχείο ήχου: Εισαγωγή παραμέτρων

Αντίστοιχα, στην επόμενη εικόνα παρουσιάζεται το ιστόγραμμα του συγκεκριμένου αρχείου ήχου. Όπως φαίνεται, πρόκειται για ένα αρχείο ήχου με ιδιαίτερα απλή κατανομή, πολύ κοντά στις μηδενικές τιμές.

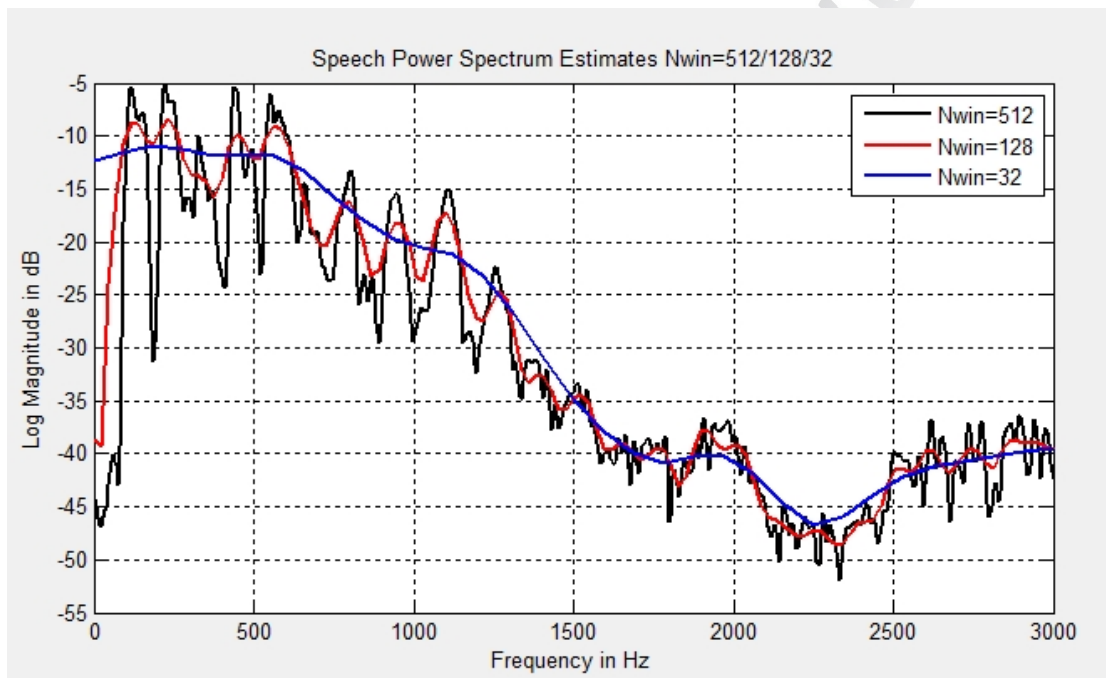


Εικόνα 7: Αποτελέσματα για το 1ο αρχείο ήχου: Ιστόγραμμα



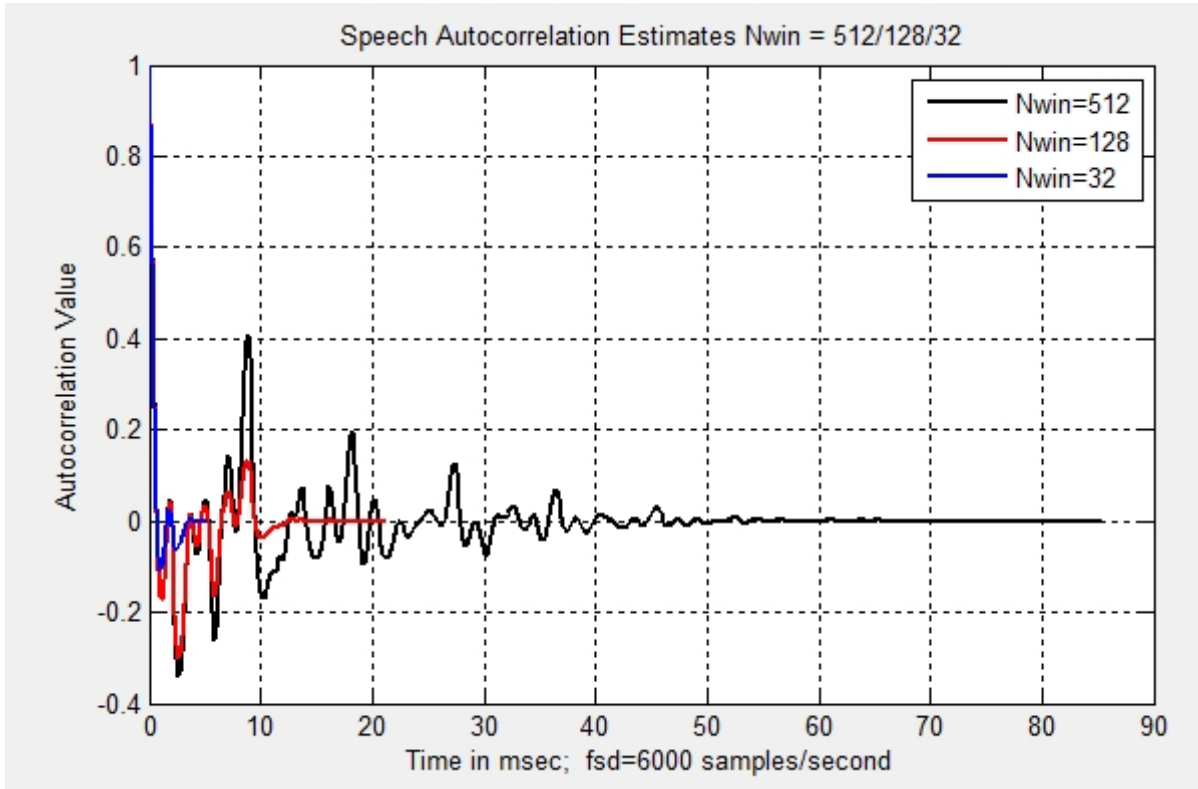
Στη συνέχεια παρουσιάζεται το φάσμα ισχύος για το εν λόγω σήμα, για τρεις περιπτώσεις μεγέθους παραθύρου: 32, 128 και 512. Όπως φαίνεται από την εικόνα, το φάσμα ισχύος ακολουθεί την ίδια περίπου μορφή και για τις τρεις περιπτώσεις, δηλαδή έχει υψηλή τιμή ισχύος στις χαμηλές συχνότητες (περίπου  $-10$  dB) και χαμηλότερες τιμές ισχύος στις υψηλές συχνότητες (κάτω από  $-30$  dB για συχνότητες μεγαλύτερες των 1500 Hz).

Από την άλλη μεριά, η βασική διαφοροποίηση που εισάγεται για τις διαφορετικές τιμές μεγέθους παραθύρου, αφορά στις διακυμάνσεις του φάσματος ισχύος. Στην περίπτωση που το παράθυρο είναι μικρό, δεν γίνονται αντιληπτές κάποιες εκ των διακυμάνσεων του αρχείου, με αποτέλεσμα η γραμμή να είναι σαφώς πιο ομαλή. Αντίθετα, όσο μεγαλώνει το μέγεθος παραθύρου, καταγράφονται και μεγαλύτερες διακυμάνσεις στην ισχύ του σήματος συναρτήσει της συχνότητας, και τελικά το αποτέλεσμα είναι μια γραμμή με σαφώς περισσότερες αυξομειώσεις στην ισχύ.



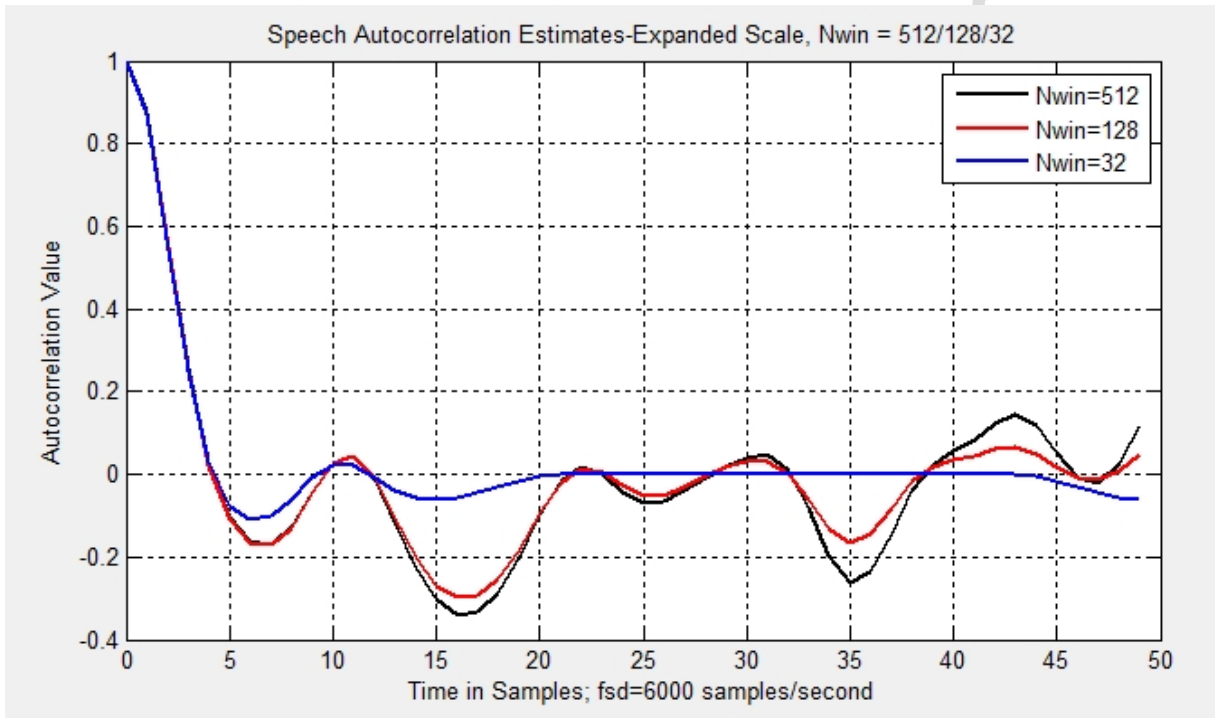
Εικόνα 8: Αποτελέσματα για το 1ο αρχείο ήχου: Φάσμα ισχύος

Αντίστοιχα, η Εικόνα 9 παρουσιάζει τα αποτελέσματα της αυτοσυσχέτισης σε σχέση με το χρόνο για το 1<sup>ο</sup> αυτό υπό εξέταση σήμα. Ανάλογα με το μέγεθος παραθύρου, η αυτοσυσχέτιση λαμβάνει τιμές για διαφορετικό εύρος τιμών του χρόνου. Επιπλέον, όσο μεγαλύτερο είναι το μέγεθος παραθύρου, τόσο περισσότερο καθυστερεί η αυτοσυσχέτιση να μηδενιστεί.



Εικόνα 9: Αποτελέσματα για το 1<sup>ο</sup> αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε second)

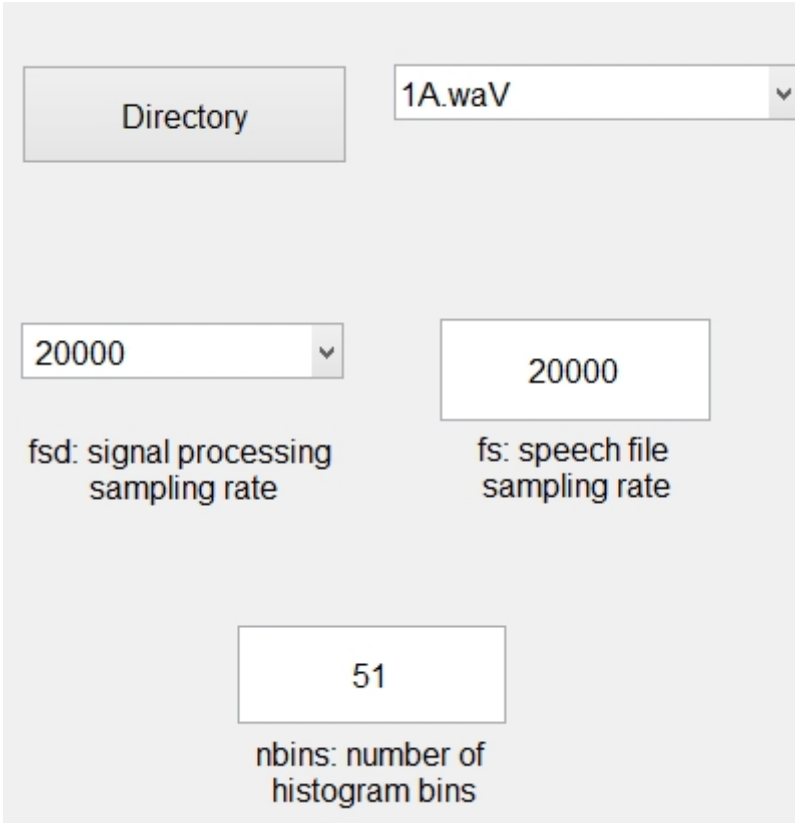
Για να φανεί λίγο καλύτερα η διαφορά που υπεισέρχεται με την αλλαγή της τιμής του παραθύρου, το παραπάνω γράφημα παρουσιάζεται σε μεγαλύτερη ανάλυση, εστιάζοντας στις μικρότερες τιμές του χρόνου και έχοντας αυτή τη φορά στον άξονα των τετμημένων δείγματα αντί για χρόνο. Το αποτέλεσμα φαίνεται στην επόμενη εικόνα, όπου βλέπουμε ότι ενώ για μικρό μέγεθος παραθύρου η τιμή έχει μηδενιστεί σχεδόν εξαρχής, η τιμή της αυτοσυσχέτισης για τις δύο μεγαλύτερες τιμές μεγέθους παραθύρου παρουσιάζει αυξομειώσεις.



Εικόνα 10: Αποτελέσματα για το 1ο αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε samples)

## 5.2 Αποτελέσματα για 1<sup>ο</sup> αρχείο ήχου – υψηλό sampling rate

Αντίστοιχα, σε αυτή την ενότητα επιλέγεται το ίδιο μεν αρχείο με την προηγούμενη ενότητα, αλλά η ανάλυση γίνεται με διαφορετικό ρυθμό δειγματοληψίας. Συγκεκριμένα, ενώ στην προηγούμενη ενότητα η ανάλυση έγινε με ρυθμό 6,000 Hz, τώρα γίνεται με ρυθμό 20,000 Hz. Όπως και πριν, από την οθόνη διεπαφής με τον χρήστη, επιλέγονται τα χαρακτηριστικά της ανάλυσης, όπως φαίνεται στην εικόνα που ακολουθεί.



The image shows a software interface for audio analysis. It features several input fields and labels:

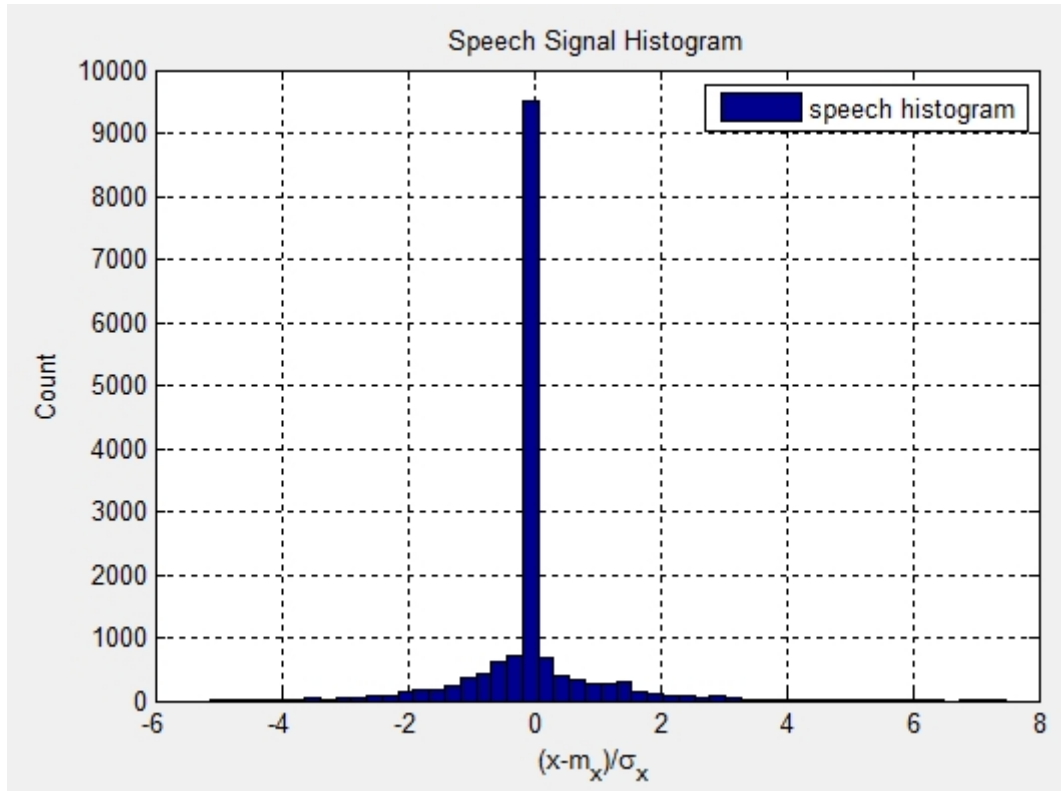
- A "Directory" button.
- A dropdown menu showing "1A.waV".
- A dropdown menu showing "20000".
- A text box containing "20000".
- A text box containing "51".

Labels for the parameters are:

- fsd: signal processing sampling rate
- fs: speech file sampling rate
- nbins: number of histogram bins

Εικόνα 11: 1ο Αποτελέσματα για το 1ο αρχείο ήχου: Εισαγωγή παραμέτρων

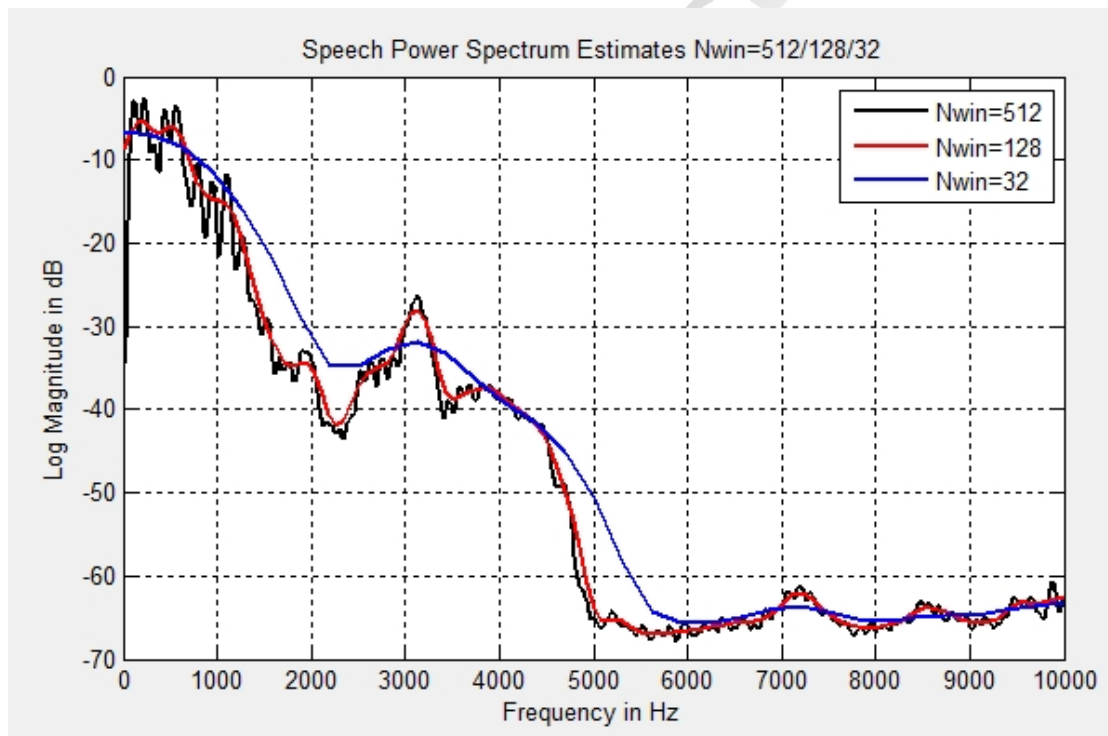
Αντίστοιχα, στην επόμενη εικόνα παρουσιάζεται το ιστόγραμμα του συγκεκριμένου αρχείου ήχου. Όπως φαίνεται, πρόκειται για ένα αρχείο ήχου με ιδιαίτερα απλή κατανομή, πολύ κοντά στις μηδενικές τιμές. Εδώ δεν βλέπουμε ιδιαίτερες διαφορές σε σύγκριση με την ανάλυση που έγινε με τον χαμηλότερο ρυθμό δειγματοληψίας.



Εικόνα 12: Αποτελέσματα για το 1ο αρχείο ήχου: Ιστόγραμμα

Στη συνέχεια παρουσιάζεται το φάσμα ισχύος για το εν λόγω σήμα, με τον υψηλό ρυθμό δειγματοληψίας, για τρεις περιπτώσεις μεγέθους παραθύρου: 32, 128 και 512. Όπως φαίνεται από την εικόνα, το φάσμα ισχύος ακολουθεί την ίδια περίπου μορφή και για τις τρεις περιπτώσεις, δηλαδή έχει υψηλή τιμή ισχύος στις χαμηλές συχνότητες (περίπου  $-10$  dB για συχνότητες κάτω των  $1000^{\omega v}$  Hz). Αντίθετα, κυμαίνεται μεταξύ των  $-30$  και  $-40$  dB για συχνότητες από  $1500$  μέχρι  $4000$  Hz, και για ακόμα μεγαλύτερες συχνότητες πέφτει κάτω από τα  $-60$  dB.

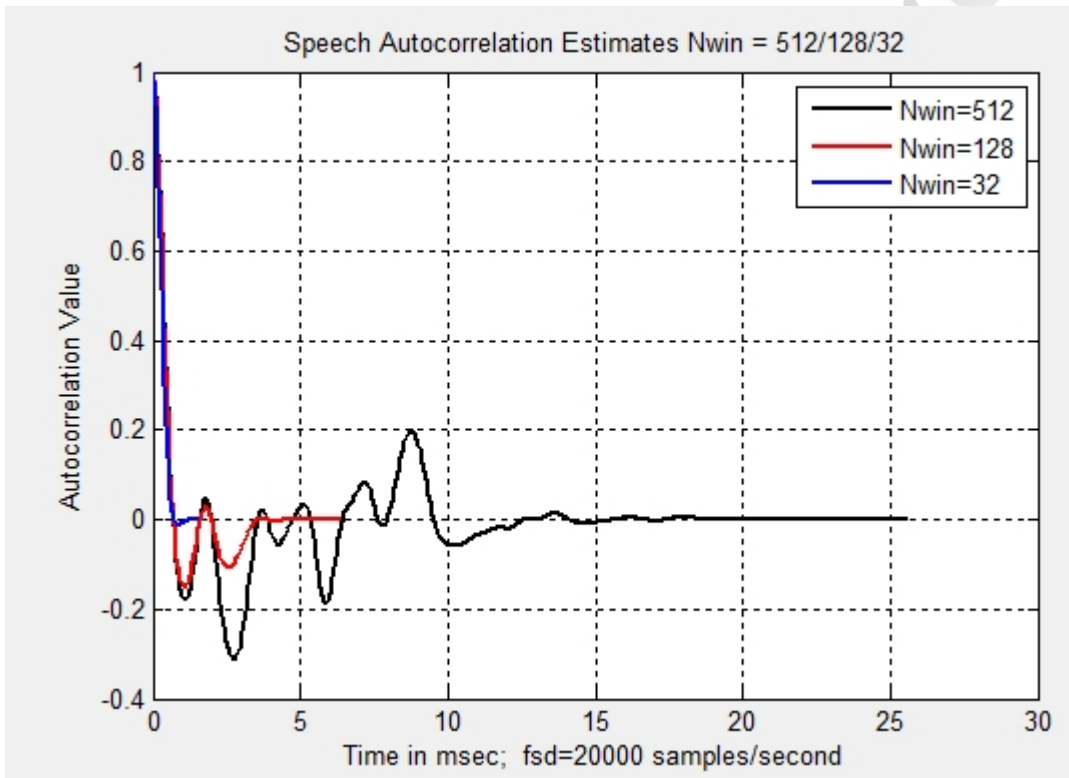
Από την άλλη μεριά, η βασική διαφοροποίηση που εισάγεται για τις διαφορετικές τιμές μεγέθους παραθύρου, αφορά στις διακυμάνσεις του φάσματος ισχύος, όπως και στην περίπτωση του χαμηλού ρυθμού δειγματοληψίας. Στην περίπτωση που το παράθυρο είναι μικρό, δεν γίνονται αντιληπτές κάποιες εκ των διακυμάνσεων του αρχείου, με αποτέλεσμα η γραμμή να είναι σαφώς πιο ομαλή. Αντίθετα, όσο μεγαλώνει το μέγεθος παραθύρου, καταγράφονται και μεγαλύτερες διακυμάνσεις στην ισχύ του σήματος συναρτήσει της συχνότητας, και τελικά το αποτέλεσμα είναι μια γραμμή με σαφώς περισσότερες αυξομειώσεις στην ισχύ. Παρόλα αυτά, σε σύγκριση με την περίπτωση του χαμηλού ρυθμού δειγματοληψίας που μελετήθηκε στην προηγούμενη ενότητα, οι διαφορές μεταξύ των φασμάτων ισχύος είναι αρκετά μικρότερες για τις διαφορετικές τιμές παραθύρου.



Εικόνα 13: Αποτελέσματα για το 1ο αρχείο ήχου: Φάσμα ισχύος

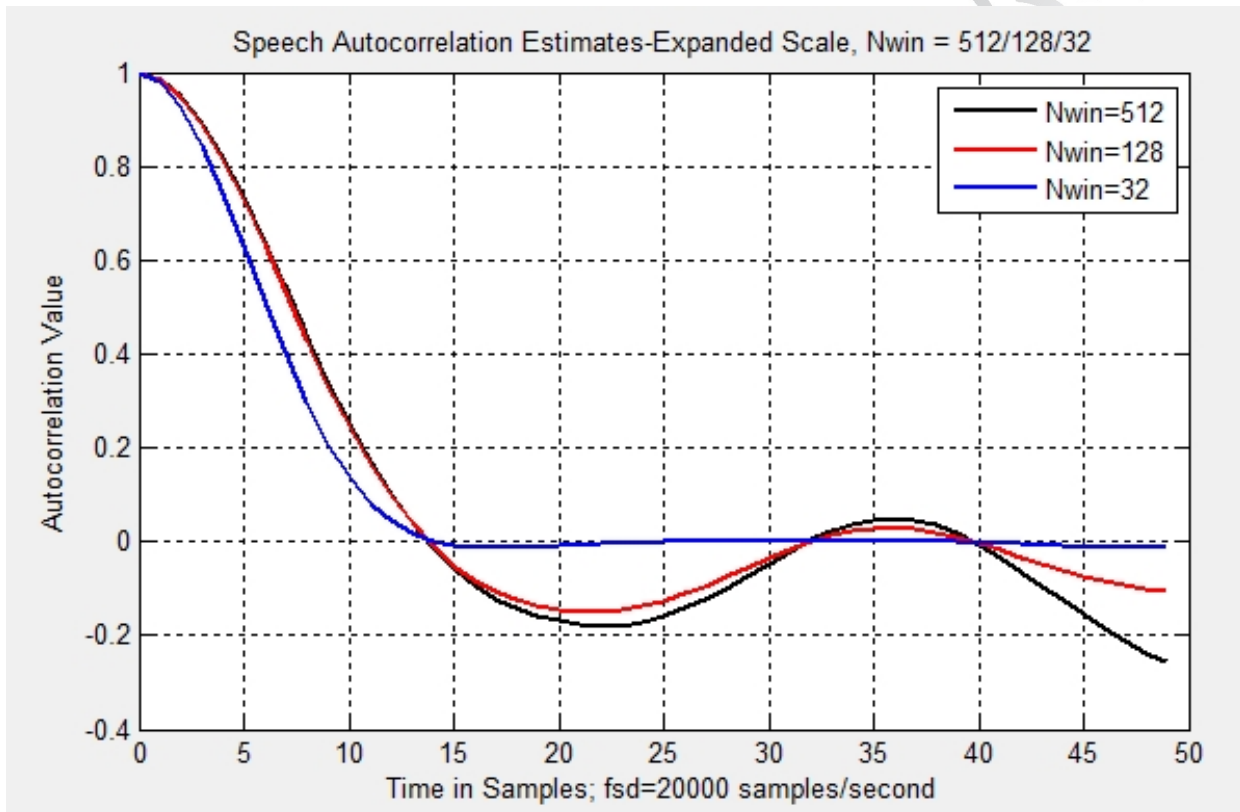
Αντίστοιχα, η Εικόνα 14 παρουσιάζει τα αποτελέσματα της αυτοσυσχέτισης σε σχέση με το χρόνο για το 1<sup>ο</sup> αυτό υπό εξέταση σήμα, και την περίπτωση του υψηλού ρυθμού δειγματοληψίας. Ανάλογα με το μέγεθος παραθύρου, η αυτοσυσχέτιση λαμβάνει τιμές για διαφορετικό εύρος τιμών του χρόνου. Επιπλέον, όσο μεγαλύτερο είναι το μέγεθος παραθύρου, τόσο περισσότερο καθυστερεί η αυτοσυσχέτιση να μηδενιστεί.

Παρόλα αυτά, είναι εμφανής η διαφορά σε σχέση με την περίπτωση του χαμηλού ρυθμού μετάδοσης. Συγκεκριμένα, φαίνεται πως οι διακυμάνσεις του σήματος αυτοσυσχέτισης είναι σαφώς μικρότερες και πιο ομαλές στην περίπτωση του υψηλού ρυθμού δειγματοληψίας.



Εικόνα 14: Αποτελέσματα για το 1ο αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε second)

Για να φανεί λίγο καλύτερα η διαφορά που υπεισέρχεται με την αλλαγή της τιμής του παραθύρου, το παραπάνω γράφημα παρουσιάζεται σε μεγαλύτερη ανάλυση, εστιάζοντας στις μικρότερες τιμές του χρόνου, και έχοντας αυτή τη φορά στον άξονα των τετμημένων δείγματα αντί για χρόνο, όπως έγινε και στην προηγούμενη ενότητα. Το αποτέλεσμα φαίνεται στην επόμενη εικόνα, όπου βλέπουμε ότι ενώ για μικρό μέγεθος παραθύρου η τιμή έχει μηδενιστεί σχεδόν εξαρχής, η τιμή της αυτοσυσχέτισης για τις δύο μεγαλύτερες τιμές μεγέθους παραθύρου παρουσιάζει αυξομειώσεις. Οι αυξομειώσεις όμως αυτές, είναι σαφώς βραδύτερες από την περίπτωση που ο ρυθμός δειγματοληψίας ήταν 6000 Hz.

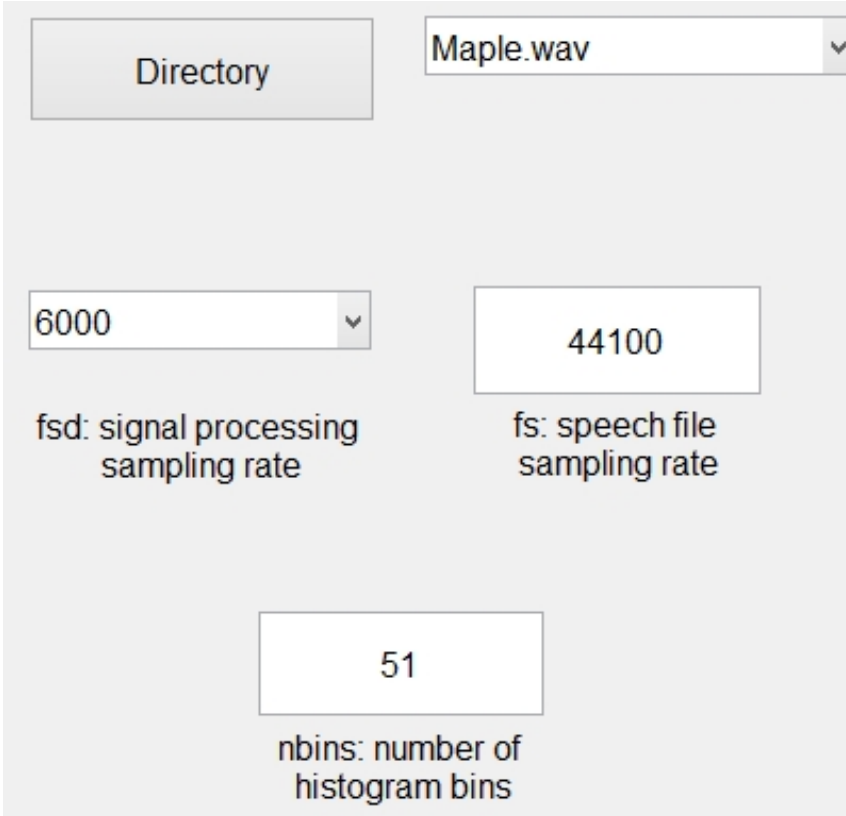


Εικόνα 15: Αποτελέσματα για το 1ο αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε samples)



### 5.3 Αποτελέσματα για 2<sup>ο</sup> αρχείο ήχου – χαμηλό sampling rate

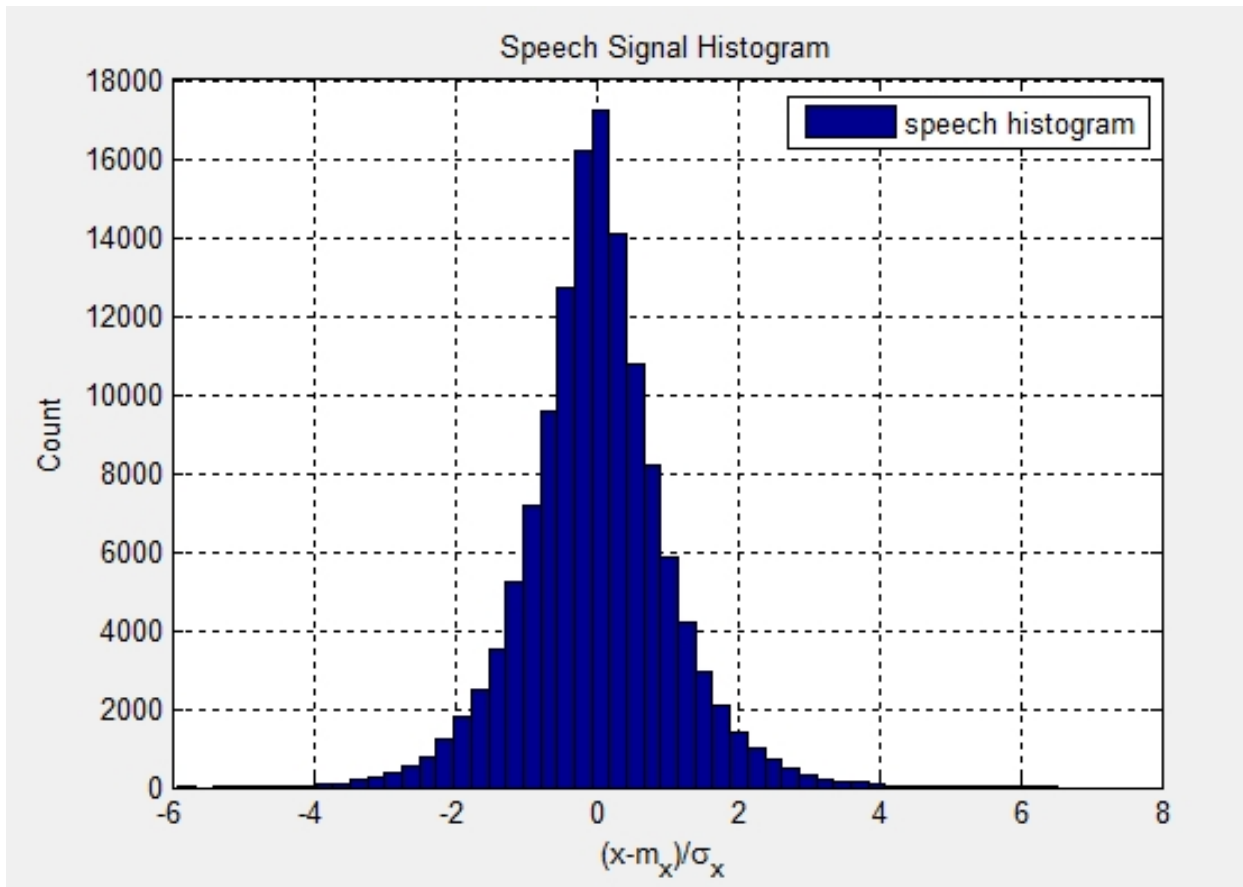
Σε αυτή την ενότητα επιλέγεται ένα διαφορετικό από μια σειρά αρχείων ήχου, τα οποία χρησιμοποιούνται για να μελετηθεί ο αλγόριθμος ανάλυσης που παρουσιάστηκε στο προηγούμενο κεφάλαιο. Αρχικά, όπως και προηγουμένως, από την οθόνη διεπαφής με τον χρήστη, επιλέγονται τα χαρακτηριστικά της ανάλυσης, όπως φαίνεται στην εικόνα που ακολουθεί. Το όνομα του αρχείου είναι «maple.wav».



The image shows a software interface for audio analysis. It features a 'Directory' field containing 'Maple.wav'. Below this, there are three input fields: 'fsd: signal processing sampling rate' with a dropdown menu set to '6000', 'fs: speech file sampling rate' with a text box containing '44100', and 'nbins: number of histogram bins' with a text box containing '51'.

Εικόνα 16: 2ο Αποτελέσματα για το 1ο αρχείο ήχου: Εισαγωγή παραμέτρων

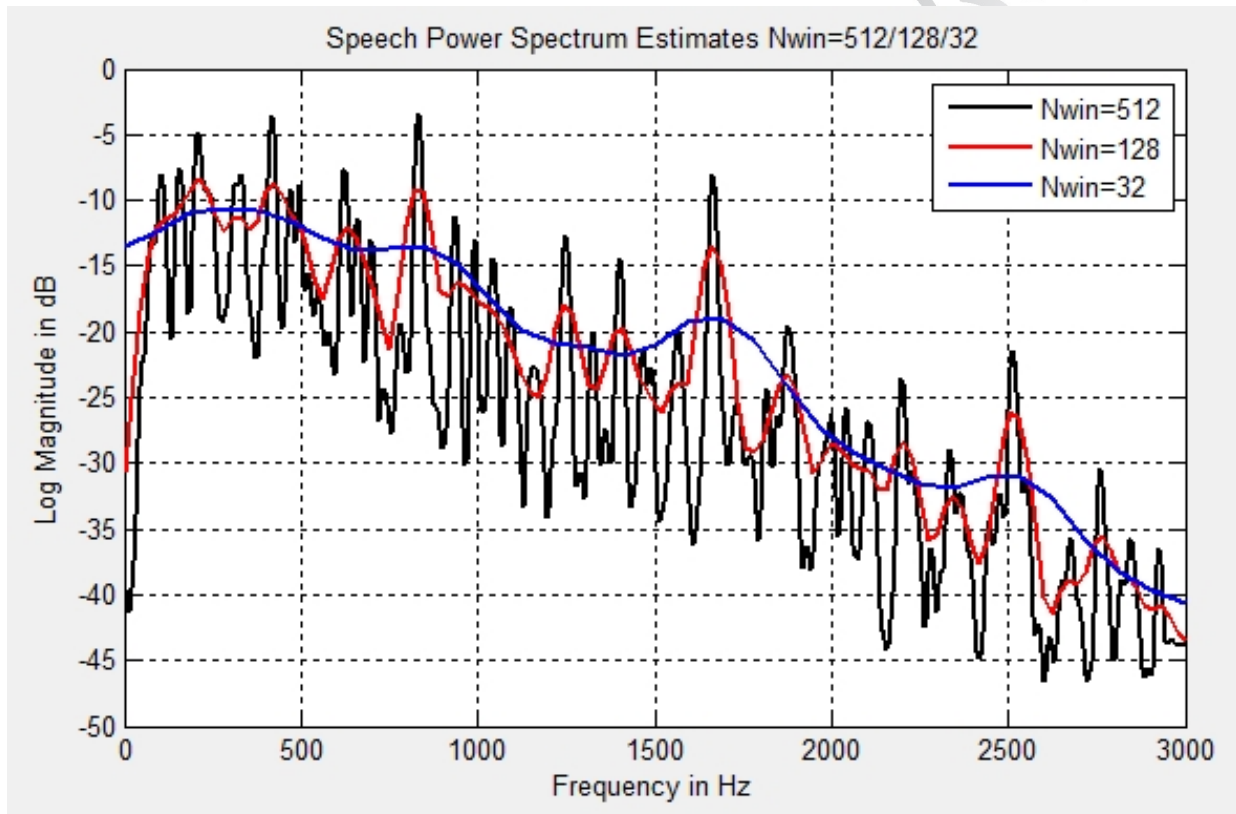
Αντίστοιχα, όπως και στις προηγούμενες περιπτώσεις, στην επόμενη εικόνα παρουσιάζεται το ιστόγραμμα του συγκεκριμένου αρχείου ήχου. Όπως φαίνεται, πρόκειται για ένα αρχείο ήχου με αρκετά πιο σύνθετη κατανομή σε ένα σαφώς μεγαλύτερο εύρος τιμών από το προηγούμενο αρχείο που εξετάστηκε.



Εικόνα 17: Αποτελέσματα για το 2ο αρχείο ήχου: Ιστόγραμμα

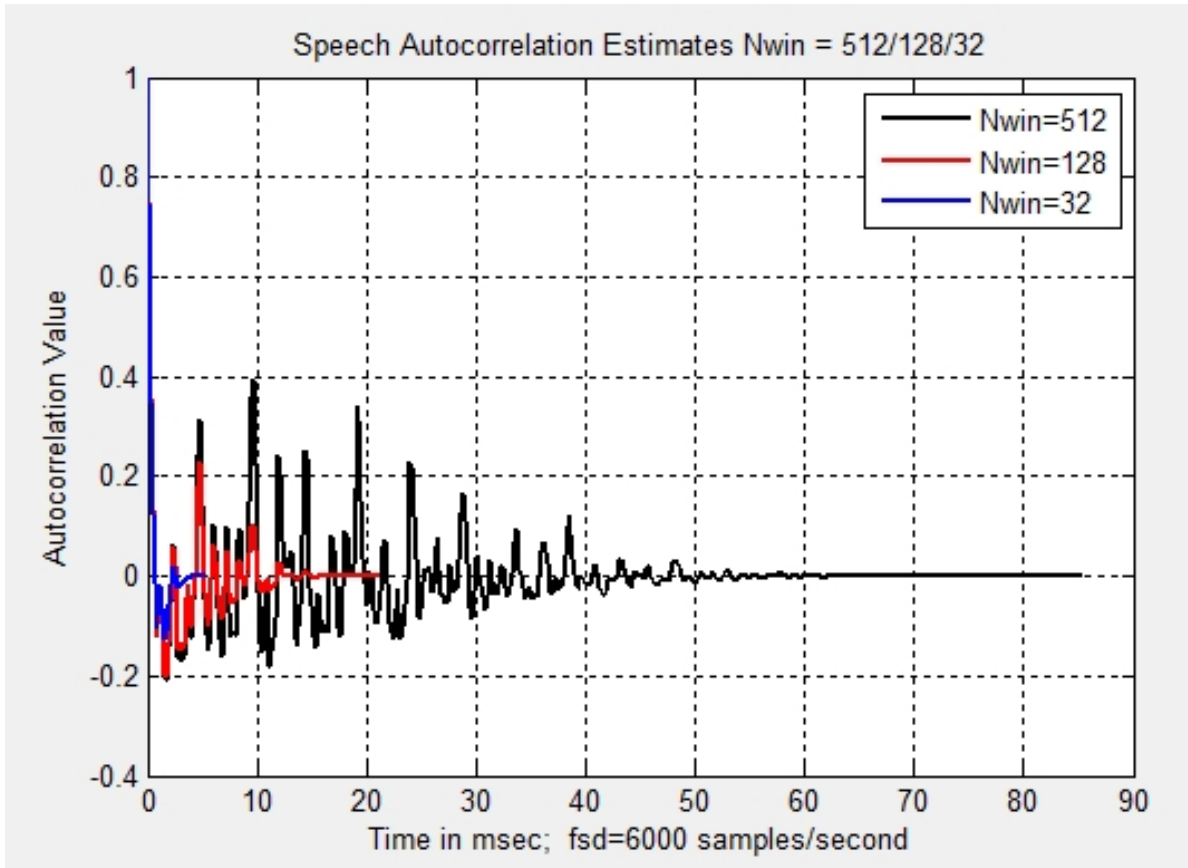
Στη συνέχεια παρουσιάζεται το φάσμα ισχύος για το εν λόγω σήμα, για τρεις περιπτώσεις μεγέθους παραθύρου, όπως ακριβώς έγινε και με το προηγούμενο σήμα: 32, 128 και 512. Όπως φαίνεται από την εικόνα, το φάσμα ισχύος ακολουθεί την ίδια περίπου μορφή και για τις τρεις περιπτώσεις, και μειώνεται σχεδόν γραμμικά από τα  $-10$  dB ως τα  $-40$  dB, τιμή την οποία φτάνει στα 3,000 Hz.

Παράλληλα, όπως και στην περίπτωση του προηγούμενου σήματος που μελετήθηκε, η βασική διαφοροποίηση που εισάγεται για τις διαφορετικές τιμές μεγέθους παραθύρου, αφορά στις διακυμάνσεις του φάσματος ισχύος. Στην περίπτωση που το παράθυρο είναι μικρό, δεν γίνονται αντιληπτές κάποιες εκ των διακυμάνσεων του αρχείου, με αποτέλεσμα η γραμμή να είναι σαφώς πιο ομαλή. Αντίθετα, όσο μεγαλώνει το μέγεθος παραθύρου, καταγράφονται και μεγαλύτερες διακυμάνσεις στην ισχύ του σήματος συναρτήσει της συχνότητας, και τελικά το αποτέλεσμα είναι μια γραμμή με σαφώς περισσότερες αυξομειώσεις στην ισχύ.



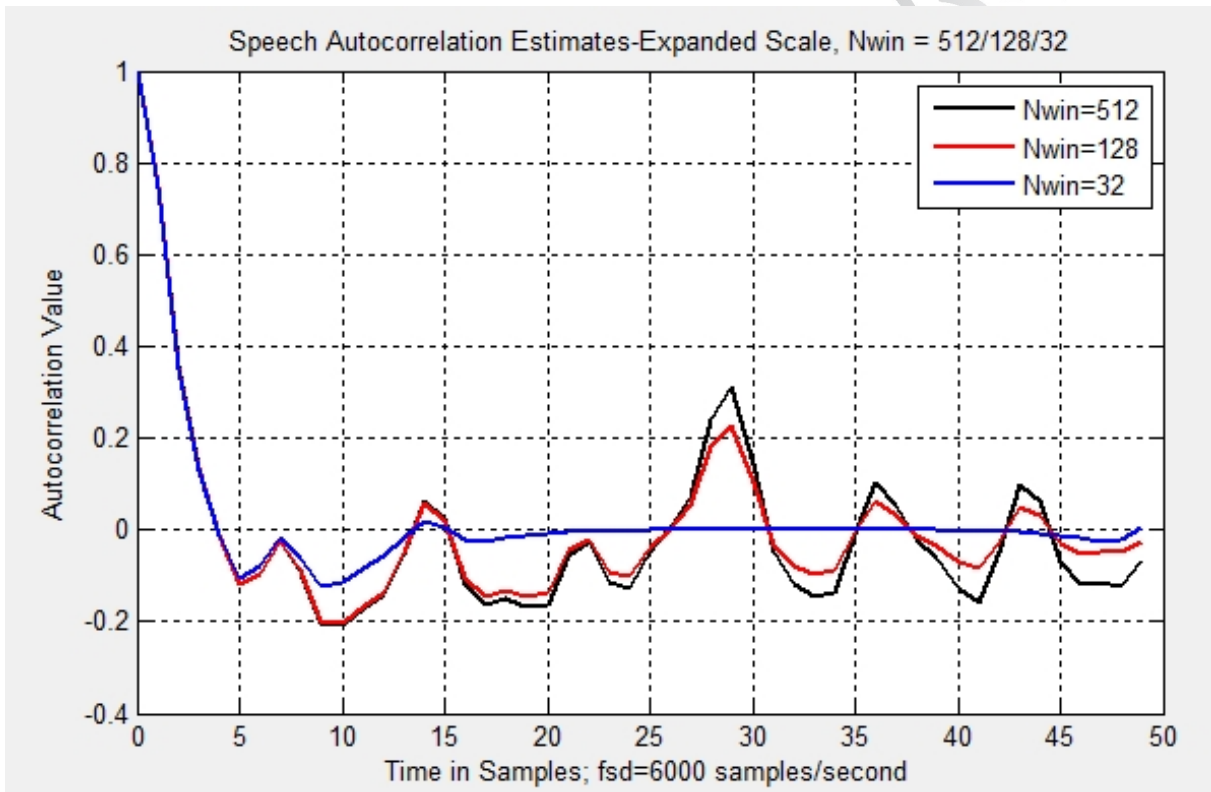
Εικόνα 18: Αποτελέσματα για το 2ο αρχείο ήχου: Φάσμα ισχύος

Αντίστοιχα, η Εικόνα 19 παρουσιάζει τα αποτελέσματα της αυτοσυσχέτισης σε σχέση με το χρόνο για το 2<sup>ο</sup> αυτό υπό εξέταση σήμα. Ανάλογα με το μέγεθος παραθύρου, η αυτοσυσχέτιση λαμβάνει τιμές για διαφορετικό εύρος τιμών του χρόνου. Επιπλέον, όσο μεγαλύτερο είναι το μέγεθος παραθύρου, τόσο περισσότερο καθυστερεί η αυτοσυσχέτιση να μηδενιστεί, κάτι που είχε παρατηρηθεί και από το προηγούμενο σήμα.



Εικόνα 19: Αποτελέσματα για το 2ο αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε second)

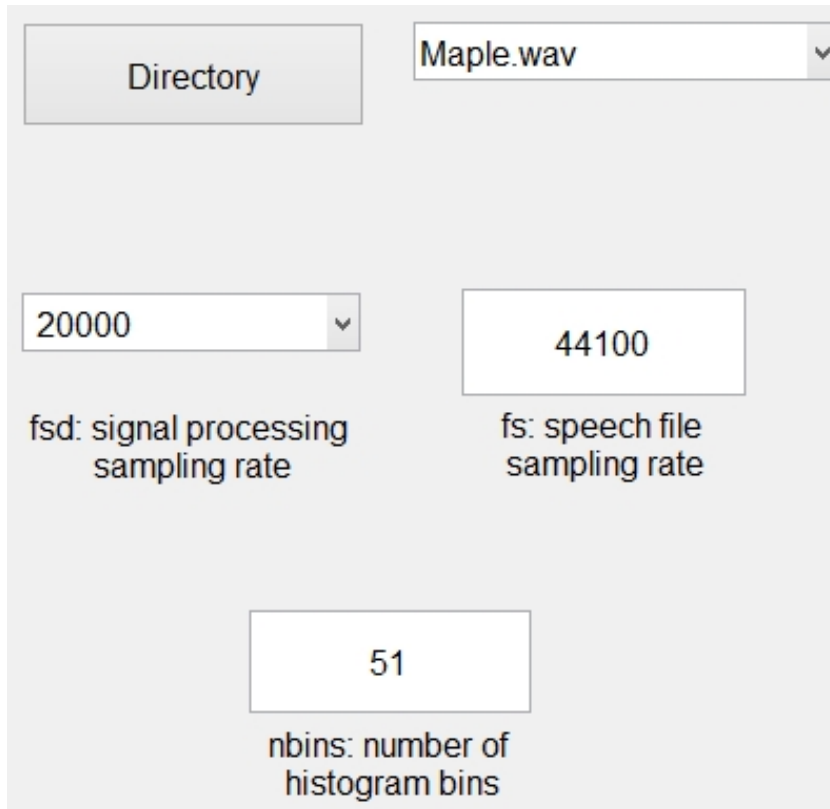
Για να φανεί λίγο καλύτερα η διαφορά που υπεισέρχεται με την αλλαγή της τιμής του παραθύρου, το παραπάνω γράφημα παρουσιάζεται σε μεγαλύτερη ανάλυση, εστιάζοντας στις μικρότερες τιμές του χρόνου και έχοντας αυτή τη φορά στον άξονα των τετμημένων δείγματα αντί για χρόνο. Το αποτέλεσμα φαίνεται στην επόμενη εικόνα, όπου βλέπουμε ότι ενώ για μικρό μέγεθος παραθύρου η τιμή έχει μηδενιστεί σχεδόν εξαρχής, η τιμή της αυτοσυσχέτισης για τις δύο μεγαλύτερες τιμές μεγέθους παραθύρου παρουσιάζει αυξομειώσεις, οι οποίες είναι παρόμοιες για τις περιπτώσεις μεγέθους παραθύρου 128 και 512.



Εικόνα 20: Αποτελέσματα για το 2ο αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε samples)

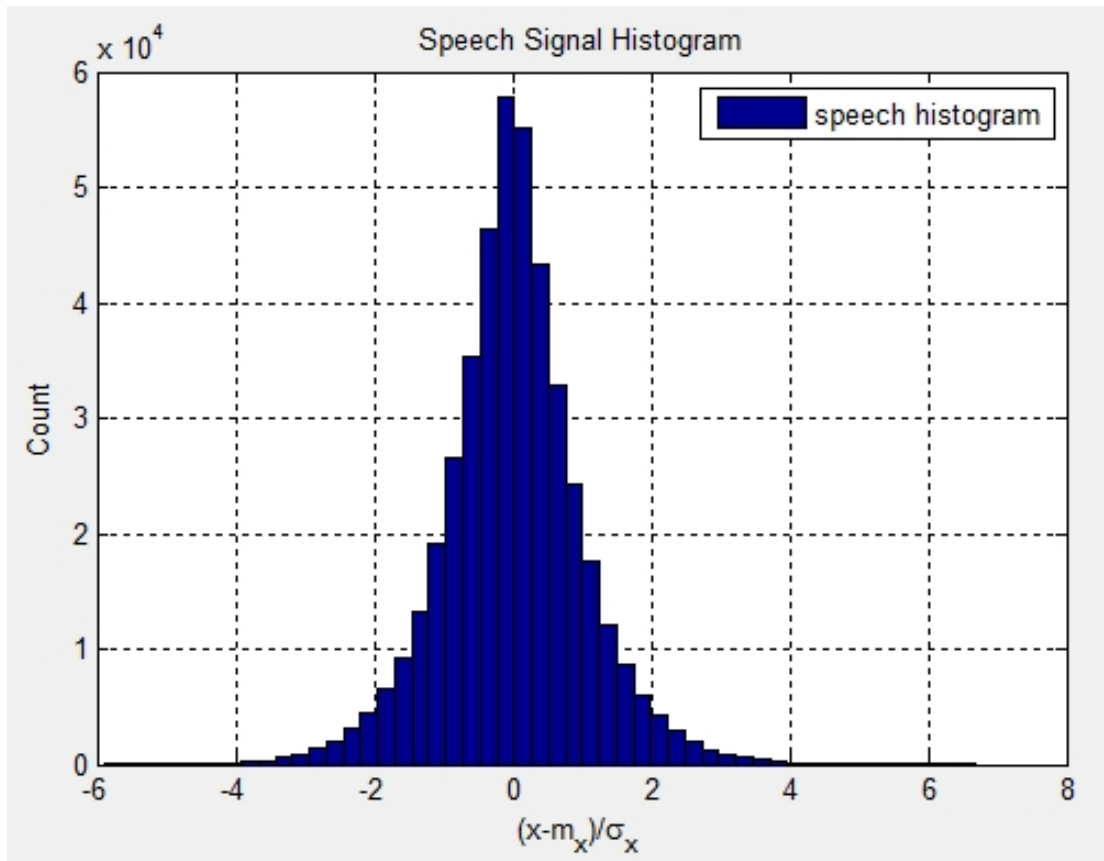
#### 5.4 Αποτελέσματα για 2<sup>ο</sup> αρχείο ήχου – υψηλό sampling rate

Όπως και στο προηγούμενο αρχείο, έτσι και σε αυτό, η ανάλυση τώρα γίνεται με διαφορετικό ρυθμό δειγματοληψίας. Συγκεκριμένα, ενώ στην προηγούμενη ενότητα η ανάλυση έγινε με ρυθμό 6,000 Hz, τώρα γίνεται με ρυθμό 20,000 Hz. Όπως και πριν, από την οθόνη διεπαφής με τον χρήστη, επιλέγονται τα χαρακτηριστικά της ανάλυσης, όπως φαίνεται στην εικόνα που ακολουθεί.



Εικόνα 21: 1ο Αποτελέσματα για το 2ο αρχείο ήχου: Εισαγωγή παραμέτρων (υψηλός ρυθμός δειγματοληψίας)

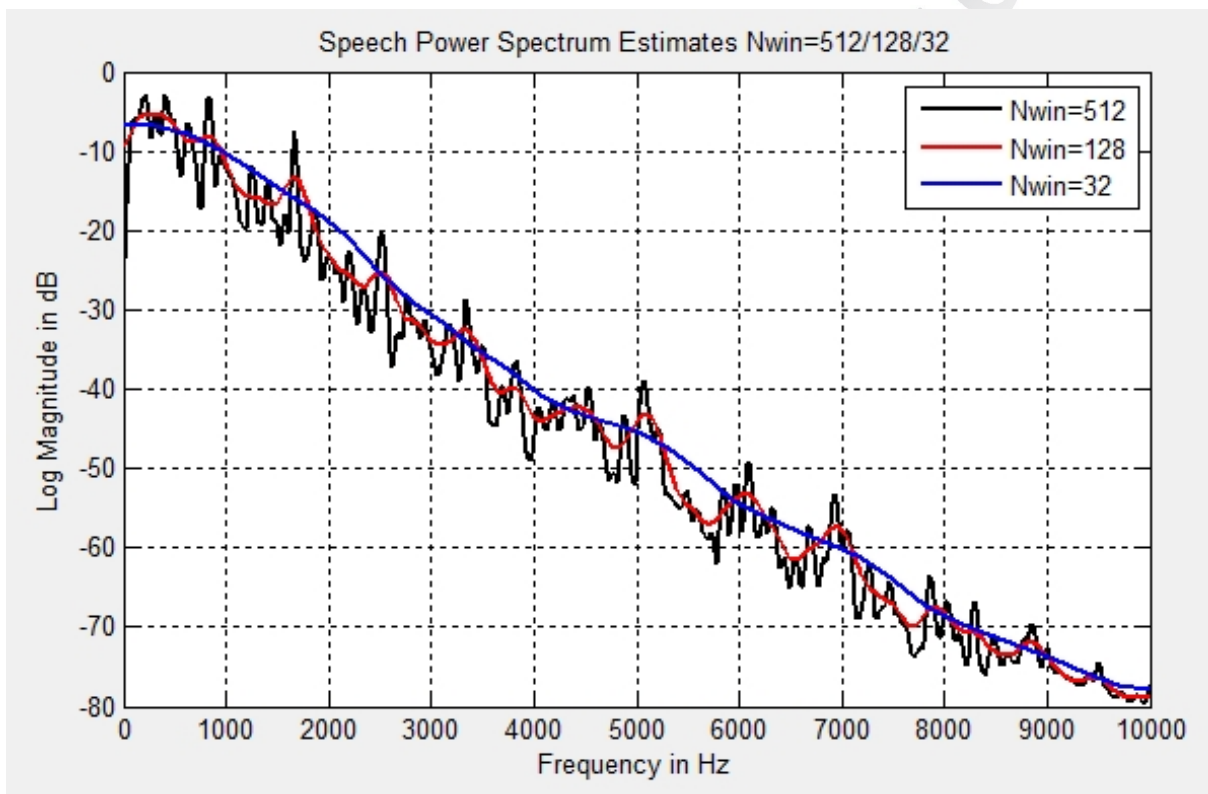
Αντίστοιχα, στην επόμενη εικόνα παρουσιάζεται το ιστόγραμμα του συγκεκριμένου αρχείου ήχου. Η κατανομή του καλύπτει αρκετά περισσότερες τιμές από το 1<sup>ο</sup> αρχείο, όπως είχε φανεί εξάλλου και από την προηγούμενη ενότητα. Εδώ δεν βλέπουμε ιδιαίτερες διαφορές σε σύγκριση με την ανάλυση που έγινε με τον χαμηλότερο ρυθμό δειγματοληψίας, κάτι που είχε παρατηρηθεί και στο 1<sup>ο</sup> αρχείο.



Εικόνα 22: Αποτελέσματα για το 2ο αρχείο ήχου: Ιστόγραμμα (υψηλός ρυθμός δειγματοληψίας)

Στο επόμενο σχήμα όπως και πριν παρουσιάζεται το φάσμα ισχύος για το εν λόγω σήμα, με τον υψηλό ρυθμό δειγματοληψίας, για τρεις περιπτώσεις μεγέθους παραθύρου: 32, 128 και 512. Όπως φαίνεται από την εικόνα, το φάσμα ισχύος ακολουθεί την ίδια περίπου μορφή και για τις τρεις περιπτώσεις, και μειώνεται σχεδόν γραμμικά από τα  $-10$  dB ως τα  $-80$  dB, τιμή την οποία φτάνει στα 10,000 Hz.

Η ουσιαστική διαφορά που εισάγεται για τις διαφορετικές τιμές μεγέθους παραθύρου, αφορά στις διακυμάνσεις του φάσματος ισχύος, όπως και στην περίπτωση του χαμηλού ρυθμού δειγματοληψίας. Στην περίπτωση που το παράθυρο είναι μικρό, δεν γίνονται αντιληπτές κάποιες εκ των διακυμάνσεων του αρχείου, με αποτέλεσμα η γραμμή να είναι σαφώς πιο ομαλή. Αντίθετα, όσο μεγαλώνει το μέγεθος παραθύρου, καταγράφονται και μεγαλύτερες διακυμάνσεις στην ισχύ του σήματος συναρτήσει της συχνότητας, και τελικά το αποτέλεσμα είναι μια γραμμή με σαφώς περισσότερες αυξομειώσεις στην ισχύ. Παρόλα αυτά, σε σύγκριση με την περίπτωση του χαμηλού ρυθμού δειγματοληψίας που μελετήθηκε στην προηγούμενη ενότητα, οι διαφορές μεταξύ των φασμάτων ισχύος είναι αρκετά μικρότερες για τις διαφορετικές τιμές παραθύρου. Το συμπέρασμα αυτό είναι αντίστοιχο με αυτό που προέκυψε και για το 1<sup>ο</sup> σήμα που μελετήθηκε.

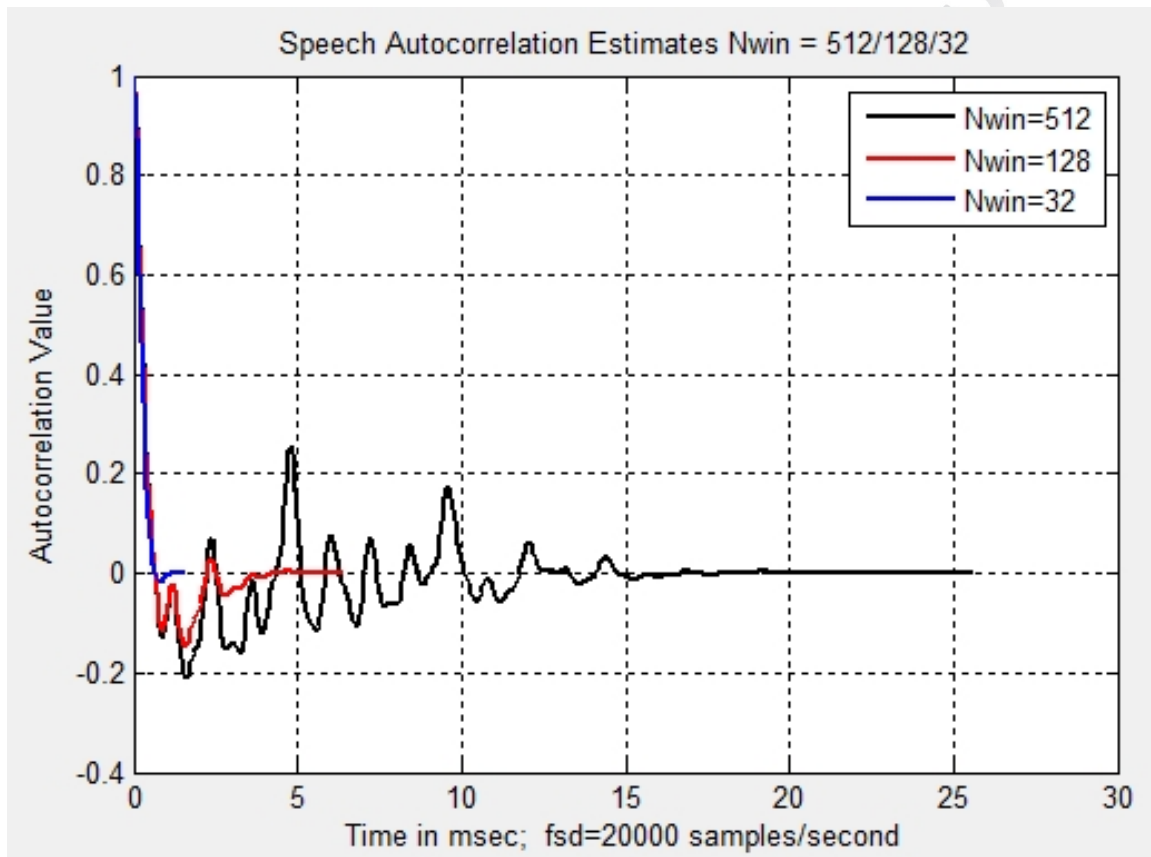


Εικόνα 23: Αποτελέσματα για το 2ο αρχείο ήχου: Φάσμα ισχύος (υψηλός ρυθμός δειγματοληψίας)



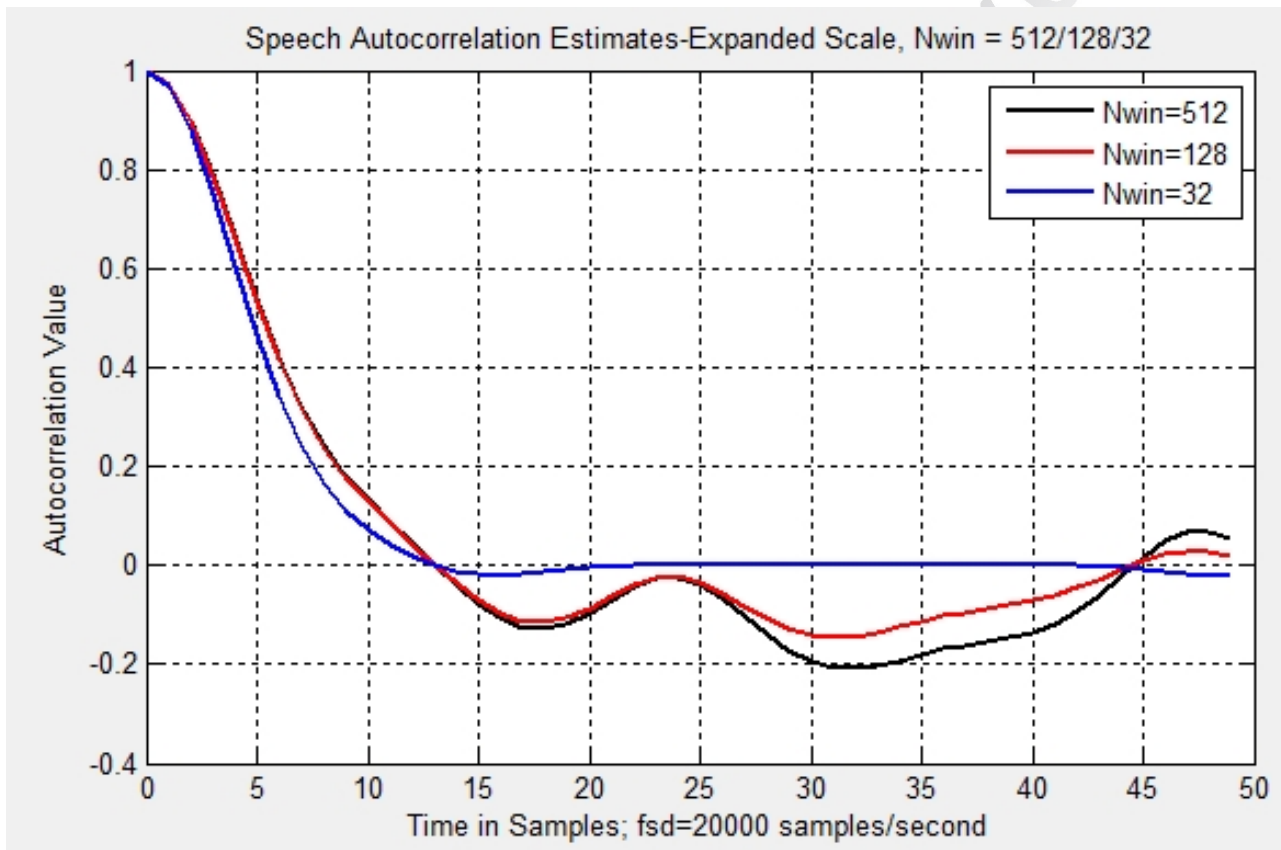
Στο ίδιο πλαίσιο, η Εικόνα 24 παρουσιάζει τα αποτελέσματα της αυτοσυσχέτισης σε σχέση με το χρόνο για το 2<sup>ο</sup> αυτό υπό εξέταση σήμα, και την περίπτωση του υψηλού ρυθμού δειγματοληψίας. Ανάλογα με το μέγεθος παραθύρου, η αυτοσυσχέτιση λαμβάνει τιμές για διαφορετικό εύρος τιμών του χρόνου. Επιπλέον, όσο μεγαλύτερο είναι το μέγεθος παραθύρου, τόσο περισσότερο καθυστερεί η αυτοσυσχέτιση να μηδενιστεί.

Όπως και με το 1<sup>ο</sup> σήμα, έτσι και εδώ είναι εμφανής η διαφορά σε σχέση με την περίπτωση του χαμηλού ρυθμού μετάδοσης. Συγκεκριμένα, φαίνεται πως οι διακυμάνσεις του σήματος αυτοσυσχέτισης είναι σαφώς μικρότερες και πιο ομαλές στην περίπτωση του υψηλού ρυθμού δειγματοληψίας.



Εικόνα 24: Αποτελέσματα για το 2ο αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε second) (υψηλός ρυθμός δειγματοληψίας)

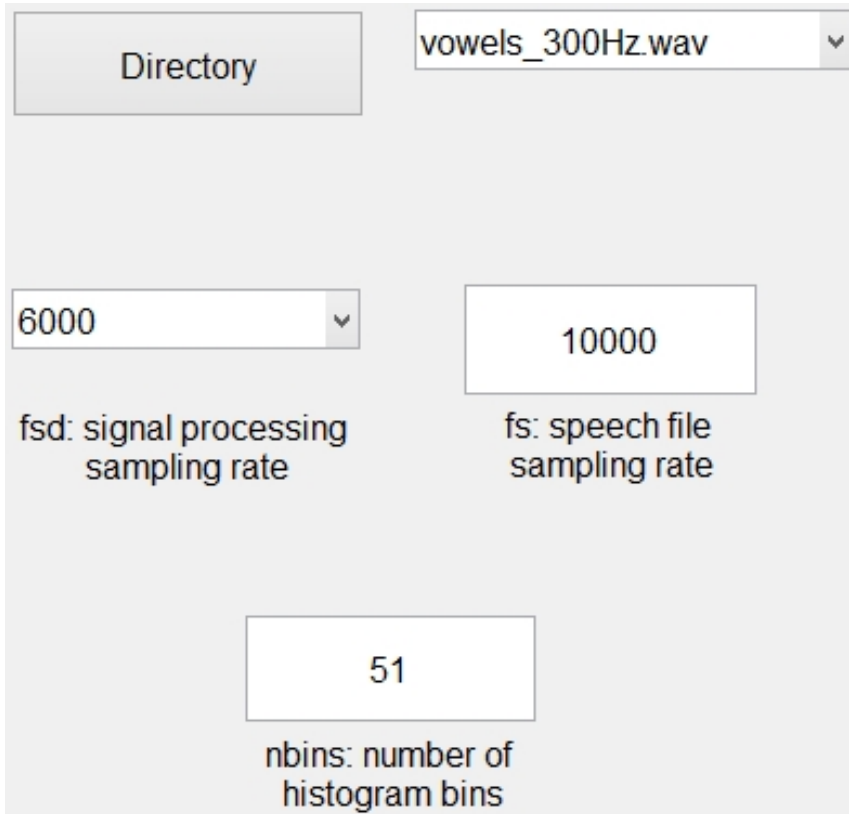
Όπως και με το προηγούμενο αρχείο, έτσι και εδώ για να φανεί λίγο καλύτερα η διαφορά που υπεισέρχεται με την αλλαγή της τιμής του παραθύρου, το παραπάνω γράφημα παρουσιάζεται σε μεγαλύτερη ανάλυση, εστιάζοντας στις μικρότερες τιμές του χρόνου, και έχοντας αυτή τη φορά στον άξονα των τετμημένων δείγματα αντί για χρόνο, όπως έγινε και στην προηγούμενη ενότητα. Το αποτέλεσμα φαίνεται στην επόμενη εικόνα, όπου βλέπουμε ότι ενώ για μικρό μέγεθος παραθύρου η τιμή έχει μηδενιστεί σχεδόν εξαρχής, η τιμή της αυτοσυσχέτισης για τις δύο μεγαλύτερες τιμές μεγέθους παραθύρου παρουσιάζει αυξομειώσεις. Οι αυξομειώσεις όμως αυτές, είναι σαφώς βραδύτερες από την περίπτωση που ο ρυθμός δειγματοληψίας ήταν 6000 Hz. Μάλιστα, και σε σχέση με το 1<sup>ο</sup> αρχείο, εδώ βλέπουμε πως ειδικά για τα μεγέθη παραθύρου 128 και 512 οι διαφορές είναι πολύ μικρές.



Εικόνα 25: Αποτελέσματα για το 2ο αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε samples) (υψηλός ρυθμός δειγματοληψίας)

### 5.5 Αποτελέσματα για 3<sup>ο</sup> αρχείο ήχου – χαμηλό sampling rate

Τέλος, σε αυτή την ενότητα επιλέγεται ένα τρίτο διαφορετικό από μια σειρά αρχείων ήχου, τα οποία χρησιμοποιούνται για να μελετηθεί ο αλγόριθμος ανάλυσης που παρουσιάστηκε στο προηγούμενο κεφάλαιο. Αρχικά, όπως και προηγουμένως, από την οθόνη διεπαφής με τον χρήστη, επιλέγονται τα χαρακτηριστικά της ανάλυσης, όπως φαίνεται στην εικόνα που ακολουθεί. Το όνομα του αρχείου είναι «vowels\_300Hz.wav». Ουσιαστικά πρόκειται τώρα για ένα αρχείο ήχου που περιλαμβάνει απλά φωνήεντα στα 300 Hz.



Directory: vowels\_300Hz.wav

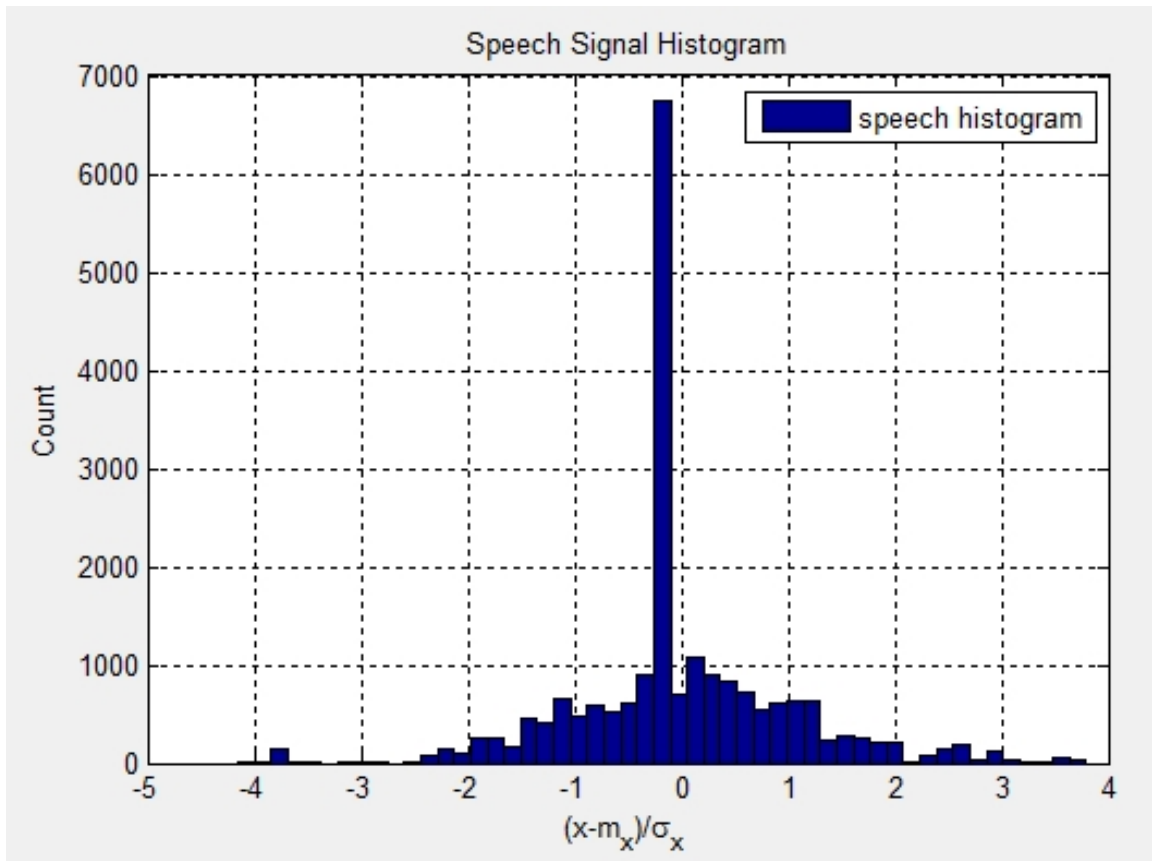
fsd: signal processing sampling rate: 6000

fs: speech file sampling rate: 10000

nbins: number of histogram bins: 51

Εικόνα 26: 2ο Αποτελέσματα για το 3ο αρχείο ήχου: Εισαγωγή παραμέτρων

Αντίστοιχα, όπως και στις προηγούμενες περιπτώσεις, στην επόμενη εικόνα παρουσιάζεται το ιστόγραμμα του συγκεκριμένου αρχείου ήχου. Όπως φαίνεται, πρόκειται για ένα αρχείο ήχου με αρκετά πιο σύνθετη κατανομή σε ένα σαφώς μεγαλύτερο εύρος τιμών από το πρώτο αρχείο που εξετάστηκε, ενώ μάλιστα βλέπουμε πως η μεγαλύτερη συχνότητα πλέον δεν είναι στο μηδέν, όπως στα προηγούμενα αρχεία, αλλά έχει αρνητική τιμή.

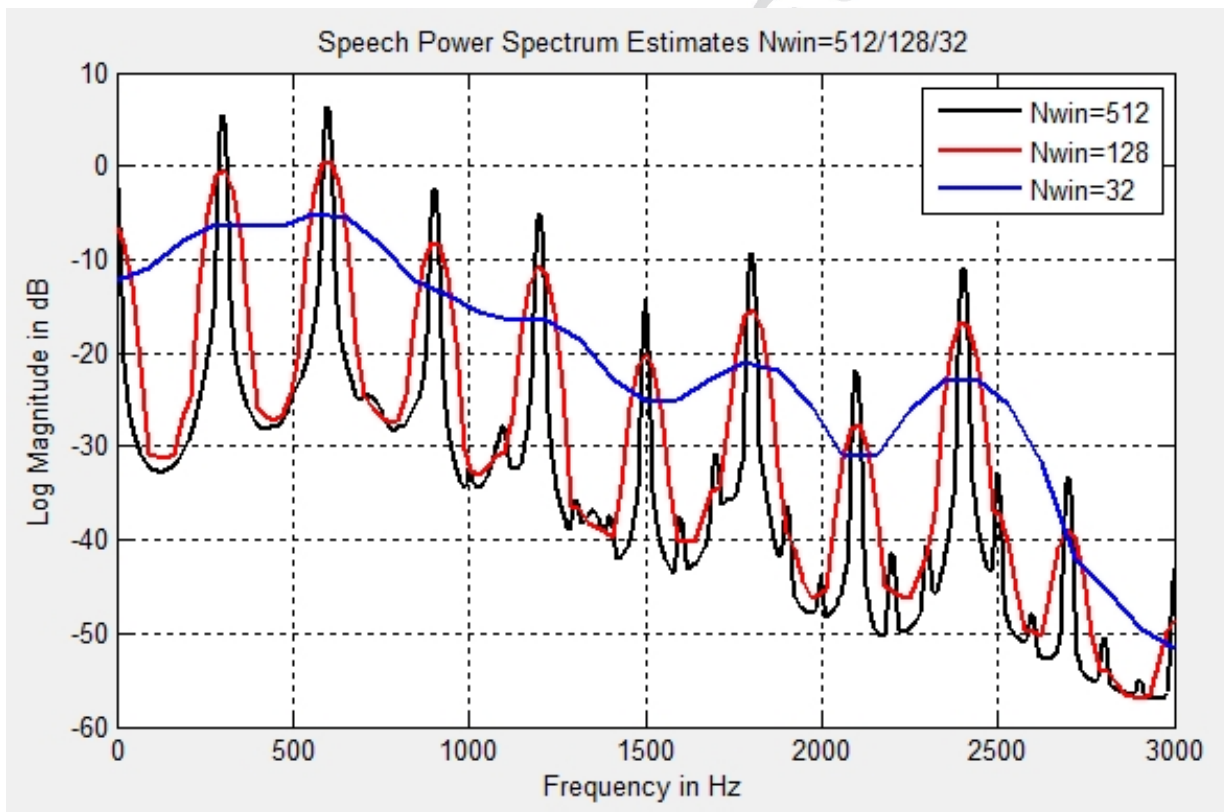


Εικόνα 27: Αποτελέσματα για το 3ο αρχείο ήχου: Ιστόγραμμα

Στη συνέχεια παρουσιάζεται το φάσμα ισχύος για το εν λόγω σήμα, για τρεις περιπτώσεις μεγέθους παραθύρου, όπως ακριβώς έγινε και με τα δύο προηγούμενα σήματα: 32, 128 και 512. Όπως φαίνεται από την εικόνα, το φάσμα ισχύος ακολουθεί την ίδια περίπου μορφή για τις περιπτώσεις του παραθύρου 512 και 128, με συγκεκριμένες αυξομειώσεις, ενώ όταν το παράθυρο έχει τιμή 32, οι μεταβολές είναι σαφώς ομαλότερες.

Ουσιαστικά δηλαδή, η βασική διαφοροποίηση που εισάγεται για τις διαφορετικές τιμές μεγέθους παραθύρου, αφορά στις διακυμάνσεις του φάσματος ισχύος. Στην περίπτωση που το παράθυρο είναι μικρό, δεν γίνονται αντιληπτές κάποιες εκ των διακυμάνσεων του αρχείου, με αποτέλεσμα η γραμμή να είναι σαφώς πιο ομαλή. Αντίθετα, όσο μεγαλώνει το μέγεθος παραθύρου, καταγράφονται και μεγαλύτερες διακυμάνσεις στην ισχύ του σήματος συναρτήσει της συχνότητας, και τελικά το αποτέλεσμα είναι μια γραμμή με σαφώς περισσότερες αυξομειώσεις στην ισχύ.

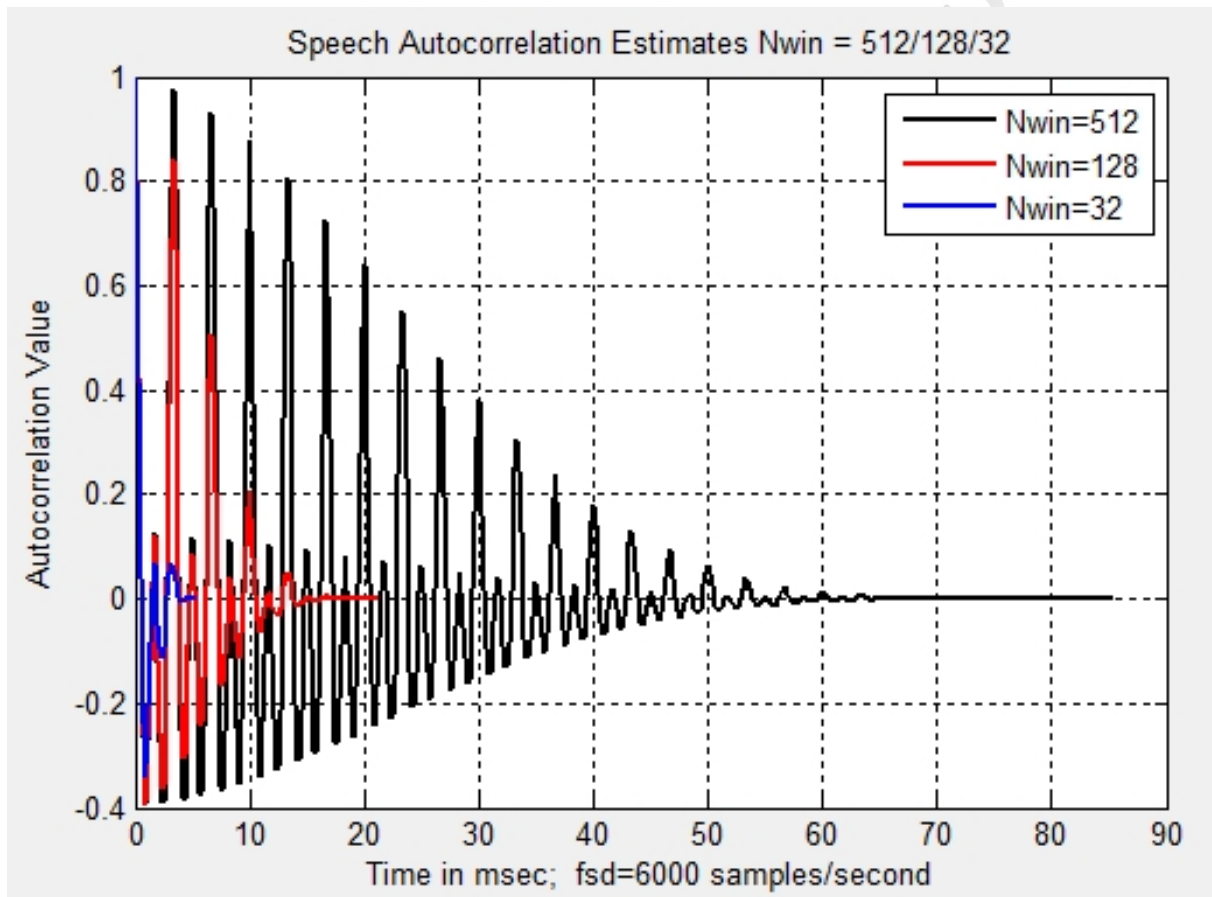
Μάλιστα, οι αυξομειώσεις δεν είναι φυσικά τυχαίες, αλλά παρατηρούμε πως η μέγιστη ισχύς εμφανίζεται ακριβώς κάθε 300 Hz. Πιο συγκεκριμένα, οι δύο πρώτες κορυφές (300 και 600 Hz) ξεπερνούν τα 0 dB ισχύος, ενώ οι δύο επόμενες (900 και 1200 Hz) είναι περίπου στα - 5 dB. Από τα 1500 Hz και πάνω οι μέγιστες τιμές δεν ξεπερνούν τα - 10 dB. Σε κάθε περίπτωση, οι ελάχιστες τιμές είναι περίπου 30 - 40 dB πιο κάτω από τις μέγιστες σε κάθε συχνότητα.



Εικόνα 28: Αποτελέσματα για το 3ο αρχείο ήχου: Φάσμα ισχύος

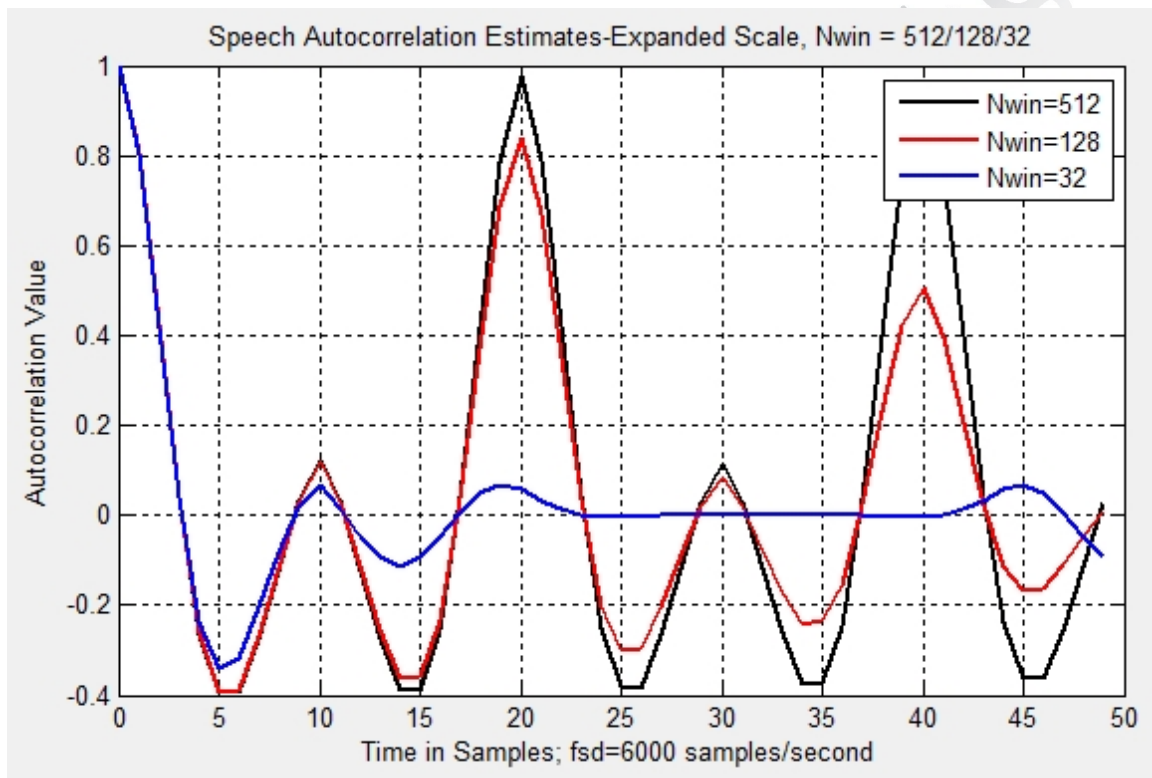
Αντίστοιχα, η Εικόνα 29 παρουσιάζει τα αποτελέσματα της αυτοσυσχέτισης σε σχέση με το χρόνο για το 3<sup>ο</sup> αυτό υπό εξέταση σήμα με τα φωνήεντα. Ανάλογα με το μέγεθος παραθύρου, η αυτοσυσχέτιση λαμβάνει τιμές για διαφορετικό εύρος τιμών του χρόνου. Επιπλέον, όσο μεγαλύτερο είναι το μέγεθος παραθύρου, τόσο περισσότερο καθυστερεί η αυτοσυσχέτιση να μηδενιστεί, κάτι που είχε παρατηρηθεί και από το προηγούμενο σήμα.

Παρατηρούμε επίσης πως η αυτοσυσχέτιση λαμβάνει τιμές (ειδικά για μεγάλα μεγέθη παραθύρου) με συγκεκριμένες αυξομειώσεις, που λαμβάνουν χώρα κάθε 3.33 msec, κάτι που προκύπτει φυσιολογικά με βάση τα 300 Hz του αρχικού σήματος και τον ρυθμό δειγματοληψίας των 6000<sup>ωv</sup> samples ανά δευτερόλεπτο.



Εικόνα 29: Αποτελέσματα για το 3ο αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε second)

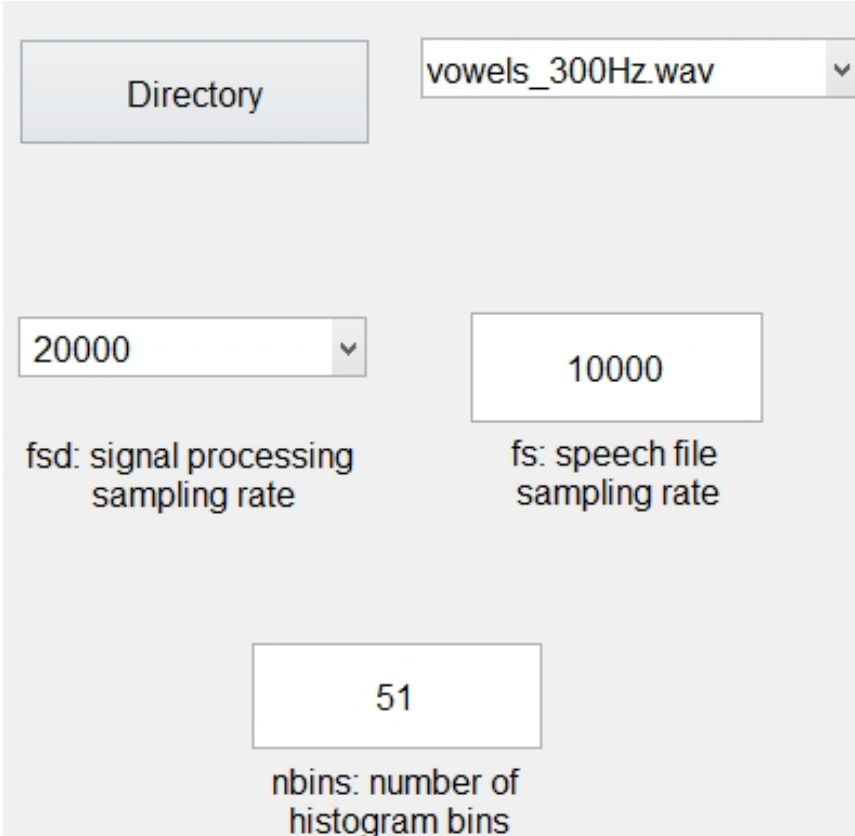
Όπως έγινε και για τα δύο προηγούμενα σήματα, με στόχο να φανεί λίγο καλύτερα η διαφορά που υπεισέρχεται με την αλλαγή της τιμής του παραθύρου, το παραπάνω γράφημα παρουσιάζεται σε μεγαλύτερη ανάλυση, εστιάζοντας στις μικρότερες τιμές του χρόνου και έχοντας αυτή τη φορά στον άξονα των τετμημένων δείγματα αντί για χρόνο. Το αποτέλεσμα φαίνεται στην επόμενη εικόνα, όπου βλέπουμε ότι ενώ για μικρό μέγεθος παραθύρου η τιμή έχει μηδενιστεί σχεδόν εξαρχής, η τιμή της αυτοσυσχέτισης για τις δύο μεγαλύτερες τιμές μεγέθους παραθύρου παρουσιάζει αυξομειώσεις, οι οποίες είναι παρόμοιες για τις περιπτώσεις μεγέθους παραθύρου 128 και 512.



Εικόνα 30: Αποτελέσματα για το 3ο αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε samples)

### 5.6 Αποτελέσματα για 3<sup>ο</sup> αρχείο ήχου – υψηλό sampling rate

Τέλος, όπως και με τα δύο προηγούμενα σήμερα, σε αυτή την ενότητα επιλέγεται το ίδιο μεν αρχείο με την προηγούμενη ενότητα, αλλά η ανάλυση γίνεται με διαφορετικό ρυθμό δειγματοληψίας. Συγκεκριμένα, ενώ στην προηγούμενη ενότητα η ανάλυση έγινε με ρυθμό 6,000 Hz, τώρα γίνεται με ρυθμό 20,000 Hz. Όπως και πριν, από την οθόνη διεπαφής με τον χρήστη, επιλέγονται τα χαρακτηριστικά της ανάλυσης, όπως φαίνεται στην εικόνα που ακολουθεί.



The image shows a software interface for audio analysis. It features several input fields and labels:

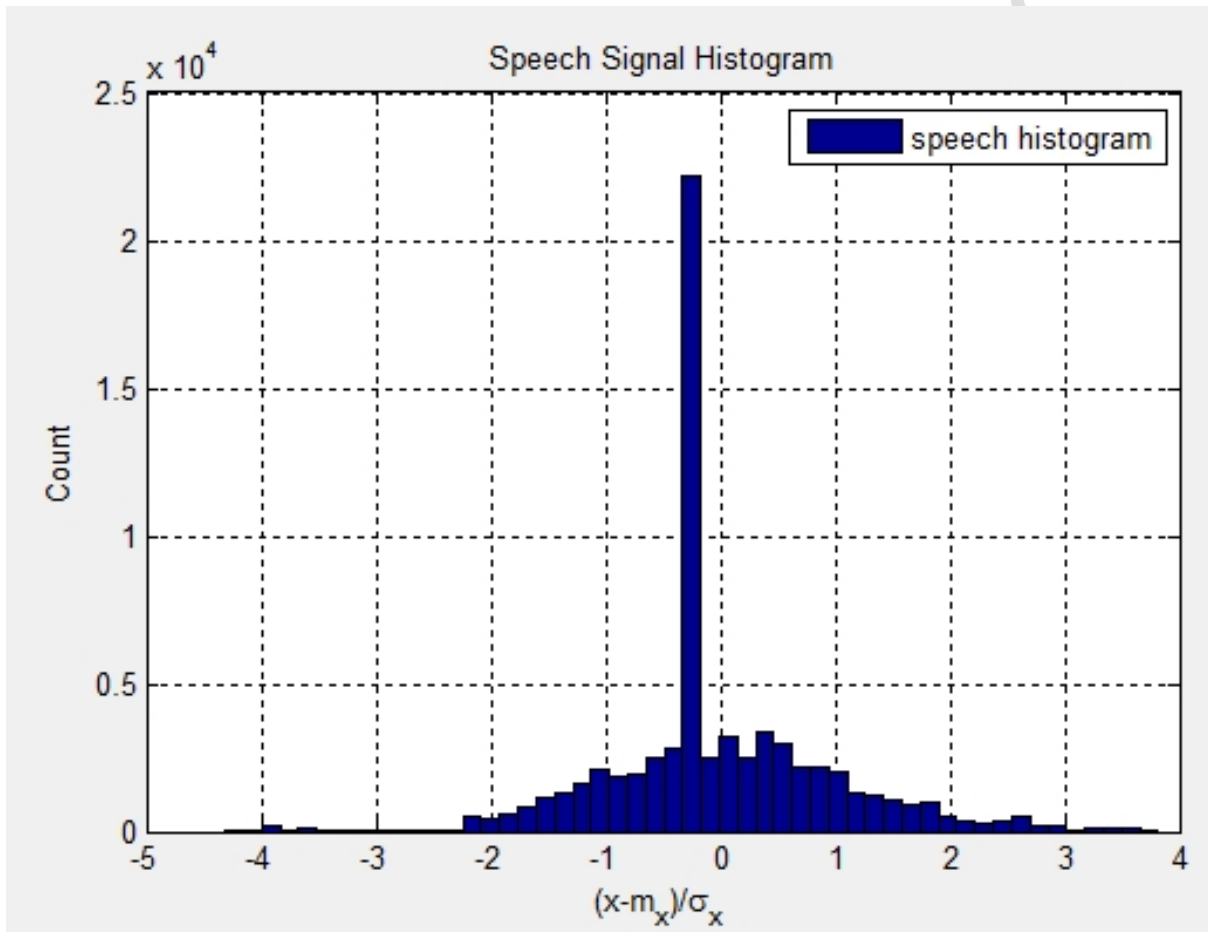
- A "Directory" button.
- A dropdown menu showing the file name "vowels\_300Hz.wav".
- A dropdown menu for "fsd: signal processing sampling rate" with the value "20000".
- A text box for "fs: speech file sampling rate" with the value "10000".
- A text box for "nbins: number of histogram bins" with the value "51".

Εικόνα 31: 1ο Αποτελέσματα για το 3ο αρχείο ήχου: Εισαγωγή παραμέτρων (υψηλός ρυθμός δειγματοληψίας)



Αντίστοιχα, στην επόμενη εικόνα παρουσιάζεται το ιστόγραμμα του συγκεκριμένου αρχείου ήχου. Όπως φαίνεται, πρόκειται για ένα αρχείο ήχου με αρκετά πιο σύνθετη κατανομή σε ένα σαφώς μεγαλύτερο εύρος τιμών από το πρώτο αρχείο που εξετάστηκε, ενώ μάλιστα βλέπουμε πως η μεγαλύτερη συχνότητα πλέον δεν είναι στο μηδέν, όπως στα προηγούμενα αρχεία, αλλά έχει αρνητική τιμή.

Εδώ δεν βλέπουμε ιδιαίτερες διαφορές σε σύγκριση με την ανάλυση που έγινε με τον χαμηλότερο ρυθμό δειγματοληψίας.



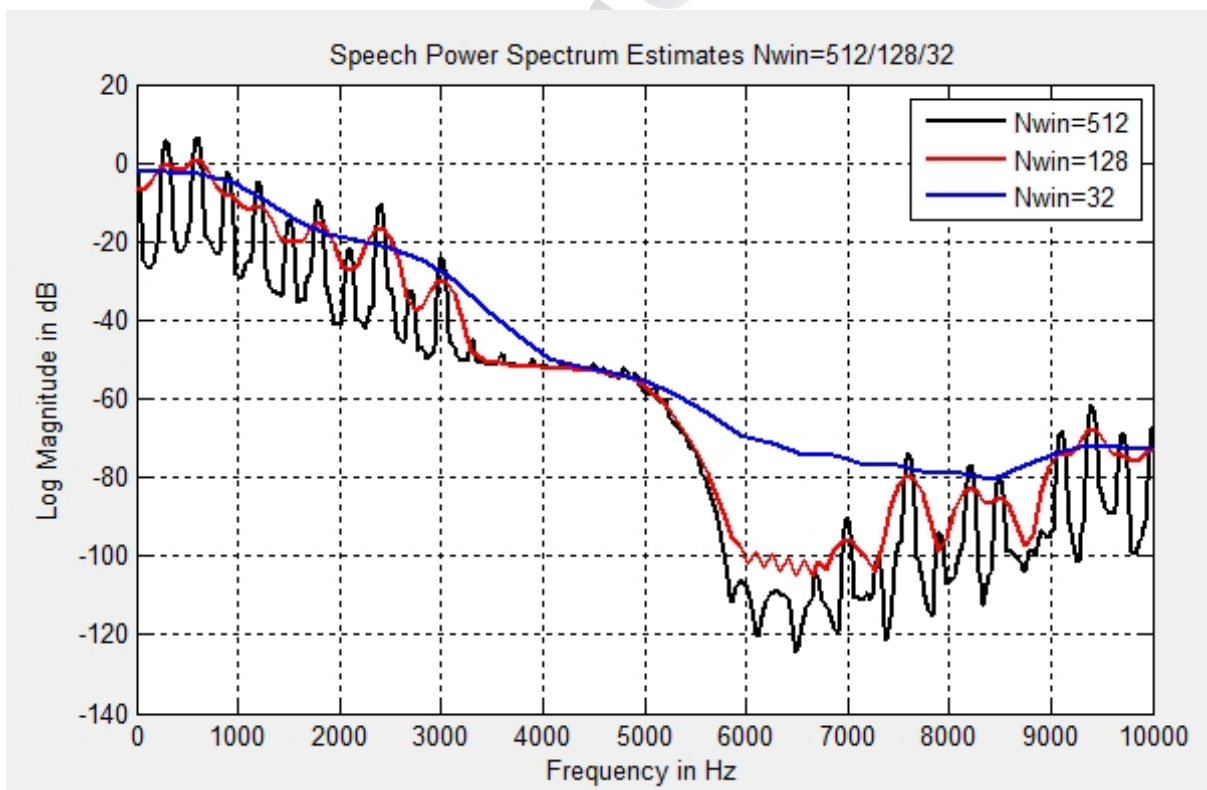
Εικόνα 32: Αποτελέσματα για το 3ο αρχείο ήχου: Ιστόγραμμα (υψηλός ρυθμός δειγματοληψίας)

Στη συνέχεια παρουσιάζεται το φάσμα ισχύος για το εν λόγω σήμα, με τον υψηλό ρυθμό δειγματοληψίας, για τρεις περιπτώσεις μεγέθους παραθύρου: 32, 128 και 512. Όπως φαίνεται από την εικόνα, το φάσμα ισχύος ακολουθεί την ίδια περίπου μορφή για τις περιπτώσεις του παραθύρου 512 και 128, με συγκεκριμένες αυξομειώσεις, ενώ όταν το παράθυρο έχει τιμή 32, οι μεταβολές είναι σαφώς ομαλότερες, κάτι που είδαμε για το ίδιο σήμα και με τον χαμηλότερο ρυθμό δειγματοληψίας.

Ουσιαστικά δηλαδή, η βασική διαφοροποίηση που εισάγεται για τις διαφορετικές τιμές μεγέθους παραθύρου, αφορά στις διακυμάνσεις του φάσματος ισχύος. Στην περίπτωση που το παράθυρο είναι μικρό, δεν γίνονται αντιληπτές κάποιες εκ των διακυμάνσεων του αρχείου, με αποτέλεσμα η γραμμή να είναι σαφώς πιο ομαλή. Αντίθετα, όσο μεγαλώνει το μέγεθος παραθύρου, καταγράφονται και μεγαλύτερες διακυμάνσεις στην ισχύ του σήματος συναρτήσει της συχνότητας, και τελικά το αποτέλεσμα είναι μια γραμμή με σαφώς περισσότερες αυξομειώσεις στην ισχύ.

Μάλιστα, οι αυξομειώσεις δεν είναι φυσικά τυχαίες, αλλά παρατηρούμε πως η μέγιστη ισχύς εμφανίζεται ακριβώς κάθε 300 Hz. Πιο συγκεκριμένα, οι δύο πρώτες κορυφές (300 και 600 Hz) ξεπερνούν τα 0 dB ισχύος, ενώ οι δύο επόμενες (900 και 1200 Hz) είναι περίπου στα - 5 dB. Από τα 1500 Hz και πάνω οι μέγιστες τιμές δεν ξεπερνούν τα - 10 dB. Σε κάθε περίπτωση, οι ελάχιστες τιμές είναι περίπου 30 - 40 dB πιο κάτω από τις μέγιστες σε κάθε συχνότητα.

Παρόλα αυτά, σε σύγκριση με την περίπτωση του χαμηλού ρυθμού δειγματοληψίας που μελετήθηκε στην προηγούμενη ενότητα, βλέπουμε πως τώρα μπορούμε να παρατηρήσουμε το φάσμα του σήματός μας και για μεγαλύτερες συχνότητες. Εκεί, η ισχύς του σήματος είναι σαφώς μικρότερη, καθώς κάνει μια απότομη βύθιση της τάξης των 40 dB στα 5000 Hz. Εκεί, από τα 6000 ως τα 10000 Hz εμφανίζει και πάλι αυξομειώσεις περίπου κάθε 300 Hz αλλά σε σαφώς χαμηλότερα επίπεδα ισχύος.

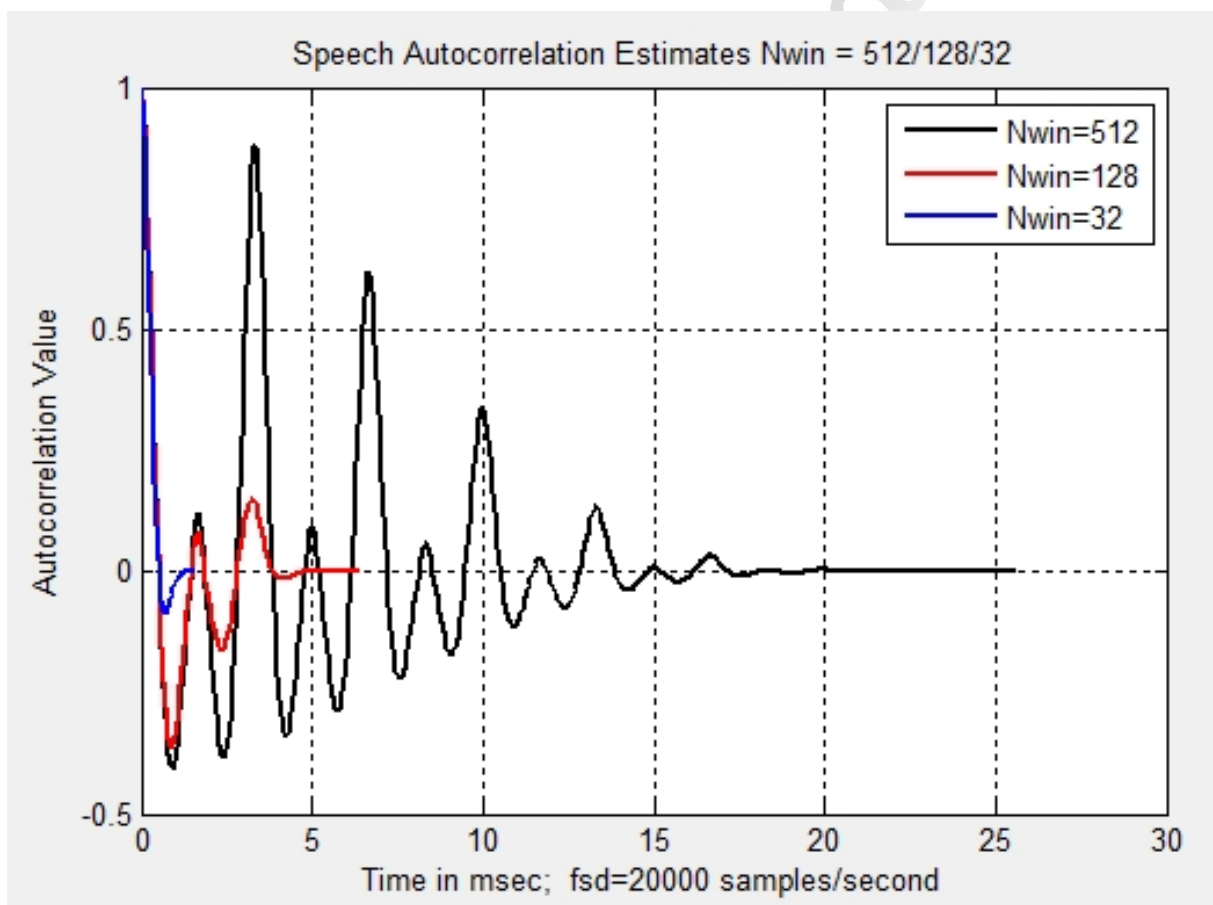


Εικόνα 33: Αποτελέσματα για το 3ο αρχείο ήχου: Φάσμα ισχύος (υψηλός ρυθμός δειγματοληψίας)

Αντίστοιχα, η Εικόνα 34 παρουσιάζει τα αποτελέσματα της αυτοσυσχέτισης σε σχέση με το χρόνο για το 3<sup>ο</sup> αυτό υπό εξέταση σήμα, και την περίπτωση του υψηλού ρυθμού δειγματοληψίας. Ανάλογα με το μέγεθος παραθύρου, η αυτοσυσχέτιση λαμβάνει τιμές για διαφορετικό εύρος τιμών του χρόνου. Επιπλέον, όσο μεγαλύτερο είναι το μέγεθος παραθύρου, τόσο περισσότερο καθυστερεί η αυτοσυσχέτιση να μηδενιστεί.

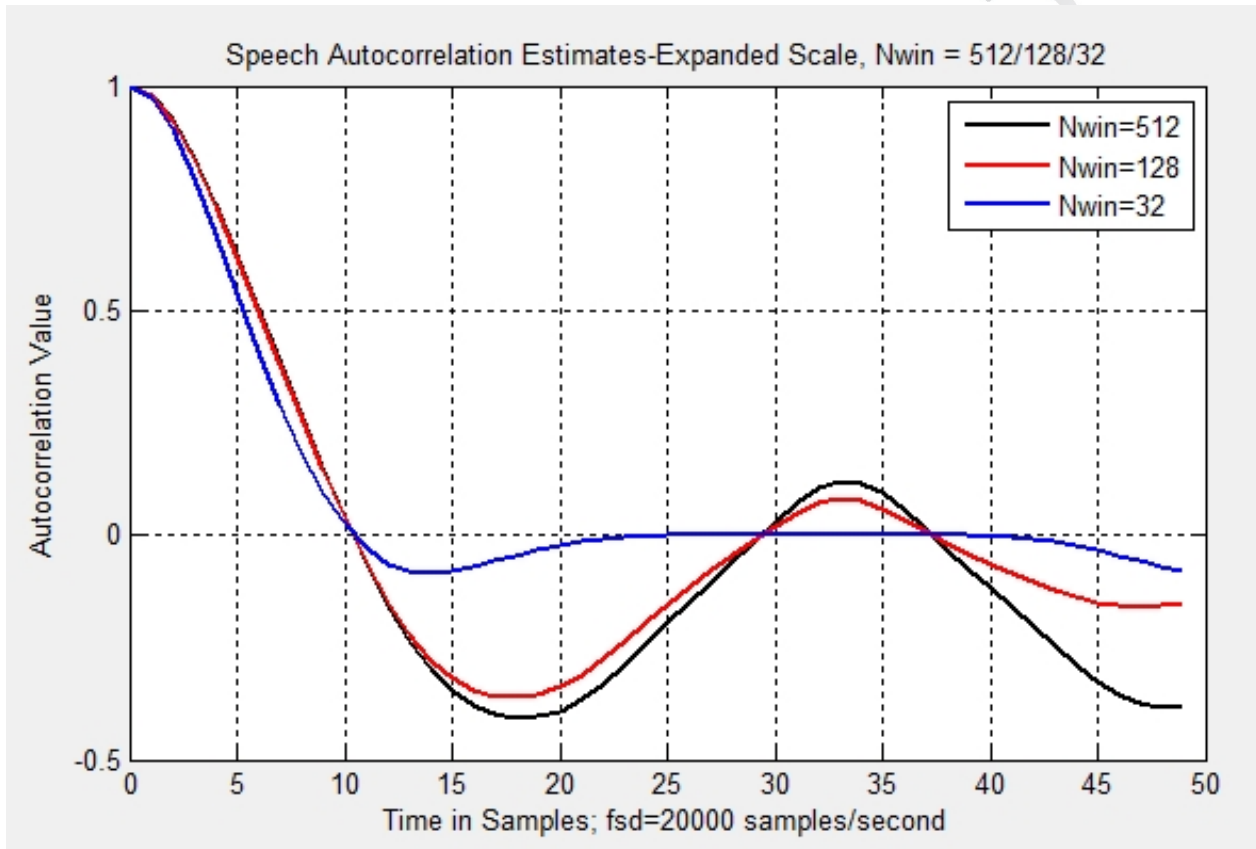
Παρατηρούμε επίσης πως η αυτοσυσχέτιση λαμβάνει τιμές (ειδικά για μεγάλα μεγέθη παραθύρου) με συγκεκριμένες αυξομειώσεις, που λαμβάνουν χώρα κάθε 3.33 msec, κάτι που προκύπτει φυσιολογικά με βάση τα 300 Hz του αρχικού σήματος και τον ρυθμό δειγματοληψίας των 20000<sup>ωv</sup> samples ανά δευτερόλεπτο.

Παρόλα αυτά, είναι εμφανής η διαφορά σε σχέση με την περίπτωση του χαμηλού ρυθμού μετάδοσης. Συγκεκριμένα, φαίνεται πως οι διακυμάνσεις του σήματος αυτοσυσχέτισης είναι σαφώς λιγότερες και πιο ομαλές στην περίπτωση του υψηλού ρυθμού δειγματοληψίας.



Εικόνα 34: Αποτελέσματα για το 3ο αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε second) (υψηλός ρυθμός δειγματοληψίας)

Τέλος, για να φανεί λίγο καλύτερα η διαφορά που υπεισέρχεται με την αλλαγή της τιμής του παραθύρου, το παραπάνω γράφημα παρουσιάζεται σε μεγαλύτερη ανάλυση, εστιάζοντας στις μικρότερες τιμές του χρόνου, και έχοντας αυτή τη φορά στον άξονα των τετμημένων δείγματα αντί για χρόνο, όπως έγινε και στις προηγούμενες ενότητες. Το αποτέλεσμα φαίνεται στην επόμενη εικόνα, όπου βλέπουμε ότι ενώ για μικρό μέγεθος παραθύρου η τιμή έχει μηδενιστεί σχεδόν εξαρχής, η τιμή της αυτοσυσχέτισης για τις δύο μεγαλύτερες τιμές μεγέθους παραθύρου παρουσιάζει αυξομειώσεις. Οι αυξομειώσεις όμως αυτές, είναι σαφώς βραδύτερες από την περίπτωση που ο ρυθμός δειγματοληψίας ήταν 6000 Hz.



Εικόνα 35: Αποτελέσματα για το 3ο αρχείο ήχου: Αυτοσυσχέτιση σε σχέση με τον χρόνο (σε samples) (υψηλός ρυθμός δειγματοληψίας)

## 6 Συμπεράσματα

Η εργασία αυτή αποτέλεσε μια σημαντική προσπάθεια προς την αποτύπωση των βασικών χαρακτηριστικών και των ιδιοτήτων των σημάτων φωνής, με απώτερο στόχο να καταδείξει τις προκλήσεις που εμφανίζονται στην επεξεργασία αυτών των σημάτων, και να αποτυπώσει μερικές από τις ιδιότητές δείχνοντας έτσι με αυτόν τον τρόπο την ανάγκη εφαρμογής βασικών τεχνικών της επεξεργασίας σημάτων σε αυτά. Είναι εξάλλου γενικότερα αποδεκτό, πως τα σήματα ήχου και δη τα σήματα φωνής είναι ιδιαίτερα πολύπλοκα και σύνθετα σήματα, και η πλήρης επεξεργασία τους αποτελεί μια ιδιαίτερα σημαντική πρόκληση.

Σε αυτό το πλαίσιο, μέσω της εργασίας αυτής έγινε μια προσπάθεια να μελετηθούν κάποια βασικά στατιστικά χαρακτηριστικά τέτοιων σημάτων, τα οποία είναι απαραίτητα για την μετέπειτα επεξεργασία τους. Τέτοια στατιστικά χαρακτηριστικά είναι η probability density function (pdf, συνάρτηση πυκνότητας πιθανότητας), το power density spectrum (φάσμα πυκνότητας ισχύος) και το autocorrelation (αυτοσυσχέτιση).

Πριν την παρουσίαση του πειραματικού μέρους της εργασίας, έγινε πρώτα μια περιγραφή των βασικών χαρακτηριστικών της θεωρίας, σχετικά με έννοιες όπως οι τυχαίες μεταβλητές και η συνάρτηση κατανομής, αλλά και η συνάρτηση πυκνότητας πιθανότητας. Επίσης, κρίθηκε σκόπιμο να παρουσιαστούν τα βασικά χαρακτηριστικά του φάσματος πυκνότητας ισχύος και της αυτοσυσχέτισης, με αναφορά σε βασικούς ορισμούς εννοιών όπως οι σειρές Fourier, τα αιτιοκρατικά και τα τυχαία / στοχαστικά σήματα, το θεώρημα Parseval και άλλα. Επιπλέον, έγινε παρουσίαση των βασικών χαρακτηριστικών των σημάτων φωνής, εστιάζοντας στην χρόνο – συχνотική ανάλυση των σημάτων ομιλίας, στην επίδραση του μήκους παραθύρου αλλά και σε βασικά χαρακτηριστικά της ανάλυσης τέτοιων σημάτων.

Ακολούθησε το πειραματικό μέρος, όπου με τη βοήθεια του πακέτου λογισμικού Matlab μελετήθηκαν τα τρία επιμέρους στατιστικά χαρακτηριστικά σημάτων φωνής, δηλαδή η probability density function, το power density spectrum και η autocorrelation function. Παρουσιάστηκαν τα αποτελέσματα της εφαρμογής του κώδικα για τρία διαφορετικά αρχεία ήχου με διαφορετικά χαρακτηριστικά, και για δύο διαφορετικές τιμές της συχνότητας δειγματοληψίας. Τα συμπεράσματα που προέκυψαν είναι ιδιαίτερα ενδιαφέροντα, και η ανάλυση αυτή μπορεί να γίνει για μια πληθώρα σημάτων ήχου με διαφορετικά χαρακτηριστικά, ώστε να αποκτήσει ο ερευνητής πλήρη γνώση των βασικών χαρακτηριστικών του εκάστοτε σήματος προτού προχωρήσει σε περαιτέρω επεξεργασία και ανάλυσή του.

Η παρούσα ερευνητική προσπάθεια μπορεί να αποτελέσει τη βάση και για περαιτέρω αναλύσεις, που μπορούν να συνδέονται παραδείγματος χάριν με την κωδικοποίηση και την αποκωδικοποίηση των σημάτων φωνής ή την αναγνώριση των ιδιαίτερων χαρακτηριστικών των σημάτων με βάση τον ομιλητή (παραδείγματος χάριν αναγνώριση φύλου ή και ηλικίας με βάση τις κατανομές συχνότητων και του φάσματος πυκνότητας ισχύος).

Πανεπιστήμιο Πειραιώς

## Βιβλιογραφία

- [1]. C.S. Burrus, J.H. McClellan, A.V. Oppenheim, T.W. Parks, R.W. Schafer, and H.W. Schussler "Computer-Based Exercises for Signal Processing Using MATLAB," Prentice-Hall, 1994.
- [2]. B. Gold and N. Morgan, "Speech and Audio Signal Processing," John Wiley & Sons, Inc., 2000.
- [3]. M. H. Hayes, "Statistical Digital Signal Processing and Modeling," John Wiley & Sons, Inc., 1996.
- [4]. Γ. Β. Μουστακίδης, "Βασικές Τεχνικές Ψηφιακής Επεξεργασίας Σημάτων," Εκδόσεις Τζιόλα, 2004.
- [5]. Sanjit K. Mitra, "Digital Signal Processing, A Computer-Based Approach," McGraw-Hill, 1998.
- [6]. P. E. Papamichalis, "Practical Approaches to Speech Coding," Prentice-Hall 1987.
- [7]. Papoulis, "Signal Analysis," McGraw-Hill, 1985.
- [8]. M. Vetterli and J. Kovacevic, "Wavelets and Subband Coding," Prentice Hall Inc. 1995.
- [9]. Ευσταθίου, σημειώσεις,  
[http://www.chem.uoa.gr/courses/organologia/PDF/Ch05\\_1xy.pdf](http://www.chem.uoa.gr/courses/organologia/PDF/Ch05_1xy.pdf)
- [10]. Μπαρτζελιώτης Κ. (2009). Φασματική Ανάλυση: Φασματική ανάλυση και μέθοδοι ανάλυσης. Πανεπιστήμιο Πελοποννήσου, 2009.
- [11]. Διακολουκάς, σημειώσεις.  
[http://www.telecom.tuc.gr/courses/statdsp/docs/class\\_notes/03\\_notes.pdf](http://www.telecom.tuc.gr/courses/statdsp/docs/class_notes/03_notes.pdf)
- [12]. ΕΑΠ, σημειώσεις [http://edu.eap.gr/pli/pli12/shmeiwseis/Fourier\\_series.pdf](http://edu.eap.gr/pli/pli12/shmeiwseis/Fourier_series.pdf)
- [13]. Matlab's Tutorial, Mathworks Inc
- [14]. G.B. Thomas, R.L. Finney, M.D. Weir, F.R. Giordano, "Thomas' Calculus", Addison-Wesley, 2000
- [15]. Χ. Ιωάννου, "Μελέτη Αλγόριθμων Αποθορυβοποίησης Σήματος Ομιλίας", Πανεπιστήμιο Πατρών, 2010
- [16]. Δαμιανού, Παπαδάτος, Χαραλαμπίδης, σημειώσεις.  
<https://mathbooksg.files.wordpress.com/2011/08/pithstatopen.pdf>

Πανεπιστήμιο Πειραιώς



## Παράρτημα

### Ο Κώδικας

```

% Modifiable runGUI file
clc;
clear all;

path_to_speech=pwd;

% USER - ENTER FILENAME
fileName = 'statistical_analysis.mat';
fileData=load(fileName);
temp=fileData(1).temp;

f = figure('Visible','on',...
'Units','normalized',...
'Position',[0,0,1,1],...
'MenuBar','none',...
'NumberTitle','off');

%SENSE COMPUTER AND SET FILE DELIMITER
switch(computer)
    case 'MACI64', char= '/';
    case 'GLNX86', char=' /';
    case 'PCWIN', char= '\';
    case 'PCWIN64', char=' \';
    case 'GLNXA64', char= '/';
end

% find speech files directory by going up one level and down one level
% on the directory chain; as follows:
dir_cur=pwd; % this is the current Matlab exercise directory path
s=regexp(dir_cur,char); % find the last '\' for the current directory
s1=s(length(s)); % find last '\' character; this marks upper level
directory
dir_fin=strcat(dir_cur(1:s1),'speech_files'); % create new
directory
start_path=dir_fin; % save new directory for speech files location

% USER - ENTER PROPER CALLBACK FILE
Callbacks_statistical_analysis_GUI25(f,temp,start_path);
%panelAndButtonEdit(f, temp); % Easy access to Edit Mode

% Note comment PanelandBUttonCallbacks(f,temp) if panelAndButtonEdit is
to
% be uncommented and used

```

```

function Callbacks_statistical_analysis_GUI25(f,C,start_path)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%SENSE COMPUTER AND SET FILE DELIMITER
switch(computer)
    case 'MACI64',      char= '/';
    case 'GLNX86',     char= '/';
    case 'PCWIN',      char= '\';
    case 'PCWIN64',    char= '\';
    case 'GLNXA64',    char= '/';
end

x=C{1,1};
y=C{1,2};
a=C{1,3};
b=C{1,4};
u=C{1,5};
v=C{1,6};
m=C{1,7};
n=C{1,8};
lengthbutton=C{1,9};
widthbutton=C{1,10};
enterType=C{1,11};
enterString=C{1,12};
enterLabel=C{1,13};
noPanels=C{1,14};
noGraphicPanels=C{1,15};
noButtons=C{1,16};
labelDist=C{1,17};%distance that the label is below the button
noTitles=C{1,18};
buttonTextSize=C{1,19};
labelTextSize=C{1,20};
textboxFont=C{1,21};
textboxString=C{1,22};
textboxWeight=C{1,23};
textboxAngle=C{1,24};
labelHeight=C{1,25};
fileName=C{1,26};
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%PANELS
for j=0:noPanels-1
    uipanel('Parent',f,...
    'Units','Normalized',...
    'Position',[x(1+4*j) y(1+4*j) x(2+4*j)-x(1+4*j) y(3+4*j)-y(2+4*j)]);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%GRAPHIC PANELS
for i=0:noGraphicPanels-1
    switch (i+1)
    case 1
        graphicPanell = axes('parent',f,...
        'Units','Normalized',...
        'Position',[a(1+4*i) b(1+4*i) a(2+4*i)-a(1+4*i) b(3+4*i)-b(2+4*i)],...

```

```

'GridLineStyle', '--');
case 2
graphicPanel2 = axes('parent',f,...
'Units','Normalized',...
'Position',[a(1+4*i) b(1+4*i) a(2+4*i)-a(1+4*i) b(3+4*i)-b(2+4*i)],...
'GridLineStyle', '--');
case 3
graphicPanel3 = axes('parent',f,...
'Units','Normalized',...
'Position',[a(1+4*i) b(1+4*i) a(2+4*i)-a(1+4*i) b(3+4*i)-b(2+4*i)],...
'GridLineStyle', '--');
case 4
graphicPanel4 = axes('parent',f,...
'Units','Normalized',...
'Position',[a(1+4*i) b(1+4*i) a(2+4*i)-a(1+4*i) b(3+4*i)-b(2+4*i)],...
'GridLineStyle', '--');
end
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%
%TITLES BOXES
for k=0:noTitles-1
switch (k+1)
case 1
titleBox1 = uicontrol('parent',f,...
'Units','Normalized',...
'Position',[u(1+4*k) v(1+4*k) u(2+4*k)-u(1+4*k) v(3+4*k)-v(2+4*k)],...
'Style','text',...
'FontSize',textboxFont{k+1},...
'String',textboxString(k+1),...
'FontWeight',textboxWeight{k+1},...
'FontAngle',textboxAngle{k+1});
end
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%
%BUTTONS
for i=0:(noButtons-1)
enterColor='w';
if strcmp(enterType{i+1}, 'pushbutton')==1
|| strcmp(enterType{i+1}, 'text')==1
enterColor='default';
end
if (strcmp(enterLabel{1,(i+1)}, '')==0 &&...
strcmp(enterLabel{1,(i+1)}, '...')==0) %i.e. there is a label
%creating a label for some buttons
uicontrol('Parent',f,...
'Units','Normalized',...
'Position',[m(1+2*i) n(1+2*i)-labelDist-labelHeight(i+1) ...
(m(2+2*i)-m(1+2*i)) labelHeight(i+1)],...
'Style','text',...
'String',enterLabel{i+1},...
'FontSize', labelTextSize(i+1),...
'HorizontalAlignment','center');
end
switch (i+1)

```

```

case 1
button1=uicontrol('Parent',f,...
'Units','Normalized',...
'Position',[m(1+2*i)      n(1+2*i)      (m(2+2*i)-m(1+2*i))      (n(2+2*i)-
n(1+2*i))],...
'Style',enterType{i+1},...
'String',enterString{i+1},...
'FontSize', buttonTextSize(1+i),...
'BackgroundColor',enterColor,...
'HorizontalAlignment','center',...
'Callback',@button1Callback);
case 2
button2=uicontrol('Parent',f,...
'Units','Normalized',...
'Position',[m(1+2*i)      n(1+2*i)      (m(2+2*i)-m(1+2*i))      (n(2+2*i)-
n(1+2*i))],...
'Style',enterType{i+1},...
'String',enterString{i+1},...
'FontSize', buttonTextSize(1+i),...
'BackgroundColor',enterColor,...
'HorizontalAlignment','center',...
'Callback',@button2Callback);
case 3
button3=uicontrol('Parent',f,...
'Units','Normalized',...
'Position',[m(1+2*i)      n(1+2*i)      (m(2+2*i)-m(1+2*i))      (n(2+2*i)-
n(1+2*i))],...
'Style',enterType{i+1},...
'String',enterString{i+1},...
'FontSize', buttonTextSize(1+i),...
'BackgroundColor',enterColor,...
'HorizontalAlignment','center',...
'Callback',@button3Callback);
case 4
button4=uicontrol('Parent',f,...
'Units','Normalized',...
'Position',[m(1+2*i)      n(1+2*i)      (m(2+2*i)-m(1+2*i))      (n(2+2*i)-
n(1+2*i))],...
'Style',enterType{i+1},...
'String',enterString{i+1},...
'FontSize', buttonTextSize(1+i),...
'BackgroundColor',enterColor,...
'HorizontalAlignment','center',...
'Callback',@button4Callback);
case 5
button5=uicontrol('Parent',f,...
'Units','Normalized',...
'Position',[m(1+2*i)      n(1+2*i)      (m(2+2*i)-m(1+2*i))      (n(2+2*i)-
n(1+2*i))],...
'Style',enterType{i+1},...
'String',enterString{i+1},...
'FontSize', buttonTextSize(1+i),...
'BackgroundColor',enterColor,...
'HorizontalAlignment','center',...
'Callback',@button5Callback);
case 6

```

```

button6=uicontrol('Parent',f,...
'Units','Normalized',...
'Position',[m(1+2*i)      n(1+2*i)      (m(2+2*i)-m(1+2*i))      (n(2+2*i)-
n(1+2*i))],...
'Style',enterType{i+1},...
'String',enterString{i+1},...
'FontSize', buttonTextSize(1+i),...
'BackgroundColor',enterColor,...
'HorizontalAlignment','center',...
'Callback',@button6Callback);
case 7
button7=uicontrol('Parent',f,...
'Units','Normalized',...
'Position',[m(1+2*i)      n(1+2*i)      (m(2+2*i)-m(1+2*i))      (n(2+2*i)-
n(1+2*i))],...
'Style',enterType{i+1},...
'String',enterString{i+1},...
'FontSize', buttonTextSize(1+i),...
'BackgroundColor',enterColor,...
'HorizontalAlignment','center',...
'Callback',@button7Callback);
end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%
%USER CODE FOR THE VARIABLES AND CALLBACKS
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Initialize Variables
curr_file=1;
fs=8000;
fsd=10000;
directory_name='abcd';
wav_file_names='abce';
fin_path='filename';
fname='output';
nsamp=1;
xin=[];
nbins=25;

% Name the GUI
set(f,'Name','statistical_analysis');

% CALLBACKS
% Callback for button1 -- Get Speech Files Directory
function button1Callback(h,eventdata)
directory_name=uigetdir(start_path,'dialog_title');
A=strvcat(strcat((directory_name),'\*.wav'));
struct_filenames=dir(A);
wav_file_names={struct_filenames.name};
set(button2,'String',wav_file_names);
set(button2,'val',1);

% once the popupmenu/drop down menu is created, by default, the first
% selection from the popupmenu/drop down menu id not called
indexOfDrpDwnMenu=1;

```

```

% by default first option from the popupmenu/dropdown menu will be loaded

[curr_file,fs]=loadSelection(directory_name,wav_file_names,indexOfDrpD
wnMenu);
end

% Callback for button2 -- Choose speech file for play and plot
function button2Callback(h,eventdata)
    indexOfDrpDwnMenu=get(button2,'val');

[curr_file,fs]=loadSelection(directory_name,wav_file_names,indexOfDrpD
wnMenu);
end

%*****
****
% function -- load selection from designated directory and file
%
function [curr_file,fs]=loadSelection(directory_name,wav_file_names,...
    indexOfDrpDwnMenu);
%
% read in speech/audio file
% fin_path is the complete path of the .wav file that is selected

fin_path=strcat(directory_name,'\',strvcat(wav_file_names(indexOfDrpDw
nMenu)));

% clear speech/audio file
clear curr_file;

% read in speech/audio signal into curr_file; sampling rate is fs
[curr_file,fs]=wavread(fin_path);
xin=curr_file;
xinmax=max(max(xin),-min(xin));
xin=xin/xinmax;

% create title information with file, sampling rate, number of samples
fname=wav_file_names(indexOfDrpDwnMenu);
FS=num2str(fs);

% display speech file sampling rate on button 7; rate to be changed
% to fsd if fs not equal to fsd
set(button7,'string',FS);

    nsamp=num2str(length(curr_file));
    file_info_string=strcat('    file: ',fname,', fs: ',FS,' Hz,
nsamp:',nsamp);

% load filename (fname) from cell array
fname=wav_file_names{indexOfDrpDwnMenu};
end

% Callback for button3 -- fsd: signal processing rate

```

```

function button3Callback(h,eventdata)
    fsd=str2num(get(button3,'string'));
    a = [6000 8000 10000 16000 20000];
    fsd = a(get(button3,'val'));
end

% Callback for button4 -- nbins: number of bins in signal histogram
function button4Callback(h,eventdata)
    nbins=str2num(get(button4,'string'));
    if ~(nbins >= 25 && nbins <= 101)
        waitfor(errordlg('nbins must be a positive integer between 25
and 101'))
        return;
    end
    nbins=round(nbins);
    set(button4,'string',num2str(nbins));
end

% Callback for button5 -- Run statistical analysis
function button5Callback(h,eventdata)

% check editable buttons for changes
    button3Callback(h,eventdata);
    button4Callback(h,eventdata);

% setup and perform statistical analysis of speech file
    set_statistical_analysis(xin,fs,fsd,nbins,fname);
end

%*****
function set_statistical_analysis(xin,fsi,fsd,nbins,filein)
%
% function to perform statistical analysis of speech file
%
% Inputs:
%   xin: speech signal
%   fs: speech sampling rate
%   fsd: signal processing rate
%   nbins: number of bins for speech histogram
%   filein: speech filename

% resample if fsi ~= fsd
    if (fsi ~= fsd)
        fprintf('speech rate:%d, processing rate:%d \n',fsi,fsd);
        xin=srconv(xin,fsi,fsd);
    end

% process entire file from ss=1 to es=length(xin)
    ss=1;
    es=length(xin);

% determine minimum, maximum, mean and variance, and plot histogram of
% normalized samples, xinn=(xin-xmean)/xsigma
    xmin=min(xin(ss:es));
    xmax=max(xin(ss:es));

```

```

    xmean=mean(xin(ss:es));
    xsigma=std(xin(ss:es));

% normalize by subtracting the mean and dividing by the standard
deviation
    xinn=(xin-xmean)/xsigma;

% clear graphics Panel 3
    reset(graphicPanel3);
    axes(graphicPanel3);
    cla;

% plot speech signal histogram on graphics Panel 3
    hist(xinn(ss:es),nbins);
    string1=sprintf(' file:%s, ss,es:%d %d',filein,ss,es);
    string2=sprintf(' nbins:%d',nbins);
    string3=sprintf(' min,max,mean,sigma:%d %d %7.2f %7.0f',....
        xmin,xmax,xmean,xsigma);
    stitle=strcat(string1,string2,string3);
    grid on, xlabel('(x-m_{x})/\sigma_{x}'),ylabel('Count');
    legend('speech histogram'); title('Speech Signal Histogram');

% display fname and signal processing parameters on titleBox1
    stitle1=strcat('Statistical Analysis of Speech -- Speech
Histogram,',...
        ' Power Spectrum Estimates, Autocorrelation Estimates,',...
        ' Expanded AC Estimate: ',string1,string2);
    set(titleBox1,'string',stitle1);
    set(titleBox1,'FontSize',15);

% obtain estimate of long-time average power spectrum using a range of
% window lengths from 32 to 512 samples
    Nwinr=[512, 128, 32];
    color=['k','r','b','g','m'];
    index=1;
    for Nwin=[512,128,32]
        Nfft=2*Nwin;

% estimate power density spectrum in series of overlapped windows;
convert
% to estimated autocorrelation function
        [P,F,R,T]=pd_spect_U(xin(ss:es),fsd,Nfft,Nwin);

% plot power density spectral estimates for the
% different window lengths on graphics Panel 4
        if (Nwin == 512)
            reset(graphicPanel4); % clear graphics Panel 4
            axes(graphicPanel4);
            cla;
        else
            axes(graphicPanel4);
        end

% plot series of power density spectral estimates on graphics Panel 4
    plot(F,10*log10(P),color(index),'LineWidth',2),grid on;

```



```

ylabel('Log Magnitude in dB');
xlabel('Frequency in Hz');
title(['Speech Power Spectrum Estimates Nwin=512/128/32']);
hold on;

% clear graphics Panel 1
if (Nwin == 512)
    reset(graphicPanel1);
    axes(graphicPanel1);
    cla;
else
    axes(graphicPanel1);
end

% plot sequence of autocorrelation functions on graphics Panel 1
plot(T(1:Nfft/2+1)*1000,R(1:Nfft/2+1),color(index),'LineWidth',2);
grid on; ylabel('Autocorrelation Value');
xpp=['Time in msec; fsd=',num2str(fsd),' samples/second'];
xlabel(xpp);
title(['Speech Autocorrelation Estimates Nwin = 512/128/32']);
hold on;

% clear graphics Panel 2
if (Nwin == 512)
    reset(graphicPanel2);
    axes(graphicPanel2);
    cla;
else
    axes(graphicPanel2);
end

% plot first 50 lags of sequence of autocorrelation functions on graphics
% Panel 2
plot(0:49,R(1:50),color(index),'LineWidth',2),grid on;
ylabel('Autocorrelation Value');
xpp=['Time in Samples; fsd=',num2str(fsd),' samples/second'];
xlabel(xpp);
title(['Speech Autocorrelation Estimates-Expanded Scale, Nwin =
512/128/32']);
hold on;

% iterate to next window size
index=index+1;
end

% add legends to plots
axes(graphicPanel4);
legend('Nwin=512','Nwin=128','Nwin=32');
axes(graphicPanel1);
legend('Nwin=512','Nwin=128','Nwin=32');
axes(graphicPanel2);
legend('Nwin=512','Nwin=128','Nwin=32');
end

```

```
% Callback for button6 -- close GUI
function button6Callback(h,eventdata)
    close(gcf);
end

% Callback for button7 -- speech file sampling rate
function button7Callback(h,eventdata)
    set(button7,'string',num2str(fs));
end
end

function [y] = srconv(x,fsin,fsout)
%
% function to convert sampling rate from one sampling rate to another
% so long as the sampling rates have an integer least common multiple
%
% Inputs:
%   x: input signal at rate fsin
%   fsin: sampling rate on input
%   fsout: new sampling rate on output
%
% Output:
%   y: output signal at sampling rate fsout
%
% determine m, the least common multiple (lcm) of fsin and fsout
    m=lcm(fsin,fsout);

% determine the up and down sampling rates
    up=m/fsin;
    down=m/fsout;

% resample the input using the computed up/down rates
    y=resample(x,up,down);

end
```

Πανεπιστήμιο Πειραιώς