



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ  
ΕΡΓΑΣΤΗΡΙΟ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ ΚΑΙ ΟΡΑΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

**Στατιστική Πολυζωνική Αποδιαμόρφωση και Ανάλυση  
Φωνής με φίλτρα Kalman και ενεργειακές μεθόδους**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

**Αθανασίου Α. Κατσαμάνη**

**Επιβλέπων:** Πέτρος Α. Μαραγκός  
Καθηγητής

Αθήνα, 29 Οκτωβρίου 2003





**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ**  
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗ-  
ΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙ-  
ΚΗΣ  
ΕΡΓΑΣΤΗΡΙΟ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ ΚΑΙ Ο-  
ΡΑΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

## **Στατιστική Πολυζωνική Αποδιαμόρφωση και Ανάλυση Φωνής με φίλτρα Kalman και ενεργειακές μεθόδους**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

**Αθανασίου Α. Κατσαμάνη**

**Επιβλέπων:** Πέτρος Α. Μαραγκός  
Καθηγητής

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 29η Οκτωβρίου 2003.

.....  
Πέτρος Μαραγκός  
Καθηγητής

.....  
Γιώργος Καραγιάννης  
Καθηγητής

.....  
Τρύφων Κουσιουρής  
Καθηγητής

Αθήνα, 29 Οκτωβρίου 2003.

.....  
**(Αθανάσιος Α. Κατσαμάνης)**

(Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Ηλεκτρονικών Υπολογιστών Ε.Μ.Π.)

© Αθανάσιος Α. Κατσαμάνης, 2003

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

# Περίληψη

Με την παρούσα εργασία προτείνεται μια νέα προσέγγιση στο πρόβλημα της ανάλυσης σήματος που αποτελείται από πολλές συνιστώσες οι οποίες είναι διαμορφωμένες τόσο κατά πλάτος όσο και κατά συχνότητα μέσα σε θόρυβο. Η ανάλυση συνίσταται στο διαχωρισμό των επιμέρους συνιστωσών και την επιτυχή αποδιαμόρφωσή τους. Η προτεινόμενη προσέγγιση, KESDA, συνδυάζει ενεργειακές μεθόδους με μεθόδους εκτίμησης με τη χρήση παραλλαγών του φίλτρου Kalman για μη γραμμικά συστήματα (Kalman Energy Separation Demodulation Algorithm) και παρουσιάζει κάποια ιδιαίτερα ενδιαφέροντα χαρακτηριστικά. Παρουσιάζονται διάφορες παραλλαγές της και γίνεται κάποια προσπάθεια συγκριτικής αξιολόγησης τους ενώ συγκρίνονται τα αποτελέσματά τους σε διάφορες περιπτώσεις και με έναν άλλον αλγόριθμο που αντιμετωπίζει το ίδιο πρόβλημα, τον Kalman-MBDA, (Model Based Demodulation Algorithm). Τα αποτελέσματα είναι αρκετά ικανοποιητικά.

Βασικό κίνητρο για την ανάπτυξη της νέας προσέγγισης αποτελεί η ανάγκη για ανάλυση σημάτων φωνής με βάση ένα νέο μη γραμμικό μοντέλο, σύμφωνα με το οποίο κάθε σήμα φωνής μπορεί να μοντελοποιηθεί ως υπέρθεση AM-FM σημάτων. Απώτερος στόχος είναι η αναζήτηση νέων ενδιαφερόντων χαρακτηριστικών του σήματος φωνής. Παρουσιάζονται σχετικά πειράματα και γίνεται κάποια πρώτη διερεύνηση. Υπάρχουν ενδείξεις ότι ο νέος αλγόριθμος μπορεί τελικά να δώσει καλά αποτελέσματα και σε αυτό το χώρο.

## Λέξεις Κλειδιά

Ενεργειακός τελεστής, αποδιαμόρφωση, σήματα AM-FM, ανάλυση φωνής, μη γραμμικό μοντέλο φωνής, αλγόριθμος διαχωρισμού με υπολογισμό της ενέργειας, εκτίμηση, εκτεταμένο φίλτρο Kalman, φίλτρο Kalman χωρίς καθοδήγηση, ενεργειακοί παλμοί, στατιστική πολυζωνική αποδιαμόρφωση

# Abstract

This thesis basically addresses the problem of demodulation of multicomponent AM-FM signals in noise. Speech is considered lately to have the structure of such signals, according to a recently proposed model. This consideration and the need for relevant speech analysis motivated the proposal of a new approach to the problem. The new approach is actually a demodulation algorithm that combines both estimation methods based on different variations of Kalman filters and methods based on energy calculation. It is shortly called KESDA after the initials of its full name, which is: Kalman Energy Separation Demodulation Algorithm. Four different variations of KESDA are presented and applied to different cases of modulated signals. The results are quite satisfying. Especially, in comparison to the recently introduced Kalman-MBDA (Model Based Demodulation Algorithm) the new algorithm seems to have upgraded performance in many situations.

Special interest has been shown in the analysis of speech signals. Some first experiments are presented along with basic conclusions to which the relevant research using the new algorithm, has driven.

## **Keywords**

ESA, Kalman, Energy Separation Algorithm, Speech Analysis, AM-FM Speech model, KESDA, Demodulation, extended Kalman filter, unscented Kalman filter, multicomponent AM-FM signals, nonlinear speech processing, multiband demodulation

# Περιεχόμενα

<b>1</b>	<b>Εισαγωγή</b>	<b>12</b>
1.1	Γραμμικό μοντέλο φωνής και γραμμική πρόβλεψη . . . . .	12
1.2	Μη γραμμικό μοντέλο και μη γραμμικά συστήματα επεξεργασίας φωνής	14
1.3	Σκοπός της εργασίας . . . . .	15
<b>2</b>	<b>Προκαταρκτικά</b>	<b>17</b>
2.1	Εισαγωγή . . . . .	17
2.2	Το μοντέλο διαμόρφωσης AM-FM . . . . .	17
2.2.1	Περιγραφή . . . . .	17
2.2.2	Σύγκριση με το κλασσικό μοντέλο . . . . .	18
2.2.3	Ενδείξεις . . . . .	18
2.3	Αλγόριθμος διαχωρισμού με υπολογισμό της ενέργειας . . . . .	19
2.3.1	Ο ενεργειακός τελεστής . . . . .	19
2.3.2	Περιγραφή του ESA . . . . .	22
2.3.3	Παραλλαγές του διακριτού αλγορίθμου . . . . .	23
2.3.4	Παραδείγματα εφαρμογής του ESA . . . . .	23
2.3.5	Επέκταση του αλγορίθμου διαχωρισμού με υπολογισμό της ε- νέργειας σε πολλές ζώνες συχνοτήτων . . . . .	25
2.4	Εκτίμηση και φίλτρο <i>Kalman</i> . . . . .	30
2.4.1	Το πρόβλημα της εκτίμησης . . . . .	30
2.4.2	Το φίλτρο <i>Kalman</i> . . . . .	30
2.4.3	Το εκτεταμένο φίλτρο <i>Kalman-EKF</i> . . . . .	32
2.4.4	Το χωρίς καθοδήγηση φίλτρο <i>Kalman</i> -( <i>Unscented Kalman Filter</i> )	35
<b>3</b>	<b>Αποδιαμόρφωση βασισμένη σε μοντέλο</b>	<b>38</b>
3.1	Εισαγωγή . . . . .	38
3.2	Περιγραφή του μοντέλου . . . . .	38
3.3	Στόχος του αλγορίθμου . . . . .	40
3.4	Ταυτοποίηση του συστήματος . . . . .	40
3.5	Εκτίμηση με χρήση του εκτεταμένου φίλτρου <i>Kalman</i> . . . . .	42
3.6	Εφαρμογές . . . . .	42
3.6.1	Εφαρμογή σε συνθετικά σήματα . . . . .	42
3.6.2	Εφαρμογή σε φωνή . . . . .	43

3.7	Παραλλαγές . . . . .	45
3.8	Συμπεράσματα . . . . .	48
<b>4</b>	<b>Kalman-ESA Αποδιαμόρφωση</b>	<b>49</b>
4.1	Εισαγωγή . . . . .	49
4.2	ESA και σήμα με πολλές συνιστώσες . . . . .	50
4.3	Αλγόριθμος ανάλυσης με βάση το μοντέλο <i>bKES</i> . . . . .	52
4.3.1	Μέθοδος υπολογισμού του στιγμιαίου πλάτους και της στιγμιαίας συχνότητας μιας από τις AM-FM συνιστώσες (IFAE) . . . . .	53
4.3.2	Αρχικοποίηση και εκτίμηση με χρήση φίλτρου Kalman . . . . .	54
4.3.3	Εφαρμογή σε συνθετικά σήματα . . . . .	55
4.4	Αλγόριθμος ανάλυσης με βάση το μοντέλο <i>JKES</i> . . . . .	57
4.4.1	Αρχικοποίηση και εκτίμηση με χρήση του εκτεταμένου φίλτρου Kalman . . . . .	60
4.4.2	Εφαρμογή σε συνθετικά σήματα . . . . .	60
4.5	Αλγόριθμος ανάλυσης με βάση το μοντέλο <i>iKES</i> . . . . .	61
4.5.1	Αρχικοποίηση και εκτίμηση με χρήση του εκτεταμένου φίλτρου Kalman . . . . .	63
4.5.2	Εφαρμογή σε συνθετικά σήματα . . . . .	63
4.6	Αλγόριθμος ανάλυσης με βάση το μοντέλο <i>mKES</i> . . . . .	64
4.6.1	Αρχικοποίηση και εκτίμηση με χρήση φίλτρου Kalman . . . . .	66
4.6.2	Εφαρμογή σε συνθετικά σήματα . . . . .	67
4.7	Επίδραση του θορύβου . . . . .	70
4.8	Σύγκριση και Αξιολόγηση . . . . .	70
4.8.1	Επιδράσεις των χαρακτηριστικών διαμόρφωσης . . . . .	70
4.8.2	Σύγκριση με τον Kalman-MBDA . . . . .	75
<b>5</b>	<b>Εφαρμογή της προσέγγισης KESDA στη φωνή</b>	<b>76</b>
5.1	Εισαγωγή . . . . .	76
5.2	Εφαρμογή σε συνθετικά σήματα ομοιάζοντα φωνής . . . . .	76
5.2.1	Εφαρμογή σε συνθετικά σήματα ομοιάζοντα φωνής με βάση το γραμμικό μοντέλο . . . . .	77
5.2.2	Εφαρμογή σε συνθετικά σήματα ομοιάζοντα φωνής με βάση το μη γραμμικό μοντέλο . . . . .	81
5.3	Εφαρμογή σε πραγματικά σήματα φωνής . . . . .	86
5.3.1	Ανάλυση φωνής . . . . .	86
5.3.2	Παρακολούθηση κεντρικών συχνοτήτων συντονισμών φωνής . . . . .	90
5.4	Συμπεράσματα . . . . .	92
<b>6</b>	<b>Συμπεράσματα- Μελλοντική Έρευνα</b>	<b>98</b>
6.1	Ανακεφαλαίωση . . . . .	98
6.2	Συμπεράσματα . . . . .	99
6.3	Μελλοντική έρευνα . . . . .	100
6.3.1	Εφαρμογή στη φωνή . . . . .	101



6.3.2 Θεωρητική Διερεύνηση . . . . .	101
6.3.3 Μείωση πολυπλοκότητας . . . . .	101

# Κατάλογος Σχημάτων

2.1	Ενεργειακοί παλμοί ως αποτέλεσμα της εφαρμογής του ενεργειακού τελεστή σε σήμα φωνής . . . . .	20
2.2	Εφαρμογή του ενεργειακού τελεστή σε σήμα προερχόμενο από γραμμικό ταλαντωτή . . . . .	21
2.3	Εφαρμογή του ενεργειακού τελεστή σε σήμα διαμορφωμένο κατά πλάτος και κατά συχνότητα . . . . .	21
2.4	Εφαρμογή του ESA σε AM-FM σήμα . . . . .	24
2.5	Εφαρμογή του ESA σε AM-FM σήμα με γραμμικά μεταβαλλόμενη συχνότητα . . . . .	26
2.6	Εφαρμογή του ESA σε AM-FM σήμα . . . . .	27
2.7	Εφαρμογή του ESA σε AM-FM σήμα . . . . .	28
2.8	Κρουστική Απόκριση και Απόκριση Συχνότητας του φίλτρου Gabor . .	29
3.1	Αποτελέσματα εφαρμογής του Kalman-MBDA σε συνθετικό σήμα . . .	44
3.2	Σήμα φωνής και εκτιμήσεις κεντρικών συχνοτήτων του με τη χρήση του Kalman-MBDA . . . . .	45
3.3	Εκτιμήσεις στιγμιαίων πλατών και συχνοτήτων σήματος φωνής . . . .	46
3.4	Παρακολούθηση κεντρικών συχνοτήτων από τον αλγόριθμο MBDA . .	47
4.1	Εφαρμογή του ESA με ή χωρίς προσαρμοζόμενη συστοιχία φίλτρων . .	51
4.2	Εφαρμογή του μοντέλου <i>bKES</i> σε σήμα με σταθερές κεντρικές συχνοτήτες . . . . .	57
4.3	Εφαρμογή του μοντέλου <i>bKES</i> σε σήμα με γραμμικά μεταβαλλόμενες κεντρικές συχνοτήτες . . . . .	58
4.4	Εφαρμογή του μοντέλου <i>fKES</i> σε σήμα με σταθερές κεντρικές συχνοτήτες	61
4.5	Εφαρμογή του μοντέλου <i>fKES</i> σε σήμα με γραμμικά μεταβαλλόμενες κεντρικές συχνοτήτες . . . . .	62
4.6	Εφαρμογή του μοντέλου <i>iKES</i> σε σήμα με σταθερές κεντρικές συχνοτήτες	64
4.7	Εφαρμογή του μοντέλου <i>iKES</i> σε σήμα με γραμμικά μεταβαλλόμενες κεντρικές συχνοτήτες . . . . .	65
4.8	Αρχικοποίηση του <i>mKES</i> με υπολογισμό της φασματικής πυκνότητας ισχύος . . . . .	68

4.9 Εφαρμογή του μοντέλου <i>mKES</i> σε σήμα με σταθερές κεντρικές συχνότητες . . . . .	68
4.10 Εφαρμογή του μοντέλου <i>mKES</i> σε σήμα με γραμμικά μεταβαλλόμενες κεντρικές συχνότητες . . . . .	69
4.11 Η επίδραση του θορύβου συγκριτικά για όλους τους αλγορίθμους . . . . .	71
4.12 Επίδραση της μεταβολής του δείκτη διαμόρφωσης AM στην εκτίμηση με χρήση των προτεινόμενων μοντέλων ή του Kalman-MBDA . . . . .	73
4.13 Επίδραση της αύξησης του δείκτη διαμόρφωσης FM στην απόδοση των αλγορίθμων . . . . .	74
5.1 Φασματική πυκνότητα ισχύος του φωνήματος /ee/ . . . . .	77
5.2 Συνθετικό σήμα φωνής με τρεις συντονισμούς με βάση το γραμμικό μοντέλο . . . . .	78
5.3 Εφαρμογή του ESA σε αποκρίσεις συντονιστών . . . . .	79
5.4 Εφαρμογή του <i>bKES</i> σε συνθετικό σήμα φωνής προερχόμενο από γραμμικούς συντονιστές . . . . .	80
5.5 Εφαρμογή του <i>mKES</i> σε συνθετικό σήμα φωνής προερχόμενο από γραμμικούς συντονιστές . . . . .	82
5.6 Εφαρμογή του MBDA σε συνθετικό σήμα φωνής προερχόμενο από γραμμικούς συντονιστές . . . . .	83
5.7 Συνθετικό σήμα φωνής με βάση το μη γραμμικό μοντέλο . . . . .	85
5.8 Εφαρμογή του μοντέλου <i>bKES</i> σε συνθετικό σήμα που προσομοιάζει σήμα φωνής . . . . .	87
5.9 Ανακατασκευασμένο συνθετικό σήμα φωνής μετά από εκτίμηση με το <i>bKES</i> . . . . .	88
5.10 Εφαρμογή του μοντέλου <i>mKES</i> σε συνθετικό σήμα που προσομοιάζει σήμα φωνής . . . . .	89
5.11 Ανακατασκευασμένο συνθετικό σήμα φωνής μετά από εκτίμηση με το <i>mKES</i> . . . . .	90
5.12 Εφαρμογή του MBDA σε συνθετικό σήμα που προσομοιάζει σήμα φωνής . . . . .	91
5.13 Φασματική πυκνότητα ισχύος του στοχαστικού μοντέλου το οποίο χρησιμοποιείται στον MBDA και <i>mKES</i> για σήμα φωνής . . . . .	92
5.14 Ανάλυση σήματος φωνής με χρήση του <i>bKES</i> , παρουσίαση των στιγμιαίων πλατών και συχνοτήτων . . . . .	93
5.15 Ανάλυση σήματος φωνής με χρήση του <i>mKES</i> , παρουσίαση των στιγμιαίων πλατών και συχνοτήτων . . . . .	94
5.16 Ανάλυση σήματος φωνής με χρήση του MBDA, παρουσίαση των στιγμιαίων πλατών και συχνοτήτων . . . . .	95
5.17 Ανακατασκευασμένο σήμα φωνής από τις εκτιμήσεις με βάση το μοντέλο <i>bKES</i> . . . . .	96
5.18 Παρακολούθηση συντονισμών σε πραγματικό σήμα φωνής από το <i>mKES</i> . . . . .	97

# Κεφάλαιο 1

## Εισαγωγή

« Πολλές περιοχές της κλασσικής φυσικής, οι οποίες θεωρούνται ότι έχουν εξονυχιστικά κωδικοποιηθεί και εξερευνηθεί από εξέχοντες επιστήμονες του παρελθόντος, είναι ανολοκλήρωτες, αλλά αφού η τεχνολογία ξεπέρασε την ανάγκη για συγκεκριμένες σχετικές συσκευές, τα πεδία αυτά έγιναν αφηρημένες έννοιες και έχουν αφεθεί ως ενδιαφέροντα θεωρητικά προβλήματα. Αυτό ισχύει για ευρείες περιοχές της ρευστοδυναμικής και της ακουστικής υδροδυναμικής συγκεκριμένα », [15]. Με αυτά τα λόγια ξεκίνησε ο Teager την παρουσίαση μιας πρωτοποριακής προσέγγισης στο θέμα της μοντελοποίησης του συστήματος παραγωγής αλλά και αντίληψης φωνής, προκαλώντας παράλληλα, με βάση και τα εκτενή πειραματικά δεδομένα που συγκέντρωσε και παρουσίασε, ισχυρούς τριγμούς στο οικοδόμημα της γραμμικής μοντελοποίησης της φωνής, η οποία είναι μέχρι σήμερα επικρατούσα.

### 1.1 Γραμμικό μοντέλο επεξεργασίας φωνής και γραμμική πρόβλεψη

Η γραμμική μοντελοποίηση της φωνής, τουλάχιστον από φυσική άποψη, έχει τα θεμέλιά της στην κλασσική γραμμική ακουστική. Διέπεται από τη βασική ιδέα ότι η φωνητική οδός μπορεί να μοντελοποιηθεί ως μια συστοιχία σωλήνων διαφόρων διαστάσεων και ότι ο ήχος διαδίδεται ως μονοδιάστατο επίπεδο κύμα δια μέσου της συστοιχίας αυτής. Μετά από μελέτη των οριακών συνθηκών στα σημεία όπου ενώνονται οι σωλήνες αλλά και στα άκρα, την επεξεργασία διαφόρων άλλων παραμέτρων που πρέπει να ληφθούν υπόψη (π.χ. απώλειες), και την εφαρμογή των σχετικών κυματικών εξισώσεων, είναι δυνατή η δημιουργία κατάλληλου συστήματος-μοντέλου. Προκύπτει μετά και από τη σύζευξη επιμέρους συστημάτων-μοντέλων της στοματικής και φαρυγγικής κοιλότητας, της ρινικής κοιλότητας (για ένρινους ήχους), των φωνητικών χορδών, για έμφωνους ήχους, και των χειλιών. Σε κάθε μικρό χρονικό διάστημα στο οποίο παραμένει αμετάβλητο παρουσιάζει συγκεκριμένες ιδιοσυχνότητες (συντονισμούς) -εμφανίζονται ως πόλοι στη συνάρτηση μεταφοράς- που διεγείρονται

από κατάλληλη κάθε φορά είσοδο, κρουστική παλμοσειρά για έμφωνους ήχους και λευκό θόρυβο για άφωνους, ή σε άλλη προσέγγιση, πολυπαλμική κυματομορφή, [20], [2]. Η συνάρτηση μεταφοράς έχει τη μορφή:

$$H(z) = \frac{G}{\prod_i (z - z_{pi})} \quad (1.1)$$

Σε αυτό το υπόβαθρο στηρίζεται και η χρησιμοποίηση της θεωρίας της γραμμικής πρόβλεψης στην επεξεργασία φωνής, σύμφωνα με την οποία σε κάποια χρονική στιγμή  $n$ , το σήμα φωνής  $s(n)$  μπορεί να προσεγγιστεί ως γραμμικός συνδυασμός των  $p$  τελευταίων δειγμάτων έτσι ώστε:

$$s(n) \approx \alpha_1 s(n-1) + \alpha_2 s(n-2) + \dots + \alpha_p s(n-p) \quad (1.2)$$

$$s(n) = \sum_{i=1}^p \alpha_i s(n-i) + Gu(n) \quad (1.3)$$

όπου οι συντελεστές  $\alpha_1, \alpha_2, \dots, \alpha_p$  θεωρούνται σταθεροί κατά μήκος ενός πλαισίου ανάλυσης της φωνής και βρίσκονται σε άμεση συσχέτιση με σχετικές παραμέτρους του μοντέλου της φωνητικής οδού ενώ  $u(n)$  είναι το σφάλμα πρόβλεψης και συνδέεται με τη διέγερση του φωνητικού συστήματος, [20].

Όχι άδικα, η προσέγγιση αυτή έδωσε μεγάλη ώθηση στις τεχνολογίες φωνής. Προσφέροντας μαθηματική ακρίβεια αλλά και τη δυνατότητα απλής και άμεσης υλοποίησης τόσο σε επίπεδο υλικού όσο και λογισμικού έδωσε το έναυσμα αλλά και χρήσιμα εργαλεία για ανάπτυξη συστημάτων αναγνώρισης, σύνθεσης, κωδικοποίησης φωνής. Αν μη τι άλλο, ήταν μια καλή αρχή, μια αρχή πάνω στην οποία έχουν χτιστεί πολλά από τα πιο σημαντικά συστήματα επικοινωνίας ανθρώπου-μηχανής σήμερα. Η μεγάλη σημασία τους βέβαια αφορά κυρίως στις πολλές πιθανές εφαρμογές τους και όχι τελικά τόσο στην αποτελεσματικότητά τους. Γιατί οι περιορισμοί είναι πολλοί, [21].

Πράγματι, τα συστήματα αναγνώρισης φωνής, για παράδειγμα, ακόμα και τα καλύτερα, κρίνονται τουλάχιστον ανεπαρκή σε πολλές περιπτώσεις οι οποίες αποτελούν μάλλον τον κανόνα, παρά την εξαίρεση. Περιπτώσεις υψηλού περιβάλλοντος θορύβου, περιορισμένου εύρους ζώνης, μεταβολών στην ποιότητα φωνής ή κακής άρθρωσης εμφανίζονται συχνά στις σχετικές εφαρμογές και συνήθως προβληματίζουν. Η σύνθεση φωνής από την άλλη, που θα μπορούσε να είναι κατεξοχήν η τεχνολογία που θα καταξίωνε το γραμμικό μοντέλο φωνής, παρουσιάζει ανάλογες ατέλειες. Η συνθετική φωνή μπορεί να είναι σε αρκετά μεγάλο βαθμό κατανοητή αλλά συνήθως στερείται φυσικότητας και δεν παρουσιάζει ευελιξία στις περιπτώσεις που χρειάζεται να αλλάξει το φύλο ή η χροιά της ή ίσως να αποδωθούν κάποιες συγκεκριμένες συναισθηματικές καταστάσεις. Τα προβλήματα αυτά σίγουρα οδηγούν σε συμβιβασμούς όσον αφορά στις προσδοκίες.

Από την άλλη βέβαια δικαιολογούν και την έρευνα σχετικά με τις δυνατότητες βελτίωσης, έρευνα που στρέφεται σε πολλές κατευθύνσεις και που ίσως αμφισβητεί υποθέσεις που θεωρούνται δεδομένες, όπως είναι αυτές που οδηγούν στο γραμμικό μοντέλο πηγής-φίλτρου για το ανθρώπινο σύστημα παραγωγής φωνής, [16]:

1. Η διέγερση και το φίλτρο που μοντελοποιεί τη φωνητική οδό είναι ανεξάρτητα.
2. Η ροή μέσα από τη φωνητική οδό είναι ομαλή.
3. Η φωνή διαδίδεται μέσα στη φωνητική οδό ως επίπεδο μονοδιάστατο κύμα.

## 1.2 Μη γραμμικό μοντέλο και μη γραμμικά συστήματα επεξεργασίας φωνής

Αυτές ακριβώς οι υποθέσεις μπήκαν στο στόχαστρο των ερευνών του Teager ήδη από τις αρχές της δεκαετίας του 1970. Ορμώμενος από μετρήσεις της ροής του αέρα μέσα στη φωνητική κοιλότητα κατά την εκφώνηση έμφωνων ήχων αλλά και από μια πιο προσεκτική συσχέτιση του προβλήματος με το χώρο της ρευστοδυναμικής διατύπωσε βάσιμες αντιρρήσεις σχετικά με το κατά πόσο οι υποθέσεις αυτές είναι ρεαλιστικές. Κατ' επέκταση, δίνοντας έμφαση και σε όλες τις δυσκολίες που αντιμετώπιζαν τότε οι ερευνητές σε θέματα φωνής, αμφισβήτησε ανοιχτά το γραμμικό μοντέλο φωνής και προσπάθησε να θέσει τις βάσεις για ένα καινούριο μοντέλο που θα ήταν πιο κοντά στην πραγματικότητα, ένα μοντέλο που θα είχε ισχυρότερα θεμέλια από φυσικής άποψης, [16]. Αρωγό σε αυτή του την προσπάθεια βρήκε τον Kaiser, ο οποίος ασχολήθηκε και αυτός ενθουσιασμένος με «τη διαμόρφωση μιας νέας προσέγγισης που προσφέρει έναν τόσο συναρπαστικό τρόπο να δει κάποιος ένα σημαντικότερο θέμα που αφορά στην ανθρώπινη επικοινωνία», [18].

Βασικό χαρακτηριστικό της νέας θεωρίας που αναδύθηκε είναι η μη γραμμικότητα. Η ροή του αέρα μέσα στη φωνητική οδό είναι ανισοτροπική και ασταθής, όπως έδειξαν οι πειραματικές μετρήσεις, ενώ μεταβάλλεται συνεχώς με το χρόνο, συγκεντρώνεται στις επιφάνειες και μπορεί πολλές φορές να δημιουργήσει στροβίλους. Επόμενο είναι λοιπόν τα μη γραμμικά φαινόμενα να μην είναι δυνατόν να θεωρηθούν ως δεύτερης τάξης αμελητέες πηγές σφαλμάτων, όπως γίνεται με τη γραμμική προσέγγιση. Περιπλέκουν βέβαια την κατάσταση αλλά αυτό δεν φαίνεται να αποθαρρύνει τους θιασώτες της νέας προσέγγισης.

Η ανάγκη για περαιτέρω μελέτη των σχετικών φαινομένων οδήγησε την έρευνα σε δύο κυρίως κατευθύνσεις:

1. Πραγματοποίηση αριθμητικών προσομοιώσεων των μη γραμμικών διαφορικών (Navier-Stokes) εξισώσεων που γενικά ίσως θα μπορούσαν να περιγράψουν τρισδιάστατη ροή αέρα στο φωνητικό σωλήνα και έτσι θα μπορούσε συνθετικά να προκύψει ένα καλύτερο μοντέλο για τη φωνή.
2. Αντίστροφα, ανάπτυξη μη γραμμικών συστημάτων ικανών να εντοπίσουν τέτοια φαινόμενα και να εξάγουν την όποια σχετική πληροφορία. Στην ουσία δηλαδή, χρήση κάποιων υποθετικών μοντέλων με τρόπο ώστε να είναι δυνατή η εξαγωγή χρήσιμων συμπερασμάτων σχετικά με το μηχανισμό παραγωγής φωνής.

Κινούμενοι στον δεύτερο άξονα, το 1991-1993, οι Maragos, Kaiser & Quatieri, [26], προτείνουν ένα νέο μη γραμμικό μοντέλο για τη φωνή που στηρίζεται στη θεωρία της διαμόρφωσης AM-FM και εισάγουν νέα εργαλεία για τη σχετική επεξεργασία. Παράλληλα, η γενικότερη εφαρμογή των μη γραμμικών συστημάτων στη φωνή από διάφορους ερευνητές δίνει σε πολλές περιπτώσεις αρκετά καλά αποτελέσματα σε σύγκριση με τις κλασσικές μεθόδους, [21]. Τα μη γραμμικά συστήματα αναδεικνύονται ιδιαίτερα χρήσιμα, παρά τα όποια προβλήματά τους, στα οποία συμπεριλαμβάνονται η αυξημένη συνήθως πολυπλοκότητα αλλά και η αδυναμία αναλυτικής αντιμετώπισής τους λόγω απουσίας κατάλληλων εργαλείων.

Με αυτό το πνεύμα το 1996 οι Lu & Doerschuk παρουσίασαν σε ένα άρθρο τους [33] ένα νέο μη γραμμικό τρόπο προσέγγισης της επεξεργασίας φωνής βασισμένο από τη μία στο μοντέλο που πρότειναν οι Maragos et al [26] και από την άλλη σε μεθόδους μη γραμμικού φιλτραρίσματος. Το άρθρο αυτό αποτέλεσε και τη βασική έμπνευση της μεθόδου ανάλυσης φωνής που παρουσιάζεται στην παρούσα εργασία.

### 1.3 Σκοπός της εργασίας

Σκοπός της εργασίας είναι η ανάπτυξη και μελέτη μιας μεθόδου ανάλυσης φωνής με βάση το μη γραμμικό μοντέλο για το οποίο έγινε λόγος. Η μέθοδος αυτή στηρίζεται σε αλγόριθμους που έχουν ήδη παρουσιαστεί και προσπαθεί να βελτιώσει τη στατιστική προσέγγιση των Lu & Doerschuk, [33], προσδίδοντας της, όσο είναι δυνατό, μια περισσότερο φυσική διάσταση, με κατάλληλο συνδυασμό της με τον αλγόριθμο διαχωρισμού με υπολογισμό της ενέργειας, ο οποίος περιγράφεται στη συνέχεια και έχει αρκετά ισχυρό υπόβαθρο από φυσικής άποψης.

Το γενικότερο πρόβλημα που αντιμετωπίζεται αφορά σε σήματα που περιέχουν πολλές συνιστώσες διαμορφωμένες τόσο κατά πλάτος όσο και κατά συχνότητα μέσα σε θόρυβο. Τέτοια σήματα αποτελούν τη βάση γενικότερης μοντελοποίησης μη στατικών σημάτων ως υπερθέσεις διαμορφωμένων ημιτόνων, όπου κάθε συνιστώσα καταλαμβάνει μια στενή ζώνη συχνοτήτων γύρω από τη φέρουσα της συχνότητα. Βασική απαίτηση στις εφαρμογές όπου εμφανίζονται είναι ο διαχωρισμός και η αποδιαμόρφωση των επιμέρους συνιστωσών και προς αυτή την κατεύθυνση κινείται η μέθοδος που προτείνεται. Αν και βέβαια θα μπορούσε να εφαρμοστεί και σε άλλες περιοχές επεξεργασίας σημάτων, σχεδιάστηκε κυρίως για το διαχωρισμό και την αποδιαμόρφωση συνιστωσών του σήματος φωνής και σε αυτό το χώρο δοκιμάζεται κιόλας σε κάποιο βαθμό για να διαπιστωθεί κατ' αρχάς αν τα αποτελέσματά της είναι σύμφωνα με άλλες σχετικές προσεγγίσεις. Στο χωρό αυτό της φωνής πάντως, μέσω μιας τέτοιας μορφής ανάλυση με εκτίμηση των περιβαλλουσών και των στιγμιαίων συχνοτήτων των διαφόρων συνιστωσών που αντιστοιχούν στις περιοχές του φάσματος με υψηλή ενέργεια, ενδεχόμενα θα είναι δυνατή στο μέλλον η εξαγωγή χρήσιμης πληροφορίας που θα βοηθούσε στην καλύτερη κατανόησή του σήματος φωνής από τους ενδιαφερόμενους ερευνητές και σε πιο αποδοτική ίσως επεξεργασία του.

Στα κεφάλαια που ακολουθούν παρουσιάζεται όσο το δυνατόν πληρέστερα το πρόβλημα το οποίο εκτέθηκε σε συντομία αλλά κυρίως η λύση που προτείνεται. Στο

**Κεφάλαιο 2** εξετάζονται διάφορα προκαταρκτικά θέματα τόσο χάριν πληρότητας, όσο και για εύκολη αναφορά κατά την περαιτέρω ανάπτυξη. Περιγράφεται το μη γραμμικό μοντέλο φωνής αλλά και ο αλγόριθμος διαχωρισμού με υπολογισμό της ενέργειας ενώ δίνονται συνοπτικά κάποια βασικά στοιχεία από το χώρο της στατιστικής εκτίμησης στοχαστικών διαδικασιών με τη χρήση φίλτρου Kalman, και κυρίως της εκτεταμένης αλλά και της χωρίς καθοδήγηση παραλλαγής του. Το **Κεφάλαιο 3** είναι αφιερωμένο στην περιγραφή του αλγορίθμου αποδιαμόρφωσης βασισμένου σε μοντέλο (*Model Based Demodulation Algorithm*), του στατιστικού αλγορίθμου δηλαδή που πρότειναν οι Lu & Doerschuk, [33], και ο οποίος αποτελεί τη βασική έμπνευση της νέας μεθόδου που προτείνεται. Ουσιαστική περιγραφή και ανάλυση της τελευταίας, με τη μορφή διάφορων σχετικών παραλλαγών δίνεται στο **Κεφάλαιο 4**. Στο κεφάλαιο αυτό εξετάζονται επίσης σε κάποιο βαθμό συστηματικά οι δυνατότητες εφαρμογής της μεθόδου σε συνθετικά σήματα, ενώ ξεχωριστά, στο **Κεφάλαιο 5** μελετάται σε ένα πρώτο επίπεδο η εφαρμογή της προσέγγισης που προτείνεται σε σήματα φωνής. Στο **Κεφάλαιο 6**, τέλος, εκτίθενται τα συμπεράσματα που προκύπτουν στα πλαίσια της σχετικής έρευνας αλλά και οι κατευθύνσεις προς τις οποίες θα μπορούσε να κινηθεί αυτή η έρευνα στο μέλλον ώστε να συμβάλλει ακόμα περισσότερο στην « αποκρυπτογράφηση » του σήματος φωνής.



# Κεφάλαιο 2

## Προκαταρκτικά

### 2.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζεται κατ' αρχάς το μοντέλο διαμόρφωσης AM-FM για τη φωνή και στη συνέχεια περιγράφεται ένας απλός αλγόριθμος αποδιαμόρφωσης που στηρίζεται στον λεγόμενο ενεργειακό τελεστή. Δίνονται κάποια στοιχεία για τις δυνατότητες και τους περιορισμούς του αλγορίθμου και στη συνέχεια παρουσιάζεται μια παραλλαγή του, ή μάλλον επέκτασή του, για την περίπτωση σημάτων με θόρυβο, με τη χρήση κατάλληλης συστοιχίας φίλτρων. Τέλος, παρουσιάζεται το πρόβλημα της εκτίμησης σημάτων με κριτήριο την ελαχιστοποίηση του τετραγωνικού λάθους και περιγράφεται η ιδέα του φιλτραρίσματος *Kalman* ως λύση, σε δύο παραλλαγές της, την εκτεταμένη μορφή *extended Kalman Filter* και την μορφή χωρίς καθοδήγηση *unscented Kalman Filter*. Δίνονται και κάποιες πληροφορίες σχετικά με την εφαρμογή της ιδέας αυτής στην αποδιαμόρφωση σημάτων.

### 2.2 Το μοντέλο διαμόρφωσης AM-FM

#### 2.2.1 Περιγραφή

Σύμφωνα με τους Maragos et al, [26], είναι δυνατή η μοντελοποίηση κάθε συντονισμού της φωνής ως ένα εκθετικά αποσβένον AM-FM σήμα :

$$R_{afm}^c(t) = A \exp^{-\sigma t} \cos(\omega_c t + \omega_m \int_0^t q(t) dt + \theta) \quad (2.1)$$

στη συνεχή μορφή και :

$$R_{afm}^d(n) = r^n A(n) \cos(\Omega_c n + \Omega_m \sum_{k=0}^{n-1} q(k) dk + \varphi(0)) \quad (2.2)$$

στη διακριτή μορφή, όπου  $\Omega_c$  είναι η κεντρική συχνότητα του συντονισμού, ενώ ως  $\Omega_i(n) \triangleq \Omega_c + \Omega_m q(n)$  μπορεί να μοντελοποιηθεί η χρονικά μεταβαλλόμενη συχνότητα του συντονισμού, της οποίας η απόκλιση από την  $\Omega_c$  μπορεί να πάρει τιμή  $\Omega_m$  κατά το μέγιστο και ακολουθεί κάποιο διαμορφώνον σήμα  $q(n)$ , όπου θεωρείται πως  $|q(n)| \leq 1$ .

Το σήμα φωνής τότε μπορεί να θεωρηθεί ως άθροισμα τέτοιων διαμορφωμένων σημάτων:

$$S(t) = \sum_{k=1}^K \alpha_k(t) \cos[\varphi_k(t)] \quad (2.3)$$

όπου ο δείκτης  $k$  αναφέρεται στον  $k$ -οστό συντονισμό της φωνής και  $K$  είναι ο συνολικός αριθμός των συντονισμών. Με τον όρο «συντονισμοί της φωνής» (speech resonances ή formants) αναφερόμαστε γενικά στα συστήματα συντονιστών (resonators) που σχηματίζονται από τοπικές κοιλότητες της φωνητικής οδού και τα οποία ενισχύουν κάποιες συχνότητες και αποσβένουν κάποιες άλλες.

### 2.2.2 Σύγκριση με το κλασσικό μοντέλο

Στο κλασσικό μοντέλο σε κάθε τέτοιο σύστημα αντιστοιχεί ένα ζευγάρι συζυγών πόλων (συντονιστής δεύτερης τάξης) στη συνάρτηση μεταφοράς της φωνητικής οδού, δηλαδή απόκριση της μορφής :

$$R_{lin}(t) = A \exp^{-\sigma t} \cos(\omega_c t + \theta) \quad (2.4)$$

όπου  $\omega_c$  η συχνότητα του συντονισμού και  $\sigma > 0$  το εύρος ζώνης του, ενώ  $\theta$  είναι η αρχική φάση. Τόσο η συχνότητα όσο και το εύρος ζώνης υποτίθεται ότι παραμένουν σταθερά για μικρά χρονικά διαστήματα, οπότε το σήμα φωνής θεωρείται ότι είναι τοπικά στάσιμο.

Με την νέα προσέγγιση, από την άλλη, τόσο το πλάτος  $A$  όσο και η συχνότητα  $\omega_c$  αντικαθίστανται από αντίστοιχα χρονικά μεταβαλλόμενα, ακόμα και μεταξύ δύο περιόδων δειγματοληψίας, μεγέθη,  $A(t)$  και στιγμιαία συχνότητα  $\omega_i(t)$ .

Υπάρχουν πολλές πειραματικές και θεωρητικές ενδείξεις που στηρίζουν αυτό το νέο μοντέλο.

### 2.2.3 Ενδείξεις

Κατ' αρχάς, ο Teager, [15], με τα πειράματά του έδειξε ότι η ροή του αέρα μέσα στη φωνητική κοιλότητα κατά την παραγωγή φωνής είναι ασταθής και η μάζα του αέρα ταλαντώνεται μεταξύ των τοιχωμάτων, άλλες φορές προσκολλώντας σε αυτά, και άλλες απομακρυνόμενη από αυτά, γεγονός που επηρεάζει τις συχνότητες συντονισμού του συστήματος. Επίσης, κατά την διάρκεια παραγωγής φωνής εύκολα δημιουργούνται στρόβιλοι, όπως διαπίστωσε ο Teager και πάλι, οι οποίοι είναι δυνατόν να εγκλωβίσουν μάζες αέρα και να διαμορφώσουν την ενέργειά του. Ενώ υπάρχει και επιχείρημα από τη γενική θεωρία των ταλαντωτών, σύμφωνα με την

οποία οι αποκρίσεις ταλαντωτών δεύτερης τάξης, με χρονικά μεταβαλλόμενα χαρακτηριστικά, όπως μάζα, σταθερά επαναφοράς ή απόσβεση, περιέχουν AM και FM διαμόρφωση, [26]. Γι' αυτό θεωρείται ότι οι χρονικά μεταβαλλόμενες μάζες αέρα, μπορούν να προκαλέσουν διαμορφώσεις των πεδίων ταχύτητας και πίεσης. Τέλος, με την εφαρμογή του αλγορίθμου υπολογισμού της ενέργειας των Teager-Kaiser σε σήματα φωνής διαπιστώνεται η ύπαρξη παλμών στο σήμα της ενέργειας ακόμα και κατά τη διάρκεια μιας θεμελιώδους περιόδου (pitch period)(Σχήμα 2.1) κάτι που δεν δικαιολογείται αν θεωρήσουμε ότι το σήμα μας προέρχεται από έναν γραμμικό ταλαντωτή δεύτερης τάξης (Σχήμα 2.2). Αντίθετα, όπως φαίνεται και στο Σχήμα 2.3, για το μοντέλο AM-FM, ύπαρξη «ενεργειακών παλμών» είναι απόλυτα φυσική. Σύμφωνα με τους Maragos et al, [26], μάλιστα, στην περίπτωση της φωνής, οι διαμορφώσεις AM και FM πληρούν τις απαραίτητες προϋποθέσεις ώστε να είναι δυνατή η ανάκτηση της πληροφορίας την οποία περιέχουν με τη χρήση ενός απλού και αρκετά αποτελεσματικού αλγορίθμου. Πρόκειται για τον αλγόριθμο *ESA*.

## 2.3 Ο αλγόριθμος διαχωρισμού με υπολογισμό της ενέργειας - (Energy Separation Algorithm)

### 2.3.1 Ο ενεργειακός τελεστής

Ο αλγόριθμος διαχωρισμού με υπολογισμό της ενέργειας, στηρίζεται στο λεγόμενο ενεργειακό τελεστή που προτάθηκε για πρώτη φορά από τον Teager, [16], παρουσιάστηκε συστηματικά από τον Kaiser, [17], και εισήχθη ως εργαλείο για αποδιαμόρφωση από τους Maragos et al, [27]. Πρόκειται για ένα μη γραμμικό τελεστή που για διακριτά σήματα, έχει τη μορφή:

$$\Psi_d[x(n)] \triangleq x^2(n) - x(n-1)x(n+1) \quad (2.5)$$

ενώ επεκτείνεται για συνεχή σήματα παίρνοντας τη μορφή:

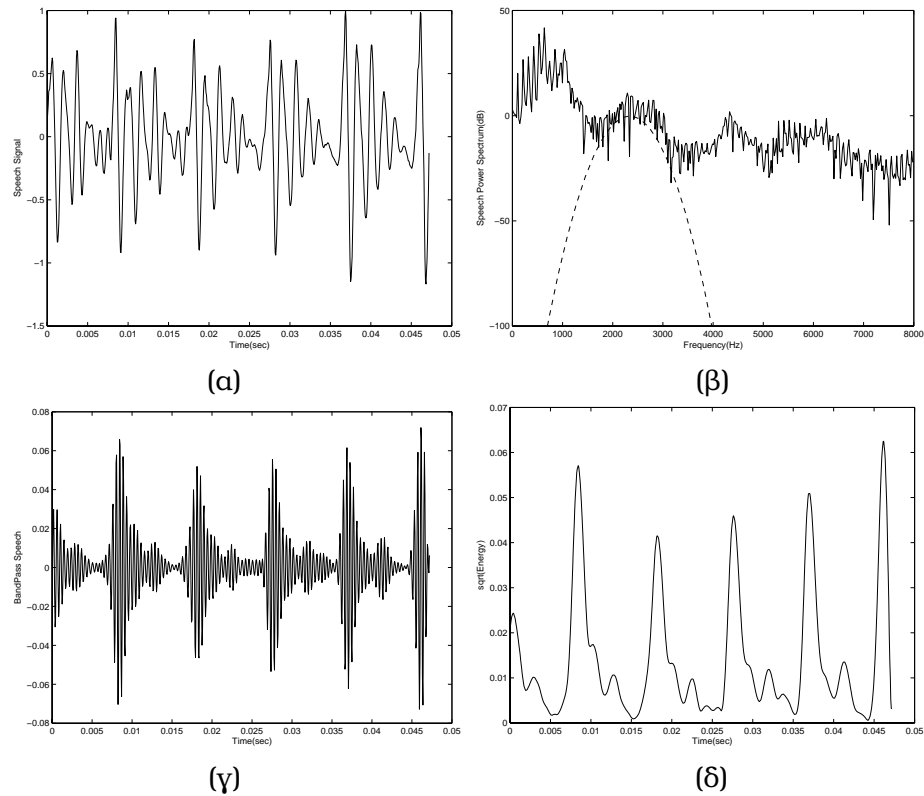
$$\Psi_c[x(t)] \triangleq [\dot{x}(t)]^2 - x(t)\ddot{x}(t) \quad (2.6)$$

Ο συγκεκριμένος τελεστής έχει την ιδιότητα να ανιχνεύει την ενέργεια ενός γραμμικού ταλαντωτή. Από εκεί προκύπτει και η θεώρησή του ως ενεργειακού τελεστή. Πιο συγκεκριμένα, για έναν ελεύθερο γραμμικό ταλαντωτή χωρίς απόσβεση με μάζα  $m$  και σταθερά επαναφοράς  $k$ , η απομάκρυνση από τη θέση ισορροπίας περιγράφεται από τη διαφορική εξίσωση κίνησης:

$$m\ddot{x} + kx = 0 \quad (2.7)$$

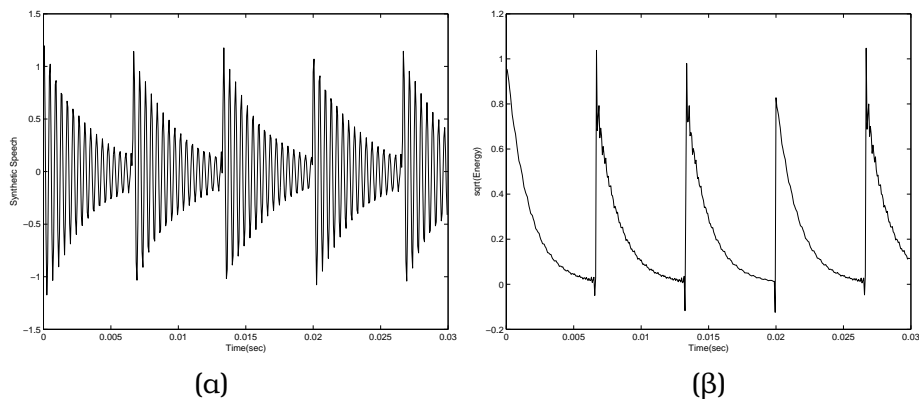
που έχει τη γενική λύση:

$$x(t) = A \cos(\omega_0 t + \theta) \quad (2.8)$$

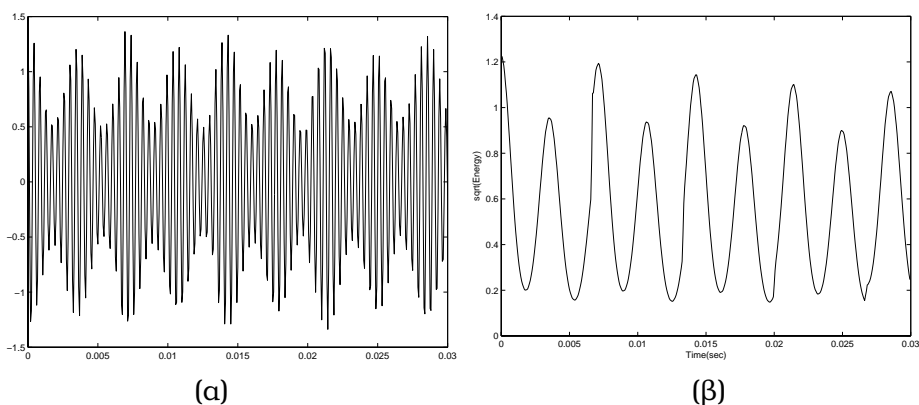


Σχήμα 2.1: (α) Φώνημα /AO/ από τη λέξη « /a/lways » στο πεδίο του χρόνου δειγματοληπτημένο με συχνότητα ίση με  $16kHz$ , (β) Φάσμα ισχύος του φωνήματος (συνεχής γραμμή). Υπερτίθεται το φάσμα ισχύος του φίλτρου Gabor (διακεκομμένη γραμμή) που χρησιμοποιείται για να απομονωθεί μια ζώνη συχνοτήτων του σήματος γύρω από τη συχνότητα  $2310Hz$  ( $f_c = 2310Hz, \alpha = 1500$ ), (γ) Το ζωνοπερατό σήμα στην έξοδο του φίλτρου Gabor, (δ) Η τετραγωνική ρίζα της ενέργειας του ζωνοπερατού σήματος όπως αυτή υπολογίζεται με τη χρήση του ενεργειακού τελεστή Teager-Kaiser.

### 2.3. ΑΛΓΟΡΙΘΜΟΣ ΔΙΑΧΩΡΙΣΜΟΥ ΜΕ ΥΠΟΛΟΓΙΣΜΟ ΤΗΣ ΕΝΕΡΓΕΙΑΣ 21



Σχήμα 2.2: (α) Σήμα προερχόμενο από γραμμικό ταλαντωτή δεύτερης τάξης με συντονισμό στα  $2310Hz$ , εύρος ζώνης ίσο με  $100Hz$  και θεμελιώδη συχνότητα ίση με  $150Hz$  (συχνότητα δειγματοληψίας ίση με  $16kHz$ ), (β) Η τετραγωνική ρίζα της ενέργειας του σήματος, όπως αυτή υπολογίζεται με τη χρήση του ενεργειακού τελεστή Teager-Kaiser



Σχήμα 2.3: (α) Το σήμα  $y(t) = 0.998^{(tF_s) \bmod (F_s/F_0)} (1 + 0.4 \cos(2\pi 280t)) \cos(2\pi 2310t + 0.4 \sin(2\pi 120t))$ , όπου  $F_s = 16000Hz$ ,  $F_0 = 150Hz$  και είναι φανερό πως έχει γίνει προσπάθεια προσομοίωσης του pitch, (β) Η τετραγωνική ρίζα της ενέργειας του σήματος, όπως αυτή υπολογίζεται με τη χρήση του ενεργειακού τελεστή Teager-Kaiser

όπου  $\omega_0 = \sqrt{\frac{k}{m}}$  και η στιγμιαία ενέργειά του είναι σταθερή και ίση με το άθροισμα της δυναμικής και της κινητικής του ενέργειας:

$$E_0 = \frac{m}{2}\dot{x}^2 + \frac{k}{2}x^2 \quad (2.9)$$

δηλαδή είναι ανάλογη τόσο με το τετράγωνο του πλάτους όσο και με το τετράγωνο της συχνότητας της ταλάντωσης. Επιπλέον:

$$\Psi_c[A \cos(\omega_0 t + \theta)] = A^2 \omega_0^2 = \frac{E_0}{m/2} \quad (2.10)$$

Έτσι, ο ενεργειακός τελεστής, εφαρμοζόμενος στην ταλάντωση  $x(t)$  ουσιαστικά δίνει την ενέργεια (ανά μισή μονάδα μάζας) της πηγής που προκαλεί την ταλάντωση. Αυτή η συγκεκριμένη ιδιότητα ενέπνευσε τους ερευνητές που εισήγαγαν τον αλγόριθμο διαχωρισμού με υπολογισμό της ενέργειας, [26].

### 2.3.2 Περιγραφή του ESA

Πρόκειται για μία απλή μέθοδο απόδιαμόρφωσης σημάτων που είναι διαμορφωμένα τόσο κατά πλάτος όσο και κατά συχνότητα. Στην περίπτωση συνεχούς σήματος:

$$x(t) = \alpha(t) \cos(\underbrace{\omega_c t + \omega_m \int_0^t q(\tau) d\tau}_{\varphi(t)} + \theta) \quad (2.11)$$

$$\omega_i(t) \triangleq \frac{d\varphi(t)}{dt} = \omega_c + \omega_m q(t) \quad (2.12)$$

με δεδομένο ότι οι διαμορφώσεις τόσο του πλάτους όσο και της συχνότητας είναι σχετικά μικρές και αργές, ισχύουν οι σχέσεις:

$$\omega_i(t) \approx \sqrt{\frac{\Psi_c[\dot{x}(t)]}{\Psi_c[x(t)]}} \quad (2.13)$$

$$|\alpha(t)| \approx \frac{\Psi_c[x(t)]}{\sqrt{\Psi_c[\dot{x}(t)]}} \quad (2.14)$$

όπου  $\omega_i(t)$  είναι η στιγμιαία συχνότητα και  $\alpha(t)$  το στιγμιαίο πλάτος.  $\Psi[\sim]$  είναι ο ενεργειακός τελεστής, όπως ορίζεται στην 2.6. Σημειώνεται ότι  $\omega_c$  είναι η φέρουσα συχνότητα ενώ  $\omega_m$  είναι η μέγιστη απόκλιση της στιγμιαίας συχνότητας από τη φέρουσα, αφού  $|q(t)| \leq 1$ . Με  $\theta$  συμβολίζεται η αρχική τιμή της φάσης. Πρόκειται για τον συνεχή αλγόριθμο διαχωρισμού με βάση την ενέργεια (CESA).

Ανάλογα, στην περίπτωση διακριτού σήματος:

$$x(n) = \alpha(n) \cos[\underbrace{\Omega_c n + \Omega_m \int_0^n q(m) dm}_{\varphi(n)} + \theta] \quad (2.15)$$

$$\Omega_i(n) \triangleq \frac{d}{dn} \varphi(n) = \Omega_c + \Omega_m q(n) \quad (2.16)$$

## 2.3. ΑΛΓΟΡΙΘΜΟΣ ΔΙΑΧΩΡΙΣΜΟΥ ΜΕ ΥΠΟΛΟΓΙΣΜΟ ΤΗΣ ΕΝΕΡΓΕΙΑΣ 23

όπου  $\Omega_c, \Omega_m$  ανάλογα των  $\omega_c, \omega_m$  ενώ  $\theta \triangleq \varphi(0)$ , ο αλγόριθμος παίρνει την παρακάτω μορφή, αν υποθέσουμε ότι η περίοδος δειγματοληψίας είναι  $T = 1$ :

$$y(n) = x(n) - x(n-1) \quad (2.17)$$

$$\Omega_i(n) \approx \arccos\left(1 - \frac{\Psi[y(n)] + \Psi[y(n+1)]}{4\Psi[x(n)]}\right) \quad (2.18)$$

$$|\alpha(n)| \approx \sqrt{\frac{\Psi[x(n)]}{1 - \left(1 - \frac{\Psi[y(n)] + \Psi[y(n+1)]}{4\Psi[x(n)]}\right)^2}} \quad (2.19)$$

όπου  $\Omega_i(n), \alpha(n)$  είναι η διακριτή στιγμιαία συχνότητα και το διακριτό στιγμιαίο πλάτος αντίστοιχα. Θα πρέπει να σημειωθεί ότι οι συχνότητες  $\Omega_c, \Omega_m, \Omega_i$  θεωρούνται μη αρνητικές και μικρότερες από  $\pi$  ενώ υποτίθεται ότι είναι σε ακτίνια/δείγμα.

### 2.3.3 Παραλλαγές του διακριτού αλγορίθμου

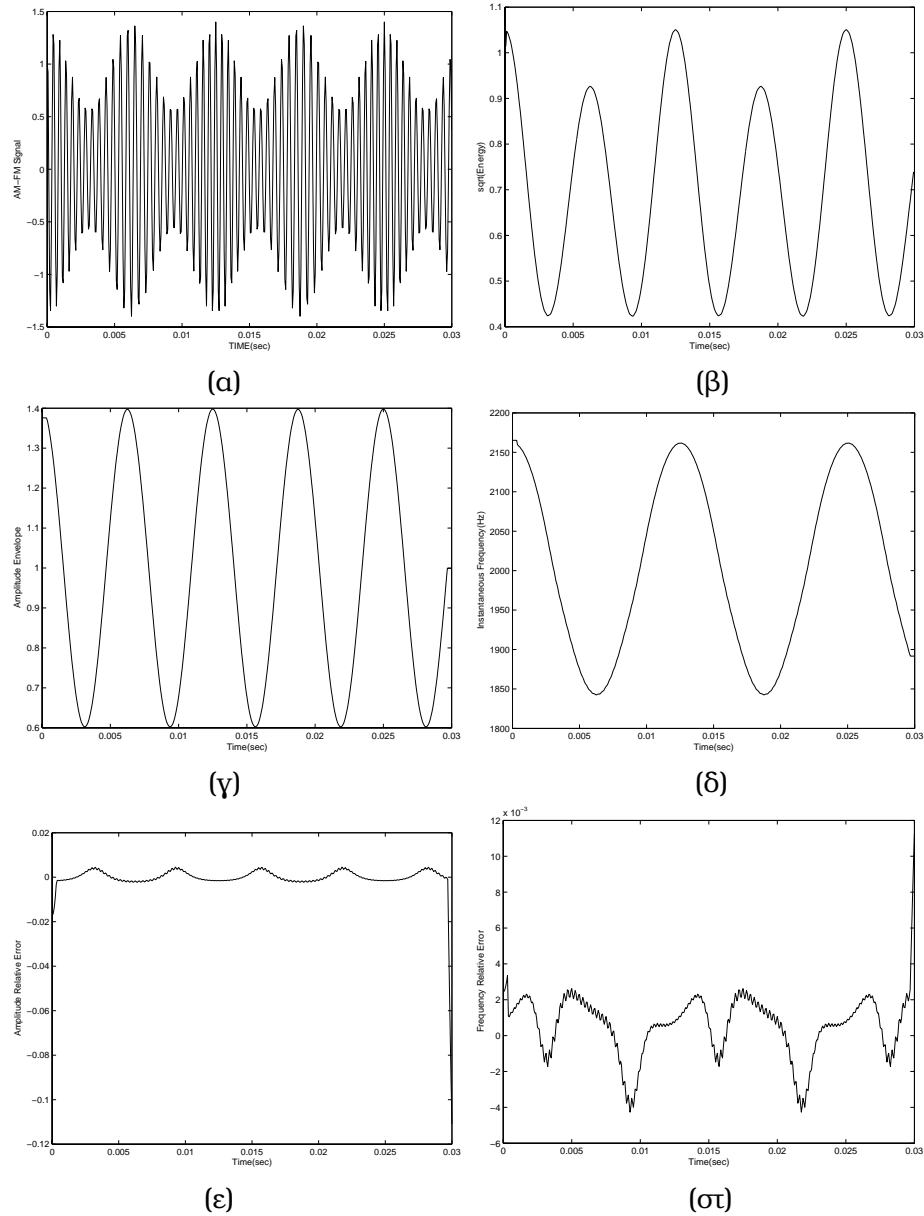
Η μορφή του αλγορίθμου που δίνεται, επωνομάζεται *DESA-1*. Ανάλογα με τον τρόπο με τον οποίο προσεγγίζεται η παράγωγος του σήματος  $x(n)$  υπάρχουν διάφορες μορφές (*DESA-1a, DESA-1, DESA-2*). Στην περίπτωση του *DESA-1*, η παράγωγος προσεγγίζεται μεν από τη διαφορά του τρέχοντος δείγματος από το προηγούμενο (όπως και στον *DESA-1a*) αλλά γίνεται μερική αναίρεση των επιπτώσεων αυτής της ασυμμετρίας στο αποτέλεσμα του ενεργειακού τελεστή με κατάλληλη εφαρμογή του τελευταίου. Από την άλλη, στην παραλλαγή *DESA-2*, η παράγωγος στην τρέχουσα περίοδο δειγματοληψίας προσεγγίζεται από τη διαφορά μεταξύ του επόμενου και του προηγούμενου δείγματος. Αναφέρεται ότι σε σχετικές δοκιμές οι επιδόσεις των τριών αυτών μορφών του αλγορίθμου είναι ανάλογες. Ανεπαίσθητα καλύτερη εμφανίζεται η παραλλαγή *DESA-1*, που είναι και αυτή που χρησιμοποιείται στη συνέχεια.

Για την ακρίβεια, χρησιμοποιείται η παραλλαγή *DESA-1* σε συνδυασμό με ομαλοποιητικό βαθυπερατό φιλτράρισμα της εξόδου του ενεργειακού τελεστή ώστε να απομονωθεί όσο είναι δυνατόν το επιθυμητό μέρος της εξόδου από το σφάλμα, αφού από σχετική μελέτη (βλ. [4]) προκύπτει ότι τα δύο μέρη είναι αρκετά καλά διαχωρισμένα στο χώρο συχνοτήτων. Το φιλτράρισμα που χρησιμοποιείται γίνεται προσπάθεια να μη ζημιώσει την πολύ καλή διακριτική ικανότητα στο χρόνο που έχει ο *ESA* και ισοδυναμεί ουσιαστικά με τρεις φορές εφαρμογή του φίλτρου τριών σημείων  $[0.25 \ 0.5 \ 0.25]$  στην έξοδο του ενεργειακού τελεστή και έχει ως συνέπεια τη σε μεγάλο βαθμό μείωση των σφαλμάτων. Εναλλακτικά, θα μπορούσε βέβαια να εφαρμοστεί και στα σήματα του πλάτους και της συχνότητας που υπολογίζονται τελικά, αλλά τα αποτελέσματα είναι παρόμοια και στις δύο περιπτώσεις.

Ακολουθούν παραδείγματα εφαρμογής του εξομαλυμένου *DESA-1*.

### 2.3.4 Παραδείγματα εφαρμογής του *ESA*

Η γραφική παράσταση 2.4 δείχνει ότι η εφαρμογή του αλγορίθμου σε ένα AM-FM σήμα, στο οποίο τα διαμορφώνοντα σήματα είναι απλά συνημίτονα οδηγεί σε



Σχήμα 2.4: (α) Το σήμα  $y(t) = [1 + 0.4 \cos(2\pi 160t)] \cos(2\pi 1600t + 2 \sin(2\pi 80t + \pi/4))$ , δειγματοληπτημένο με συχνότητα  $F_s = 16000 Hz$ , (β)  $\sqrt{\Psi}[y(n)]$ , (γ) Περιβάλλουσα του σήματος όπως υπολογίζεται με τον αλγόριθμο DESA-1, (δ) Στιγμιαία συχνότητα του σήματος όπως προκύπτει από τον DESA-1, (ε) Σχετικό σφάλμα υπολογισμού του πλάτους, (στ) Σχετικό σφάλμα υπολογισμού της στιγμιαίας συχνότητας



## 2.3. ΑΛΓΟΡΙΘΜΟΣ ΔΙΑΧΩΡΙΣΜΟΥ ΜΕ ΥΠΟΛΟΓΙΣΜΟ ΤΗΣ ΕΝΕΡΓΕΙΑΣ 25

ικανοποιητικό υπολογισμό του στιγμιαίου πλάτους και της στιγμιαίας συχνότητας με σχετικά μικρά λάθη. Αυτό συμβαίνει παρά το υψηλό ποσοστό διαμόρφωσης του σήματος, 40% AM και 10% FM. Στην περίπτωση μικρότερων διαμορφώσεων, η απόδοση του αλγορίθμου είναι ακόμα καλύτερη. Ο αλγόριθμος δουλεύει με επιτυχία και στην περίπτωση που η συχνότητα μεταβάλλεται γραμμικά. Στη γραφική παράσταση 2.5 φαίνονται τα αποτελέσματα της εφαρμογής του αλγορίθμου σε ένα AM-FM σήμα με το χαρακτηριστικό της γραμμικά μεταβαλλόμενης συχνότητας. Και σε αυτή την περίπτωση τα εκτιμώμενα σήματα του πλάτους και της συχνότητας είναι πολύ κοντά στα πραγματικά. Μεγάλα λάθη εμφανίζονται σε περιπτώσεις ύπαρξης θορύβου ( $SNR = 10dB$  για το σήμα που εμφανίζεται στη γραφική παράσταση 2.4). Ο Βονίκ και οι συνεργάτες του, [5], εντόπισαν και μελέτησαν την ευασθησία του ενεργειακού τελεστή στο θόρυβο και πρότειναν μια επέκτασή του *ESA* ώστε να αντιμετωπιστούν τα σχετικά προβλήματα.

### 2.3.5 Επέκταση του αλγορίθμου διαχωρισμού με υπολογισμό της ενέργειας σε πολλές ζώνες συχνότητων

#### Φιλτράρισμα σε μία ζώνη συχνότητων

Η πρόταση αυτή στηρίζεται στην ιδέα κατ' αρχάς και στα αναλυτικά συμπεράσματα σχετικής ανάλυσης στη συνέχεια, ότι το φιλτράρισμα του θορυβώδους AM-FM σήματος στη ζώνη συχνότητων που περιέχει τη χρήσιμη πληροφορία, δηλαδή γύρω από τη φέρουσα συχνότητα, οδηγεί σε μείωση των επιδράσεων του θορύβου, οπότε κάνει δυνατή την ικανοποιητική αποδιαμόρφωση του σήματος. Πράγματι, τα αποτελέσματα μετά την εφαρμογή φιλτραρίσματος (Σχήμα 2.7) φαίνεται ότι βελτιώνονται αρκετά σε σχέση με αυτά που παρουσιάζονται στο Σχήμα 2.6.

#### Φίλτρο Gabor

Το φιλτράρισμα στην περίπτωση αυτή έγινε με τη χρήση Gabor φίλτρου με κρουστική απόκριση (Σχήμα 2.8):

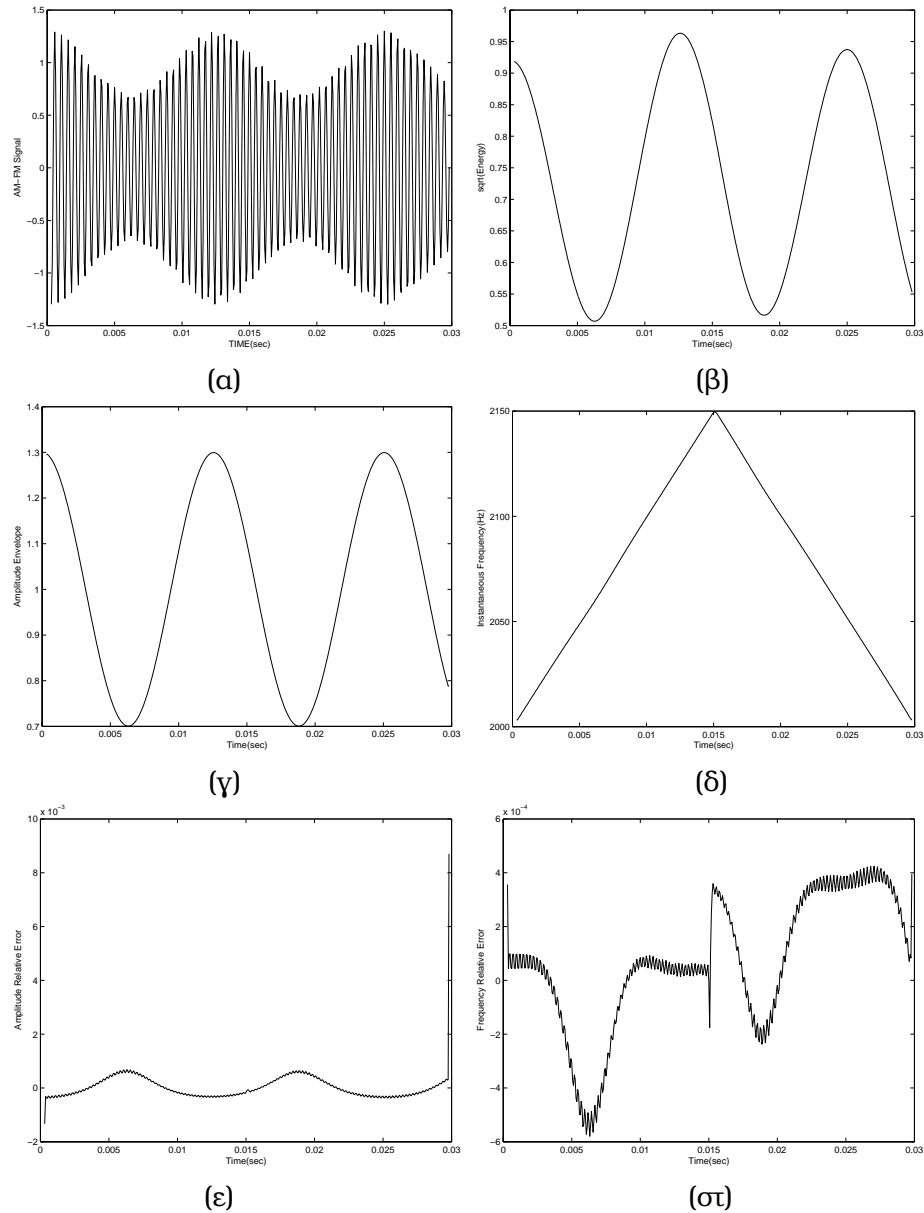
$$h(t) = \exp(-\alpha^2 t^2) \cos(\omega_c t) \quad (2.20)$$

και απόκριση συχνότητας (Σχήμα 2.8):

$$H(\omega) = \frac{\sqrt{\pi}}{2\alpha} \left( \exp\left[-\frac{(\omega - \omega_c)^2}{4\alpha^2}\right] + \exp\left[-\frac{(\omega + \omega_c)^2}{4\alpha^2}\right] \right) \quad (2.21)$$

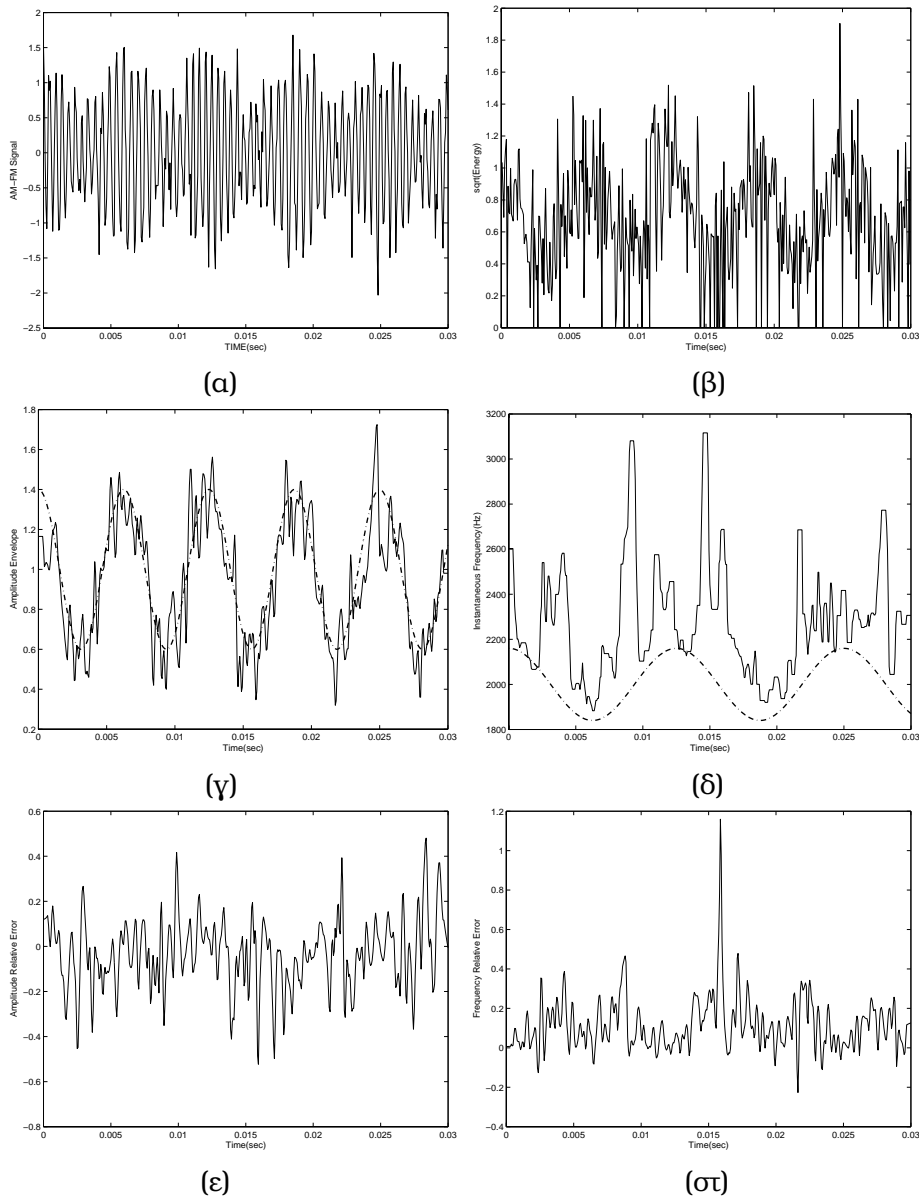
όπου  $\omega_c$  είναι η επιλεγόμενη κεντρική συχνότητα ενώ η παράμετρος  $\alpha$  ελέγχει το εύρος της απόκρισης συχνότητας αφού το αποδοτικό rms εύρος ζώνης του φίλτρου σύμφωνα με τον Gabor είναι ίσο με  $\alpha/\sqrt{(2\pi)}$ . Η διακριτή μορφή της κρουστικής απόκρισης του φίλτρου προκύπτει ύστερα από δειγματοληψία της Εξ. 2.20 και αποκοπή κάποιων δειγμάτων και είναι:

$$h(n) = \exp(-b^2 n^2) \cos(\Omega_c n) \quad (2.22)$$

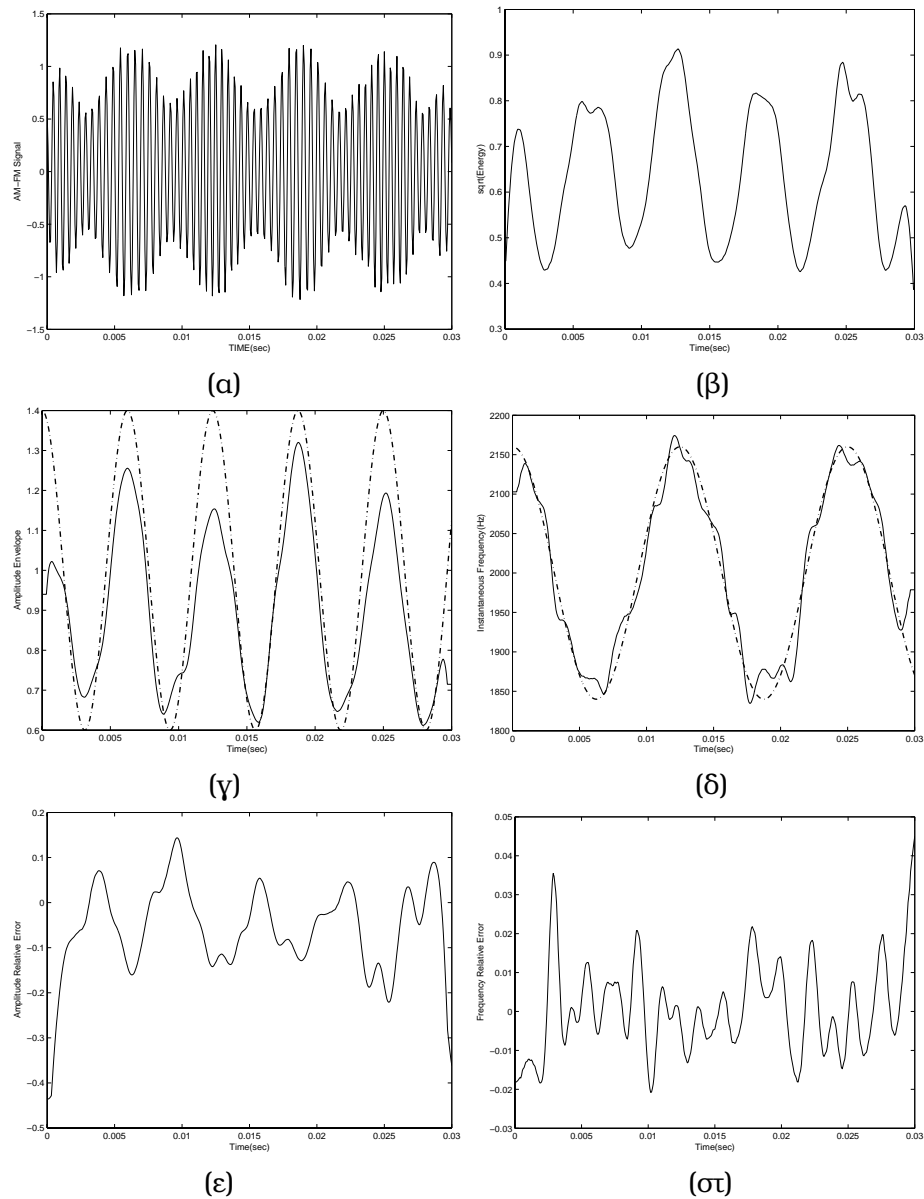


Σχήμα 2.5: (α) Το σήμα  $y(t) = [1 + 0.3 \cos(2\pi 80t)] \cos(\int_0^t IF(\tau)d\tau)$  όπου  $IF(t) = 2000 + 10^3t, t = 0 \dots 15msec$  ενώ  $IF(t) = 2015 - 10^3t$  στη συνέχεια, μέχρι τα  $30msec$ . Το σήμα είναι δειγματοληπτημένο με συχνότητα  $F_s = 16000Hz$ , (β)  $\sqrt{\Psi}[y(n)]$ , (γ) Περιβάλλουσα του σήματος όπως υπολογίζεται με τον αλγόριθμο DESA-1, (δ) Στιγμιαία συχνότητα του σήματος όπως προκύπτει από τον DESA-1, (ε) Σχετικό σφάλμα υπολογισμού του πλάτους, (στ) Σχετικό σφάλμα υπολογισμού της στιγμιαίας συχνότητας

### 2.3. ΑΛΓΟΡΙΘΜΟΣ ΔΙΑΧΩΡΙΣΜΟΥ ΜΕ ΥΠΟΛΟΓΙΣΜΟ ΤΗΣ ΕΝΕΡΓΕΙΑΣ 27

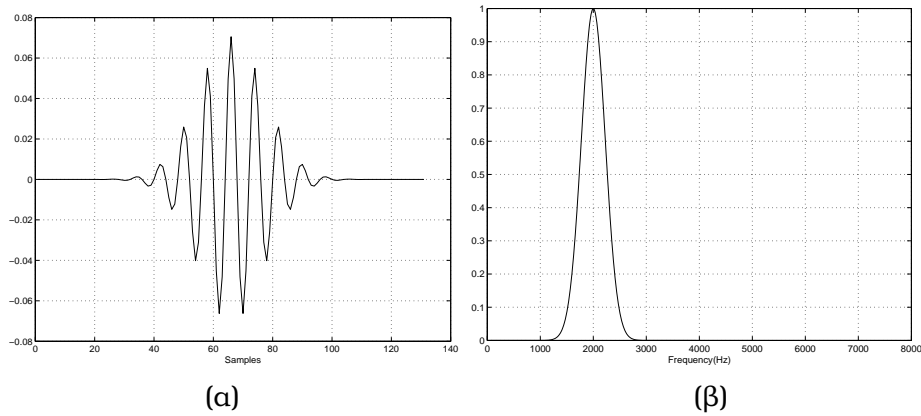


Σχήμα 2.6: (α) Το σήμα  $y(t) = [1 + 0.4 \cos(2\pi 160t)] \cos(2\pi 1600t + 4 \sin(2\pi 80t + \pi/4))$  με πρόσθετο κανονικά κατανομημένο θόρυβο ώστε ο σηματοθορυβικός λόγος να είναι περίπου ίσος με  $10dB$ . Η συχνότητα δειγματοληψίας είναι  $F_s = 16000Hz$ , (β)  $\sqrt{\Psi}[y(n)]$ , (γ) Περιβάλλουσα του σήματος όπως υπολογίζεται με τον αλγόριθμο DESA-1. Με διακεκομμένη γραμμή φαίνεται η πραγματική περιβάλλουσα του σήματος, (δ) Στιγμιαία συχνότητα του σήματος όπως προκύπτει από τον DESA-1. Σημειώνεται ότι έχει εξομαλυνθεί με τη χρήση ενός median φίλτρου 11 σημείων, ενώ με διακεκομμένη γραμμή φαίνεται η πραγματική στιγμιαία συχνότητα του σήματος, (ε) Σχετικό σφάλμα υπολογισμού του πλάτους, (στ) Σχετικό σφάλμα υπολογισμού της στιγμιαίας συχνότητας



Σχήμα 2.7: (α) Απόκριση του φίλτρου Gabor με κεντρική συχνότητα  $f_c = 1600\text{Hz}$  στο σήμα του Σχήματος 2.6. Η συχνότητα δειγματοληψίας είναι  $F_s = 16000\text{Hz}$ , (β)  $\sqrt{\Psi}[y(n)]$ , (γ) Περιβάλλουσα του σήματος όπως υπολογίζεται με τον αλγόριθμο DESA-1. Με διακεκομμένη γραμμή φαίνεται η πραγματική περιβάλλουσα του σήματος, (δ) Στιγμιαία συχνότητα του σήματος όπως προκύπτει από τον DESA-1, ενώ με διακεκομμένη γραμμή φαίνεται η πραγματική στιγμιαία συχνότητα του σήματος. (ε) Σχετικό σφάλμα υπολογισμού του πλάτους, (στ) Σχετικό σφάλμα υπολογισμού της στιγμιαίας συχνότητας

## 2.3. ΑΛΓΟΡΙΘΜΟΣ ΔΙΑΧΩΡΙΣΜΟΥ ΜΕ ΥΠΟΛΟΓΙΣΜΟ ΤΗΣ ΕΝΕΡΓΕΙΑΣ 29



Σχήμα 2.8: (α) Κρουστική απόκριση φίλτρου Gabor με  $f_c = 2000Hz$  και  $\alpha = 1000$ , (β) Απόκριση συχνότητας του φίλτρου Gabor

όπου  $-N \leq n \leq N$ ,  $b = \alpha T_s$  και  $\Omega_c = 2\pi f_c T_s$  όπου  $T_s$  η συχνότητα δειγματοληψίας ενώ ο ακέραιος  $N$  επιλέγεται έτσι ώστε να μηδενίσει ουσιαστικά την  $h(n)$ , να ισχύει για παράδειγμα  $\exp(-b^2 N^2) = 10^{-5}$ . Αυτό είναι το φίλτρο που χρησιμοποιείται και στη συνέχεια, συνελισσόμενο ουσιαστικά με το σήμα υπό ανάλυση. Η επιλογή του έγινε για δύο κυρίως λόγους:

1. Είναι βέλτιστο με κριτήριο την αρχή της αβεβαιότητας αφού το γινόμενο του εύρους χρόνου επί το εύρος συχνοτήτων παίρνει ελάχιστη τιμή. Έτσι, επιτυγχάνεται βελτιωμένη διακριτική ικανότητα τόσο στο χρόνο όσο και στη συχνότητα.
2. Λόγω του κανονικού (Gaussian) σχήματος της απόκρισης συχνότητας αποφεύγονται οι δευτερεύοντες λοβοί που θα μπορούσαν να αλλοιώσουν το αποτέλεσμα του ενεργειακού τελεστή.

### Φιλτράρισμα σε πολλές ζώνες συχνοτήτων

Το πρόβλημα βέβαια που παραμένει είναι η κατάλληλη επιλογή της κεντρικής συχνότητας  $\omega_c$  ώστε να απομονωθεί η συνιστώσα που περιέχει τη χρήσιμη πληροφορία. Η επιλογή αυτή είναι πολύ σημαντική, ιδιαίτερα στην περίπτωση που η στιγμιαία συχνότητα της συνιστώσας αυτής μπορεί να βρίσκεται σε οποιαδήποτε ζώνη του πεδίου συχνοτήτων. Θα μπορούσε να αντιμετωπιστεί είτε με υπολογισμό της κεντρικής της συχνότητας με κατάλληλη μέθοδο ανίχνευσης είτε με τη χρήση συστοιχίας φίλτρων τα οποία θα καλύπτουν πυκνά όλο το πεδίο συχνοτήτων. Στη συνέχεια ακολουθείται στην ουσία η πρώτη κατεύθυνση αλλά αξίζει να σημειωθούν κάποιες λεπτομέρειες που αφορούν στην προσέγγιση με τη συστοιχία φίλτρων που είναι αυτή που προτείνεται από τους Bovik, Maragos & Quatieri [5]. Συγκεκριμένα,

προτείνεται το φιλτράρισμα του σήματος με κατάλληλα σχεδιασμένη ακολουθία φίλτρων  $\{g_i\}$ ,  $i = 1 \dots M$  και στη συνέχεια η εφαρμογή του ενεργειακού τελεστή στην έξοδο κάθε φίλτρου  $f_i$ ,  $i = 1 \dots M$ . Η έξοδος με τη μεγαλύτερη κανονικοποιημένη ενέργεια:

$$\Psi^*(t) = \max_{1 \leq m \leq M} \frac{\Psi[f_m(t)]}{|G_m(\omega_m)|^2} \quad (2.23)$$

θεωρείται ότι είναι αυτή που περιέχει τη χρήσιμη πληροφορία και σε αυτήν εφαρμόζεται στο τέλος ο αλγόριθμος διαχωρισμού με υπολογισμό της ενέργειας.

Εναλλακτικά, όπως προαναφέρθηκε, αντί της συστοιχίας φίλτρων, είναι δυνατή η χρησιμοποίηση μεθόδου ανίχνευσης και παρακολούθησης της φέρουσας συχνότητας του καθαρού σήματος. Το πρόβλημα τότε μπορεί να θεωρηθεί ως πρόβλημα εκτίμησης.

## 2.4 Εκτίμηση και φίλτρο Kalman

### 2.4.1 Το πρόβλημα της εκτίμησης

Το γενικότερο πρόβλημα της εκτίμησης συνίσταται στον όσο το δυνατόν καλύτερο προσδιορισμό της τιμής ενός μεγέθους δεδομένων κάποιων μετρήσεων διαφόρων σχετικών μεγεθών. Έτσι, ο προσδιορισμός ενός σήματος  $x(t)$  βυθισμένου σε θόρυβο  $n(t)$  από τις τιμές του σήματος  $y(t) = x(t) + n(t)$  είναι ένα πρόβλημα εκτίμησης. Ανάλογα με το κριτήριο βελτιστότητας που χρησιμοποιείται μπορούν να υπάρξουν διάφορες μέθοδοι εκτίμησης [7], [29]. Η εκτίμηση κατά Bayes, η εκτίμηση μέγιστης πιθανότητας εκ των υστέρων (MAP) αλλά και η εκτίμηση μέγιστης πιθανοφάνειας (MLE) είναι κάποιες από αυτές. Η πρώτη είναι βέλτιστη με την έννοια ότι ελαχιστοποιεί κάποιο κόστος που ορίζεται για το πρόβλημα, η δεύτερη μεγιστοποιεί την υπό συνθήκη πιθανότητα να έχει το μέγεθος υπό εκτίμηση κάποια τιμή, δεδομένων των παρατηρήσεων ενός σχετικού μεγέθους, ενώ η τελευταία δεν λαμβάνει υπόψη την εκ των προτέρων πυκνότητα κατανομής πιθανότητας του εκτιμώμενου μεγέθους και μεγιστοποιεί την συνάρτηση πιθανοφάνειας των δεδομένων.

Υπάρχουν κάποιες καταστάσεις βέβαια όπου, λόγω δυσκολιών, το πρόβλημα αντιμετωπίζεται με την γραμμική μέθοδο ελαχιστοποίησης της μεταβλητότητας του λάθους (LMSEE). Αυτή υπαγορεύει την προσέγγιση του μεγέθους υπό εκτίμηση από μία γραμμική συνάρτηση των δεδομένων και στη συνέχεια ελαχιστοποίηση της μεταβλητότητας του λάθους της εκτίμησης. Από αυτή την οπτική γωνία μπορεί να θεωρηθεί και το φίλτρο Kalman (βλ. [23]).

### 2.4.2 Το φίλτρο Kalman

Το φίλτρο Kalman ([24], [3], [23], [13]) είναι ουσιαστικά μια ακολουθιακή μορφή της μεθόδου ελαχιστοποίησης της μεταβλητότητας του λάθους, [29]. Συγκεκριμένα, το πρόβλημα που επιλύει είναι η εκτίμηση της κατάστασης  $x \in R^n$  μιας διαδικασίας

διακριτού χρόνου που περιγράφεται από τη γραμμική στοχαστική διαφορική εξίσωση:

$$x_k = Ax_{k-1} + Bu_k + Gw_{k-1} \quad (2.24)$$

με δεδομένο-μέτρηση  $y_k \in R^m$  που σχετίζεται με την κατάσταση ως εξής:

$$y_k = Cx_k + Hv_k \quad (2.25)$$

Οι τυχαίες μεταβλητές  $w_k$  και  $v_k$  αντιπροσωπεύουν τον θόρυβο που υπεισέρχεται στη διαδικασία και στη μέτρηση αντίστοιχα. Υποτίθεται ότι είναι ανεξάρτητες μεταξύ τους, λευκές και με κανονικές (Gaussian) κατανομές πιθανότητας:

$$p(w) \sim N(0, Q^w) \quad (2.26)$$

$$p(v) \sim N(0, Q^v) \quad (2.27)$$

όπου  $Q^w$  και  $Q^v$  οι συμμεταβλητότητες του θορύβου της διαδικασίας και της μέτρησης αντίστοιχα, ενώ οι μέσες τιμές τους είναι μηδενικές.

Σημειώνεται πως οι πίνακες  $A$ ,  $B$ ,  $C$ ,  $G$  και  $H$  μπορεί να είναι και χρονικά μεταβαλλόμενοι.

Το φίλτρο *Kalman* εκτιμά την κατάσταση του συστήματος χρησιμοποιώντας μια μορφή ελέγχου με ανατροφοδότηση. Εκτιμάται η κατάσταση του συστήματος από την εξίσωση κατάστασης και στη συνέχεια υπάρχει ανατροφοδότηση με τη μορφή θορυβωδών μετρήσεων. Για αυτό, οι εξισώσεις θα μπορούσαν να χωριστούν σε εξισώσεις χρονικής ενημέρωσης (*time update*) και σε εξισώσεις ενημέρωσης από τις μετρήσεις (*measurement update*). Οι πρώτες είναι υπεύθυνες για την προβολή προς τα μπρος (στο χρόνο) της τρέχουσας κατάστασης και της συμμεταβλητότητας του λάθους ώστε να αποκτηθούν οι εκ των προτέρων (*a priori*) εκτιμήσεις για την επόμενη χρονική στιγμή. Οι εξισώσεις ενημέρωσης από τις μετρήσεις είναι υπεύθυνες για την ανατροφοδότηση - δηλαδή για την ενσωμάτωση της πληροφορίας που δίνει μια νέα μέτρηση στην εκ των προτέρων εκτίμηση ώστε να αποκτηθεί μία βελτιωμένη εκ των υστέρων (*a posteriori*) εκτίμηση. Οι εξισώσεις του φίλτρου παρουσιάζονται στη συνέχεια:

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_k \quad (2.28)$$

$$P_{x_k}^- = AP_{x_{k-1}}A^T + GQ^wG^T \quad (2.29)$$

όπου  $P_{x_k}$  είναι η συμμεταβλητότητα της εκτίμησης της κατάστασης,  $\hat{x}_k$ :

$$P_{x_k} = \mathbb{E}[(\hat{x}_k - x_k)(\hat{x}_k - x_k)^T] \quad (2.30)$$

ενώ  $P_{x_k}^-$  είναι η συμμεταβλητότητα της εκ των προτέρων εκτίμησης της κατάστασης,  $\hat{x}_k^-$ :

$$P_{x_k}^- = \mathbb{E}[(\hat{x}_k^- - x_k)(\hat{x}_k^- - x_k)^T] \quad (2.31)$$

Με τις εξισώσεις αυτές λαμβάνεται εκ των προτέρων εκτίμηση της κατάστασης ( $\hat{x}_k^-$ ) και υπολογίζεται η συμμεταβλητότητά της ( $P_{x_k}^-$ ) για τη χρονική στιγμή  $k$  από τις εκτιμήσεις για τη χρονική στιγμή  $k-1$ . Στη συνέχεια, για διευκόλυνση, εισάγεται το

λεγόμενο διάνυσμα των ανανεώσεων (*innovations vector*) που σχετίζεται με την νέα μέτρηση  $y_k$  και είναι:

$$i_k = y_k - C\hat{x}_k^- = C(x_k - \hat{x}_k^-) + Hv_k \quad (2.32)$$

με συμμεταβλητότητα:

$$P_{i_k} = CP_{x_k}^- C^T + HQ^v H^T \quad (2.33)$$

ενώ η συμμεταβλητότητα των  $i_k, x_k$  είναι:

$$P_{x_k i_k} = P_{x_k}^- C^T \quad (2.34)$$

οπότε οι εξισώσεις μέτρησης παίρνουν τη μορφή:

$$K_k = P_{x_k i_k} P_{i_k}^{-1} = P_{x_k}^- C^T (CP_{x_k}^- C^T + HQ^v H^T)^{-1} \quad (2.35)$$

$$\hat{x}_k = \hat{x}_k^- + K_k i_k = \hat{x}_k^- + K_k (y_k - C\hat{x}_k^-) \quad (2.36)$$

$$P_{x_k} = (I - K_k C) P_{x_k}^- \quad (2.37)$$

Αρχικά υπολογίζεται ο πίνακας κερδών Kalman  $K_k$  και ακολουθεί χρήση της μέτρησης  $y_k$  για τη λήψη μιας εκ των υστέρων εκτίμησης της κατάστασης  $\hat{x}_k$  ενώ στο τέλος λαμβάνεται και η συμμεταβλητότητα  $P_{x_k}$  της εκ των υστέρων εκτίμησης της κατάστασης. Στη συνέχεια η διαδικασία επαναλαμβάνεται για την επόμενη χρονική στιγμή οπότε οι παλαιότερες εκ των υστέρων εκτιμήσεις χρησιμοποιούνται για τη λήψη νέων εκ των προτέρων εκτιμήσεων κ.ο.κ., η εφαρμογή των εξισώσεων συνεχίζεται αναδρομικά.

Το φίλτρο Kalman υπό ορισμένες προϋποθέσεις είναι ένας βέλτιστος αναδρομικός αλγόριθμος επεξεργασίας σημάτων. Η βελτιστότητα είναι υπό την έννοια ότι επιτυγχάνεται η διάδοση, στο χρόνο και για όλο το πλήθος των μετρήσεων, της υπό συνθήκη πιθανότητας της κατάστασης, με δεδομένες τις μετρήσεις, οπότε είναι δυνατή η εκτίμηση της κατάστασης εκ των υστέρων. Απαιτείται βέβαια οι διαδικασίες θορύβου που υπεισέρχονται τόσο στην εξίσωση παρατήρησης, όσο και στην εξίσωση μέτρησης να είναι λευκές και Gaussian, ενώ το σύστημα του οποίου η κατάσταση είναι υπό εκτίμηση πρέπει να μπορεί να περιγραφεί από ένα γραμμικό μοντέλο.

Για μη γραμμικά μοντέλα χρησιμοποιείται το εκτεταμένο φίλτρο Kalman (*Extended Kalman Filter*), [23].

### 2.4.3 Το εκτεταμένο φίλτρο *Kalman-EKF*

#### Περιγραφή του αλγορίθμου

Το εκτεταμένο φίλτρο *Kalman* είναι ένα φίλτρο Kalman που γραμμικοποιεί το μοντέλο γύρω από την τρέχουσα εκτίμηση της κατάστασης. Σε πολλές πρακτικές εφαρμογές η διαδικασία που πρέπει να εκτιμηθεί μπορεί να περιγράφεται από μία μη γραμμική στοχαστική διαφορική εξίσωση της μορφής:

$$x_k = f(x_{k-1}, u_{k-1}, w_{k-1}) \quad (2.38)$$



ή η εξίσωση μέτρησης να είναι και αυτή μη γραμμική, της μορφής:

$$y_k = h(x_k, v_k) \quad (2.39)$$

όπου  $u_k$  είναι η είσοδος του συστήματος ενώ  $w_k$  και  $v_k$  είναι και πάλι τυχαίες μεταβλητές που αντιπροσωπεύουν το θόρυβο της διαδικασίας και της μέτρησης αντίστοιχα (υποτίθεται ότι  $w_k \sim N(0, Q^w)$  και  $v_k \sim N(0, Q^v)$ ). Στις εφαρμογές αυτές, το φίλτρο *Kalman* δεν μπορεί να δώσει βέλτιστη εκτίμηση αλλά η επέκτασή του (*EKF*) μέσω γραμμικοποίησης με τη χρήση Ιακωβιανών πινάκων, προσεγγίζει σε κάποιο βαθμό τη βελτιστότητα.

Η αρχικοποίηση του φίλτρου γίνεται ως εξής :

$$\hat{x}_0 = \mathbb{E}[x_0] \quad (2.40)$$

$$P_{x_0} = \mathbb{E}[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T] \quad (2.41)$$

όπου  $P_{x_0}$  είναι η συμμεταβλητότητα της αρχικής εκτίμησης της κατάστασης,  $\hat{x}_0$ . Στη συνέχεια οι εξισώσεις κατάστασης γραμμικοποιούνται ως:

$$x_k \approx A_{k-1}x_{k-1} + B_{k-1}u_{k-1} + G_{k-1}w_{k-1}, \quad \text{όπου} \quad (2.42)$$

$$A_{k-1} \triangleq \left. \frac{\partial f(x, u_{k-1}, 0)}{\partial x} \right|_{\hat{x}_{k-1}}, \quad B_{k-1} \triangleq \left. \frac{\partial f(x_{k-1}^-, u, 0)}{\partial u} \right|_{u_{k-1}} \quad (2.43)$$

$$G_{k-1} \triangleq \left. \frac{\partial f(x_{k-1}^-, u_{k-1}, w)}{\partial w} \right|_{w_k=0} \quad (2.44)$$

και οι εξισώσεις χρονικής ενημέρωσης είναι:

$$\hat{x}_k^- = f(\hat{x}_{k-1}, u_{k-1}, 0) \quad (2.45)$$

$$P_{x_k}^- = A_{k-1}P_{x_{k-1}}A_{k-1}^T + G_{k-1}Q^wG_{k-1}^T \quad (2.46)$$

όπου με  $A^T$  συμβολίζεται γενικά ο ανάστροφος πίνακας του πίνακα  $A$ . Με  $P_{x_k}$  συμβολίζεται η συμμεταβλητότητα της εκτίμησης  $\hat{x}_k$  ενώ με  $P_{x_k}^-$  η συμμεταβλητότητα της εκ των προτέρων εκτίμησης,  $\hat{x}_k^-$ . Στην περίπτωση του εκτεταμένου φίλτρου είναι δυνατή απλά η εκτίμησή των εν λόγω μεγεθών και όχι ο ακριβής υπολογισμός τους.

Στη συνέχεια γραμμικοποιείται η εξίσωση μέτρησης ως  $y_k \approx C_k x_k + H_k v_k$ , όπου:

$$C_k \triangleq \left. \frac{\partial h(x, 0)}{\partial x} \right|_{\hat{x}_k^-}, \quad H_k \triangleq \left. \frac{\partial h(\hat{x}_k^-, v)}{\partial v} \right|_{v_k=0} \quad (2.47)$$

Για διευκόλυνση, ορίζεται και πάλι το διάνυσμα των ανανεώσεων:

$$i_k = y_k - h(\hat{x}_k^-, 0) \quad (2.48)$$

με συμμεταβλητότητα που εκτιμάται ως:

$$P_{i_k} = C_k P_{x_k}^- C_k^T + H_k Q^v H_k^T \quad (2.49)$$

ενώ η συμμεταβλητότητα των  $i_k$  και  $x_k$  εκτιμάται ότι είναι:

$$P_{x_k i_k} = P_{x_k}^- C_k^T \quad (2.50)$$

Τελικά εφαρμόζονται οι εξισώσεις ενημέρωσης από τις μετρήσεις, με βάση τις σχέσεις που δόθηκαν και τη γραμμικοποίηση:

$$K_k = P_{x_k, i_k} P_{i_k}^{-1} = P_{x_k}^- C_k^T (C_k P_{x_k}^- C_k^T + H_k Q^v H_k^T)^{-1} \quad (2.51)$$

$$\hat{x}^k = \hat{x}_k^- + K_k (y_k - h(\hat{x}_k^-, 0)) \quad (2.52)$$

$$P_{x_k} = (I - K_k C_k) P_{x_k}^- \quad (2.53)$$

Η ιδέα στην οποία στηρίζεται η επέκταση του φίλτρου Kalman είναι ιδιαίτερα απλή και η εφαρμογή της σε διάφορα πρακτικά προβλήματα έχει αρκετά ικανοποιητικά αποτελέσματα.

### Εφαρμογές του Εκτεταμένου φίλτρου Kalman

Ο κλασικός χώρος εφαρμογών του εκτεταμένου φίλτρου Kalman είναι ο χώρος του αυτομάτου ελέγχου αλλά έχει γενικότερα χρησιμοποιηθεί και σε ανίχνευση συχνότητας σημάτων αλλά και διαφόρων άλλων χαρακτηριστικών τους, π.χ. [10], [19]. Το 1994 παρουσιάστηκε από τους Niranjani και Cox, [22], μία προσπάθεια αναδρομικής παρακολούθησης των συντονισμών της φωνής με τη χρήση του εκτεταμένου Kalman φίλτρου με καλά αποτελέσματα σε κάποιες περιπτώσεις ενώ και ο αλγόριθμος για ανάλυση φωνής που θα παρουσιαστεί στο επόμενο κεφάλαιο χρησιμοποιεί το ίδιο φίλτρο.

Η αρκετά διαδεδομένη χρήση του φίλτρου έχει αναδείξει και τις σημαντικότερες ατέλειές του.

### Μειονεκτήματα του Εκτεταμένου φίλτρου Kalman

Το εκτεταμένο φίλτρο Kalman έχει κάποιες εγγενείς ατέλειες που κάνουν την εφαρμογή του προβληματική σε κάποιες περιπτώσεις, [32]:

1. Η γραμμικοποίηση μπορεί να προκαλέσει ιδιαίτερα ασταθή φίλτρα αν δεν ισχύουν υποθέσεις τοπικής γραμμικότητας.
2. Η εξαγωγή των Ιακωβιανών πινάκων δεν είναι απλή στις περισσότερες των περιπτώσεων και συχνά προκαλεί πολλές δυσκολίες κατά την υλοποίηση.

Επιπλέον, δεν είναι δυνατό να θεωρήσει κανείς ότι οι διάφορες τυχαίες μεταβλητές, μετά το μη γραμμικό μετασχηματισμό συνεχίζουν να ακολουθούν κανονική κατανομή, η οποία θα μπορούσε να προσεγγιστεί από τη μέση τιμή και τη μεταβλητότητά της, αλλά αντίθετα υπάρχουν περιπτώσεις που αυτή η προσέγγιση μπορεί να οδηγήσει σε σημαντικά λάθη. Οπότε, το EKF, αντίθετα με την απλή μορφή του φίλτρου που είναι βέλτιστη, είναι απλά ένας εκτιμητής που προσπαθεί να προσεγγίσει τη βελτιστότητα μέσω γραμμικοποίησης.

Ενδιαφέρουσα παραλλαγή του φίλτρου Kalman για μη γραμμικά συστήματα, ως εναλλακτική του EKF, είναι το λεγόμενο φίλτρο Kalman χωρίς καθοδήγηση (*Unscented Kalman Filter*), το οποίο αντιμετωπίζει πιο αποτελεσματικά την απουσία γραμμικότητας.

#### 2.4.4 Το χωρίς καθοδήγηση φίλτρο Kalman-(*Unscented Kalman Filter*)

Η συγκεκριμένη παραλλαγή του φίλτρου Kalman, παρουσιάστηκε από τους Julier & Uhlmann, [32], [31], για πρώτη φορά και στη συνέχεια μελετήθηκε συστηματικά από τους Wan & Merwe, οι οποίοι πρότειναν και σειρά διαφορετικών υλοποιήσεων για καλύτερη απόδοση, [11], [30], [36]. Στο UKF η κατανομή του μεγέθους υπό εκτίμηση προσεγγίζεται ως μία Gaussian τυχαία μεταβλητή και αντιπροσωπεύεται από μια σειρά κατάλληλα επιλεγμένων σημείων-διανυσμάτων του χώρου καταστάσεων, τα οποία ονομάζονται **σίγμα** σημεία (*sigma points*). Αυτά τα σημεία αντιπροσωπεύουν πλήρως την κατανομή που είναι υπό εκτίμηση και περνώντας μέσα από το μη γραμμικό σύστημα μπορούν να περιγράψουν την εκ των υστέρων μέση τιμή και συμμεταβλητότητα της κατανομής με πολύ καλή προσέγγιση.

Σε σύγκριση με το EKF το UKF στηρίζεται στη διαίσθηση ότι είναι ευκολότερο να προσεγγιστεί μια κανονική κατανομή από ότι είναι να προσεγγιστεί μια τυχαία μη γραμμική συνάρτηση ή μετασχηματισμός μέσω γραμμικοποίησης. Στην περίπτωση που ο θόρυβος στις εξισώσεις κατάστασης ( $w \sim N(0, Q^w)$ ) και μέτρησης ( $v \sim N(0, Q^v)$ ) του συστήματος είναι μηδενικής μέσης τιμής και αθροιστικός, όπως συμβαίνει και στην περίπτωση που το σύστημα είναι της μορφής:

$$x_k = f(x_{k-1}) + Gw_{k-1} \quad (2.54)$$

$$y_k = h(x_k) + Hv_k \quad (2.55)$$

το UKF περιγράφεται από τις εξισώσεις που δίνονται στη συνέχεια. Αρχικοποίηση:

$$\hat{x}_0 = \mathbb{E}[x_0] \quad (2.56)$$

$$P_{x_0} = \mathbb{E}[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T] \quad (2.57)$$

Το  $k$  θεωρείται ότι παίρνει ακέραιες τιμές μεγαλύτερες του 0, δηλαδή  $k = 1, 2, 3 \dots$

Αρχικά υπολογίζονται τα **σίγμα** διανύσματα  $\mathcal{X}_i$ . Το πλήθος των διανυσμάτων είναι ίσο με  $2L + 1$ , όπου  $L$  η διάσταση του εκτιμώμενου μεγέθους  $x_k$  (κατάσταση του παρατηρούμενου συστήματος):

$$\mathcal{X}_{0,k-1} = \hat{x}_{k-1} \quad (2.58)$$

$$\mathcal{X}_{i,k-1} = \hat{x}_{k-1} + \gamma (\sqrt{P_{x_{k-1}}})_i, \quad i = 1, \dots, L \quad (2.59)$$

$$\mathcal{X}_{i,k-1} = \hat{x}_{k-1} - \gamma (\sqrt{P_{x_{k-1}}})_i, \quad i = L + 1, \dots, 2L \quad (2.60)$$

όπου  $\gamma \triangleq \sqrt{L + \lambda}$ .  $\lambda$  είναι μια παράμετρος κλιμάκωσης (*scaling parameter*) που ισούται με  $\lambda \triangleq \alpha^2(L + \kappa) - L$ . Η σταθερά  $\alpha$  καθορίζει το εύρος των σημείων **σίγμα** γύρω από την εκτίμηση  $\hat{x}_{k-1}$  και συνήθως λαμβάνει μια μικρή θετική τιμή. Η σταθερά  $\kappa$  είναι μια δευτερεύουσα παράμετρος κλιμάκωσης που συνήθως τίθεται ίση με 0 ή  $\kappa = 3 - L$ . Σημειώνεται ότι  $P_{x_{k-1}}$  συμβολίζεται ο πίνακας συμμεταβλητότητας της εκτίμησης  $x_{k-1}$  ενώ  $(\sqrt{P_{x_{k-1}}})_i$  είναι η  $i$ -οστή στήλη του πίνακα  $\sqrt{P_{x_{k-1}}}$  που προκύπτει μετά από παραγοντοποίηση Cholesky.

Οι εξισώσεις χρονικής ενημέρωσης είναι:

$$\mathcal{X}_{i,k|k-1} = f(\mathcal{X}_{i,k-1}, u_{k-1}, 0), \quad i = 0, \dots, 2L \quad (2.61)$$

$$\hat{x}_k^- = \sum_{i=0}^{2L} W_i^{(m)} \mathcal{X}_{i,k|k-1} \quad (2.62)$$

$$P_{x_k}^- = \sum_{i=0}^{2L} W_i^{(c)} [\mathcal{X}_{i,k|k-1} - \hat{x}_k^-] [\mathcal{X}_{i,k|k-1} - \hat{x}_k^-]^T + GQ^w G^T \quad (2.63)$$

όπου τα βάρη  $W_i$ , βάσει του μετασχηματισμού χωρίς καθοδήγηση (Unscented Transform), [31] και δεδομένων των παραμέτρων κλιμάκωσης  $\alpha, \beta, \lambda$ , δίνονται από τις σχέσεις:

$$W_0^{(m)} = \frac{\lambda}{L + \lambda} \quad (2.64)$$

$$W_0^{(c)} = \frac{\lambda}{L + \lambda} + (1 - \alpha^2 + \beta) \quad (2.65)$$

$$W_i^{(m)} = W_i^{(c)} = \frac{1}{2L + \lambda} \quad (2.66)$$

Η παράμετρος  $\beta$  χρησιμοποιείται για να ενσωματώσει πρότερη (*prior*) γνώση για την κατανομή του  $x_k$ . Για Gaussian κατανομές, συνήθως επιλέγεται  $\beta = 2$ .

Στη συνέχεια, εισάγεται και πάλι το διάνυσμα των ανανεώσεων:

$$i_k = y_k - \hat{y}_k^- \quad (2.67)$$

όπου, αυτή τη φορά, η τιμή του  $y_k^-$  προκύπτει από τα **σίγμα** διανύσματα  $\mathcal{Y}_{i,k|k-1}$  που υπολογίζονται από τη διάδοση των  $\mathcal{X}_{i,k|k-1}$  μέσω του μη γραμμικού συστήματος. Συγκεκριμένα, είναι:

$$\mathcal{Y}_{i,k|k-1} = h(\mathcal{X}_{i,k|k-1}) \quad (2.68)$$

$$\hat{y}_k^- = \sum_{i=0}^{2L} W_i^{(m)} \mathcal{Y}_{i,k|k-1} \quad (2.69)$$

Η συμμεταβλητότητα του  $i_k$  είναι:

$$P_{i_k} = \sum_{i=0}^{2L} W_i^{(c)} [\mathcal{Y}_{i,k|k-1} - \hat{y}_k^-] [\mathcal{Y}_{i,k|k-1} - \hat{y}_k^-]^T + HQ^v H^T \quad (2.70)$$

ενώ η συμμεταβλητότητα των  $x_k, i_k$  είναι:

$$P_{x_k i_k} = \sum_{i=0}^{2L} W_i^{(e)} [\mathcal{X}_{i,k|k-1} - \hat{x}_k^-][\mathcal{Y}_{i,k|k-1} - \hat{y}_k^-]^T \quad (2.71)$$

Τελικά, εφαρμόζονται οι εξισώσεις ενημέρωσης από τις μετρήσεις:

$$K_k = P_{x_k i_k} P_{i_k}^{-1} \quad (2.72)$$

$$\hat{x}_k = \hat{x}_k^- + K_k i_k \quad (2.73)$$

$$P_k = P_{x_k}^- - K_k P_{i_k} K_k^T \quad (2.74)$$

και υπολογίζεται η εκ των υστέρων εκτίμηση  $x_k$  η οποία χρησιμοποιείται στην νέα επανάληψη του αλγορίθμου.

Το θετικό με τον αλγόριθμο που περιγράφηκε είναι ότι έχει γενικά πολυπλοκότητα ίδιας τάξης με το *EKF* και δίνει καλύτερη εκτίμηση από αυτό σε πολλές εφαρμογές. Στα παραδείγματα εφαρμογών βέβαια που παρουσιάζονται στη συνέχεια, σημειώνεται ότι οι αντίστοιχες διαφορές είναι πολύ μικρές.

## Κεφάλαιο 3

# Αλγόριθμος αποδιαμόρφωσης βασισμένος σε μοντέλο με εκτίμηση Kalman

### 3.1 Εισαγωγή

Ο αλγόριθμος αποδιαμόρφωσης βασισμένος σε μοντέλο (**Model Based Demodulation Algorithm-MBDA**) με εκτίμηση Kalman προτάθηκε το 1996 από τους Lu & Doerschuk ως μία μέθοδος ανάλυσης φωνής βασισμένη στο μη γραμμικό μοντέλο που εισηγήθηκε ο Maragos και οι συνεργάτες του (Ενότητα 2.2). Είναι μια στατιστική προσέγγιση που αντιμετωπίζει συνολικά το πρόβλημα του διαχωρισμού και της αποδιαμόρφωσης της φωνής μέσα σε θόρυβο. Δεν συμπεριλαμβάνει τη χρήση φίλτρων, όπως οι προηγούμενες προσεγγίσεις που προτάθηκαν από Maragos et al [26] και Bovic et al [5] και χρησιμοποιεί μεθόδους εκτίμησης για την εξαγωγή των ζητούμενων μεγεθών. Δίνεται περιγραφή του αλγορίθμου καθώς και κάποιες εφαρμογές του σε φωνή και σε συνθετικά σήματα.

### 3.2 Περιγραφή του μοντέλου

Με βάση το μη γραμμικό μοντέλο φωνής του Maragos και των συνεργατών του, (βλ. ενότητα 2.2) κάθε συντονισμός της φωνής για χρονικές κλίμακες μικρότερες της θεμελιώδους περιόδου (pitch) μπορεί να παρασταθεί από ένα AM-FM σήμα της μορφής:

$$r_i(k) = \alpha_i(k) \cos(\varphi(k)) = \alpha_i(k) \cos(2\pi T_s F_c k + 2\pi T_s F_m \sum_0^{k-1} q(m) + \varphi(0)) \quad (3.1)$$

όπου  $T_s$  είναι η περίοδος δειγματοληψίας,  $F_c$  είναι η κεντρική συχνότητα του συντονισμού και  $F_m$  είναι η μέγιστη απόκλιση της στιγμιαίας συχνότητας από την κεντρική

λόγω διαμόρφωσης ( $|q(m)| < 1$ ). Επιτρέποντας περαιτέρω αργή μεταβολή της συχνότητας  $F_c$  και ίσως της  $F_m$  είναι δυνατή η μοντελοποίηση του συντονισμού και της εξέλιξής του σε ένα μεγαλύτερο τμήμα φωνής από  $10msec$ . Έτσι, το μοντέλο, σε μεγαλύτερη χρονική κλίμακα θα μπορούσε να οριστεί ως εξής :

$$r_i(k) = \alpha_i(k) \cos(\varphi(k)), \quad \text{όπου} \quad (3.2)$$

$$\varphi_i(k) = \varphi_i0 + 2\pi T_s \sum_{m=0}^{k-1} [f_i(m) + v_i(m)] \quad (3.3)$$

και το σήμα  $f_i$  περιγράφει την αργή μεταβολή της κεντρικής συχνότητας του συντονισμού ενώ το σήμα  $v_i$  περιγράφει τις γρήγορες μεταβολές γύρω από την κεντρική συχνότητα που μπορεί να έχουν ακόμα και αργά μεταβαλλόμενο εύρος.

Πάνω σε αυτή την ιδέα οι ερευνητές πρότειναν ένα δυναμικό σύστημα που θα μπορούσε να περιγράψει συνολικά τη φωνή. Πιο συγκεκριμένα, στο δυναμικό αυτό σύστημα, με δεδομένο ότι η κεντρική συχνότητα αναμένεται να μένει σταθερή για διαστήματα των λίγων χιλιοστών του δευτερολέπτου, το σήμα  $f_i$  περιγράφεται από έναν τυχαίο περίπατο (random walk), αφού αυτός αποτελεί το μοναδικό πρώτης τάξης μοντέλο Gauss-Markov όπου μια τέτοια συμπεριφορά έχει μεγάλη πιθανότητα να συμβεί. Όσο για το στιγμιαίο πλάτος  $\alpha_i$  και το σήμα  $v_i$ , αυτά περιγράφονται από αυτοσυρρικνούμενες διαδικασίες πρώτης τάξης, ώστε να υπάρχει δυνατότητα ελέγχου της ισχύος τους και του εύρους ζώνης τους ανεξάρτητα. Τέλος, η δυναμική της στιγμιαίας φάσης περιγράφεται από τη σχέση που την ορίζει, ενώ το σήμα φωνής εμφανίζεται ως γραμμική υπέρθεση των επιμέρους συνεισφορών κάθε συντονισμού μαζί με πρόσθετο θόρυβο μέτρησης. Τυπικά, για κάθε συντονισμό το σύστημα είναι :

$$\alpha_i(k+1) = \alpha_{\alpha_i} \alpha_i(k) + q_{\alpha_i} w_{\alpha_i}(k) \quad (3.4)$$

$$v_i(k+1) = \alpha_{v_i} v_i(k) + q_{v_i} w_{v_i}(k) \quad (3.5)$$

$$f_i(k+1) = f_i(k) + q_{f_i} w_{f_i}(k) \quad (3.6)$$

$$\varphi_i(k+1) = \varphi_i(k) + 2\pi T_s f_i(k) + 2\pi T_s v_i(k) \quad (3.7)$$

ενώ η έξοδος του συστήματος είναι :

$$y(k) = \sum_i \alpha_i(k) \cos(\varphi_i(k)) + ru(k) \quad (3.8)$$

όπου οι διαδικασίες θορύβου  $w_{\alpha_i}, w_{v_i}, w_{f_i}$  στις εξισώσεις κατάστασης και  $u$  στην εξίσωση εξόδου, είναι ανεξάρτητες στοχαστικές διαδικασίες με κανονική κατανομή πυκνότητας πιθανότητας με μέση τιμή 0 και μεταβλητότητα 1 ( $N(0, 1)$ ). Όσο για τις αρχικές συνθήκες, αυτές είναι τυχαίες μεταβλητές ανεξάρτητες μεταξύ τους και με το

θόρυβο και ακολουθούν τις κατανομές:

$$\alpha_i(0) \sim N\left(0, \frac{q_{\alpha_i}^2}{1 - \alpha_{\alpha_i}^2}\right) \quad (3.9)$$

$$v_i(0) \sim N\left(0, \frac{q_{v_i}^2}{1 - \alpha_{v_i}^2}\right) \quad (3.10)$$

$$f_i(0) \sim N(m_{f_i,0}, p_{f_i,0}^2) \quad (3.11)$$

$$\varphi_i(0) \sim N(0, p_{\varphi_i,0}^2) \quad (3.12)$$

Σημειώνεται ότι οι αρχικές συνθήκες απαιτούν  $|\alpha_{\alpha_i}| < 1$  και  $|\alpha_{v_i}| < 1$ , αλλιώς οι μεταβλητότητες στις παραπάνω κατανομές θα ήταν αρνητικές. Οπότε, οι διαδικασίες  $\alpha_i$  και  $v_i$  θεωρούνται στάσιμες με την ευρεία έννοια.

### 3.3 Στόχος του αλγορίθμου

Στο πλαίσιο που μόλις περιγράφηκε, ο στόχος της αποδιαμόρφωσης του παρατηρούμενου σήματος αντιστοιχεί στον υπολογισμό του διανύσματος

$$z_i(k) = (\alpha_i(k), f_i(k), v_i(k), \varphi_i(k))^T \quad (3.13)$$

για κάθε συντονισμό της φωνής με δεδομένες τις τιμές του σήματος που παρατηρείται  $y(0), \dots, y(k)$ . Έστω ότι  $\hat{z}_i(k|k)$ , μια συνάρτηση της ακολουθίας  $y(0), \dots, y(k)$ , είναι ένας εκτιμητής του  $z_i(k)$  βασισμένος σε δεδομένα μέχρι τη στιγμή  $k$ . Έστω ότι:

$$\epsilon(k|k) \triangleq \mathbb{E}[(z_i(k) - \hat{z}_i(k|k))(z_i(k) - \hat{z}_i(k|k))^T] \quad (3.14)$$

είναι το μέσο τετραγωνικό λάθος της εκτίμησης. Ως βέλτιστος εκτιμητής,  $\hat{z}_i^*(k|k)$ , θεωρείται αυτός που ελαχιστοποιεί το  $\epsilon(k|k)$  και είναι  $\hat{z}_i^*(k|k) = \mathbb{E}[z_i(k)|y(0), \dots, y(k)]$  ενώ το ελάχιστο τετραγωνικό λάθος είναι:

$$\epsilon^*(k|k) = \mathbb{E}[(z_i(k) - \hat{z}_i^*(k|k))(z_i(k) - \hat{z}_i^*(k|k))^T] \quad (3.15)$$

Στόχος του αλγορίθμου είναι η όσο το δυνατόν καλύτερη εκτίμηση σε αυτά τα πλαίσια του διανύσματος  $z_i(k)$ . Ο εκτιμητής που τελικά προκύπτει, στηρίζεται στο εκτεταμένο φίλτρο *Kalman* και είναι υποβέλτιστος.

### 3.4 Ταυτοποίηση του συστήματος

Πριν βέβαια είναι δυνατή η όποια επεξεργασία πρέπει να προσδιοριστούν οι άγνωστες παράμετροι του δυναμικού συστήματος, δηλαδή το διάνυσμα

$$\theta_i = (\alpha_{\alpha_i}, q_{\alpha_i}, \alpha_{v_i}, q_{v_i}, q_{f_i}, m_{f_i,0}, p_{f_i,0}, p_{\varphi_i,0})$$



για κάθε συντονισμό και η παράμετρος  $r$  που αφορά στην ισχύ του θορύβου στην εξίσωση παρατήρησης. Η τελευταία προσδιορίζεται άμεσα με τη χρήση εκ των προτέρων γνώσης για το θόρυβο παρατήρησης. Έτσι, για παράδειγμα για σήματα από την, κατά τεκμήριο αθόρυβη, βάση εκφωνήσεων TIMIT ο θόρυβος παρατήρησης είναι απλά ο θόρυβος λόγω κβαντοποίησης ο οποίος έχει τυπική απόκλιση  $r = \sqrt{\frac{1}{12}}$ , με την προϋπόθεση ενός ομοιόμορφου μοντέλου κβαντοποίησης  $\pm \frac{1}{2}$  bit, αν δηλαδή θεωρήσουμε ότι το σφάλμα κβαντοποίησης ακολουθεί ομοιόμορφη κατανομή μεταξύ  $-1/2$  και  $+1/2$ , [1].

Οι υπόλοιπες παράμετροι, σύμφωνα με τους ερευνητές, καθορίζονται με τη βοήθεια της φασματικής πυκνότητας ισχύος του σήματος που αντιστοιχεί στο σήμα  $y$  (Εξ.3.8) και η οποία μπορεί να υπολογιστεί αριθμητικά.

Έστω οι  $\alpha_i(k)$ ,  $\varphi_i(k)$  τυχαίες διαδικασίες, όπου η  $\varphi$  ακολουθεί κανονική κατανομή. Έστω  $\psi_i$  μια τυχαία μεταβλητή η οποία είναι κατανομημένη ομοιόμορφα στο  $[-\pi, \pi]$  και  $\alpha_i, \varphi_i, \psi_i$  ανεξάρτητες μεταξύ τους. Τότε για το σήμα  $y_i(k) = \alpha_i(k) \cos(\varphi_i(k) + \psi)$  ισχύει ότι  $m_{y_i}(k) = 0$  και η αυτοσυσχετίσή του είναι ίση με τη συμμεταβλητότητά του και είναι:

$$\begin{aligned} R_{y_i}(k_1, k_2) &= P_{y_i}(k_1, k_2) \\ &= R_{\alpha_i}(k_1, k_2) \frac{\cos(m_{\varphi_i}(k_1) - m_{\varphi_i}(k_2))}{2} \exp\left(-\frac{1}{2} Q_{\varphi_i}(k_1, k_2)\right) \end{aligned} \quad (3.16)$$

όπου

$$Q_{\varphi_i}(k_1, k_2) \triangleq P_{\varphi_i}(k_1, k_1) - 2P_{\varphi_i}(k_1, k_2) + P_{\varphi_i}(k_2, k_2) \quad (3.17)$$

Έτσι, είναι δυνατός ο υπολογισμός της συμμεταβλητότητας του σήματος  $y(k) = \sum_{i=1}^K y_i(k) + ru(k)$  ως:

$$P_y(k_1, k_2) = \sum_i P_{y_i}(k_1, k_2) + r^2 P_u(k_1, k_2) \quad (3.18)$$

ενώ η μέση τιμή του είναι  $m_y(k) = rm_u(k)$  ( $m_u(k)$  είναι η μέση τιμή του θορύβου παρατήρησης η οποία θεωρείται μηδενική) οπότε, κατ' επέκταση μπορεί να υπολογιστεί και η αυτοσυσχετίσή του. Από την τελευταία γίνεται εκτίμηση της φασματικής πυκνότητας ισχύος της διαδικασίας. Για λεπτομέρειες, [33].

Εδώ σημειώνεται απλά ότι οι παράμετροι  $p_{f_i,0}$ ,  $p_{\varphi_i,0}$  λαμβάνονται μηδενικές ενώ η τιμή της παραμέτρου  $q_{f_i}$  τίθεται ίση με τη μέγιστη επιθυμητή μεταβολή ανά δείγμα της κεντρικής συχνότητας του συντονισμού. Όσο για τις παραμέτρους  $\alpha_{\alpha_i}$ ,  $\alpha_{v_i}$ , για λόγους ευστάθειας απαιτείται να ισχύει  $|\alpha_{\alpha_i}| < 1$  και  $|\alpha_{v_i}| < 1$ . Επιπλέον, έτσι ώστε να ελαχιστοποιηθεί το εύρος ζώνης απαιτείται  $\alpha_{\alpha_i} \approx 1$ ,  $\alpha_{v_i} \approx 1$ . Επιλέγεται λοιπόν  $\alpha_{\alpha_i} = \alpha_{v_i} = 0.99$ . Τέλος, η επιλογή των  $m_{f_i,0}$ ,  $q_{\alpha_i}$  και  $q_{v_i}$  καθορίζεται από την επιθυμητή αρχική κεντρική συχνότητα του συντονισμού, την επιθυμητή ισχύ και το εύρος του, αντίστοιχα, με βάση τη φασματική πυκνότητα ισχύος της διαδικασίας σε σύγκριση με αυτή του σήματος υπό ανάλυση.

### 3.5 Εκτίμηση με χρήση του εκτεταμένου φίλτρου Kalman

Για την εκτίμηση χρησιμοποιείται εκτεταμένο φίλτρο Kalman. Στην περίπτωση που το στιγμιαίο πλάτος ήταν σταθερό τότε το εκτεταμένο αυτό φίλτρο θα ήταν ένας βρόχος κλειδωμένης φάσης (phase locked loop, [34]) και ένας τέτοιος βρόχος είναι πολύ καλός εκτιμητής. Για αυτό χρησιμοποιείται και στην πιο πολύπλοκη περίπτωση που το στιγμιαίο πλάτος μεταβάλλεται. Οι υπολογιστικές απαιτήσεις είναι πολύ μικρές, αφού οι εξισώσεις κατάστασης είναι γραμμικές και ο πίνακας μετάβασης  $A$  του συστήματος είναι μπλοκ διαγώνιος (ένα μπλοκ για κάθε συντονισμό) και κάθε μπλοκ είναι αραιό και έτσι ο πολλαπλασιασμός με τον  $A$  δεν έχει μεγάλο κόστος. Επίσης, η παρατήρηση είναι βαθμωτή και έτσι αντί για αντιστροφή του πίνακα εξόδου  $C$  απαιτείται απλή διαίρεση.

Το αποτέλεσμα του EKF είναι οι εκτιμήσεις  $\hat{\alpha}_i(k)$ ,  $\hat{v}_i(k)$ ,  $\hat{f}_i(k)$  και  $\hat{\varphi}_i(k)$ . Με αυτές τις εκτιμήσεις μπορούμε να υπολογίσουμε ένα ανακατασκευασμένο σήμα φωνής  $\hat{y}(k)$  ως:

$$\hat{y}(k) = \sum_i \hat{\alpha}_i(k) \cos(\hat{\varphi}_i(k)) \quad (3.19)$$

### 3.6 Εφαρμογές

Παρουσιάζονται εφαρμογές του αλγορίθμου σε αποδιαμόρφωση συνθετικών σημάτων, σε ανάλυση σημάτων φωνής, καθώς και σε παρακολούθηση συντονισμών φωνής (formant tracking).

#### 3.6.1 Εφαρμογή σε συνθετικά σήματα

Ο αλγόριθμος εφαρμόζεται κατ' αρχάς σε ένα σήμα το οποίο αποτελείται από δύο AM συνιστώσες, των οποίων οι κεντρικές συχνότητες, μεταβαλλόμενες γραμμικά, κάποια στιγμή διασταυρώνονται. Το σήμα αυτό παρουσιάζεται από τον Lu στο άρθρο [33] και είναι:

$$y(k) = \cos(2\pi f_a kT) \cos(2\pi(f_{m1} + f_c kT)kT) + 0.2 \cos(2\pi f_a kT) \cos(2\pi(f_{m2} - f_c kT)kT) \quad (3.20)$$

όπου  $T = 1/16000 \text{ sec}$ ,  $f_a = 30 \text{ Hz}$ ,  $f_{m1} = 200 \text{ Hz}$ ,  $f_{m2} = 2000 \text{ Hz}$  και  $f_c = 6000 \text{ Hz/s}$ . Το εκτεταμένο φίλτρο Kalman έχει παραμέτρους  $\alpha_{\alpha_1} = \alpha_{\alpha_2} = 0.99$ ,  $\alpha_{v_1}, \alpha_{v_2} = 0.99$ ,  $q_{\alpha_1} = \sqrt{0.001}$ ,  $q_{\alpha_2} = \sqrt{0.001 \times 0.04}$ ,  $q_{f_1} = q_{f_2} = \sqrt{12}$ ,  $q_{v_1}, q_{v_2} = \sqrt{0.01}$ ,  $r = 10^{-7}$ ,  $m_{f_1,0} = f_{m1}$ ,  $m_{f_2,0} = f_{m2}$  και όλες οι άλλες αρχικές συνθήκες και μεταβλητότητες επιλέγονται μηδενικές.

Τα αποτελέσματα του διαχωρισμού των δύο συνιστωσών, της παρακολούθησης των κεντρικών συχνοτήτων και της αποδιαμόρφωσης με τη χρήση του αλγορίθμου Kalman-MBDA φαίνονται στο Σχήμα 3.1 και είναι αρκετά ικανοποιητικά, παρά

τη διασταύρωση των συχνοτήτων. Παρατηρούμε ότι ακόμα και μετά τη στιγμή της διασταύρωσης οι δύο συνιστώσες εξακολουθούν να διαχωρίζονται χωρίς πρόβλημα. Πρέπει να παρατηρηθεί βέβαια ότι το σήμα δεν έχει άλλη διαμόρφωση FM εκτός από τη γραμμική μεταβολή των κεντρικών συχνοτήτων.

### 3.6.2 Εφαρμογή σε φωνή

#### Ανάλυση σημάτων φωνής

Στη συνέχεια ο αλγόριθμος εφαρμόζεται στην ανάλυση, με βάση το μη γραμμικό μοντέλο, ενός σήματος φωνής. Το σήμα αυτό είναι το φώνημα /ee/ στη λέξη "m/ee/ting" από τη βάση εκφωνήσεων TIMIT (test/dr2/mdbb0/sx295) (Σχήμα 3.2(a)). Το μοντέλο έχει τέσσερις συνιστώσες (συντονισμούς) με κεντρικές συχνότητες οι οποίες έχουν αρχικές τιμές  $m_{f_i,0} = [390, 2200, 2800, 3600] Hz$ . Και για τις τέσσερις συνιστώσες ισχύει  $\alpha_{\alpha_i} = \alpha_{v_i} = 0.99$ ,  $q_{f_i} = \sqrt{8}$ ,  $p_{f_i,0} = 0$  και  $p_{\varphi_i=0}$ . Οι τιμές των  $q_{a_i}$  και  $q_{v_i}$  διαφέρουν από συνιστώσα σε συνιστώσα και είναι  $q_{a_i} = 158, 20, 11, 7$  και  $q_{v_i} = 14, 14, 21, 21$  για  $i = 1, 2, 3, 4$  αντίστοιχα. Τελικά επιλέγεται  $r = \sqrt{1/12}$ .

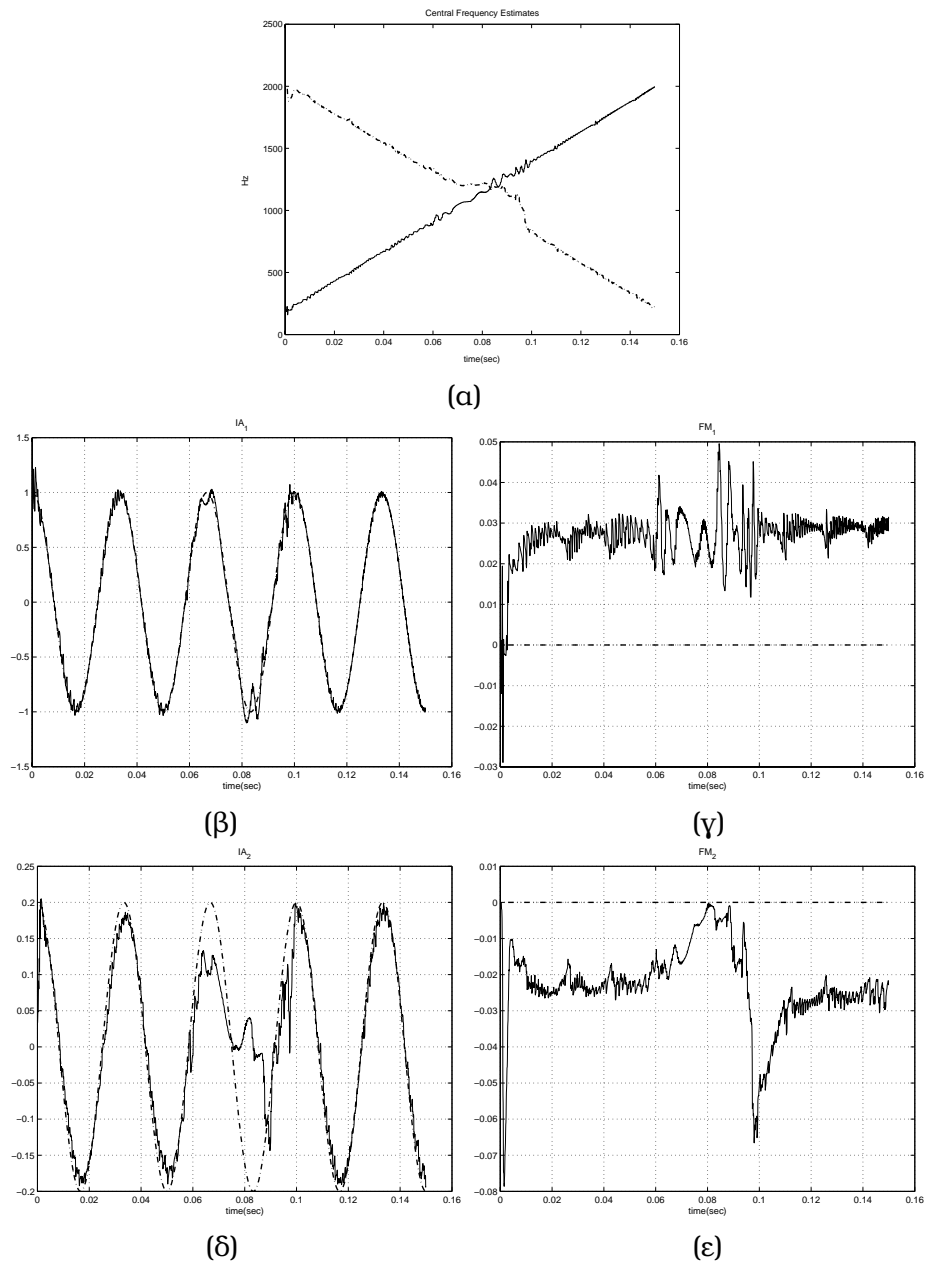
Τα αποτελέσματα παρουσιάζονται στο σχήμα 3.3 ενώ οι εκτιμήσεις των κεντρικών συχνοτήτων στο σχήμα 3.2.

Από τα διαγράμματα διαπιστώνει κανείς κατ' αρχάς ότι οι εκτιμήσεις των διαφορών μεγεθών είναι ιδιαίτερα θορυβώδεις, κάτι που θα μπορούσε να αλλάξει σε κάποιο βαθμό με αύξηση της παραμέτρου  $r$ , οπότε και θα μειωνόταν ουσιαστικά το κόστος της μη ακριβούς αποσύνθεσης (δηλαδή  $y(k) \neq \hat{y}(k)$ ). Αυτό που πάντως παρατηρείται άμεσα εκτός των άλλων, είναι η ύπαρξη των ενεργειακών παλμών για τους οποίους έγινε λόγος στο προηγούμενο κεφάλαιο, και οι οποίοι συνηγορούν για την ορθότητα της αποδοχής του μοντέλου AM-FM για τη φωνή. Βέβαια οι παλμοί αυτοί δεν περιορίζονται προς τις θετικές τιμές μόνο αλλά μπορεί να πάρουν και αρνητικές τιμές, όπως εκτιμώνται από τον αλγόριθμο, κάτι που προκαλεί προβλήματα στην εφαρμογή της θεωρίας της διαμόρφωσης AM-FM αφού εμφανίζεται αμφισημία στην ερμηνεία της στιγμιαίας συχνότητας.

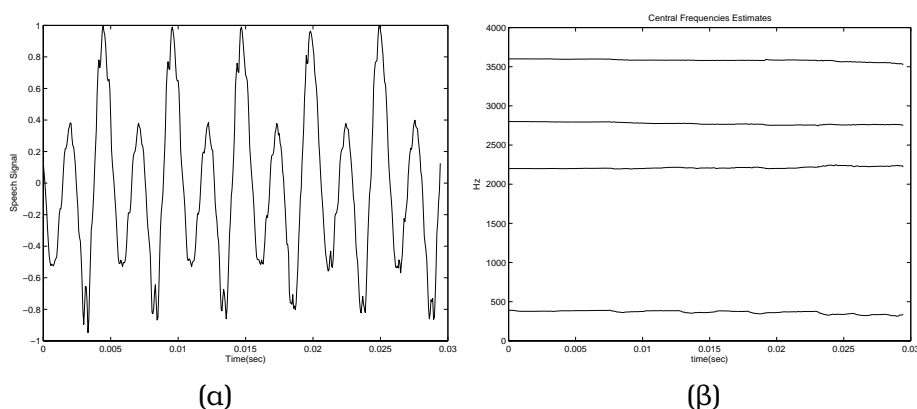
#### Παρακολούθηση συντονισμών φωνής

Ως τελευταία ενδιαφέρουσα εφαρμογή παρουσιάζεται η παρακολούθηση των συντονισμών (formants) ενός σήματος φωνής μεγάλης διάρκειας. Μια τέτοια εφαρμογή θεωρείται πολύ σημαντική στο χώρο της επεξεργασίας φωνής και έχει υπάρξει πολλή σχετική έρευνα από διάφορες οπτικές γωνίες, π.χ. [12]. Μια τέτοια, που στηρίζεται σε μη γραμμική μοντελοποίηση της φωνής αλλά και τον αλγόριθμο διαχωρισμού με υπολογισμό της ενέργειας είναι του Potamianos et al, [6], ή της Hanson et al, [14].

Το σήμα φωνής που χρησιμοποιείται για την εφαρμογή αντιστοιχεί στην πρόταση "Where were you while we were away" από τη βάση TIMIT. Πρόκειται για ένα σήμα το οποίο είναι έμφωνο σε όλο το μήκος του και έτσι δεν υπάρχει, διαισθητικά τουλάχιστον, πρόβλημα στην αντίληψη των formants. Πρόκειται για τις περιοχές



Σχήμα 3.1: Αποτελέσματα εφαρμογής τους Kalman-MBDA στο σήμα της εξίσωσης 3.20 (α) Οι κεντρικές συχνότητες των δύο συνιστωσών του συνθετικού σήματος όπως εκτιμώνται από τον Kalman-MBDA (β),(δ) Το στιγμιαίο πλάτος της 1ης και της 2ης συνιστώσας αντίστοιχα του συνθετικού σήματος, (γ),(ε) Η στιγμιαία διαμόρφωση συχνότητας εκτός της μεταβολής των κεντρικών συχνοτήτων για την 1η και τη 2η συνιστώσα. Στα διαγράμματα, με διακεκομμένες γραμμές είναι τα πραγματικά σήματα.



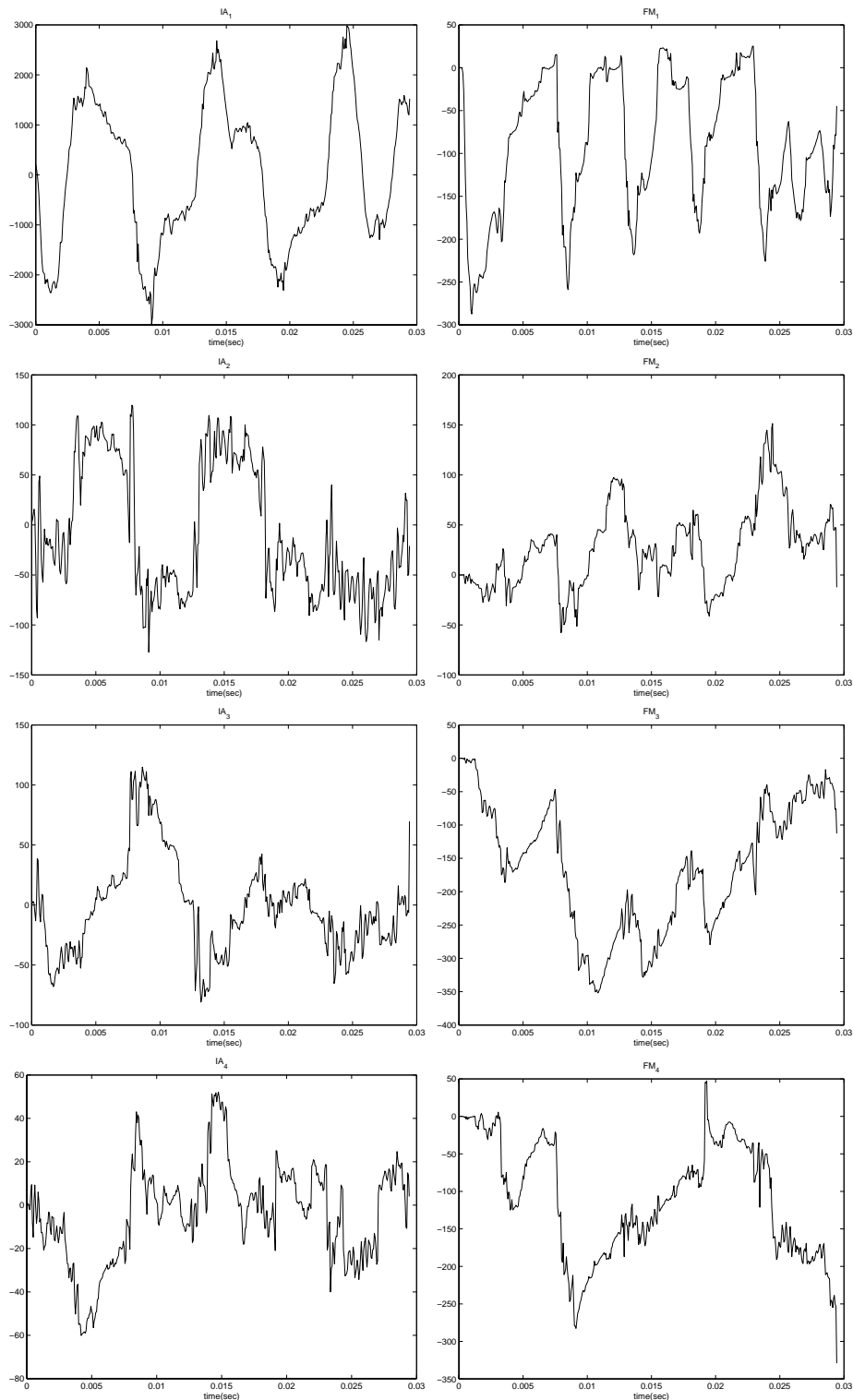
Σχήμα 3.2: Εκτίμηση κεντρικών συχνοτήτων σήματος φωνής με τη χρήση του Kalman-MBDA (α) Το φώνημα /ee/ από τη λέξη m/ee/ting (β) Οι κεντρικές συχνοτήτες των τεσσάρων συνιστωσών του φωνήματος, όπως εκτιμώνται με τη βοήθεια του MBDA

συχνοτήτων όπου η φασματική πυκνότητα ισχύος εμφανίζει τοπικά μέγιστα. Ο αλγόριθμος αρχικοποιείται σύμφωνα με τις υποδείξεις των ερευνητών και τα αποτελέσματα φαίνονται στο σπεκτρογράφημα 3.4 όπου έχουν υπερτεθεί οι τροχιές των κεντρικών συχνοτήτων. Σημειώνεται ότι σε κάθε πλαίσιο, λαμβάνεται η μέση τιμή της κεντρικής συχνότητας κατά μήκος όλου του πλαισίου και αυτή είναι που αναπαρίσταται στο σχήμα. Διαπιστώνουμε ότι η παρακολούθηση είναι αρκετά ικανοποιητική παρά τις γρήγορες και πολλές μεταβολές των formants με το μόνο πρόβλημα ίσως να εντοπίζεται για  $100msec$ , από τα 300 έως τα 400 περίπου, οπότε και διασταυρώνονται ο δεύτερος με τον τρίτο συντονισμό.

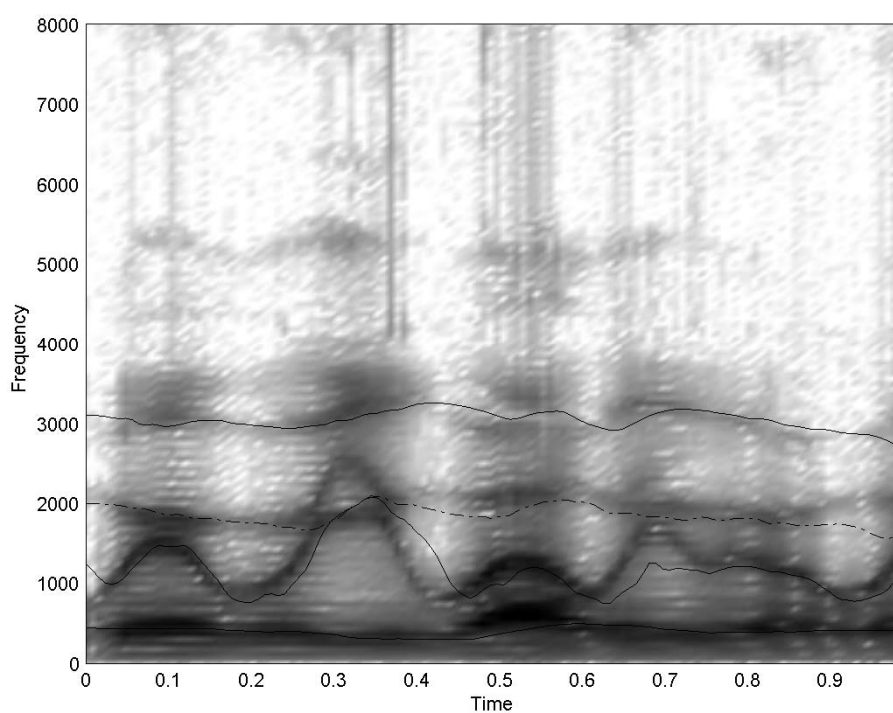
### 3.7 Παραλλαγές

Σίγουρα είναι δυνατές διάφορες παραλλαγές του αλγορίθμου που παρουσιάστηκε, απλά με την πραγματοποίηση μικρών τροποποιήσεων στο μοντέλο του συστήματος, για καλύτερη ίσως προσαρμογή σε διαφορετικά σήματα. Μια ενδιαφέρουσα πρόταση, βασισμένη και αυτή στην ιδέα της αποδιαμόρφωσης με τη χρήση μοντέλου και εκτίμησης Kalman παρουσιάστηκε από τους Doerschuk και Pai, [35]. Πρόκειται για τον αλγόριθμο AEKF ο οποίος είναι εμπνευσμένος από την ημιτονοειδή αναπαράσταση της φωνής που προτείνεται από τον Quatieri, [28], και τους συνεργάτες του.

Ο αλγόριθμος αυτός δεν ασχολείται βέβαια με το πρόβλημα του διαχωρισμού των συνιστωσών, αφού στην είσοδο του θεωρείται ότι φθάνει ζωνοπερατό σήμα και μάλιστα η μιγαδική βαθυπερατή του αναπαράσταση. Το στιγμιαίο πλάτος και η στιγμιαία συχνότητα υπολογίζονται με γραμμική παρεμβολή μεταξύ ακραίων σημείων που εκτιμώνται με τη χρήση εκτεταμένου φίλτρου Kalman, στο οποίο, σημειώνεται ότι



Σχήμα 3.3: Τα σιγμιαία πλάτη  $\hat{a}_i$  ( $IA_i$ ) και οι συχνότητες διαμόρφωσης  $\hat{\nu}_i$  ( $FM_i$ ) των συνιστωσών του σήματος φωνής του σχήματος 3.2 (α), όπως υπολογίζονται με τον αλγόριθμο Kalman-MBDA.



Σχήμα 3.4: Σπεκτρογράφημα του σήματος φωνής "Where were you while we were away" από τη βάση TIMIT με υπέρθεση των εκτιμήσεων των κεντρικών συχνοτήτων των συντονισμών από τον αλγόριθμο Kalman-MBDA

οι ισχείς των θορύβων είναι προσαρμοζόμενες στο σήμα. Σε αυτή την προσέγγιση τα στιγμιαία πλάτη περιορίζονται ώστε να παίρνουν μόνο θετικές τιμές. Η μέθοδος παρουσιάζει ιδιαίτερο ενδιαφέρον και σε κάποιο βαθμό επηρέασε μια από τις παραλλαγές της νέας μεθόδου που προτείνεται στο επόμενο κεφάλαιο.

### 3.8 Συμπεράσματα

Ο αλγόριθμος Kalman-MBDA που παρουσιάστηκε έχει κάποια ιδιαίτερα ενδιαφέροντα χαρακτηριστικά. Κατ' αρχάς, μπορεί να αναλύσει σήματα με πολλές συνιστώσες άμεσα, χωρίς να χρειάζεται να προηγηθεί κάποιας μορφής διαχωρισμός. Από την άλλη, η παρουσία παρεμβαλλόμενων σημάτων, όπως θορύβου για παράδειγμα, συμπεριλαμβάνεται με φυσικό τρόπο στην προσέγγιση, τόσο στις εξισώσεις κατάστασης όσο και στη μέτρηση. Επιπλέον, τα σήματα υπό ανάλυση μπορούν να περιγραφούν με ακρίβεια. Για παράδειγμα, τα εύρη ζώνης των σημάτων AM και FM μπορούν να ελεγχθούν ανεξάρτητα όπως επίσης και ο ρυθμός μεταβολής των κεντρικών συχνοτήτων.

Σίγουρα πρόκειται για μια πολύ ενδιαφέρουσα μέθοδο με βασικό της βέβαια μειονέκτημα να είναι το γεγονός ότι δεν είναι εύκολο να αυτοματοποιηθεί αυτόματα και να είναι ιδιαίτερα ευαίσθητη σε μικρές αλλαγές των παραμέτρων. Η παραλλαγή που προτείνεται στο επόμενο κεφάλαιο προσπαθεί να βελτιώσει τα πράγματα, χρησιμοποιώντας πληροφορία από τον αλγόριθμο διαχωρισμού με υπολογισμό της ενέργειας.



## Κεφάλαιο 4

# Αλγόριθμος αποδιαμόρφωσης με χρήση φίλτρου Kalman και του διαχωρισμού με υπολογισμό της ενέργειας

### 4.1 Εισαγωγή

Εμπνευσμένη από τον αλγόριθμο αποδιαμόρφωσης βασισμένο σε μοντέλο (Kalman-MBDA), σε αυτό το κεφάλαιο παρουσιάζεται μια καινούρια προσέγγιση στο πρόβλημα του διαχωρισμού και αποδιαμόρφωσης σημάτων με πολλές συνιστώσες AM-FM μέσα σε θόρυβο. Χαρακτηριστικό του MBDA είναι ότι θεωρεί τα στιγμιαία πλάτη και τις στιγμιαίες συχνότητες των συνιστώσων του σήματος αυτοσυρρικνούμενες στοχαστικές διαδικασίες χωρίς κάποια ιδιαίτερη φυσική ερμηνεία. Η ιδέα στην οποία στηρίζεται η νέα μέθοδος που προτείνεται είναι ότι ο αλγόριθμος διαχωρισμού με υπολογισμό της ενέργειας θα μπορούσε να δώσει περισσότερες πληροφορίες για τα στιγμιαία πλάτη και τις συχνότητες στην περίπτωση που το σήμα προέρχεται από ένα φυσικό δυναμικό σύστημα, όπως είναι το σύστημα παραγωγής της φωνής. Επιπλέον, θα μπορούσε να κάνει ευκολότερη την προσαρμογή του μοντέλου σε διάφορα σήματα, μειώνοντας τον αριθμό των παραμέτρων που πρέπει να προκαθοριστούν.

Ακολουθεί περιγραφή της ιδέας της προσέγγισης (**Kalman Energy Separation Demodulation Algorithm-KESDA**) καθώς και διαφόρων μοντέλων που την υλοποιούν και δοκιμάστηκαν. Συνολικά δοκιμάστηκαν τέσσερα τέτοια μοντέλα. Το πρώτο, το βασικό, υλοποιεί τη βασική ιδέα της νέας προσέγγισης. Επονομάζεται *bKES* από τα αρχικά των λέξεων **b**asic **K**alman **E**nergy **S**eparation. Το δεύτερο, *fKES*, (**f**rame **K**alman **E**nergy **S**eparation), είναι μια παραλλαγή του πρώτου που εφαρμόζεται σε πλαίσια του σήματος. Το τρίτο, *iKES* (**i**nterpolated **K**alman **E**nergy **S**eparation), είναι παραλλαγή του δεύτερου, στην οποία χρησιμοποιείται γραμμική παρεμβολή για την εκτίμηση των στιγμιαίων πλατών. Τέλος, το τέταρτο μοντέλο,

*mKES* (**m**easurement **K**alman **E**nergy **S**eparation), είναι σε διαφορετικό πνεύμα από τα υπόλοιπα τρία και εισάγει το διαχωρισμό με υπολογισμό της ενέργειας στις εξισώσεις μέτρησης, αντί για τις εξισώσεις κατάστασης.

Λεπτομέρειες, σχόλια και εφαρμογές δίνονται στις ενότητες που ακολουθούν.

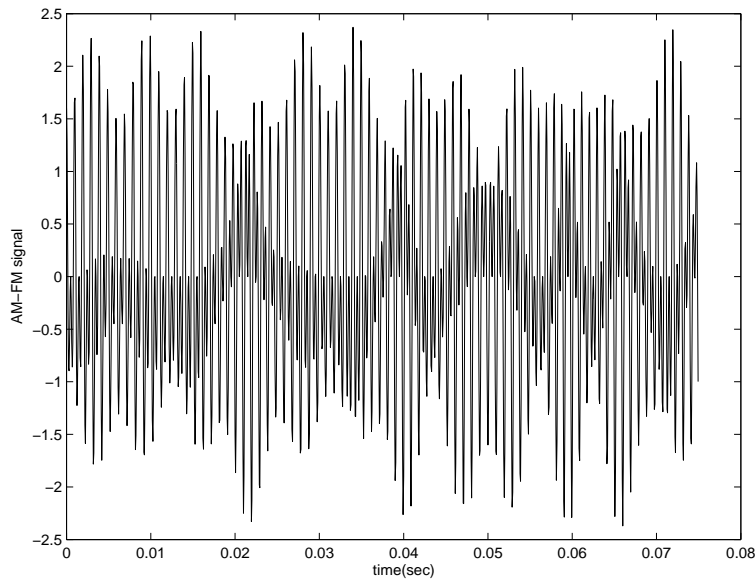
## 4.2 ESA και σήμα με πολλές συνιστώσες

Η εφαρμογή του *ESA* σε σήματα που αποτελούνται από πολλές συνιστώσες και θόρυβο, χωρίς να προηγηθεί κάποιου είδους επεξεργασία είναι γνωστό πως δεν είναι δυνατό να δώσει αξιόπιστα αποτελέσματα. Η επεξεργασία αυτή μπορεί να πάρει διάφορες μορφές. Για την ανάλυση σημάτων φωνής, αρχικά χρησιμοποιήθηκαν, για παράδειγμα, φίλτρα Gabor των οποίων οι παράμετροι καθορίζονταν κάθε φορά μετά από εποπτεία του φάσματος του σήματος υπό ανάλυση, [26] ή μετά από διαδοχικές εφαρμογές του *ESA*, [14]. Χρησιμοποιήθηκε επίσης σταθερή συστοιχία φίλτρων και στη συνέχεια επιλογή της περισσότερο ενεργής εξόδου για αποθορυβοποίηση, [5], ή εφαρμογή του *ESA* σε κάθε έξοδο για τον υπολογισμό των συντονισμών της φωνής, [6]. Τέλος, έχουν χρησιμοποιηθεί και μέθοδοι διαχωρισμού των διαφόρων συνιστωσών με τη χρήση ενεργειακών τελεστών υψηλότερης τάξης, [25], όπως στον αλγόριθμο EDM, [9], ή αλγεβρικών μεθόδων όπως στον αλγόριθμο PASED, [8].

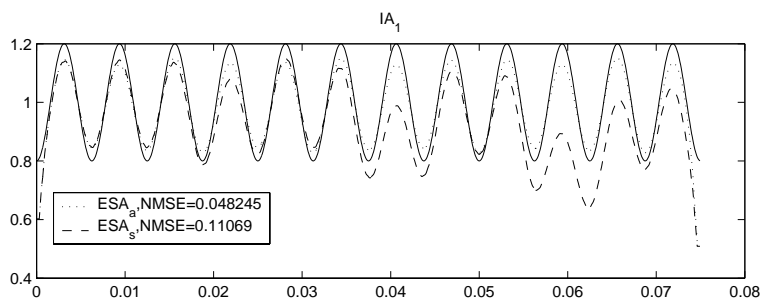
Στόχος του αλγορίθμου που προτείνεται είναι κυρίως η ανάλυση σημάτων φωνής, η οποία ερμηνεύεται κατ' αρχάς ως αποδιαμόρφωση των συνιστωσών που αντιστοιχούν στους συντονισμούς, με βάση το μοντέλο των Maragos et al. Η ανάλυση αυτή απαιτεί, πριν την εφαρμογή οποιουδήποτε αλγορίθμου αποδιαμόρφωσης, το διαχωρισμό των συνιστωσών αυτών. Η απαίτηση αυτή αντιμετωπίζεται με την προσπάθεια παρακολούθησης των κεντρικών συχνοτήτων των συντονισμών. Στη συνέχεια, στις συχνότητες αυτές εφαρμόζονται φίλτρα Gabor και έτσι απομονώνονται οι επιθυμητές συνιστώσες και είναι δυνατή η αποδιαμόρφωσή τους με τη χρήση του αλγορίθμου διαχωρισμού με υπολογισμό της ενέργειας. Παράλληλα, με τη χρησιμοποίηση φίλτρου *Kalman* είναι δυνατή η συμμετοχή των υπολογισμένων στιγμιαίων πλατών και συχνοτήτων σε κατάλληλη ανατροφοδότηση ώστε να διορθωθεί, αν χρειάζεται, η τιμή των κεντρικών συχνοτήτων και να γίνει σωστή εφαρμογή των φίλτρων Gabor την επόμενη χρονική στιγμή.

Το πόσο σημαντικό είναι να προσαρμόζονται οι θέσεις των φίλτρων Gabor κάθε χρονική στιγμή στην περίπτωση που οι κεντρικές συχνότητες των διαφόρων συνιστωσών είναι μεταβαλλόμενες, γίνεται φανερό στις γραφικές παραστάσεις που ακολουθούν (Σχ. 4.1). Σε αυτές επιδεικνύονται τα αποτελέσματα εφαρμογής του *ESA* σε ένα σήμα με δύο συνιστώσες στην περίπτωση όπου προηγείται ένα ζευγάρι σταθερών φίλτρων Gabor (με σταθερές κεντρικές συχνότητες  $1000\text{Hz}$  και  $3000\text{Hz}$  και αποδοτικό rms εύρος ζώνης το καθένα ίσο με  $bw = 1000/\sqrt{2\pi}$ ) ή ένα προσαρμοζόμενο ζευγάρι (με κεντρικές συχνότητες γραμμικά μεταβαλλόμενες όπως και οι αντίστοιχες συχνότητες του σήματος) αντίστοιχα.

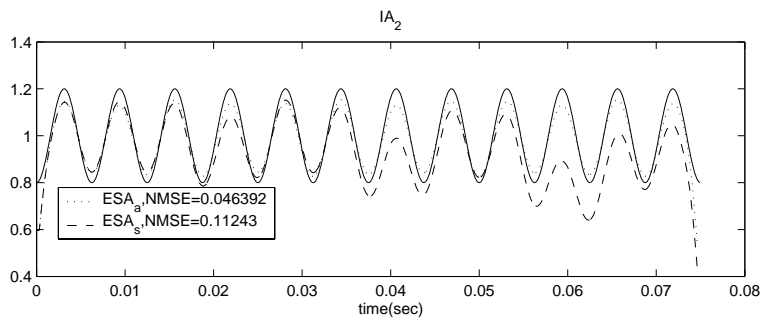
Σημειώνεται ότι σε ολόκληρο το κεφάλαιο στη συνέχεια, αλλά και στο επόμενο, γίνεται εκτεταμένη χρήση του κανονικοποιημένου μέσου τετραγωνικού λάθους ως



(α)



(β)



(γ)

Σχήμα 4.1: Εφαρμογή του ESA στις εξόδους συστοιχίας φίλτρων Gabor τα οποία είναι σταθερών (*steady*) ή προσαρμοζόμενων (*adaptive*) κεντρικών συχνοτήτων(α) Το AM-FMσήμα  $y(t) = [1 - 0.2 \cos(2\pi 50t)][\cos(2\pi 1000t + 2\pi \int_0^t 2000\tau + 50 \cos(2\pi 50\tau) d\tau) + \cos(2\pi 2000t + 2\pi \int_0^t 2000\tau + 50 \cos(2\pi 50\tau) d\tau)]$ , στο οποίο οι κεντρικές συχνοτήτες των δύο συνιστωσών μεταβάλλονται γραμμικά, (β) Τα στιγμιαία πλάτη των δύο συνιστωσών ( $IA_1, IA_2$ ) όπως υπολογίζονται με τον ESA και ενώ έχει προηγηθεί σταθερή συστοιχία φίλτρων Gabor ( $ESA_s$ ) ή προσαρμοζόμενη αντίστοιχα συστοιχία ( $ESA_a$ ). Με συνεχή γραμμή είναι τα πραγματικά σήματα. Στα υπομνήματα φαίνεται και το αντίστοιχο κανονικοποιημένο μέσο τετραγωνικό λάθος

κριτηρίου σύγκρισης και αξιολόγησης, το οποίο, για μία διαδικασία  $y(k)$  της οποίας η εκτίμηση είναι  $\hat{y}(k)$ , ορίζεται ως εξής:

$$NMSE = \frac{\sum_k [y(k) - \hat{y}(k)]^2}{\sum_k y^2(k)} \quad (4.1)$$

Πιο λεπτομερής περιγραφή του νέου αλγορίθμου αλλά και διαφόρων παραλλαγών γίνεται στις ενότητες που ακολουθούν.

### 4.3 Αλγόριθμος ανάλυσης με βάση το μοντέλο *bKES*

Προτείνεται ένα νέο δυναμικό σύστημα, το οποίο έχει λίγο διαφορετικές μεταβλητές κατάστασης από αυτό των Lu & Doerschuk (Κεφάλαιο 3). Συγκεκριμένα, περιλαμβάνει ως μεταβλητές κατάστασης τις κεντρικές συχνότητες των συνιστωσών οι οποίες και πάλι δεχόμαστε ότι εκτελούν τυχαίο περίπατο, τα στιγμιαία πλάτη, τις στιγμιαίες συχνότητες των συνιστωσών αυτών και τις αντίστοιχες φάσεις, ενώ έχει την ίδια εξίσωση μέτρησης. Για κάθε χρονική στιγμή  $n$  ισχύει:

$$f_i(n) = f_i(n-1) + q_{f_i} w_{f_i}(n-1) \quad (4.2)$$

$$\begin{pmatrix} \alpha_i(n) \\ IF_i(n) \end{pmatrix} = \mathbf{IFAE}(s, f_i(n-1), bw_i, n) + \begin{pmatrix} q_{\alpha_i} & 0 \\ 0 & q_{IF_i} \end{pmatrix} \begin{pmatrix} w_{\alpha_i}(n) \\ w_{IF_i}(n) \end{pmatrix} \quad (4.3)$$

$$\Phi_i(n) = \Phi_i(n-1) + 2\pi T_s IF_i(n-1) \quad (4.4)$$

όπου  $w_{f_i}$  είναι κανονικά κατανομημένη στοχαστική διαδικασία θορύβου με μέση τιμή 0 και μεταβλητότητα 1 ( $N(0, 1)$ ) και  $T_s$  είναι η συχνότητα δειγματοληψίας. Οι διαδικασίες  $w_{IF_i}$  και  $w_{\alpha_i}$  είναι και αυτές  $N(0, 1)$  στοχαστικές διαδικασίες με ισχύ που καθορίζεται από τις παραμέτρους  $q_{v_i}$  και  $q_{\alpha_i}$  αντίστοιχα και οι οποίες αντιπροσωπεύουν οποιεσδήποτε αποκλίσεις από τις τιμές που επιστρέφει η μέθοδος **IFAE**. Θεωρούνται ότι είναι ανεξάρτητες μεταξύ τους. Η μέθοδος **IFAE** (Instantaneous Frequency & Amplitude Estimation), έχοντας ως δεδομένα το σήμα υπό ανάλυση  $s(n)$ , την τιμή της κεντρικής συχνότητας  $f_i$  της  $i$ -συνιστώσας του  $s(n)$ , καθώς και το εύρος ζώνης  $bw_i$  του φίλτρου Gabor που πρέπει να χρησιμοποιηθεί, επιστρέφει τις τιμές του πλάτους και της συχνότητας της  $i$ -συνιστώσας τη δεδομένη στιγμή  $n$ . Περισσότερες λεπτομέρειες δίνονται στη συνέχεια. Η εξίσωση μέτρησης, από την άλλη, είναι :

$$y(k) = \sum_{i=1}^K \alpha_i(n) \cos(\varphi_{f_i}(n)) + ru(n) \quad (4.5)$$

όπου  $K$  είναι ο συνολικός αριθμός των συνιστωσών που περιέχει το σήμα, ο οποίος υποτίθεται ότι είναι γνωστός και  $u(n)$  είναι και αυτή  $N(0, 1)$  στοχαστική διαδικασία που παριστάνει τον θόρυβο στη μέτρηση.

Στο νέο μοντέλο η μεταβλητή κατάσταση  $v_i$ , που αντιπροσώπευε τις γρήγορες μεταβολές της συχνότητας γύρω από την αργά μεταβαλλόμενη  $f_i$ , (βλ. ενότητα 3.2), έχει αντικατασταθεί από τη μεταβλητή  $IF_i$ , η οποία αντιπροσωπεύει συνολικά τη στιγμιαία συχνότητα, δηλαδή ισχύει  $IF_i = f_i + v_i$ . Η συχνότητα αυτή όπως και το πλάτος υπολογίζονται με κατάλληλη εφαρμογή φίλτρου Gabor εύρους ζώνης  $bw_i$  και του ESA ή, όπως σημειώνεται στις εξισώσεις, με τη μέθοδο υπολογισμού στιγμιαίας συχνότητας και στιγμιαίου πλάτους (IFAE).

Λόγω της ιδιαίτερης σημασίας που έχει η συγκεκριμένη μέθοδος στην περαιτέρω ανάπτυξη, σημειώνονται κάποιες σχετικές λεπτομέρειες.

#### 4.3.1 Μέθοδος υπολογισμού του στιγμιαίου πλάτους και της στιγμιαίας συχνότητας μιας από τις AM-FM συνιστώσες (IFAE)

Ουσιαστικά, πρόκειται για « κατάλληλη εφαρμογή φίλτρου Gabor και του ESA » ώστε να είναι δυνατός ο όσο το δυνατόν αποδοτικότερος υπολογισμός του πλάτους και της συχνότητας μιας συνιστώσας του σήματος κάποια συγκεκριμένη χρονική στιγμή  $n$ . Αβίαστα κάποιος μπορεί να σκεφτεί ότι χρειάζεται απλά συνέλιξη του δεδομένου σήματος  $s$  με κατάλληλο φίλτρο Gabor και στη συνέχεια εφαρμογή του ESA στην έξοδο και υπολογισμός των σημάτων του πλάτους και της συχνότητας σε όλο το μήκος του σήματος. Οι ζητούμενες τιμές του πλάτους και της συχνότητας θα είναι απλά οι τιμές των παραπάνω σημάτων αντίστοιχα για τη στιγμή  $n$ .

Η προσέγγιση αυτή να μεν δίνει τα επιθυμητά αποτελέσματα, αλλά δεν αξιοποιεί ορισμένες χαρακτηριστικές ιδιότητες τόσο του φίλτρου Gabor όσο και του ESA με αποτέλεσμα να αυξάνεται αδικαιολόγητα η πολυπλοκότητα.

Ένα από τα θετικά χαρακτηριστικά του ESA είναι ότι έχει μεγάλη διακριτική ικανότητα στο χρόνο. Ειδικά, η παραλλαγή DESA-1 για τον υπολογισμό του πλάτους και της συχνότητας ενός σήματος κάποια χρονική στιγμή χρειάζεται συνολικά τις τιμές του σήματος σε ένα παράθυρο 5 δειγμάτων κεντραρισμένο στη δεδομένη χρονική στιγμή, οπότε δεν υπάρχει λόγος να εφαρμοστεί ο ESA σε όλο το μήκος της εξόδου του φίλτρου, που μετά την πράξη της συνέλιξης έχει μήκος μεγαλύτερο από το μήκος του σήματος.

Βέβαια, από τον ορισμό της συνέλιξης δύο σημάτων  $h$  και  $x$ :

$$y(n) \triangleq h(n) * x(n) \triangleq \sum_{k=-\infty}^{k=\infty} h(k)x(n-k) \quad (4.6)$$

και με δεδομένο ότι τα σήματα στην περίπτωση που μελετάται έχουν πεπερασμένο μήκος, η τιμή της εξόδου του φίλτρου Gabor μια συγκεκριμένη χρονική στιγμή, αν η κρουστική απόκρισή του έχει μήκος  $M$  εξαρτάται το πολύ από ένα παράθυρο  $M$  δειγμάτων του σήματος εισόδου. Αφού, εντωμεταξύ, ο ESA απαιτεί τουλάχιστον ένα παράθυρο 5 δειγμάτων της εξόδου του φίλτρου, αυτό που χρειάζεται είναι ένα παράθυρο  $M + 5$  δειγμάτων του αρχικού σήματος.

Αυτές ακριβώς οι ιδέες υλοποιούνται με τη μέθοδο υπολογισμού του στιγμιαίου πλάτους και της στιγμιαίας συχνότητας μιας από τις AM-FM συνιστώσες του σήματος. Υποτίθεται ότι δίνεται το σήμα με τις πολλές συνιστώσες, η κεντρική συχνότητα της συνιστώσας στην οποία θα κεντραριστεί το φίλτρο Gabor καθώς και το εύρος ζώνης του φίλτρου, ενώ επιστρέφονται οι τιμές του πλάτους και της συχνότητας της συνιστώσας με τη δεδομένη κεντρική συχνότητα, τη δεδομένη χρονική στιγμή.

Σημειώνεται πάντως ότι ενδέχεται σε κάποιες περιπτώσεις ο απαιτούμενος αριθμός δειγμάτων να είναι μεγαλύτερος ανάλογα με το αν γίνεται εφαρμογή του ομαλοποιημένου ESA ή όχι, (βλ. υποενότητα 2.3.3) και με το βαθμό της ομαλοποίησης, δηλαδή πόσο μεγάλο είναι το αντίστοιχο ομαλοποιητικό φίλτρο.

Οι τιμές βέβαια που επιστρέφονται είναι αυτές που παίρνουν κατ' αρχάς οι αντίστοιχες μεταβλητές κατάστασης (εκτιμήσεις εκ των προτέρων των στιγμιαίων πλατών και των στιγμιαίων συχνοτήτων) και οι οποίες ανανεώνονται στη συνέχεια με κατάλληλη αξιοποίηση της τιμής του παρατηρούμενου σήματος από το φίλτρο *Kalman*.

### 4.3.2 Αρχικοποίηση και εκτίμηση με χρήση φίλτρου Kalman

Είναι φανερό πως, σε σύγκριση με το μοντέλο των Lu και Doerschuk, το νέο μοντέλο που προτείνεται έχει το πλεονέκτημα ότι έχει μικρότερο αριθμό παραμέτρων που πρέπει να προκαθοριστούν. Οι παράμετροι αυτές είναι οι τιμές των  $q_{f_i}$ , που καθορίζονται ανάλογα με την επιθυμητή μεταβολή των κεντρικών συχνοτήτων ανά δείγμα, των  $q_{\alpha_i}, q_{IF_i}$  οι οποίες επιλέγονται ανάλογα με το πόσο ακριβή αναμένονται να είναι τα αποτελέσματα της IFAE και η τιμή  $r$  που καθορίζει την ισχύ του θορύβου μέτρησης και ουσιαστικά επιλέγεται με βάση εκ των προτέρων γνώση για το θόρυβο που υπάρχει στο σήμα υπό ανάλυση.

Η εκτίμηση, με δεδομένο ότι το μοντέλο που περιγράφηκε είναι μη γραμμικό, μπορεί να γίνει με το εκτεταμένο φίλτρο Kalman ή με το φίλτρο Kalman χωρίς καθοδήγηση. Όπως έχει προαναφερθεί βέβαια η εκτίμηση αυτή δεν είναι βέλτιστη αλλά είναι απλό να υλοποιηθεί και με κάποιο τρόπο προσεγγίζει τη βελτιστότητα. Στην περίπτωση του εκτεταμένου φίλτρου, ο Ιακωβιανός πίνακας  $G$  είναι μπλοκ διαγώνιος (ένα μπλοκ  $G_i$  για κάθε συνιστώσα) και ο υπολογισμός του είναι άμεσος:

$$G_i = \begin{pmatrix} q_{f_i} & 0 & 0 \\ 0 & q_{\alpha_i} & 0 \\ 0 & 0 & q_{IF_i} \\ 0 & 0 & 0 \end{pmatrix} \quad (4.7)$$

ενώ ο πίνακας  $C$  είναι στην ουσία διάνυσμα που αποτελείται από τμήματα της μορφής  $C_i$  για κάθε συνιστώσα:

$$C_i = (0 \quad \cos(\varphi_i(k)) \quad 0 \quad -\alpha_i(k) \sin(\varphi_i(k))) \quad (4.8)$$

Τέλος, ο πίνακας  $A$  υπολογίζεται προσεγγιστικά με τη χρήση μικρών διαταραχών της κατάστασης του συστήματος κάθε φορά, αφού δεν είναι προφανής ο αναλυτικός υπολογισμός της παραγώγου. Συγκεκριμένα δεν είναι άμεσα, αναλυτικά υπολογίσιμη

η σχέση των μεταβλητών κατάστασης με τις κεντρικές συχνότητες. Αν θεωρήσουμε ότι  $\mathbf{x}_n = g(\mathbf{x}_{n-1}, \mathbf{w}_{n-1})$  η αναλυτική έκφραση της παραγώγου  $\frac{dg}{df_i}$  δεν είναι εύκολη υπόθεση και για αυτό προσεγγίζεται ως :

$$\frac{dg}{df_i} \approx \frac{\Delta g}{\Delta f_i} \quad (4.9)$$

Το γεγονός αυτό βέβαια έχει ως αποτέλεσμα να επιβαρύνεται κάπως υπολογιστικά ο αλγόριθμος αφού θα πρέπει κάθε φορά που η  $f_i$  υπόκειται μια έστω και πολύ μικρή μεταβολή να ξαναχρησιμοποιείται η μέθοδος **IFAE** για να υπολογιστούν τα νέα πλάτη και οι νέες στιγμιαίες συχνότητες και να μπορεί έτσι να προσδιοριστεί η μεταβολή  $\Delta g$ .

Όσο για την επιλογή των αρχικών κεντρικών συχνοτήτων, αυτή στα συνθετικά σήματα γίνεται με τη χρήση εκ των προτέρων γνώσης για τις κεντρικές συχνότητες ενώ οι στιγμιαίες συχνότητες αρχικοποιούνται με την ίδια τιμή που έχουν οι κεντρικές και οι φάσεις και τα στιγμιαία πλάτη αρχικοποιούνται ακολουθώντας κανονική κατανομή με μέση τιμή στο μηδέν και μεταβλητότητα μεγάλη για τα πλάτη (τουλάχιστον μεγαλύτερη από  $q_{\alpha_i}^2$ ) και μικρότερη από  $\pi^2$  για τις φάσεις.

Από την άλλη, για το φίλτρο Kalman χωρίς καθοδήγηση αφού ο θόρυβος τόσο στην εξίσωση παρατήρησης όσο και στην εξίσωση μέτρησης είναι προσθετικός, είναι δυνατή η χρήση περιορισμένου αριθμού  $\sigma$ -σημείων όπως περιγράφηκε στην υποενοότητα 2.4.4 και έτσι η πολυπλοκότητα της εκτίμησης είναι ανάλογη αυτής που έχει η εκτίμηση με το εκτεταμένο φίλτρο. Τα αποτελέσματα πάντως σε διάφορα σχετικά πειράματα και με τη χρήση των δύο φίλτρων, διαφέρουν ελάχιστα και μάλλον δεν δικαιολογείται επαρκώς, λόγω αυξημένης, έστω και λίγο, πολυπλοκότητας, η χρήση του unscented Kalman Filter.

### 4.3.3 Εφαρμογή σε συνθετικά σήματα

Ο αλγόριθμος εφαρμόστηκε με επιτυχία σε διάφορες περιπτώσεις συνθετικών AM-FM σημάτων. Ενδεικτικά θα δοθούν δύο, οι οποίες είναι ίσως και οι πιο ενδιαφέρουσες με γνώμονα πάντα την επικείμενη εφαρμογή του μοντέλου σε σήματα φωνής. Οι ίδιες δύο περιπτώσεις είναι αυτές που προσεγγίζονται και με τις παραλλαγές που παρουσιάζονται στη συνέχεια ώστε να είναι δυνατή η σύγκριση με βάση και τις μορφές των κυματομορφών που προκύπτουν αλλά και των κανονικοποιημένων μέσω των τετραγωνικών λαθών που δίνονται κατά περίπτωση.

#### Εφαρμογή σε σήμα με σταθερές κεντρικές συχνότητες

Το πρώτο σήμα το οποίο αναλύεται είναι το :

$$y_s(t) = [1 - 0.5 \cos(2\pi 160t)] [\cos(2\pi 1000t + 2\pi 80 \int_0^t \cos(2\pi 80\tau) d\tau) + \cos(2\pi 3000t + 2\pi 80 \int_0^t \cos(2\pi 80\tau) d\tau)] \quad (4.10)$$

το οποίο αποτελείται από δύο συνιστώσες διαμορφωμένες κατά πλάτος και συχνότητα από τα ίδια σήματα και διαφέρουν μόνο στην τιμή των κεντρικών συχνοτήτων οι οποίες παραμένουν σταθερές στο χρόνο. Η μία κεντρική συχνότητα είναι  $1000Hz$  ενώ η άλλη  $3000Hz$ .

Οι τιμές των  $q_{IF_i}$  και  $q_{\alpha_i}$  επιλέγονται μικρές ( $q_{IF_i} = 1$  και  $q_{\alpha_i} = 0.001$ ) με δεδομένο ότι η μέθοδος  $IFAE$  αναμένεται να είναι αξιόπιστη ενώ  $q_{f_i} = 0.001$  αφού η κεντρική συχνότητα παραμένει αμετάβλητη. Επιπλέον, τα δύο Gabor φίλτρα που χρησιμοποιούνται έχουν αποδοτικό rms εύρος ζώνης ίσο με  $\alpha/\sqrt{2\pi}$ , όπου  $\alpha = 1500$ . Η αρχική κατάσταση θεωρείται ότι είναι:

$$x_0 = \begin{pmatrix} f_1(0) \\ \alpha_1(0) \\ IF_1(0) \\ \varphi_1(0) \\ f_2(0) \\ \alpha_2(0) \\ IF_2(0) \\ \varphi_2(0) \end{pmatrix} = \begin{pmatrix} 1000 \\ 0 \\ 1000 \\ 0 \\ 3000 \\ 0 \\ 3000 \\ 0 \end{pmatrix} \quad (4.11)$$

$$P_{x_0} = \mathbf{diag}((0.1 \quad 10000 \quad 2 \quad 0 \quad 0.1 \quad 10000 \quad 2 \quad 0)) \quad (4.12)$$

όπου  $\mathbf{diag}(c)$  είναι διαγώνιος πίνακας με κύρια διαγώνιο το διάνυσμα  $c$ . Τα αποτελέσματα φαίνονται στη γραφική παράσταση 4.2 για χρονικό διάστημα ίσο με  $0.0375sec$  θεωρώντας συχνότητα δειγματοληψίας  $F_s = 16000Hz$ .

### Εφαρμογή σε σήμα με γραμμικώς μεταβαλλόμενες κεντρικές συχνότητες

Το επόμενο παράδειγμα αφορά σε ένα σήμα με δύο συνιστώσες πάλι, μόνο που τώρα οι κεντρικές τους συχνότητες είναι γραμμικά μεταβαλλόμενες. Συγκεκριμένα, το σήμα  $y(t)$  υπό ανάλυση είναι:

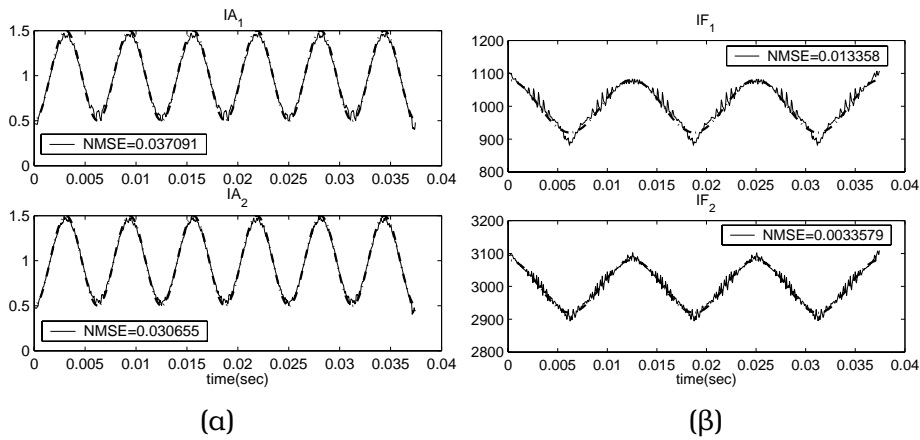
$$y_v(t) = [1 - 0.5 \cos(2\pi 160t)] [\cos(2\pi 1000t + 2\pi \int_0^t [3000\tau + 80 \cos(2\pi 80\tau) d\tau]) \\ + \cos(2\pi 2400t + 2\pi \int_0^t (-3000)\tau + \cos(2\pi 80\tau) d\tau)] \quad (4.13)$$

Στην περίπτωση αυτή, εκτός από τη διαμόρφωση συχνότητας που είχε και το προηγούμενο σήμα, υπάρχει και επιπλέον διαμόρφωση λόγω της γραμμικής μεταβολής της κεντρικής συχνότητας. Οι αργά μεταβαλλόμενες κεντρικές συχνότητες των δύο συνιστωσών είναι:

$$f_1 = 1000 + 3000t \quad (4.14)$$

$$f_2 = 2400 - 3000t \quad (4.15)$$



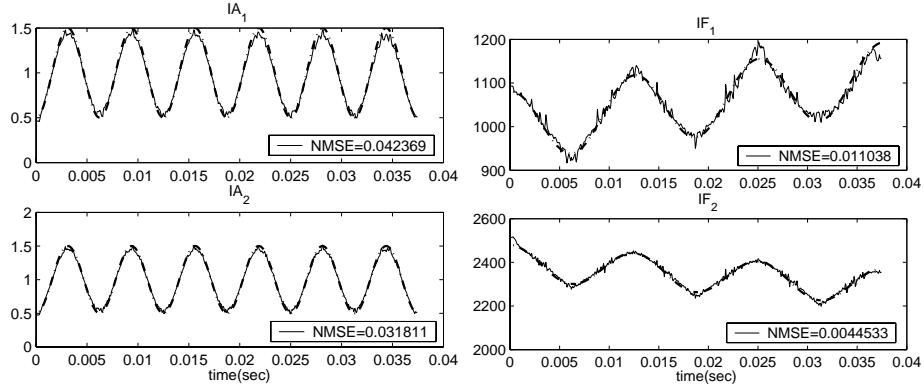


Σχήμα 4.2: Αποτελέσματα εφαρμογής της εκτίμησης Kalman με το μοντέλο *bKES* στο σήμα  $y(t) = [1 - 0.5 \cos(2\pi 160t)][\cos(2\pi 1000t + 2\pi 80 \int_0^t \cos(2\pi 80\tau) d\tau) + \cos(2\pi 3000t + 2\pi 80 \int_0^t \cos(2\pi 80\tau) d\tau)]$ . (α) Στιγμιαία πλάτη των δύο συνιστωσών του  $y$ , (β) στιγμιαίες συχνότητες των δύο συνιστωσών σε  $Hz$ . Με διακεκομμένες γραμμές είναι οι πραγματικές τιμές των σημάτων, ενώ στα υπομνήματα αναγράφεται και το κανονικοποιημένο μέσο τετραγωνικό λάθος κατά περίπτωση.

Οι διάφορες παράμετροι είναι όπως και προηγουμένως, εκτός από τις παραμέτρους  $q_{f_i}$  οι οποίες επιλέγονται ίσες με τη μεταβολή που αναμένεται να έχουν οι κεντρικές συχνότητες ανά δείγμα, δηλαδή:  $q_{f_i} = 3000/16000 = 0.1875$ . Η αρχική κατάσταση είναι η ίδια όπως και προηγουμένως με εξαίρεση τις αρχικές τιμές των κεντρικών συχνοτήτων των δύο συνιστωσών που τώρα είναι  $1000Hz$  και  $2400Hz$  αντίστοιχα. Στη γραφική παράσταση 4.3 φαίνονται τα αποτελέσματα του αλγορίθμου τα οποία κρίνονται αρκετά ικανοποιητικά. Ουσιαστικά η απόδοση του αλγορίθμου δεν άλλαξε, παρά τη μεταβολή των κεντρικών συχνοτήτων.

## 4.4 Αλγόριθμος ανάλυσης με βάση το μοντέλο *fKES*

Το γεγονός ότι το μοντέλο που προτάθηκε στην προηγούμενη ενότητα επιτρέπει τη μεταβολή των κεντρικών συχνοτήτων ακόμα και μεταξύ δύο διαδοχικών δειγμάτων, δεν είναι πάντα επιθυμητό, ειδικά αν οι κεντρικές συχνότητες μεταβάλλονται τόσο αργά ώστε το μόνο ουσιαστικό αποτέλεσμα να είναι η αύξηση της πολυπλοκότητας της εκτίμησης. Το μοντέλο που περιγράφεται σε αυτή την ενότητα επιτρέπει την αντιμετώπιση αυτής της κατάστασης ανανεώνοντας τις τιμές των κεντρικών συχνοτήτων ανά κάποιο συγκεκριμένο αριθμό δειγμάτων  $N$ . Στις μεταβλητές κατάστασης του μοντέλου συμπεριλαμβάνονται, εκτός από τις κεντρικές συχνότητες, τα στιγμιαία



Σχήμα 4.3: Αποτελέσματα εφαρμογής της εκτίμησης Kalman με το μοντέλο *bKES* στο σήμα  $y(t) = [1 - 0.5 \cos(2\pi 160t)][\cos(2\pi 1000t + 2\pi \int_0^t 3000\tau + 80 \cos(2\pi 80\tau)d\tau) + \cos(2\pi 2400t + 2\pi \int_0^t (-3000)\tau + \cos(2\pi 80\tau)d\tau)]$ . (α) Στιγμιαία πλάτη των δύο συνιστωσών του  $y$ , (β) στιγμιαίες συχνότητες των δύο συνιστωσών σε  $Hz$ . Με διακεκομμένες γραμμές είναι οι πραγματικές τιμές των σημάτων, ενώ στα υπομνήματα αναγράφεται και το κανονικοποιημένο μέσο τετραγωνικό λάθος κατά περίπτωση.

πλάτη, οι στιγμιαίες συχνότητες και οι φάσεις της κάθε συνιστώσας για όλο το μήκος του πλαισίου. Πιο συγκεκριμένα, οι εξισώσεις κατάστασης για κάθε συνιστώσα είναι:

$$f_i(k) = f_i(k-1) + q_{f_i} w_{f_i}(k) \quad (4.16)$$

$$\begin{pmatrix} A_i^k \\ IF_i^k \end{pmatrix} = \mathbf{IFAE}_{frame}(s, f_i(k-1), bw_i, k) + \begin{pmatrix} Q_{A_i} \\ Q_{IF_i} \end{pmatrix} \begin{pmatrix} W_{A_i}^k \\ W_{IF_i}^k \end{pmatrix} \quad (4.17)$$

$$\varphi_i(j) = \varphi(j-1) + 2\pi T_s IF(j-1), \quad j = (k-1)N + 1 \dots kN \quad (4.18)$$

όπου  $k$  είναι ο αύξων αριθμός του πλαισίου ανάλυσης μήκους  $N$  και

$$A_i^k \triangleq \begin{pmatrix} \alpha_i((k-1)N + 1) \\ \dots \\ \alpha_i(kN) \end{pmatrix} \quad IF_i^k \triangleq \begin{pmatrix} IF_i((k-1)N + 1) \\ \dots \\ IF_i(kN) \end{pmatrix} \quad (4.19)$$

$$\Phi \triangleq \begin{pmatrix} \varphi_i((k-1)N + 1) \\ \dots \\ \varphi_i(kN) \end{pmatrix} \quad (4.20)$$

είναι διανύσματα μήκους  $N$  με τις τιμές των στιγμιαίων πλατών, στιγμιαίων συχνοτήτων και φάσεων αντίστοιχα για όλο το μήκος του πλαισίου  $k$ . Ο συνολικός αριθμός των μεταβλητών κατάστασης για κάθε συνιστώσα είναι ίσος με  $3N + 1$  όπως είναι φανερό. Σημειώνεται ότι  $w_{f_i}(k)$  είναι στοχαστική διαδικασία  $N(0, 1)$  από την οποία είναι ανεξάρτητες, όπως και μεταξύ τους, οι στοχαστικές διαδικασίες που παριστάνουν ενδεχόμενα σφάλματα στις τιμές που επιτρέφει η μέθοδος **IFAE** και που

συμπαγώς εμφανίζονται στις εξισώσεις με τη μορφή του διανύσματος  $W_i^k$ :

$$W_i^k \triangleq \begin{pmatrix} W_{A_i} \\ W_{IF_i} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} w_{\alpha_i,1}(k) \\ w_{\alpha_i,2}(k) \\ \dots \\ w_{\alpha_i,N}(k) \end{pmatrix} \\ \begin{pmatrix} w_{IF_i,1}(k) \\ w_{IF_i,2}(k) \\ \dots \\ w_{IF_i,N}(k) \end{pmatrix} \end{pmatrix} \quad (4.21)$$

Σημειώνεται επίσης ότι:

$$Q_i \triangleq \begin{pmatrix} Q_{A_i} \\ Q_{IF_i} \end{pmatrix} = \begin{pmatrix} q_{A_i} I_{N \times N} \\ q_{IF_i} I_{N \times N} \end{pmatrix} \quad (4.22)$$

Οι εξισώσεις μέτρησης του συστήματος είναι :

$$Y^k = \sum_{i=1}^K A_i^k \cdot * \cos(\Phi_i^k) + RU^k \quad (4.23)$$

όπου  $K$  είναι ο γνωστός εκ των προτέρων συνολικός αριθμός των συνιστωσών στο σήμα και :

$$Y^k \triangleq \begin{pmatrix} y((k-1)N+1) \\ \dots \\ y(kN) \end{pmatrix} \quad (4.24)$$

είναι το διάνυσμα που περιλαμβάνει τις αναμενόμενες τιμές του σήματος για το μήκος  $N$  του πλαισίου  $k$  και  $\cos(\Phi_i^k)$  είναι απλώς ένας συμπαγής συμβολισμός για το διάνυσμα :

$$\cos(\Phi_i^k) \triangleq [\cos(\varphi_i(j))], j = (k-1)N+1 \dots kN \quad (4.25)$$

ενώ έχει υιοθετηθεί ο συμβολισμός  $'.*'$  της *MATLAB* για πολλαπλασιασμό διανυσμάτων σημείο προς σημείο, δηλαδή:

$$(a_1, a_2 \dots a_n) \cdot * (b_1, b_2 \dots) = (a_1 b_1, a_2 b_2, \dots a_n b_n) \quad (4.26)$$

Με  $U^k$  παριστάνονται συμπαγώς οι διαδικασίες που παριστάνουν το θόρυβο στις μετρήσεις. Πρόκειται για  $N(0,1)$  διαδικασίες οι οποίες είναι ανεξάρτητες μεταξύ τους και με τις διαδικασίες που εμφανίζονται στις εξισώσεις κατάστασης. Όσο για τον πίνακα  $R$ , αυτός ρυθμίζει την αυτοσυσχέτιση των θορυβών στις μετρήσεις και είναι  $R = r I_{N \times N}$  αφού κατ' αρχάς δεν υπάρχει κάποιος λόγος να θεωρηθεί μια μέτρηση περισσότερο ή λιγότερο θορυβώδης από τις υπόλοιπες ενώ δεν υπάρχει και συσχέτιση μεταξύ των μετρήσεων.

Σημειώνεται επίσης ότι με  $\mathbf{IFAE}_{frame}$  συμβολίζεται μια παραλλαγή της μεθόδου υπολογισμού που περιγράφηκε στην προηγούμενη ενότητα η οποία επιστρέφει  $N$  το πλήθος τιμές του στιγμιαίου πλάτους και της στιγμιαίας συχνότητας. Οι τιμές αυτές στη συγκεκριμένη περίπτωση αντιστοιχούν στις χρονικές στιγμές  $(k-1)N+1$  έως και  $kN$ .

#### 4.4.1 Αρχικοποίηση και εκτίμηση με χρήση του εκτεταμένου φίλτρου Kalman

Όσον αφορά στην επιλογή των διαφόρων παραμέτρων του φίλτρου, δεν αλλάζει ουσιαστικά η νοοτροπία που περιγράφηκε στην αντίστοιχη παράγραφο στην παρουσίαση του προηγούμενου μοντέλου. Αυτό που σημειώνεται στη συγκεκριμένη περίπτωση είναι ότι λόγω αυξημένου αριθμού μεταβλητών κατάστασης και της φύσης της μεθόδου **IFAE** δεν είναι δυνατή η χρησιμοποίηση του φίλτρου Kalman χωρίς καθοδήγηση για την επίτευξη της εκτίμησης των ζητούμενων σημάτων. Έτσι, η εκτίμηση γίνεται μόνο με τη χρήση του εκτεταμένου φίλτρου Kalman, που με το προτεινόμενο μοντέλο είναι αρκετά γρηγορότερη από ότι προηγουμένως, κύριως λόγω του ότι υπολογίζονται πολλές ( $N$ ) τιμές του πλάτους και της συχνότητας με την εφαρμογή της μεθόδου **IFAE** μόνο μια φορά, που σημαίνει μόνο μια φορά φιλτράρισμα και εφαρμογή του **ESA**, ενώ με το **bKES** χρειάζεται αντίστοιχα να εφαρμοστεί  $N$  φορές η μέθοδος **IFAE**, για τον υπολογισμό του ίδιου αριθμού πλατών και συχνοτήτων.

#### 4.4.2 Εφαρμογή σε συνθετικά σήματα

Μελετάται η εφαρμογή του αλγορίθμου στα δύο σήματα που παρουσιάστηκαν κατά την παρουσίαση του μοντέλου **bKES**.

##### Εφαρμογή σε σήμα δύο συνιστωσών με σταθερές κεντρικές συχνότητες

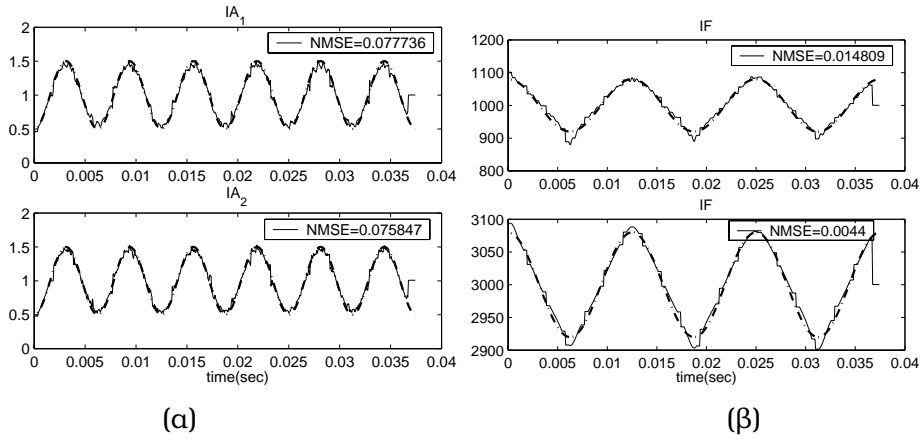
Για την ανάλυση του σήματος  $y_s$  (Εξίσωση 4.10), λόγω του ότι οι κεντρικές συχνότητες των συνιστωσών του παραμένουν σταθερές θεωρήθηκε, όπως και στην περίπτωση του **bKES**, ότι  $q_{f_i} = 10^{-4}$ . Όσο για τις τιμές των πινάκων  $Q_i$  αυτές θεωρήθηκαν και πάλι σχετικά μικρές, λαμβάνοντας τις τιμές που επιστρέφει η μέθοδος **IFAE** ως ιδιαίτερα αξιόπιστες. Συγκεκριμένα:

$$q_{A_i} = 0.0001, \quad q_{IF_i} = 1; \quad (4.27)$$

ενώ ο πίνακας  $R$  είναι  $R = 10^{-7} I_{N \times N}$  αφού δεν υπάρχει θόρυβος στο σήμα που παρατηρείται. Τα φίλτρα Gabor επιλέγονται με  $\alpha = 1500$  ενώ η αρχική κατάσταση του συστήματος είναι η ίδια με αυτή που δόθηκε για το μοντέλο **bKES** στην περίπτωση του ίδιου σήματος. Το μήκος  $N$  της περιόδου επιλέχθηκε ίσο με 31, πλαίσιο που με δεδομένη συχνότητα δειγματοληψίας ίση με  $F_s = 16000$  αντιστοιχεί σε περίπου 20ms. Τα αποτελέσματα φαίνονται στη γραφική παράσταση 4.4.

##### Εφαρμογή σε σήμα δύο συνιστωσών με γραμμικά μεταβαλλόμενες κεντρικές συχνότητες

Για την ανάλυση του σήματος  $y_v$  4.13 τιμή των  $q_{f_i}$  θεωρήθηκε ίση με τη μεταβολή των κεντρικών συχνοτήτων των συνιστωσών του σήματος ανά ένα πλαίσιο, δηλαδή



Σχήμα 4.4: Αποτελέσματα εφαρμογής της εκτίμησης Kalman με το μοντέλο *fKES* στο σήμα  $y(t) = [1 - 0.5 \cos(2\pi 160t)][\cos(2\pi 1000t + 2\pi 80 \int_0^t \cos(2\pi 80\tau) d\tau) + \cos(2\pi 3000t + 2\pi 80 \int_0^t \cos(2\pi 80\tau) d\tau)]$ . (α) Στιγμιαία πλάτη των δύο συνιστωσών του  $y$ , (β) στιγμιαίες συχνότητες των δύο συνιστωσών σε  $Hz$ . Με διακεκομμένες γραμμές είναι οι πραγματικές τιμές των σημάτων, ενώ στα υπομνήματα αναγράφεται και το κανονικοποιημένο μέσο τετραγωνικό λάθος κατά περίπτωση.

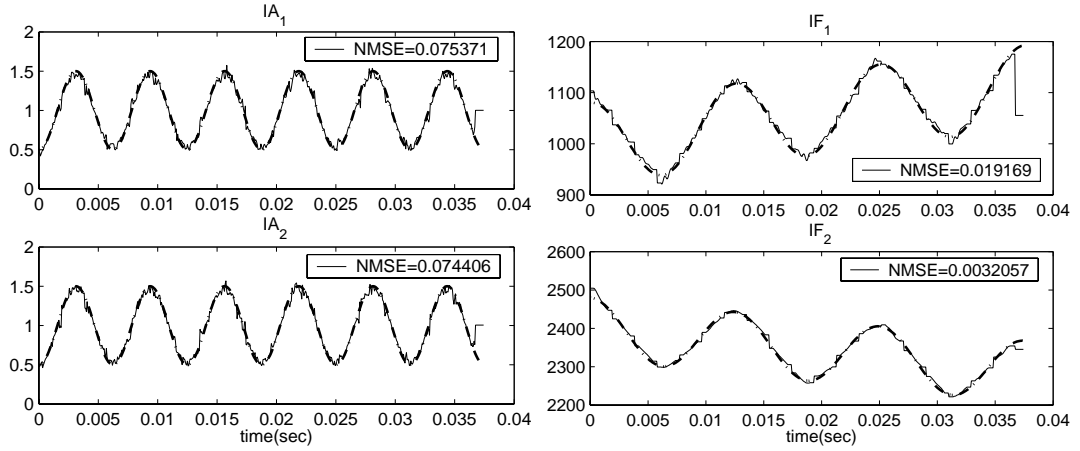
$q_{f_i} = 31 \times 3000 / 16000 = 5.8125$ . Ο πίνακας  $Q$  είναι λίγο διαφορετικός από ότι προηγουμένως με δεδομένο ότι γενικά αναμένεται η μέθοδος *IFAE* να έχει μεγαλύτερο σφάλμα όσο οι κεντρικές συχνότητες πλησιάζουν. Γι' αυτό επιλέχθηκαν:

$$q_{A_i} = 0.01, \quad q_{IF_i} = 10 \quad (4.28)$$

ενώ ο πίνακας  $R$  είναι και πάλι  $R = 10^{-7} I_{31 \times 31}$ , όπου  $I_{31 \times 31}$  ο μοναδιαίος πίνακας διάστασης 31. Οι αρχικές κεντρικές συχνότητες των δύο συνιστωσών θεωρούνται αντίστοιχα ότι είναι  $1000 Hz$  και  $2400 Hz$  ενώ τα στιγμιαία πλάτη και οι φάσεις θεωρούνται ότι είναι μηδενικά στην αρχή. Τα αποτελέσματα του αλγορίθμου φαίνονται στη γραφική παράσταση 4.5. Άμεσα διαπιστώνεται ότι τα αποτελέσματα της εκτίμησης με το *fKES* είναι χειρότερα αυτής που γίνεται με το *bKES*. Παρ' όλ' αυτά το γεγονός ότι τα λάθη είναι ανεκτά και η μορφή των κυματομορφών παραμένει αναλλοίωτη καθώς και η αρκετά μειωμένη πολυπλοκότητα του δεύτερου μοντέλου το καθιστούν ιδιαίτερα ελκυστικό.

## 4.5 Αλγόριθμος ανάλυσης με βάση το μοντέλο *iKES*

Προκειμένου να αποφευχθεί η θορυβώδης εκτίμηση του πλάτους που προκύπτει από το προηγούμενο μοντέλο πραγματοποιώντας μια μορφή εξομάλυνσης, προτείνε-



Σχήμα 4.5: Αποτελέσματα εφαρμογής της εκτίμησης Kalman με το μοντέλο *fKES* στο σήμα  $y(t) = [1 - 0.5 \cos(2\pi 160t)][\cos(2\pi 1000t + 2\pi \int_0^t 3000\tau + 80 \cos(2\pi 80\tau)d\tau) + \cos(2\pi 2400t + 2\pi \int_0^t (-3000)\tau + \cos(2\pi 80\tau)d\tau)]$ . (α) Στιγμαιαία πλάτη των δύο συνιστωσών του  $y$ , (β) στιγμιαίες συχνότητες των δύο συνιστωσών σε  $Hz$ . Με διακεκομμένες γραμμές είναι οι πραγματικές τιμές των σημάτων, ενώ στα υπομνήματα αναγράφεται και το κανονικοποιημένο μέσο τετραγωνικό λάθος κατά περίπτωση.

ται μια παραλλαγή με μειωμένο αριθμό μεταβλητών κατάστασης. Η μείωση γίνεται στον αριθμό των μεταβλητών κατάστασης που παριστάνουν το στιγμιαίο πλάτος. Συγκεκριμένα, από αυτές κρατώνται μόνο δύο για κάθε συνιστώσα, αυτές που παριστάνουν το στιγμιαίο πλάτος στην αρχή ( $\alpha_i^L$ ) και στο τέλος του πλαισίου ( $\alpha_i^H$ ) αντίστοιχα, πλαισίο το οποίο έχει μήκος  $N$ . Οι ενδιάμεσες τιμές του πλάτους προκύπτουν από γραμμική παρεμβολή μεταξύ των δύο αυτών τιμών και έτσι το νέο μοντέλο, με  $2N + 3$  εξισώσεις κατάστασης, έχει ως εξής :

$$f_i(k) = f_i(k-1) + q_{f_i} w_{f_i}(k) \quad (4.29)$$

$$\alpha_i^L(k) = \mathbf{IFAE}(s, f_i(k-1), bw_i, (k-1)N+1) + q_{\alpha_i^L} w_{\alpha_i^L} \quad (4.30)$$

$$\alpha_i^H(k) = \mathbf{IFAE}(s, f_i(k-1), bw_i, kN) + q_{\alpha_i^H} w_{\alpha_i^H} \quad (4.31)$$

$$IF_i^k = \mathbf{IFAE}_{frame}(s, f_i(k-1), bw_i, k) + Q_{IF_i} W_{IF_i}^k \quad (4.32)$$

$$\begin{aligned} \varphi_i(j) &= \varphi_i(j-1) + 2\pi T_s IF_i(j-1), \\ j &= (k-1)N+1 \dots kT \end{aligned} \quad (4.33)$$

όπου  $\alpha_i^L(k)$  είναι η τιμή του πλάτους στην αρχή του πλαισίου  $k$ ,  $\alpha_i^H(k)$  η αντίστοιχη τιμή στο τέλος του πλαισίου  $k$  ενώ οι υπόλοιποι συμβολισμοί είναι όπως περιγράφηκαν στο προηγούμενο μοντέλο. Σημειώνεται ότι οι διαδικασίες  $w_{\alpha_i^L}, w_{\alpha_i^H}$  είναι ανεξάρτητες  $N(0, 1)$  και πολλαπλασιασμένες με τις αντίστοιχες παραμέτρους  $q_{\alpha_i^L}, q_{\alpha_i^H}$  εκφράζουν τα όποια σφάλματα της μεθόδου **IFAE**.

Οι εξισώσεις μέτρησης (πλήθους  $N$ ) έχουν τη μορφή:

$$Y^k = \sum_{i=1}^K A_i^k \cdot * \cos(\Phi_i^k) + RU^k \quad (4.34)$$

όπου

$$A_i^k = (\alpha_{i,k}(j)) \quad \text{με} \quad (4.35)$$

$$\alpha_{i,k}(j) = \alpha_i^L(k) + \frac{\alpha_i^H(k) - \alpha_i^L(k)}{N-1} (j - [(k-1)N + 1]), \quad (4.36)$$

$$j = (k-1)N + 1 \cdots kN$$

#### 4.5.1 Αρχικοποίηση και εκτίμηση με χρήση του εκτεταμένου φίλτρου Kalman

Και με το συγκεκριμένο μοντέλο, παρά τη μείωση των μεταβλητών κατάστασης είναι ιδιαίτερα δαπανηρή η εκτίμηση με χρήση του φίλτρου Kalman χωρίς καθοδήγηση και για αυτό τα πειράματα πραγματοποιούνται με το εκτεταμένο φίλτρο. Όσο για την αρχικοποίηση του αλγορίθμου, τα πράγματα δεν αλλάζουν και ισχύουν όσα αναφέρθηκαν και για το προηγούμενο μοντέλο.

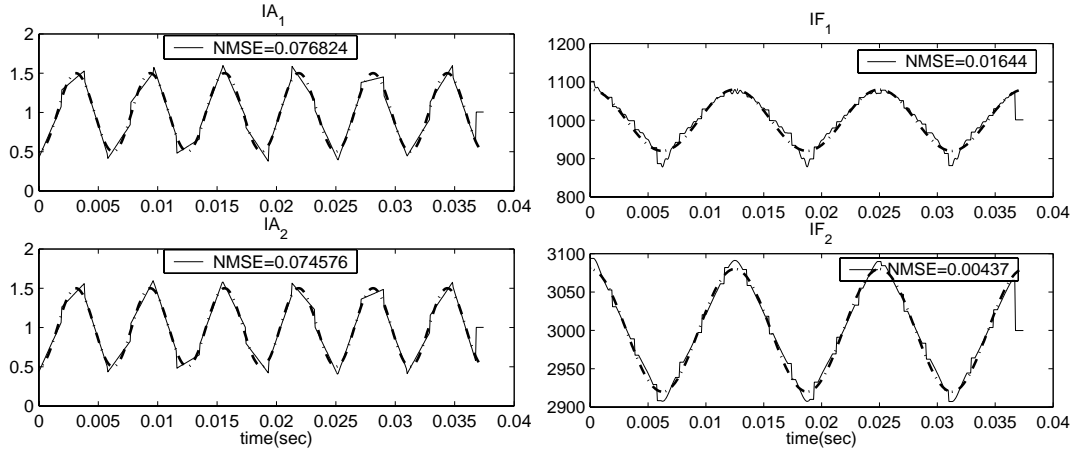
#### 4.5.2 Εφαρμογή σε συνθετικά σήματα

##### Εφαρμογή σε σήμα με δύο συνιστώσες με σταθερές γραμμικές συχνότητες

Στο σχήμα 4.6 φαίνονται τα στιγμιαία πλάτη και οι στιγμιαίες συχνότητες των συνιστωσών του σήματος  $y_s$  (Εξίσωση 4.10), όπως προκύπτουν με βάση το μοντέλο *iKES*. Είναι φανερή η εξομάλυνση που έχει επιτευχθεί, χωρίς όμως να έχει μειωθεί το κανονικοποιημένο μέσο τετραγωνικό λάθος. Δοκιμάστηκε και πλαίσιο με μήκος μικρότερο από  $N = 31$ , χωρίς όμως να παρατηρηθούν σημαντικές διαφορές στα αποτελέσματα.

##### Εφαρμογή σε σήμα με δύο συνιστώσες με γραμμικά μεταβαλλόμενες κεντρικές συχνότητες

Μετά από την παρατήρηση και των εκτιμήσεων των διαφόρων χαρακτηριστικών του σήματος  $y_v$  (Εξ. 4.13) με το μοντέλο *iKES* (Σχήμα 4.7) οδηγείται κανείς σε ανάλογα με τα προηγούμενα συμπεράσματα. Η εξομάλυνση που επιτυγχάνεται δεν μειώνει επαρκώς το σφάλμα εκτίμησης ενώ μπορεί και να αλλοιώσει τη μορφή των κυματομορφών σε κάποιο βαθμό. Από την άλλη, είναι ευκολότερη η ταυτοποίηση συστήματος για το τελευταίο μοντέλο, λόγω του ότι οι παράμετροι  $\alpha_i$ , οι οποίες επηρεάζουν σε μεγάλο βαθμό όλα τα μοντέλα, είναι ουσιαστικά μόνο δύο για κάθε



Σχήμα 4.6: Αποτελέσματα εφαρμογής της εκτίμησης Kalman με το μοντέλο *iKES* στο σήμα  $y(t) = [1 - 0.5 \cos(2\pi 160t)][\cos(2\pi 1000t + 2\pi 80 \int_0^t \cos(2\pi 80\tau) d\tau) + \cos(2\pi 3000t + 2\pi 80 \int_0^t \cos(2\pi 80\tau) d\tau)]$ . (α) Στιγμαίαια πλάτη των δύο συνιστωσών του  $y$ , (β) στιγμιαίες συχνότητες των δύο συνιστωσών σε  $Hz$ . Με διακεκομμένες γραμμές είναι οι πραγματικές τιμές των σημάτων, ενώ στα υπομνήματα αναγράφεται και το κανονικοποιημένο μέσο τετραγωνικό λάθος κατά περίπτωση.

συνιστώσα και έτσι υπό μία έννοια με το *iKES* συνδυάζεται η ευκολία στην αρχικοποίηση του *bKES* με τη μειωμένη πολυπλοκότητα του *fKES* με τα δεδομένα βέβαια αποτελέσματα, που σχολιάστηκαν προηγουμένως.

## 4.6 Αλγόριθμος ανάλυσης με βάση το μοντέλο *mKES*

Η τελευταία παραλλαγή που παρουσιάζεται είναι αρκετά διαφορετική από τις προηγούμενες αλλά είναι πιο κοντά στον αλγόριθμο MBDA. Είναι μάλλον η πιο φυσική επέκταση του αλγορίθμου αυτού ώστε να συμπεριλάβει πληροφορία από τον αλγόριθμο διαχωρισμού με υπολογισμό της ενέργειας. Το μοντέλο για το οποίο γίνεται λόγος, έχει τις ίδιες εξισώσεις κατάστασης με αυτό του MBDA αλλά διαφορετικές εξισώσεις μέτρησης. Συγκεκριμένα, για την κατάσταση είναι:

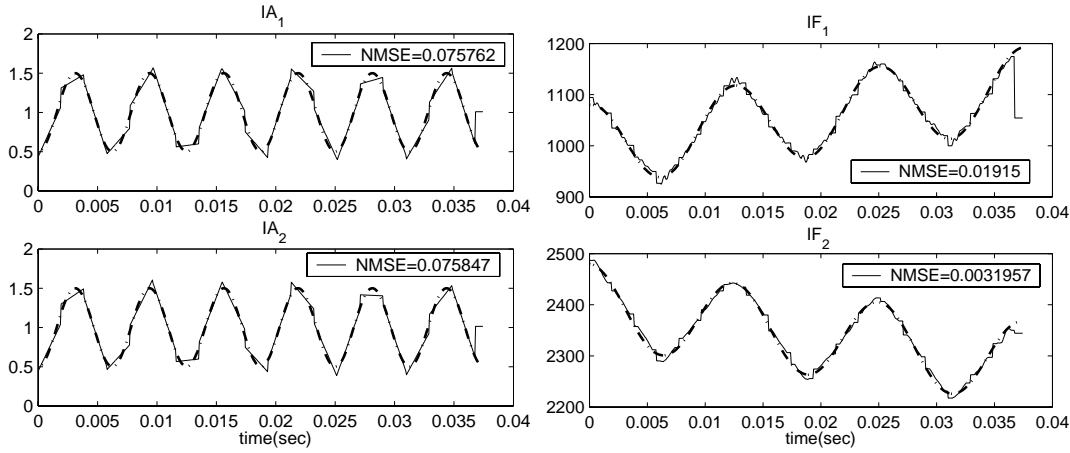
$$f_i(k) = f_i(k-1) + q_{f_i} w_{f_i}(k-1) \quad (4.37)$$

$$\alpha_i(k) = \alpha_{\alpha_i} \alpha_i(k-1) + q_{\alpha_i} w_{\alpha_i}(k-1) \quad (4.38)$$

$$v_i(k) = \alpha_{v_i} v_i(k-1) + q_{v_i} w_{v_i}(k-1) \quad (4.39)$$

$$\varphi_i(k) = \varphi_i(k-1) + 2\pi T_s f_i(k-1) + 2\pi T_s v_i(k-1) \quad (4.40)$$





Σχήμα 4.7: Αποτελέσματα εφαρμογής της εκτίμησης Kalman με το μοντέλο *iKES* στο σήμα  $y(t) = [1 - 0.5 \cos(2\pi 160t)][\cos(2\pi 1000t + 2\pi \int_0^t 3000\tau + 80 \cos(2\pi 80\tau)d\tau) + \cos(2\pi 2400t + 2\pi \int_0^t (-3000)\tau + \cos(2\pi 80\tau)d\tau)]$ . (α) Στιγμιαία πλάτη των δύο συνιστωσών του  $y$ , (β) στιγμιαίες συχνότητες των δύο συνιστωσών σε  $Hz$ . Με διακεκομμένες γραμμές είναι οι πραγματικές τιμές των σημάτων, ενώ στα υπομνήματα αναγράφεται και το κανονικοποιημένο μέσο τετραγωνικό λάθος κατά περίπτωση.

ενώ προστίθενται δύο νέες εξισώσεις μέτρησης για κάθε συνιστώσα, μία που αφορά στο άθροισμα των συχνοτήτων  $f_i$  και  $v_i$  και μία για την απόλυτη τιμή του πλάτους  $\alpha_i$ :

$$y(k) = \sum_{i=1}^K \alpha_i \cos(\varphi_i(k)) + ru(k) \quad (4.41)$$

$$y_{i,1}(k) = f_i(k) + v_i(k) + r_{i,1}u_{i,1}(k) \quad (4.42)$$

$$y_{i,2}(k) = |\alpha_i(k)| + r_{i,2}u_{i,2}(k) \quad (4.43)$$

όπου οι διαδικασίες  $u_{i,1}$ ,  $u_{i,2}$  παριστάνουν τον θόρυβο παρατήρησης στις αντίστοιχες εξισώσεις, ενώ  $r_{i,1}$ ,  $r_{i,2}$  καθορίζουν τη μεταβλητότητα του θορύβου αυτού. Οι υπόλοιποι συμβολισμοί είναι ίδιοι με αυτούς στην υποενότητα 3.2

Αυτό που χαρακτηρίζει το μοντέλο, είναι ότι οι μεταβλητές κατάστασης που μας ενδιαφέρουν εμφανίζονται στην έξοδο σε μια πιο άμεση μορφή από αυτή της πρώτης εξίσωσης μέτρησης. Οι τιμές για τις μετρήσεις αυτές προέρχονται από τη μέθοδο υπολογισμού του στιγμιαίου πλάτους και της στιγμιαίας συχνότητας (IFAE) με τα φίλτρα Gabor να είναι κεντραρισμένα στις εκ των προτέρων εκτιμήσεις των  $f_i$ , δηλαδή πριν αυτές ενημερωθούν από τις μετρήσεις. Πρόκειται αναμφισβήτητα για θορυβώδεις μετρήσεις αλλά αυτό είναι κάτι το οποίο λαμβάνεται υπόψη μέσω των διαδικασιών  $u_{i,1}$ ,  $u_{i,2}$ .

Είναι αρκετά σημαντικό το γεγονός ότι μπορεί να ελεγχθεί άμεσα η επίδραση του ESA στο μοντέλο, απλά με τον καθορισμό κατάλληλων τιμών για τις παραμέτρους  $r_{i,1}$  και  $r_{i,2}$ . Για παράδειγμα, σχετικά μικρή τιμή της δεύτερης και μεγάλη της πρώτης θα σημαίνει ότι έχει μεγάλη επίδραση η συχνότητα που μετράται μέσω του *IFAE* στην ανανέωση των εκ των προτέρων εκτιμήσεων των μεταβλητών κατάστασης ενώ θα αγνοηθεί σε μεγάλο βαθμό η μέτρηση του πλάτους.

#### 4.6.1 Αρχικοποίηση και εκτίμηση με χρήση φίλτρου Kalman

Η επιλογή των διαφόρων παραμέτρων του φίλτρου γίνεται γενικά με τον τρόπο που ακολουθείται και από τους Doerschuk & Lu, βλ. ενότητα 3.4. Δηλαδή βρίσκεται η αυτοσυσχέτιση της διαδικασίας που περιγράφεται από το μοντέλο και υπολογίζεται η φασματική πυκνότητα της διαδικασίας και στη συνέχεια γίνεται προσπάθεια προσαρμογής της ώστε να ταιριάζει με τη φασματική πυκνότητα ισχύος του υπό ανάλυση σήματος. Οι τιμές των  $q_{f_i}$  επιλέγονται ανάλογα με την αναμενόμενη μεταβολή ανά δείγμα των κεντρικών συχνοτήτων των διαφόρων συνιστωσών του σήματος. Όσο για τις νέες παραμέτρους, αυτές δηλαδή που δεν εμφανίζονται στο μοντέλο του MBDA οι οποίες είναι οι  $r_{i,1}, r_{i,2}$  και τα εύρη ζώνης των φίλτρων Gabor, καθορίζονται διαισθητικά περισσότερο ανάλογα με το πόσο αποτελεσματικός αναμένεται να είναι ο ESA.

Για την εκτίμηση χρησιμοποιήθηκε το εκτεταμένο φίλτρο Kalman. Οι εξισώσεις κατάστασης είναι ήδη γραμμικές οπότε δεν χρειαζόταν ιδιαίτερος υπολογισμός των Ιακωβιανών πινάκων  $A, G$ . Είναι μπλοκ διαγώνιοι, με ένα μπλοκ,  $A_i, G_i$  αντίστοιχα, για κάθε συνιστώσα:

$$A_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha_{\alpha_i} & 0 & 0 \\ 0 & 0 & \alpha_{v_i} & 0 \\ 2\pi T_s & 0 & 2\pi T_s & 1 \end{pmatrix} \quad G_i = \begin{pmatrix} q_{f_i} & 0 & 0 \\ 0 & q_{\alpha_i} & 0 \\ 0 & 0 & q_{v_i} \\ 0 & 0 & 0 \end{pmatrix} \quad (4.44)$$

ενώ οι πίνακες  $C$  και  $H$  είναι και αυτοί εύκολα προσδιορίσιμοι. Σημειώνεται ότι ο πίνακας  $C$  είναι διαστάσεων  $3 \times 4K$  ή  $C = (C_i), i = 1 \dots K$ , όπου  $K$  ο συνολικός αριθμός των συνιστωσών στο σήμα:

$$C_i = \begin{pmatrix} 0 & \cos(\varphi_i(k)) & 0 & -\alpha_i(k) \sin(\varphi_i(k)) \\ 1 & 0 & 1 & 0 \\ 0 & \pm 1 & 0 & 0 \end{pmatrix} \quad \text{ενώ} \quad (4.45)$$

$$H = \begin{pmatrix} r & 0 & 0 \\ 0 & r_{i,1} & 0 \\ 0 & 0 & r_{i,2} \end{pmatrix} \quad (4.46)$$

όπου θεωρείται ότι  $\frac{d|\alpha_i(k)|}{d\alpha_i(k)} = \pm 1$  όπου το θετικό πρόσημο ισχύει όταν  $\alpha_i(k) > 0$  ενώ το αρνητικό όταν  $\alpha_i(k) < 0$ .

Έτσι, το βάρος της πολυπλοκότητας του αλγορίθμου είναι κυρίως στις εξισώσεις μέτρησης, όπου εφαρμόζεται φιλτράρισμα Gabor και στη συνέχεια ESA.

### 4.6.2 Εφαρμογή σε συνθετικά σήματα

#### Εφαρμογή σε σήμα με δύο συνιστώσες με σταθερές κεντρικές συχνότητες

Ο αλγόριθμος εφαρμόστηκε στο σήμα  $y_s$  (Εξ. 4.10) αφού αρχικοποιήθηκαν οι παράμετροι του μοντέλου με τη μέθοδο που περιγράφηκε στην προηγούμενη ενότητα. Συγκεκριμένα, θεωρήθηκε  $\alpha_{\alpha_i} = 0.99$ ,  $q_{\alpha_i} = 0.1$  και  $\alpha_{v_i} = 0.99$ ,  $q_{v_i} = 14$ . Με τις παραμέτρους αυτές, η φασματική πυκνότητα ισχύος της στοχαστικής διαδικασίας που περιγράφεται από το μοντέλο φαίνεται στη γραφική παράσταση σε σύγκριση με την πραγματική φασματική πυκνότητα ισχύος του σήματος. Επιπλέον, επιλέχθηκε  $q_{f_i} = 0.001$  αφού οι κεντρικές συχνότητες των συνιστωσών παραμένουν σταθερές. Τέλος, επιλέχθηκε  $r_{i,1} = 0.01, r_{i,2} = 10$  και το εύρος ζώνης του φίλτρου Gabor ίσο με  $\alpha = 1500$ , όπως και σε όλες τις περιπτώσεις που παρουσιάστηκαν.

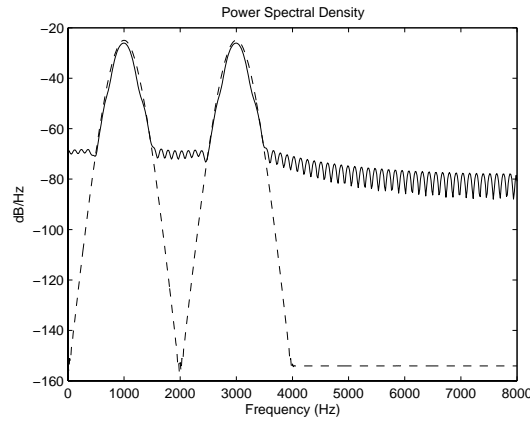
Η αρχική κατάσταση του συστήματος θεωρείται ότι είναι :

$$x_0 = \begin{pmatrix} f_1 \\ \alpha_1 \\ v_1 \\ \varphi_1 \\ f_2 \\ \alpha_2 \\ v_2 \\ \varphi_2 \end{pmatrix}_{n=0} = \begin{pmatrix} 1000 \\ 0 \\ 0 \\ 0 \\ 3000 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (4.47)$$

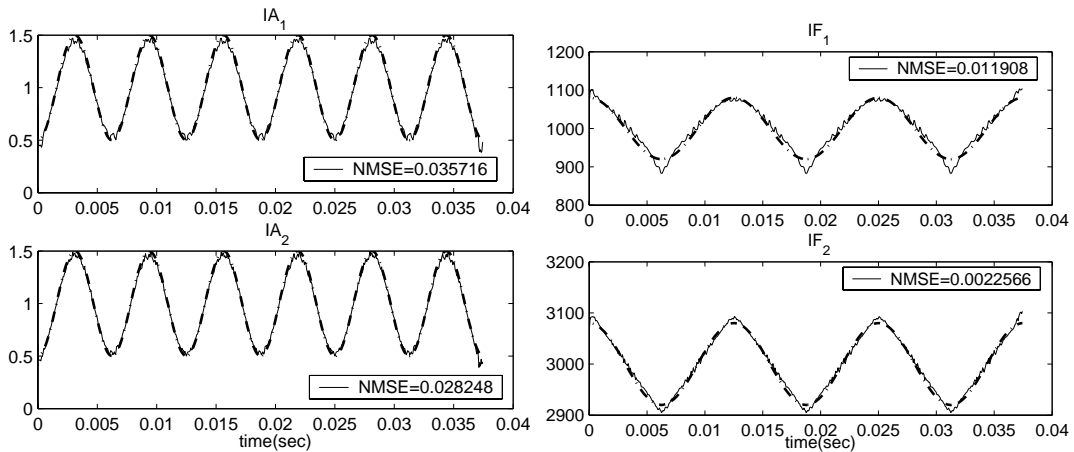
$$P_{x_0} = \text{diag} (0.1 \quad 0.01 \quad 14^2 \quad 0 \quad 0.1 \quad 0.01 \quad 14^2 \quad 0) \quad (4.48)$$

#### Εφαρμογή σε σήμα με δύο συνιστώσες με γραμμικά μεταβαλλόμενες κεντρικές συχνότητες

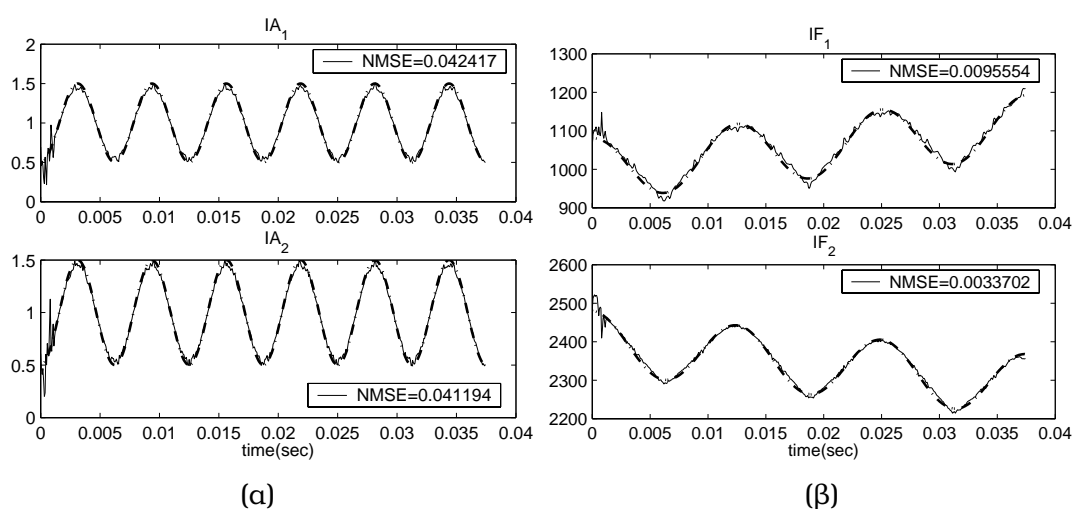
Το μοντέλο δεν αλλάζει ιδιαίτερα σε σχέση με προηγουμένως για την ανάλυση του σήματος  $y_v$ , Εξ. 4.13. Οι αλλαγές αφορούν στις παραμέτρους που εμφανίζονται στις εξισώσεις των κεντρικών συχνοτήτων. Οι τιμές των  $q_{f_i}$  μετά από εμποπεία και κάποια μικρή ρύθμιση των παραμέτρων επιλέχθηκαν τελικά ίσες με 4, δηλαδή αρκετά μεγαλύτερες από τη γνωστή μεταβολή ανά δείγμα των κεντρικών συχνοτήτων ( $\approx 0.2Hz/\text{δείγμα}$ ). Εντοπίστηκε ανάλογο θέμα και με τον Kalman-MBDA οπότε, μάλλον δεν ευθύνεται η εισαγωγή των νέων εξισώσεων μέτρησης για αυτή την απόκλιση. Αλλάζει και η αρχική τιμή της κεντρικής συχνότητας της δεύτερης συνιστώσας του σήματος και γίνεται ίση με  $2400Hz$ . Τα αποτελέσματα του αλγορίθμου φαίνονται στο Σχήμα 4.10. Τα μέσα τετραγωνικά λάθη δείχνουν ότι οι επιδόσεις του



Σχήμα 4.8: Φασματική πυκνότητα ισχύος του σήματος  $y_s$ , (Εξ. 4.10), (συνεχής γραμμή) και της στοχαστικής διαकाσίας που προκύπτει από το μοντέλο  $mKES$ . Η πρώτη υπολογίζεται με τη μέθοδο Welch ενώ η δεύτερη προκύπτει από μετασχηματισμό Fourier της αυτοσυσχέτισης για τον υπολογισμό της οποίας δίνονται κάποια στοιχεία στην ενότητα 3.4 για δεδομένες τιμές παραμέτρων



Σχήμα 4.9: Αποτελέσματα εφαρμογής της εκτίμησης Kalman με το μοντέλο  $mKES$  στο σήμα  $y(t) = [1 - 0.5 \cos(2\pi 160t)][\cos(2\pi 1000t + 2\pi 80 \int_0^t \cos(2\pi 80\tau) d\tau) + \cos(2\pi 3000t + 2\pi 80 \int_0^t \cos(2\pi 80\tau) d\tau)]$ . (α) Στιγμαίια πλάτη των δύο συνιστωσών του  $y$ , (β) στιγμιαίες συχνότητες των δύο συνιστωσών σε  $Hz$ . Με διακεκομμένες γραμμές είναι οι πραγματικές τιμές των σημάτων, ενώ στα υπομνήματα αναγράφεται και το κανονικοποιημένο μέσο τετραγωνικό λάθος κατά περίπτωση.



Σχήμα 4.10: Αποτελέσματα εφαρμογής της εκτίμησης Kalman με το μοντέλο *mKES* στο σήμα  $y(t) = [1 - 0.5 \cos(2\pi 160t)][\cos(2\pi 1000t + 2\pi \int_0^t 3000\tau + 80 \cos(2\pi 80\tau)d\tau) + \cos(2\pi 2400t + 2\pi \int_0^t (-3000)\tau + \cos(2\pi 80\tau)d\tau)]$ . (α) Στιγμιαία πλάτη των δύο συνιστωσών του  $y$ , (β) στιγμιαίες συχνότητες των δύο συνιστωσών σε  $Hz$ . Με διακεκομμένες γραμμές είναι οι πραγματικές τιμές των σημάτων, ενώ στα υπομνήματα αναγράφεται και το κανονικοποιημένο μέσο τετραγωνικό λάθος κατά περίπτωση.

τελευταίου μοντέλου είναι ανάλογες του πρώτου. Το μειονέκτημα του  $mKES$  είναι ουσιαστικά η αύξηση των παραμέτρων του μοντέλου του MBDA και κατά συνέπεια η αύξηση της δυσκολίας για την επίτευξη κατάλληλης αρχικοποίησης του συστήματος ενώ από την άλλη, το  $bKES$  μπορεί να έχει λιγότερες παραμέτρους από το μοντέλο του MBDA αλλά υστερεί από πλευράς πολυπλοκότητας. Πάντως, αν επιχειρούνταν συγκριτική αξιολόγηση των διαφόρων μοντέλων με βάση τα πειράματα που έχουν παρουσιαστεί ως τώρα, τα δύο μοντέλα θα κατατάσσονταν στην ίδια θέση. Στη συνέχεια παρουσιάζεται μια περισσότερο συστηματική έρευνα, η οποία θα μπορούσε να διαλευκάνει περισσότερο ποιες είναι οι δυνατότητες των διαφόρων μοντέλων και ποιοι οι περιορισμοί.

## 4.7 Επιδόσεις των διαφόρων παραλλαγών με την παρουσία θορύβου

Σημαντικό πλεονέκτημα του MBDA και των παραλλαγών με τη χρήση του ESA που προτάθηκαν είναι ότι ο θόρυβος εμπεριέχεται « φυσικά » στα διάφορα μοντέλα και με αυτόν τον τρόπο γίνεται εύκολα και με καλή απόδοση η προσαρμογή τους σε διάφορα περιβάλλοντα θορύβου, απλά με την αλλαγή κάποιων παραμέτρων. Αυτό επιβεβαιώθηκε και στην πράξη μετά την πραγματοποίηση κάποιων ενδεικτικών πειραμάτων, οπότε και εφαρμόστηκαν οι διάφορες προσεγγίσεις στο σήμα  $y_s(t)$  (Εξίσωση 4.10) για 5 τιμές  $SNR$ , από 0 έως  $40dB$ . Παρουσιάζονται τα αποτελέσματα στο Σχήμα 4.11 όπου φαίνονται τα μέσα τετραγωνικά λάθη στον προσδιορισμό των στιγμιαίων πλατών και συχνοτήτων των δύο συνιστωσών του σήματος.

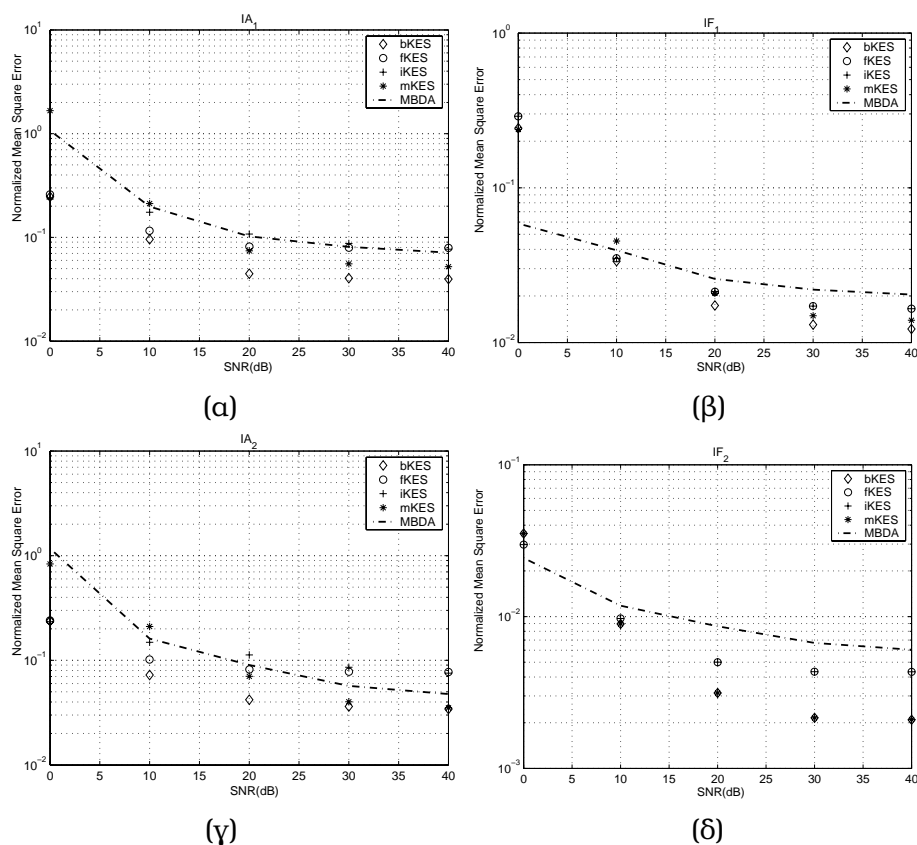
Από τα διαγράμματα είναι φανερό ότι συγκριτικά οι επιδόσεις των παραλλαγών που παρουσιάστηκαν δεν διαφέρουν ιδιαίτερα μεταξύ τους. Αξίζει να σημειωθεί πάντως ότι η καλύτερη επίδοση επιτυγχάνεται με το μοντέλο  $bKES$  γενικά για κάθε  $SNR$  ενώ πρέπει να παρατηρηθεί και η σε κάθε περίπτωση πολύ καλή εκτίμηση της στιγμιαίας συχνότητας (το σφάλμα γενικά μικρότερο του 10%).

Σημειώνεται ότι ο θόρυβος θεωρήθηκε μόνο στις εξισώσεις μέτρησης ενώ σίγουρα επηρεάζει και τις εξισώσεις κατάστασης ενώ σε καμιά περίπτωση δεν έγινε επαναπροσδιορισμός των διαφόρων παραμέτρων των μοντέλων, εκτός βέβαια από τις παραμέτρους που αφορούσαν στην ισχύ του θορύβου στις εξισώσεις μέτρησης.

## 4.8 Σύγκριση μεταξύ των προτεινόμενων παραλλαγών και αξιολόγηση

### 4.8.1 Επιδράσεις των χαρακτηριστικών διαμόρφωσης

Στη συνέχεια έγιναν κάποια ενδεικτικά πειράματα για να αναδειχτούν οι επιδράσεις που μπορεί να έχουν στα μοντέλα κάποια στοιχεία των διαμορφώσεων AM-FM.



Σχήμα 4.11: Επίδραση της αύξησης του θορύβου σε όλες τις μεθόδους εκτίμησης που έχουν παρουσιαστεί. Το σήμα που αναλύεται είναι αυτό της εξίσωσης 4.10 για διάφορες τιμές SNR, (α), (γ) Το κανονικοποιημένο μέσο τετραγωνικό λάθος στην εκτίμηση του στιγμιαίου πλάτους της πρώτης και της δεύτερης συνιστώσας του σήματος με όλες τις παραλλαγές και τον Kalman-MBDA για διάφορες τιμές του SNR, (β), (δ) Το κανονικοποιημένο μέσο τετραγωνικό λάθος στην εκτίμηση της στιγμιαίας συχνότητας των δύο συνιστωσών στις ίδιες περιπτώσεις. Για αναφορά, τα σημεία που αντιστοιχούν στον Kalman-MBDA είναι ενωμένα με διακεκομμένη γραμμή.

Έτσι εξετάστηκε σε κάποιο βαθμό η επίδραση του δείκτη διαμόρφωσης  $AM$ , του δείκτη διαμόρφωσης  $FM$  αλλά και της απόστασης μεταξύ των κεντρικών συχνοτήτων των συνιστωσών του σήματος. Το σήμα με το οποίο πραγματοποιήθηκαν οι δοκιμές είναι το :

$$y(t) = [1 - \kappa \cos(2\pi 160t)] [\cos(2\pi 1000t + 2\pi\beta \int_0^t \cos(2\pi 50\tau) d\tau) + \cos(2\pi 3000t + 2\pi\beta \int_0^t \cos(2\pi 150\tau) d\tau)] \quad (4.49)$$

Για το σήμα αυτό, ο δείκτης διαμόρφωσης  $AM$  κάθε συνιστώσας είναι ίσος με  $\kappa$  ενώ ο αντίστοιχος δείκτης διαμόρφωσης  $FM$  είναι ίσος με  $\beta$ . Υπενθυμίζεται ότι το  $\beta$  αντιπροσωπεύει το πηλίκο της μέγιστης διαμόρφωσης συχνότητας δια τη διαμορφώνουσα συχνότητα. Για τους σχετικούς ορισμούς βλ. [34].

### Μελέτη της επίδρασης του δείκτη διαμόρφωσης AM

Πάλι με κριτήριο το κανονικοποιημένο μέσο τετραγωνικό λάθος παρουσιάζονται στο Σχήμα 4.12 οι επιδόσεις των μοντέλων για διάφορες τιμές του δείκτη διαμόρφωσης  $AM$  και για δείκτη διαμόρφωσης  $FM$  σταθερό και ίσο με  $\beta = 1$ .

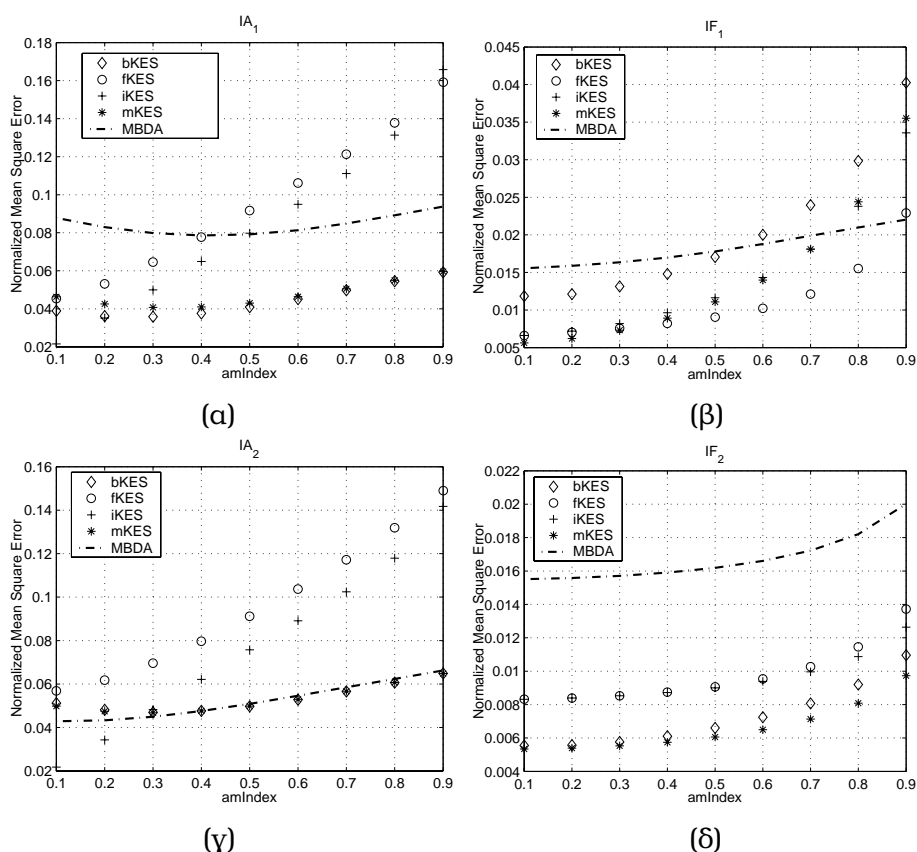
Είναι φανερό ότι μειώνεται το ποσοστό επιτυχίας όλων των αλγορίθμων καθώς αυξάνεται ο δείκτης διαμόρφωσης  $AM$  ενώ τις καλύτερες επιδόσεις φαίνεται να παρουσιάζει η χρήση του μοντέλου  $bKES$  και τις χειρότερες ο αλγόριθμος  $MBDA$ .

### Μελέτη της επίδρασης του δείκτη διαμόρφωσης FM

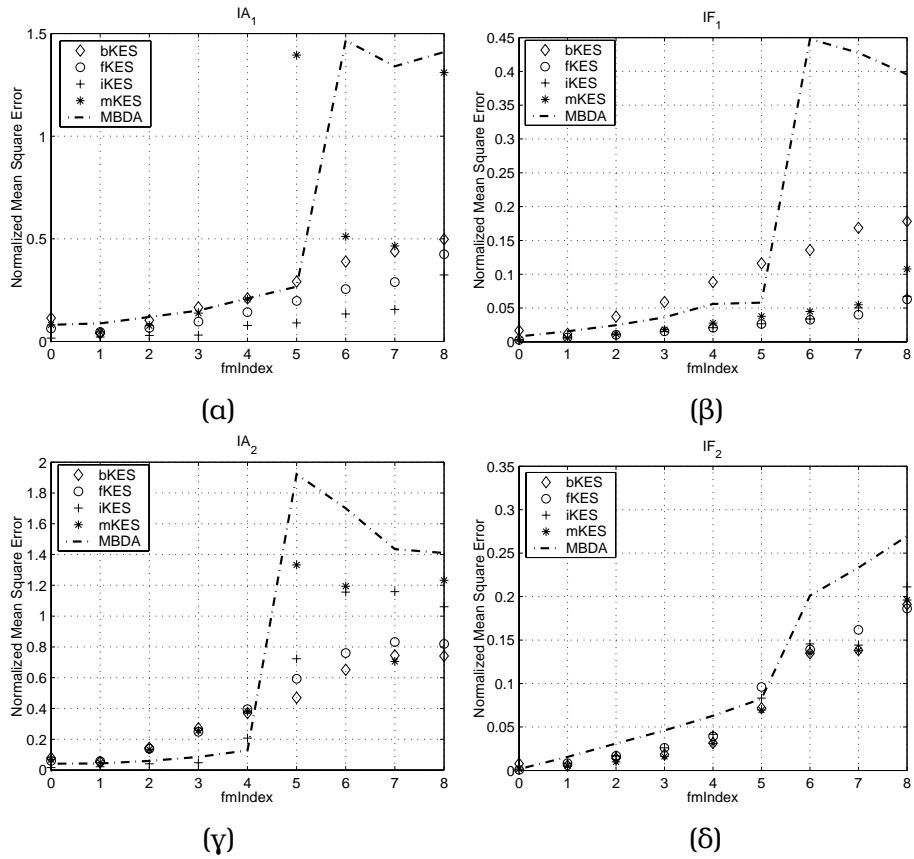
Ανάλογα πειράματα έγιναν και για τη μελέτη της επίδρασης του δείκτη διαμόρφωσης  $FM$  στις επιδόσεις των αλγορίθμων. Αυτή τη φορά ο δείκτης διαμόρφωσης  $AM$  διατηρήθηκε σταθερός και ίσος με  $\kappa = 0.1$  και έγιναν πειράματα με μεταβαλλόμενο δείκτη διαμόρφωσης  $FM$ ,  $\beta$ . Είναι φανερό ότι η επίδραση του δείκτη διαμόρφωσης  $FM$  είναι πολύ σημαντική. Υψηλός δείκτης διαμόρφωσης  $FM$  σημαίνει πολύ υψηλά λάθη στις εκτιμήσεις των αλγορίθμων. Σίγουρα ρόλο για αυτό παίζει το γεγονός ότι με την αύξηση του δείκτη διαμόρφωσης αυξάνεται σε μεγάλο βαθμό το εύρος ζώνης της κάθε συνιστώσας και υπάρχουν προβλήματα διαχωρισμού μεταξύ τους.

Σημειώνεται ότι δεν έγινε επαναπροσδιορισμός των παραμέτρων των διαφόρων παραλλαγών για κάθε διαφορετικό δείκτη διαμόρφωσης. Οι αλγόριθμοι αρχικοποιήθηκαν για το σήμα με  $\kappa = 0.1$  και  $\beta = 1$ . Ίσως με επιλογή διαφορετικού εύρους ζώνης για τα φίλτρα Gabor τα πράγματα να ήταν καλύτερα, ενώ για τον  $MBDA$  τα πράγματα ίσως βελτιώνονταν με κατάλληλο επαναπροσδιορισμό των  $q_{v_i}$ . Επαναπροσδιορισμός πάντως δεν έγινε με στόχο να εξεταστεί επιπλέον και η συμπεριφορά των αλγορίθμων σε περιβάλλοντα όπου δεν είναι εύκολη η ταυτοποίησή τους και ο επαναπροσδιορισμός των παραμέτρων τους, όπως συμβαίνει για παράδειγμα με την περίπτωση σημάτων φωνής. Τα αποτελέσματα φαίνονται στο Σχήμα 4.13.





Σχήμα 4.12: Επίδραση της μεταβολής του δείκτη διαμόρφωσης AM στην ανάλυση του σήματος που περιγράφεται στην εξίσωση 4.49 (α), (γ) Κανονικοποιημένα μέσα τετραγωνικά λάθη εκτίμησης των στιγμιαίων πλατών της πρώτης και της δεύτερης συνιστώσας του σήματος όπως προκύπτουν με τα διάφορα μοντέλα του KESDA και τον Kalman-MBDA, (β), (δ) αντίστοιχα λάθη για τις στιγμιαίες συχνότητες των δύο συνιστωσών. Σημειώνεται πως, για αναφορά, τα σημεία που αντιστοιχούν στα τετραγωνικά λάθη εκτίμησης από τον MBDA έχουν ενωθεί με διακεκομμένη γραμμή.



Σχήμα 4.13: Επίδρασης της αύξησης του δείκτη διαμόρφωσης FM με βάση τα μοντέλα που παρουσιάστηκαν (α), (γ) Μέσα τετραγωνικά λάθη της εκτίμησης των στιγμιαίων πλατών των δύο συνιστωσών του σήματος της εξίσωσης 4.49 με όλες τις νέες προτεινόμενες παραλλαγές και τον Kalman-MBDA για 9 διαφορετικές τιμές του δείκτη διαμόρφωσης  $\beta$  (β), (δ) Τα αντίστοιχα σφάλματα στις εκτιμήσεις των στιγμιαίων συχνοτήτων. Σημειώνεται ότι για αναφορά, τα σημεία που αντιστοιχούν στα σφάλματα του MBDA ενώνονται με διακεκομμένη γραμμή.

### **Μελέτη της επίδρασης της απόστασης μεταξύ των κεντρικών συχνοτήτων**

Η επίδραση του πλησιάζματος των κεντρικών συχνοτήτων των συνιστωσών είναι μάλλον αναμενόμενη. Τα αποτελέσματα χειροτερεύουν πολύ όταν αρχίζουν να συγχέονται οι φασματικές κορυφές του σήματος. Σχετικά πειράματα έδειξαν ότι για το συγκεκριμένο σήμα που παρουσιάστηκε προηγουμένως, τα σφάλματα είναι πάρα πολύ μεγάλα αν οι κεντρικές συχνότητες πλησιάσουν έτσι ώστε η απόσταση μεταξύ τους να είναι λίγο πιο μικρή από  $500\text{Hz}$ , με δεδομένο βέβαια το ότι τα φίλτρα *Gabor* που χρησιμοποιήθηκαν κατά την εκτίμηση είχαν εύρος ζώνης περίπου ίσο με  $600\text{Hz}$  ( $\alpha = 1500$ )

#### **4.8.2 Σύγκριση με τον Kalman-MBDA**

Συγκρίνοντας γενικά με τον Kalman-MBDA θα μπορούσε να σχολιαστεί ότι τουλάχιστον δύο από τα νέα μοντέλα, το *bKES* και το *mKES* έχουν καλύτερες επιδόσεις από τον αλγόριθμο του Lu, βλ. Κεφ. 3. Αυτό βέβαια έρχεται με επιβάρυνση, τουλάχιστον με τη δεδομένη υλοποίηση, της πολυπλοκότητας της εκτίμησης. Από την άλλη βέβαια τα καινούρια μοντέλα δεν είναι τόσο ευαίσθητα στις μεταβολές των παραμέτρων και μπορούν να χειριστούν διάφορες παραπλήσιες κατηγορίες σημάτων χωρίς επαναρχικοποίηση, ή τουλάχιστον χωρίς επαναρχικοποίηση στο βαθμό που τη χρειάζεται ο MBDA. Οι νέες προσεγγίσεις, όπως και να 'χει, κρίνοντας από τα αποτελέσματα που έχουν παρουσιαστεί ως τώρα, παρουσιάζουν ενδιαφέρον και αξίζει να δει κανείς τη συμπεριφορά τους στο χώρο φωνής, που σε κάποιο βαθμό εκτίθεται στο επόμενο κεφάλαιο.

## Κεφάλαιο 5

# Εφαρμογή της προσέγγισης KESDA στη φωνή

### 5.1 Εισαγωγή

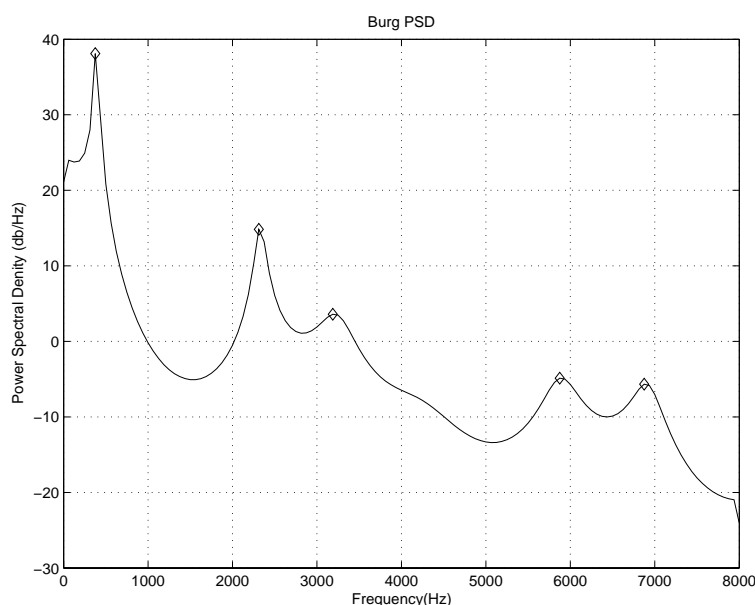
Στο προηγούμενο κεφάλαιο παρουσιάστηκαν αναλυτικά διάφορες παραλλαγές της προσέγγισης KESDA (Kalman Energy Separation Demodulation Algorithm), κάποιες εφαρμογές τους στο χώρο των συνθετικών σημάτων και μια πρώτη προσπάθεια συγκριτικής αξιολόγησής τους. Στο κεφάλαιο αυτό παρουσιάζεται μια αρχική μελέτη των επιδόσεών τους στο χώρο της φωνής, ο οποίος είναι ο κατεξοχήν χώρος για την ανάλυση του οποίου μελετώνται οι αλγόριθμοι και τα μοντέλα που έχουν περιγραφεί ως τώρα.

Η μελέτη επικεντρώνεται κυρίως στα μοντέλα *bKES* και *mKES*, επιλογή που στηρίζεται κυρίως στα συμπεράσματα του προηγούμενου κεφαλαίου για τις επιδόσεις των διαφόρων μοντέλων.

### 5.2 Εφαρμογή σε συνθετικά σήματα ομοιάζοντα φωνής

Κατ' αρχάς, κρίνεται σκόπιμο να παρουσιαστούν κάποια πειράματα που έγιναν σε συνθετικά σήματα τα οποία προσομοιάζουν σήματα φωνής. Η έμπνευση προήλθε από άρθρα του Maragos, [26], και Potamianos, [4], όπου χρησιμοποιούνται συνθετικά σήματα για την εξαγωγή χρήσιμων συμπερασμάτων. Τα σήματα αυτά σίγουρα δεν παρουσιάζουν τη συνθετικότητα των σημάτων φωνής, τα σχετικά πειράματα αποτελούν όμως ένα βήμα προς την καλύτερη κατανόηση της συμπεριφοράς του αλγορίθμου σε πολυπλοκότερα σήματα. Εξάλλου η όλη σχετική μελέτη οδήγησε σε κάποιες αποφάσεις που αφορούν στην κατάλληλη αρχικοποίηση των μοντέλων.

Το σήμα το οποίο αποτελεί το πρότυπο για τις προσομοιώσεις που παρουσιάζονται είναι το φώνημα /ee/ από τη λέξη "m/ee/ting" της βάσης TIMIT που εμφανίζεται



Σχήμα 5.1: Φασματική πυκνότητα ισχύος του φωνήματος /ee/ υπολογισμένη με τη μέθοδο Burg με μοντέλο τάξης 14

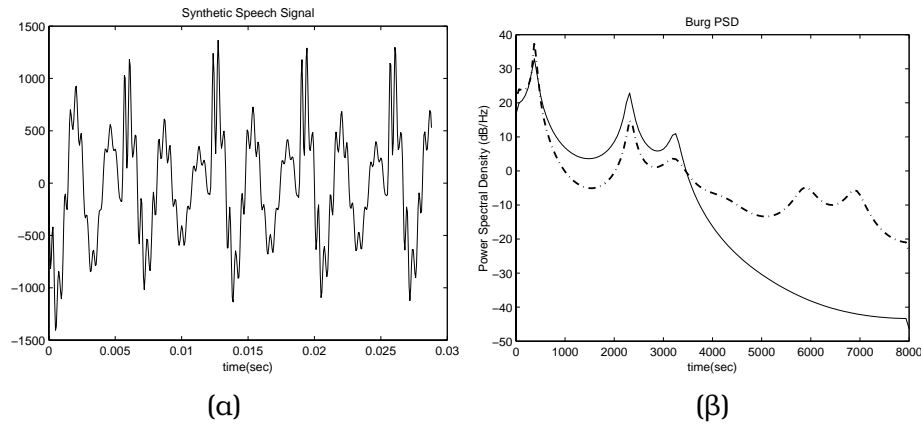
πρώτη φορά στο κεφάλαιο 3 σε εφαρμογές του MBDA σε σήματα φωνής (Σχήμα 3.2).

Το σήμα στο πεδίο του χρόνου φαίνεται στη γραφική παράσταση τάδε ενώ η φασματική πυκνότητα ισχύος του σήματος υπολογισμένη με τη μέθοδο Burg με τη χρήση μοντέλου 14ης τάξης φαίνεται στο διάγραμμα 5.1. Από τη φασματική πυκνότητα ισχύος, με έναν απλό αλγόριθμο επιλογής κορυφών, έγινε ένας υπολογισμός των θέσεων των συντονισμών του σήματος. Οι τιμές της πυκνότητας ισχύος για τους συντονισμούς αυτούς,  $fc = [375, 2312, 3250]$  είναι σημειωμένες στο διάγραμμα.

### 5.2.1 Εφαρμογή σε συνθετικά σήματα ομοιάζοντα φωνής με βάση το γραμμικό μοντέλο

Τα σήματα για τα οποία γίνεται λόγος είναι αποκρίσεις απλών γραμμικών συντονιστών, συστήματα γραμμικά χρονικά αμετάβλητα μόνο με πόλους σε κατάλληλες θέσεις, ώστε να είναι δυνατή η επίτευξη της επιθυμητής θέσης του συντονισμών αλλά και του εύρους ζώνης τους. Η διέγερση είναι περιοδική ακολουθία κρουστικών παλμών με κάποια συγκεκριμένη θεμελιώδη περίοδο  $T_0$ , ώστε να είναι δυνατή η προσομοίωση του pitch.

Θεωρήσαμε ένα τέτοιο σήμα με τρεις συντονισμούς (Σχήμα 5.2) και εφαρμόσαμε την προσέγγιση KESDA με τα μοντέλα *bKES* και *mKES*. Για να έχουμε ένα μέτρο σύγκρισης βέβαια, αρχικά θεωρήσαμε την απόκριση κάθε συντονισμού ως ξεχωριστό σήμα και εφαρμόσαμε τον αλγόριθμο διαχωρισμού με υπολογισμό της ενέργειας

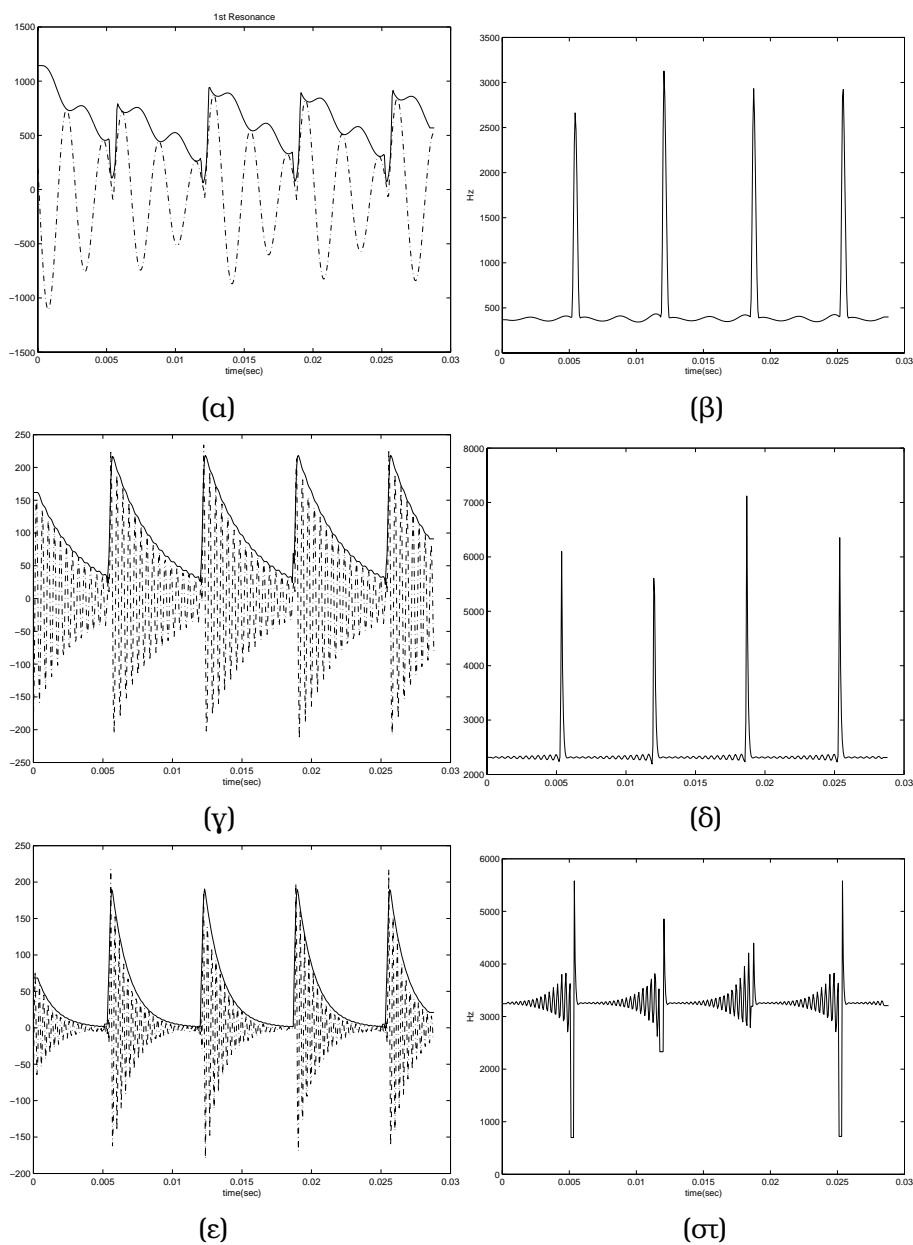


Σχήμα 5.2: Παρουσίαση του συνθετικού σήματος φωνής με συντονισμούς  $[375 \ 2312 \ 3250]$  και αντίστοιχα εύρη ζώνης  $[50 \ 100 \ 250]$  με θεμελιώδη συχνότητα ίση με  $150\text{Hz}$  (α) Το συνθετικό σήμα φωνής στο πεδίο του χρόνου, (β) φασματική πυκνότητα ισχύος του συνθετικού σήματος υπολογισμένη με τη μέθοδο Burg με μοντέλο τάξης 14 (Με διακεκομμένη γραμμή παριστάνεται η φασματική πυκνότητα ισχύος του φωνήματος /ee/)

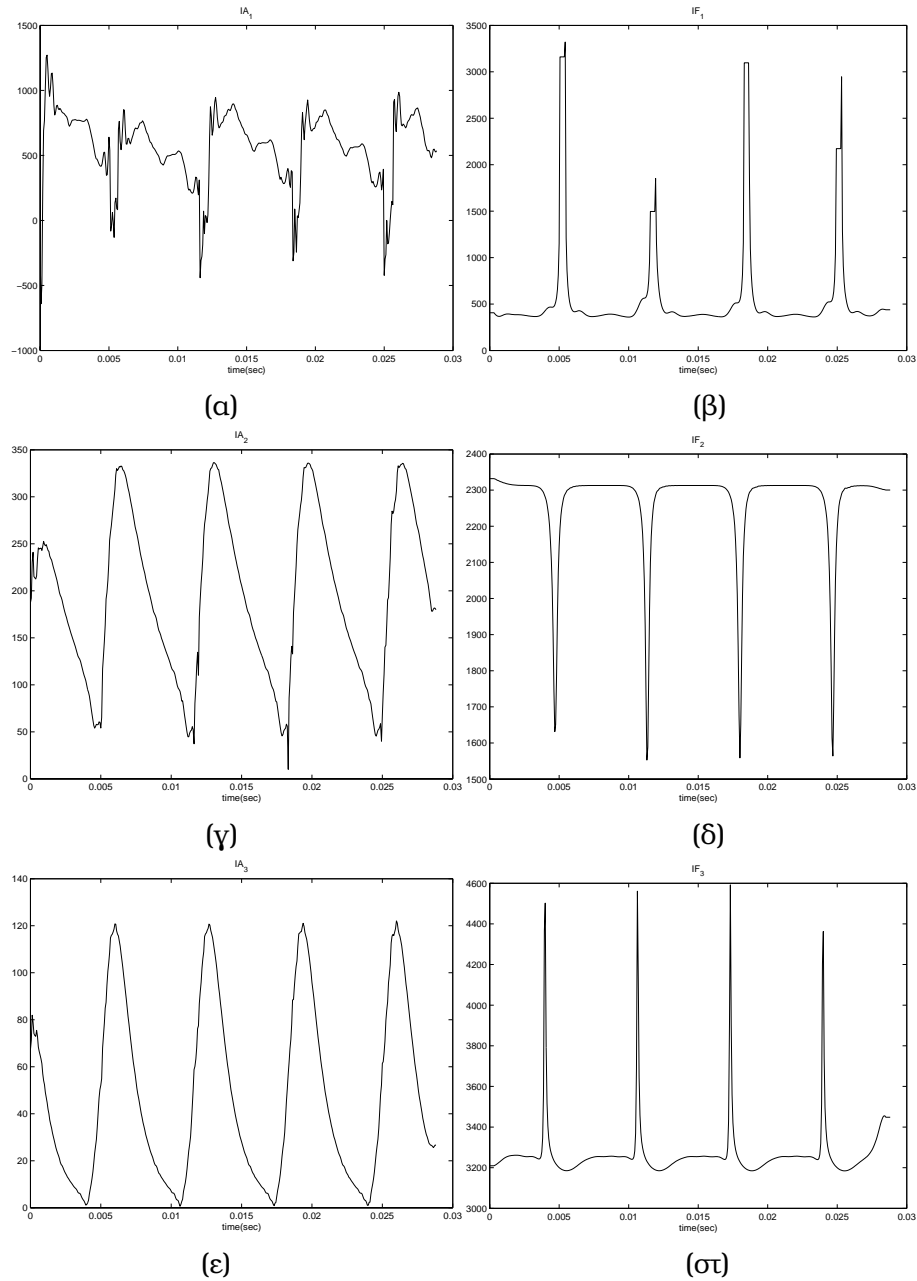
για να βρούμε τις περιβάλλουσες και τις στιγμιαίες συχνότητες κάθε συνιστώσας ξεχωριστά. Κάνοντας κάποιες απλές παρατηρήσεις στο Σχήμα 5.3 όπου φαίνονται τα σχετικά αποτελέσματα, μπορούμε να εντοπίσουμε δύο σημεία όπου ο αλγόριθμος ESA φαίνεται να αποκλίνει από την επιθυμητή συμπεριφορά, γεγονός που θα μπορούσε να επιβαρύνει τις εκτιμήσεις με βάση τα νέα μοντέλα.

Το πρώτο σημείο αφορά στον υπολογισμό των στιγμιαίων συχνοτήτων. Στα σημεία όπου η περιβάλλουσα του σήματος πλησιάζει στο μηδέν παρατηρούνται μεγάλες εκρήξεις (spikes) στην υπολογισμένη στιγμιαία συχνότητα. Οι εκρήξεις αυτές οφείλονται στο ότι ο ESA για τον υπολογισμό της συχνότητας διαιρεί με την τιμή του ενεργειακού τελεστή, η οποία στα εν λόγω σημεία είναι πολύ μικρή. Ενδεχόμενα κάτι τέτοιο, θα μπορούσε να επηρεάσει τα νέα μοντέλα που προτείνονται προκαλώντας κάποια απότομη αύξηση της φάσης που θα μπορούσε να τα οδηγήσει τελικά σε απόκλιση. Ένα άλλο πρόβλημα που φαίνεται ότι υπάρχει αφορά στην περιβάλλουσα του χαμηλόσυχνου συντονισμού η οποία, όπως υπολογίζεται, δεν έχει την αναμενόμενη μορφή. Η συχνότητα της διαμόρφωσης AM που οφείλεται στο pitch είναι περίπου ίση με τα δύο πέμπτα της συχνότητας του συντονισμού, αρκετά μεγάλη ώστε να προκαλεί, όπως φαίνεται, αποκλίσεις του ESA. Έχει σίγουρα ενδιαφέρον να δει κανείς πως μπορεί η χρήση του ESA με εκτίμηση Kalman να συμπεριφερθεί σε αυτές τις ανωμαλίες.

Στο Σχήμα (5.4) φαίνονται οι εκτιμήσεις των στιγμιαίων πλατών και συχνοτήτων όπως προκύπτουν από την εφαρμογή του μοντέλου *bKES* στο συνθετικό σήμα φωνής. Για τα φίλτρα Gabor που χρησιμοποιήθηκαν κατά την εφαρμογή, είναι  $\alpha = 1000$ . Είναι φανερό πως οι ανωμαλίες οι οποίες περιγράφηκαν προηγουμέ-



Σχήμα 5.3: (α), (γ), (ε) Με διακεκομμένες γραμμές παριστάνονται οι αποκρίσεις καθενός από τους συντονισμούς του σήματος στο σχήμα 5.2 και με συνεχείς γραμμές οι αντίστοιχες περιβάλλουσες όπως υπολογίζονται με χρήση του ESA, (β), (δ), (ε) Οι στιγμιαίες συχνότητες των αποκρίσεων όπως υπολογίζονται από τον ESA



Σχήμα 5.4: Αποτελέσματα εφαρμογής του *bKES* στο συνθετικό σήμα φωνής του σχήματος 5.2 (α), (γ), (ε) Οι εκτιμήσεις των πλατών καθεμιάς από τις τρεις συνιστώσες συνθετικού σήματος φωνής όπως υπολογίζονται με τη χρήση του *bKES* (β), (δ), (στ) Οι στιγμιαίες συχνότητες των τριών συνιστωσών όπως εκτιμώνται με βάση το ίδιο μοντέλο



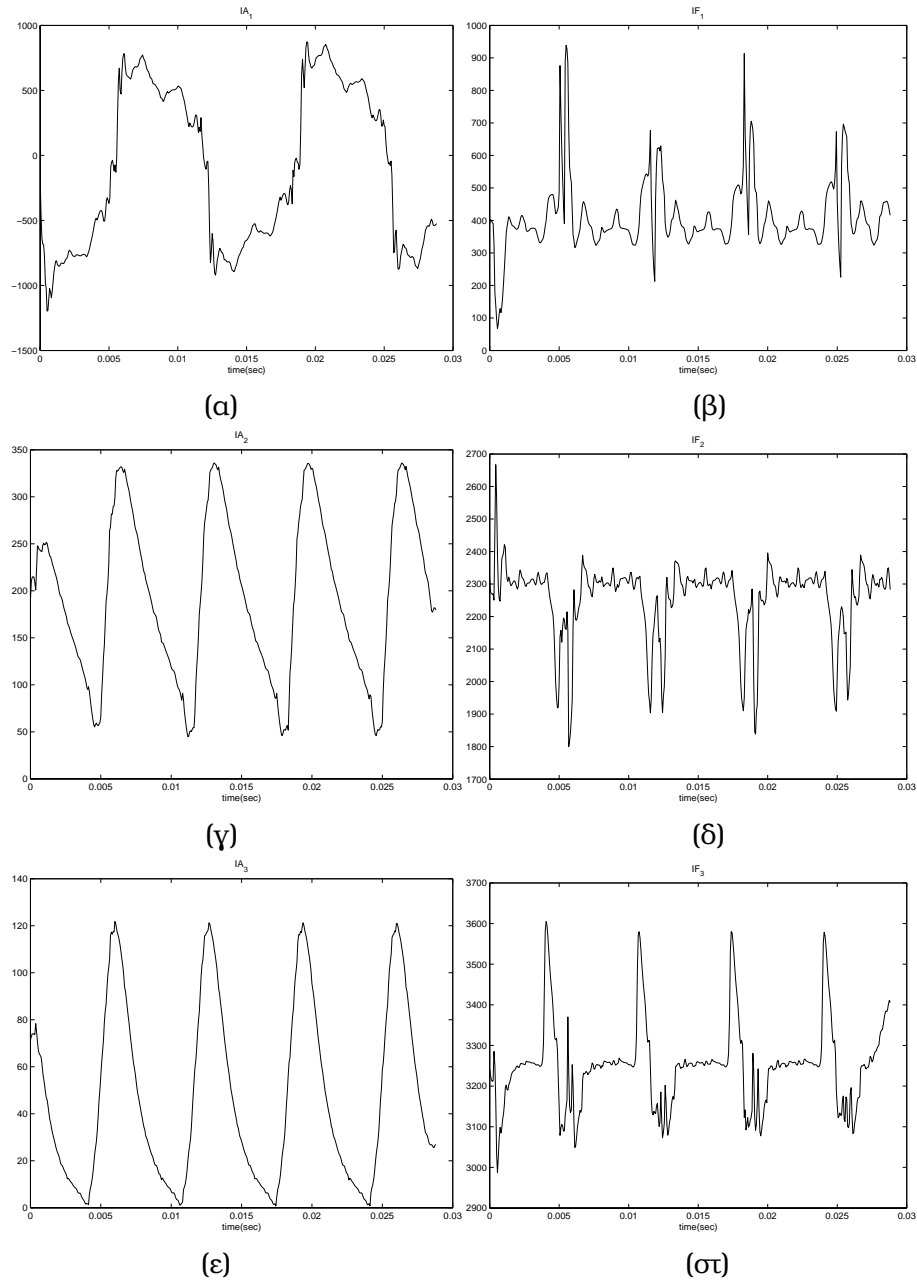
ως κληρονομήθηκαν και σε αυτή την προσέγγιση χωρίς όμως από ότι φαίνεται να δημιουργήσουν περαιτέρω προβλήματα με αποτέλεσμα τελικά η εκτίμηση των περιβαλλουσών, παρά το φιλτράρισμα με Gabor που έχει μεσολαβήσει να είναι αρκετά καλή. Για τις στιγμιαίες συχνότητες, από την άλλη, η συμπεριφορά είναι όμοια με αυτή του ESA. Τα συμπεράσματα διαφέρουν κάπως για το μοντέλο *mKES*. Όπως φαίνεται και από τις γραφικές παραστάσεις (Σχ.5.5) ενώ οι περιβάλλουσες των δύο υψίσυχων συντονισμών εκτιμώνται όπως και με τον ESA, για τον πρώτο συντονισμό η αντίστοιχη εκτίμηση είναι διαφορετική. Λόγω του ότι στο *mKES* το στιγμιαίο πλάτος δεν είναι περιορισμένο να παίρνει μόνο θετικές τιμές (όπως είναι σε μεγάλο βαθμό στο *bKES*) παρατηρείται η περιβάλλουσα να παίρνει και αρνητικές τιμές, γεγονός το οποίο βέβαια λόγω και της ιδιαιτερότητας της χαμηλόσυχνης συνιστώσας, δεν μπορεί να θεωρηθεί αρνητικό. Από την άλλη, οι εκτιμήσεις των στιγμιαίων συχνοτήτων φαίνονται να είναι περισσότερο θορυβώδεις μεν αλλά δεν παρουσιάζουν τις εκρήξεις που παρουσίαζαν τα αντίστοιχα σήματα με τις άλλες μεθόδους εκτίμησης.

Συμπερασματικά, η εφαρμογή των μοντέλων σε συνθετικό σήμα φωνής προερχόμενο από γραμμικούς συντονιστές ανέδειξε ενδεχόμενες δυσκολίες που μπορεί να αντιμετωπίσουν οι διάφορες προσεγγίσεις που χρησιμοποιούν τον ESA και την εκτίμηση Kalman για αποδιαμόρφωση. Στις δυσκολίες αυτές η συμπεριφορά και των δύο μοντέλων που εξετάστηκαν αν και διαφορετική, θα μπορούσε να χαρακτηριστεί ως ικανοποιητική σε κάποιο βαθμό. Σημειώνεται ότι για σύγκριση δίνονται και τα αποτελέσματα της εκτίμησης με χρήση του MBDA αν αυτός αρχικοποιηθεί όπως και το *mKES* (Σχήμα 5.6). Η συμπεριφορά του μοιάζει κάπως με αυτή του *mKES* ενώ αξιοσημείωτο είναι το γεγονός ότι η εκτίμηση της περιβάλλουσας της υψίσυχνης συνιστώσας παίρνει κάποια στιγμή αρνητικές τιμές, κάτι που μάλλον είναι ανεπιθύμητο και δεν μπορεί να δικαιολογηθεί, όπως γίνεται για τη χαμηλόσυχη συνιστώσα, αφού η συχνότητα της διαμόρφωσης κατά πλάτος, στην υψίσυχη συνιστώσα, είναι αρκετά μικρότερη της αντίστοιχης κεντρικής συχνότητας .

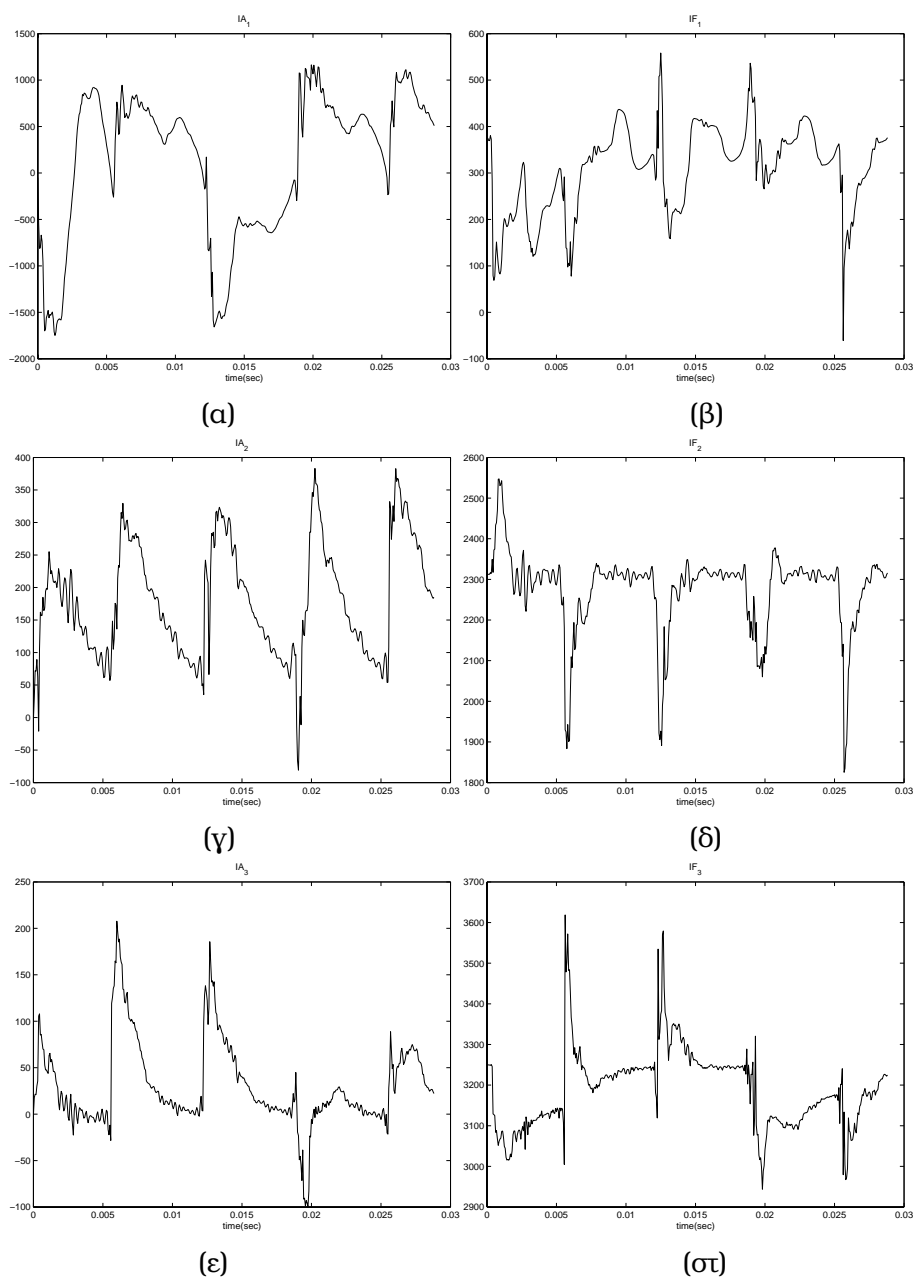
### 5.2.2 Εφαρμογή σε συνθετικά σήματα ομοιάζοντα φωνής με βάση το μη γραμμικό μοντέλο

Έγινε κάποια προσπάθεια να δημιουργηθούν προσομοιώσεις σημάτων φωνής και με βάση το μη γραμμικό μοντέλο φωνής που περιγράφηκε στην ενότητα 2.2. Συγκεκριμένα, θεωρήθηκε ότι το συνθετικό σήμα μπορεί να προκύψει ως υπέρθεση κατάλληλα διαμορφωμένων συνημιτόνων κατά πλάτος και κατά συχνότητα. Βέβαια δεδομένου ότι αυτό το μοντέλο επιτρέπει και κάποια μεταβλητότητα της κεντρικής συχνότητας του συντονισμού σε μικρή χρονική κλίμακα, θεωρήθηκε ότι θα πρέπει να ληφθεί υπόψη και μια ενδεχόμενη τέτοια μεταβολή. Για το σκοπό αυτό χρησιμοποιήθηκε ένας απλός αλγόριθμος παρακολούθησης συντονισμών ο οποίος περιγράφεται συνοπτικά στη συνέχεια.

Το σήμα χωρίζεται σε επικαλυπτόμενα πλαίσια και σε κάθε πλαίσιο υπολογίζεται η φασματική πυκνότητα ισχύος με τη μέθοδο Burg με μοντέλο τάξης 14. Στη συνέχεια, για κάθε πλαίσιο επιλέγονται με έναν απλό αλγόριθμο οι συχνότητες όπου η



Σχήμα 5.5: Αποτελέσματα εφαρμογής του μοντέλου *mKES* στο συνθετικό σήμα φωνής που παρουσιάζεται στο σχήμα 5.2 (α), (γ), (ε) Οι εκτιμήσεις των πλατών καθενιάς από τις τρεις συνιστώσες συνθετικού σήματος φωνής όπως υπολογίζονται με τη χρήση του *mKES* (β), (δ), (στ) Οι στιγμιαίες συχνότητες των τριών συνιστωσών όπως εκτιμώνται με βάση το ίδιο μοντέλο



Σχήμα 5.6: Αποτελέσματα εφαρμογής του MBDA σε συνθετικό σήμα φωνής που έχει δημιουργηθεί από γραμμικούς συντονιστές (Σχήμα 5.2), (α), (γ), (ε) Οι εκτιμήσεις των πλατών καθεμιάς από τις τρεις συνιστώσες του συνθετικού σήματος φωνής όπως υπολογίζονται με τη χρήση του MBDA (β), (δ), (στ) Οι στιγμιαίες συχνότητες των τριών συνιστωσών όπως εκτιμώνται με βάση τον ίδιο αλγόριθμο

φασματική πυκνότητα ισχύος εμφανίζει κορυφή και αυτές οι συχνότητες θεωρούνται ότι είναι οι συντονισμοί. Με αυτόν τον τρόπο, και με την επιβολή κάποιων περιορισμών για την επίτευξη της συνέχειας αλλά και την αποφυγή διασταυρώσεων μεταξύ διαφορετικών συντονισμών, είναι δυνατή η επίτευξη μιας πρόχειρης παρακολούθησης των συντονισμών του σήματος. Από τις τιμές των κεντρικών συχνοτήτων που υπολογίζονται έτσι, με γραμμική παρεμβολή προκύπτουν οι τιμές των κεντρικών συχνοτήτων των συνιστωσών του σήματος για κάθε χρονική στιγμή και όχι μόνο για κάθε πλαίσιο, και έτσι είναι δυνατή η άμεση χρησιμοποίησή τους στην κατασκευή του συνθετικού σήματος.

Για παράδειγμα, αν με τη διαδικασία που περιγράφηκε, προσδιορίζεται η εξέλιξη ενός συντονισμού του σήματος με το σήμα  $f_{ci}(t)$ , η συνιστώσα που αντιστοιχεί σε αυτόν τον συντονισμό θα είναι :

$$y_i(t) = \gamma_i r^{(tF_s) \bmod (F_s/F_0)} [1 + \kappa_i \cos(2\pi f_i^{AM} t)] \cos(2\pi \int_0^t f_{ci}(\tau) + f_i^{FM} \beta_i \cos(2\pi f_i^{FM} \tau) d\tau) \quad (5.1)$$

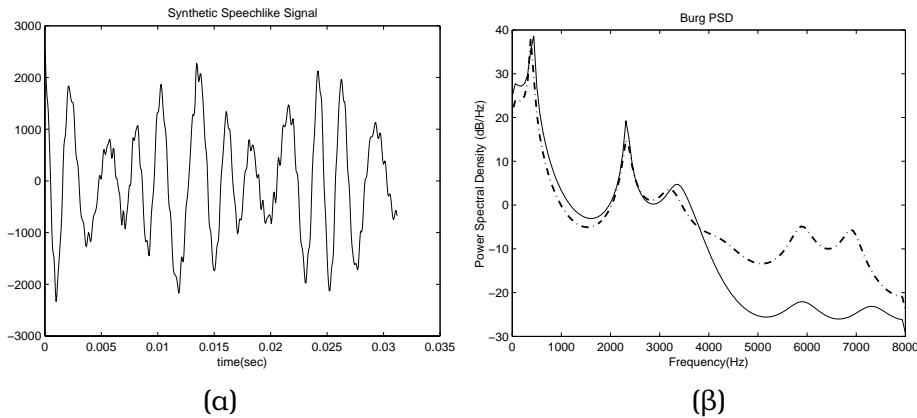
όπου  $\kappa_i$ ,  $\beta_i$  είναι οι συντελεστές διαμόρφωσης AM και FM της συνιστώσας ενώ  $\gamma_i$  είναι απλώς ένας συντελεστής που καθορίζει την ισχύ της.  $F_s$  είναι η συχνότητα με την οποία υποτίθεται ότι είναι δειγματοληπτημένο το σήμα ενώ  $F_0$  είναι η θεμελιώδης συχνότητα του σήματος (pitch). Όπως είναι φανερό, γίνεται προσπάθεια προσομοίωσης του pitch στο συνθετικό σήμα ώστε να αναδειχτούν ενδεχόμενες επιδράσεις του στον αλγόριθμο. Το συνθετικό σήμα τελικά προκύπτει ως υπέρθεση  $K$  τέτοιων συνιστωσών, όπου  $K$  είναι ο αριθμός των συντονισμών:

$$y(t) = \sum_{i=1}^K y_i(t) \quad (5.2)$$

Ένα τέτοιο σήμα, αποτελούμενο από τρεις συνιστώσες, καθώς και η φασματική πυκνότητα ισχύος του φαίνονται στο σχήμα 5.7. Δίνονται τα στοιχεία αυτού του σήματος ώστε να είναι εύκολη η αναφορά τους στην ανάλυση που ακολουθεί:

$$\begin{aligned} \mathbf{f}_c &\triangleq (f_{ci})_{i=1,2,3} = (375 \quad 2312 \quad 3250) & \gamma &\triangleq (\gamma_i)_{i=1,2,3} = (1500 \quad 150 \quad 60) \\ \kappa &\triangleq (\kappa_i)_{i=1,2,3} = (0.5 \quad 0.2 \quad 0.6) & \beta &\triangleq (\beta_i)_{i=1,2,3} = (1 \quad 1 \quad 3) \\ \mathbf{f}^{AM} &\triangleq (f_i^{AM})_{i=1,2,3} = (80 \quad 90 \quad 150) & \mathbf{f}^{FM} &\triangleq (f_i^{FM})_{i=1,2,3} = (120 \quad 120 \quad 120) \\ & & F_s &= 16000Hz \quad F_0 = 150Hz \end{aligned}$$

Με την εφαρμογή των μοντέλων *bKES* και *mKES*, έγινε προσπάθεια διαχωρισμού και αποδιαμόρφωσης των διαφόρων συνιστωσών του σήματος, με το σημαντικό πλεονέκτημα να είναι δυνατή η σύγκριση των εκτιμήσεων με τα πραγματικά σήματα των στιγμιαίων πλατών και συχνοτήτων των συνιστωσών, τα οποία είναι γνωστά. Για τη σύγκριση χρησιμοποιείται το κανονικοποιημένο μέσο τετραγωνικό λάθος ενώ με διακεκομμένες γραμμές παριστάνονται και τα πραγματικά σήματα σε όλα τα σχετικά διαγράμματα. Σημειώνεται ότι η παράμετρος  $\alpha$  όλων των φίλτρων Gabor που



Σχήμα 5.7: Παρουσίαση του σήματος της εξίσωσης 5.2 το οποίο προσομοιάζει σήμα φωνής με βάση το μη γραμμικό μοντέλο, (α) Συνθετικό σήμα φωνής στο πεδίο του χρόνου, (β) Φασματική πυκνότητα ισχύος του σήματος υπολογισμένη με τη μέθοδο Burg. Υπερτίθεται με διακεκομμένες γραμμές η φασματική πυκνότητα ισχύος του φωνήματος /ee/

χρησιμοποιούνται και στις δύο προσεγγίσεις λαμβάνεται ίση με 1500.

Από τα αποτελέσματα της εκτίμησης με το *bKES* (Σχ.5.8) θα μπορούσε ίσως κάποιος να βγάλει κάποια γενικότερα συμπεράσματα. Μια πρώτη παρατήρηση αφορά στην εκτίμηση της περιβάλλουσας της τρίτης συνιστώσας, όπου εμφανίζεται και το μεγαλύτερο λάθος. Είναι γεγονός πως η συνιστώσα αυτή έχει τη μικρότερη ισχύ από τις τρεις και η κεντρική της συχνότητα μεταβάλλεται από  $3400\text{Hz}$  έως και  $3000\text{Hz}$  ενώ λόγω της διαμόρφωσης FM έχει εύρος ζώνης τουλάχιστον ίσο με  $2(1+3)120 = 960\text{Hz}$  (κανόνας του Carson) το οποίο αυξάνεται κατά  $300\text{Hz}$  αν ληφθεί υπόψη και η διαμόρφωση AM τη στιγμή που η κεντρική συχνότητα της δεύτερης συνιστώσας, η οποία είναι αρκετά πιο ισχυρή, φτάνει μέχρι και τα  $2400\text{Hz}$  με εύρος ζώνης, μόνο λόγω της διαμόρφωσης FM, τουλάχιστον ίσο με  $2(1+1)120 = 480\text{Hz}$  ( $660\text{Hz}$  συνολικά λόγω και της διαμόρφωσης AM). Αυτό, σε συνδυασμό και με το ότι τα φίλτρα Gabor που χρησιμοποιούνται έχουν αποδοτικό rms εύρος ζώνης περίπου ίσο με  $600\text{Hz}$  δίνει μια εξήγηση στις αποκλίσεις που παρατηρούμε στην εκτίμηση του στιγμιαίου πλάτους της τρίτης συνιστώσας, ενώ με ανάλογα επιχειρήματα θα μπορούσε να εξηγηθεί και η σχετικά μεγάλη απόκλιση στον υπολογισμό της στιγμιαίας συχνότητας της ίδιας συνιστώσας, η οποία παρουσιάζει και τα spikes τα οποία σχολιάζει ο Potamianos στο άρθρο του [6] ως αποτελέσματα παρεμβολών. Όσον αφορά στην πρώτη συνιστώσα, από την άλλη, παρατηρούμε ότι τώρα που έχει εξασφαλιστεί ότι η συχνότητα διαμόρφωσης AM είναι αρκετά μικρότερη της κεντρικής συχνότητας η μορφή του εκτιμώμενου πλάτους είναι σύμφωνα με τα αναμενόμενα. Το μεγάλο σχετικά σφάλμα οφείλεται μάλλον σε μεταβατικά φαινόμενα στην αρχή.

Η εκτίμηση γενικά με τη χρήση του *bKES* κρίνεται ικανοποιητική και για του λόγου το αληθές δίνεται και το ανακατασκευασμένο σήμα καθώς και η φασματική

πυκνότητα ισχύος του (Σχ.5.9).

Βελτιωμένες, από την άλλη, με κριτήριο το μέσο τετραγωνικό λάθος, εμφανίζονται οι εκτιμήσεις με βάση το μοντέλο *mKES* (Σχήμα 5.10). Αν συγκριθούν κιόλας με τα αντίστοιχα αποτελέσματα του MBDA (Σχήμα 5.12), ο οποίος αρχικοποιείται ακριβώς όπως και το *mKES* φαίνεται ότι η χρήση του ESA οδήγησε σε αρκετά σημαντική μείωση κάποιες φορές του λάθους εκτίμησης. Χαρακτηριστικά αναφέρεται η περίπτωση της περιβάλλουσας της τρίτης συνιστώσας, όπου ο MBDA έδωσε αρνητική εκτίμηση ενώ αντίθετα, και παρά τις όποιες δυσκολίες που εμφανίζει ο διαχωρισμός της τρίτης συνιστώσας, η εκτίμηση με βάση το μοντέλο *mKES* μείωσε το λάθος σε περίπου 25%. Δίνεται το ανακατασκευασμένο σήμα και η φασματική πυκνότητα ισχύος του (Σχήμα 5.11).

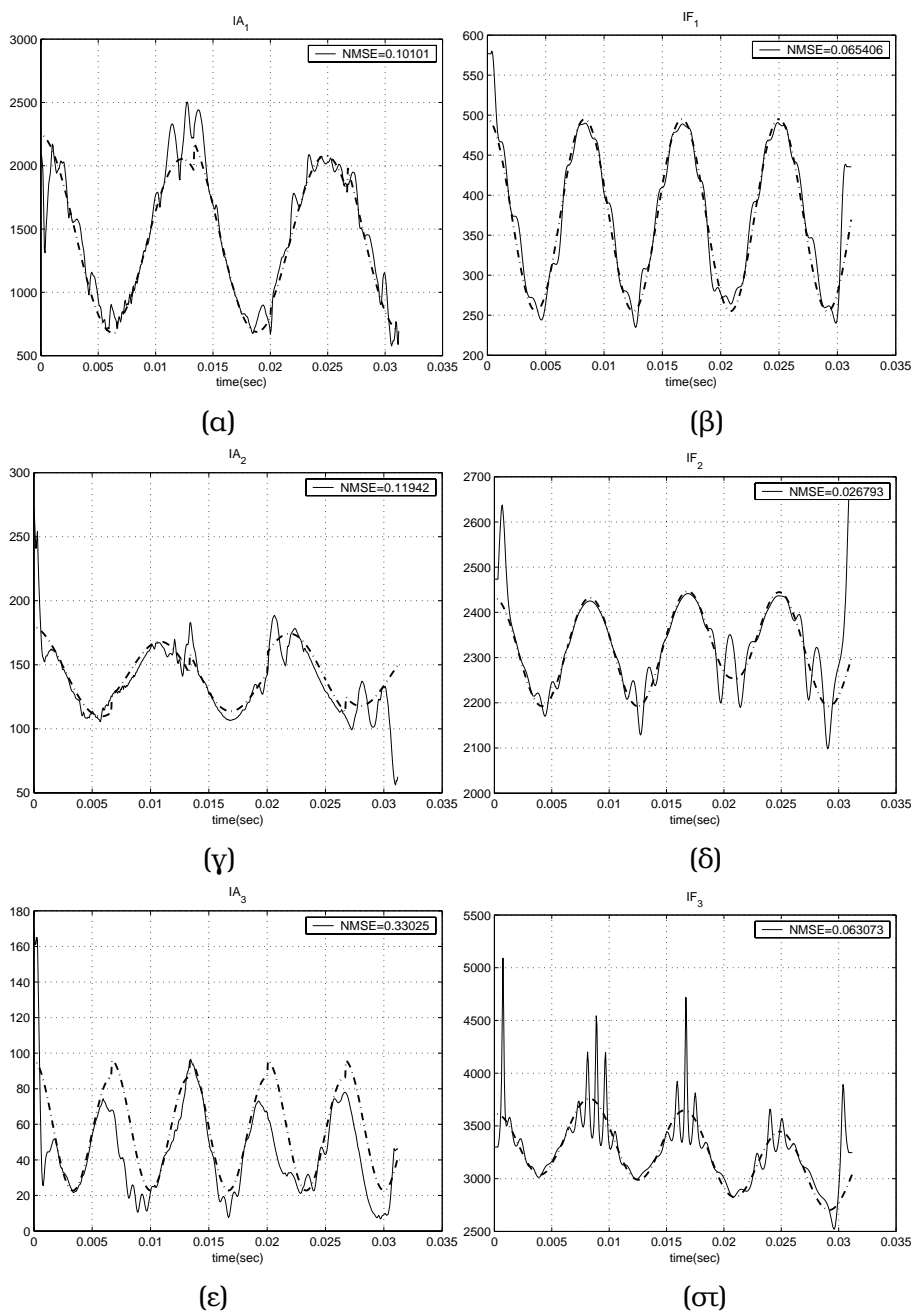
### 5.3 Εφαρμογή σε πραγματικά σήματα φωνής

Η εμπειρία που αποκομίζεται από την εφαρμογή των αλγορίθμων σε συνθετικά σήματα που προσομοιάζουν σήματα φωνής αποδεικνύεται σίγουρα χρήσιμη στην εφαρμογή των μοντέλων σε πραγματικά σήματα φωνής, ιδιαίτερα σε αποφάσεις που αφορούν στην ταυτοποίηση του συστήματος, ή αλλιώς στον προσδιορισμό των παραμέτρων του μοντέλου που χρησιμοποιείται. Η εφαρμογή των μοντέλων στη φωνή γίνεται κυρίως προς δύο κατευθύνσεις, η πρώτη από τις οποίες είναι η ανάλυση φωνής με στόχο την ανάδειξη διαφόρων χαρακτηριστικών που ενδεχόμενα φανερώνουν χρήσιμη πληροφορία για το σήμα της φωνής, και η δεύτερη είναι η παρακολούθηση των κεντρικών συχνοτήτων των συντονισμών της φωνής, εφαρμογή που, όπως έχει προαναφερθεί, έχει τύχει του ενδιαφέροντος πολλών ερευνητών λόγω της σημασίας που μπορεί να έχει στην επεξεργασία φωνής.

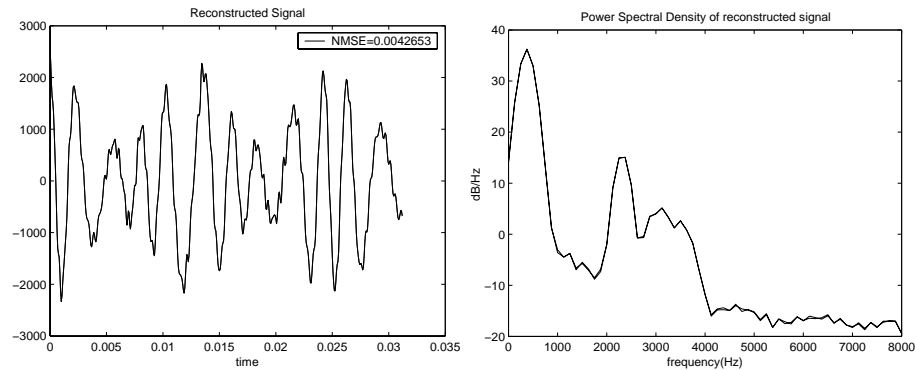
#### 5.3.1 Ανάλυση φωνής

Γίνεται ανάλυση με τη χρήση των μοντέλων *bKES*, *mKES* του φωνήματος /ee/ στη λέξη «m/ee/ting» από τη βάση TIMIT στα πλαίσια που έχουν τεθεί από τους Lu & Doerschuk όπως παρουσιάστηκαν στην ενότητα 3.6.2. Οι παράμετροι  $q_{f_i}$  αρχικοποιούνται με βάση την πρόχειρη παρακολούθηση των κεντρικών συχνοτήτων που περιγράφηκε σε προηγούμενη ενότητα. Παίρνουν την τιμή της μέγιστης μεταβολής ανά δείγμα της αντίστοιχης κεντρικής συχνότητας ενώ οι τιμές των  $q_{\alpha_i}$ ,  $q_{IF_i}$  προσδιορίζονται έχοντας σαν γνώμονα και τις επιδόσεις του μοντέλου στο συνθετικό σήμα φωνής. Στα διαγράμματα που ακολουθούν έχει γίνει υπέρθεση και των αποτελεσμάτων που εκτιμώνται με χρήση του MBDA, οποίος έχει αρχικοποιηθεί κατά τις υποδείξεις των εμπνευστών του.

Σημειώνεται ότι, ακολουθώντας το παράδειγμα των Lu & Doerschuk θεωρείται πως το συγκεκριμένο σήμα φωνής έχει 4 συντονισμούς στις συχνότητες 390, 2200, 2800 και 3600Hz αντίθετα ίσως με την εποπτεία που έχει κάποιος παρατηρώντας τη φασματική πυκνότητα ισχύος του φωνήματος (Σχήμα 5.1). Στο σχήμα 5.13 φαίνεται



Σχήμα 5.8: Ανάλυση με βάση το μοντέλο *bKES* του συνθετικού σήματος που προσομοιάζει σήμα φωνής και περιγράφεται στην εξίσωση 5.2. (α), (γ), (ε) οι εκτιμήσεις των στιγμιαίων πλατών των τριών συνιστωσών του σήματος (β), (δ), (σ) οι εκτιμήσεις των αντίστοιχων στιγμιαίων συχνοτήτων. Στα υπομνήματα φαίνονται τα αντίστοιχα μέσα τετραγωνικά λάθη ενώ με διακεκομμένες γραμμές είναι τα πραγματικά σήματα.



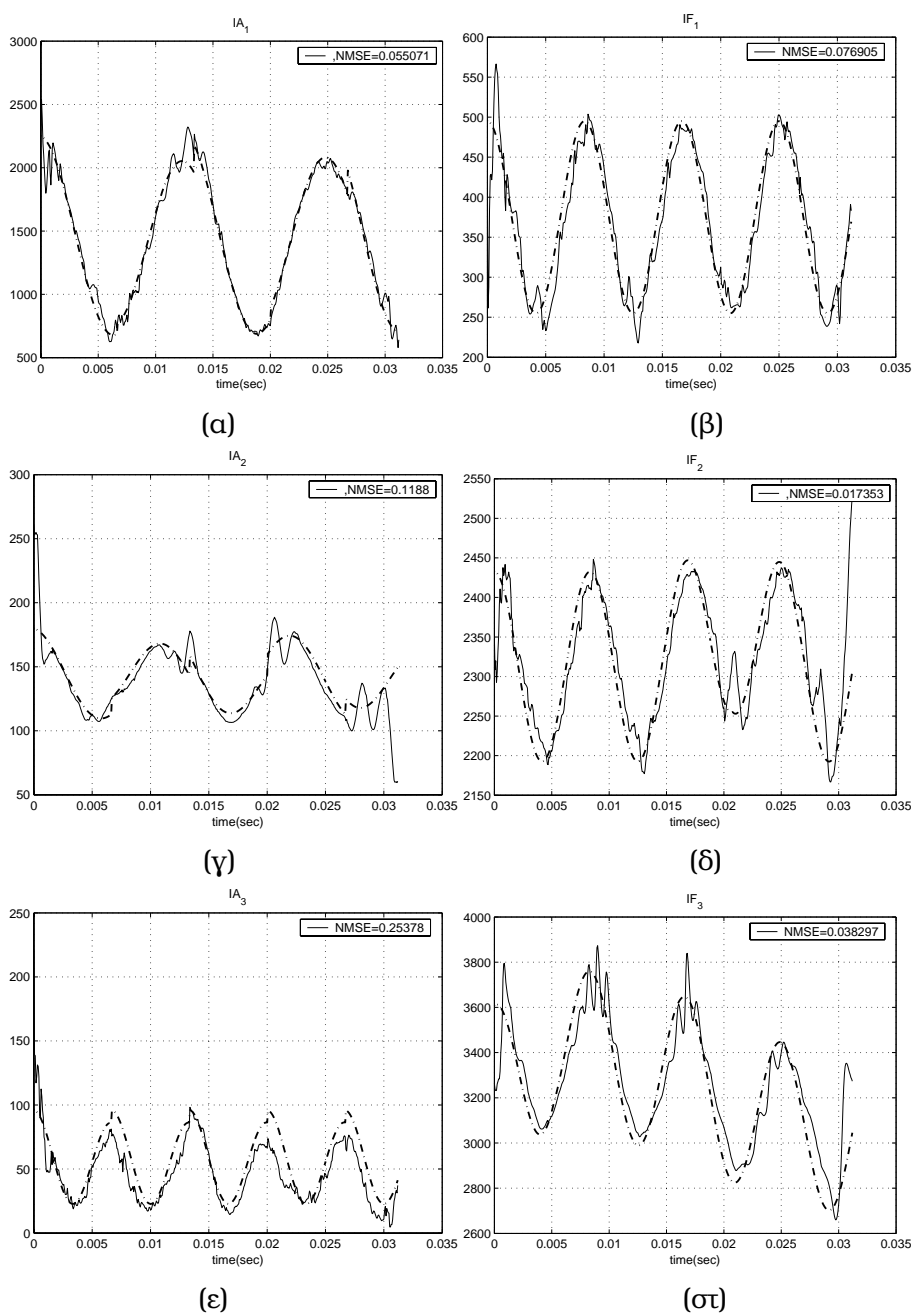
Σχήμα 5.9: Ανακατασκευή του συνθετικού σήματος φωνής της εξίσωσης 5.2 με βάση τις εκτιμήσεις με χρήση του μοντέλου *bKES* (α) Το ανακατασκευασμένο σήμα μετά την εκτίμηση με το *bKES*, (β) Η φασματική πυκνότητα ισχύος του ανακατασκευασμένου σήματος υπολογισμένη με τη μέθοδο Welch. Σημειώνεται ότι στις γραφικές παραστάσεις υπερτίθενται τα πραγματικά σήματα. Οι διαφορές όμως είναι τόσο μικρές που δεν διακρίνονται ξεχωριστές γραμμές.

πάντως η φασματική πυκνότητα ισχύος του μοντέλου που χρησιμοποιούν οι προσεγγίσεις MBDA και *mKES*, όπως αυτή υπολογίζεται με βάση τα σχετικά στοιχεία που δίνονται στην ενότητα 3.4. Για λόγους πληρότητας αλλά και για ευκολότερη σύγκριση, δίνονται ξανά τα αποτελέσματα της εφαρμογής του MBDA, μόνο που τώρα, αντί για τα σήματα  $\hat{v}_i$ , στα διαγράμματα του σχήματος 5.16 φαίνονται οι στιγμιαίες συχνότητες που προκύπτουν ως άθροισμα των σημάτων  $\hat{f}_i$  και  $\hat{v}_i$  για κάθε συνιστώσα.

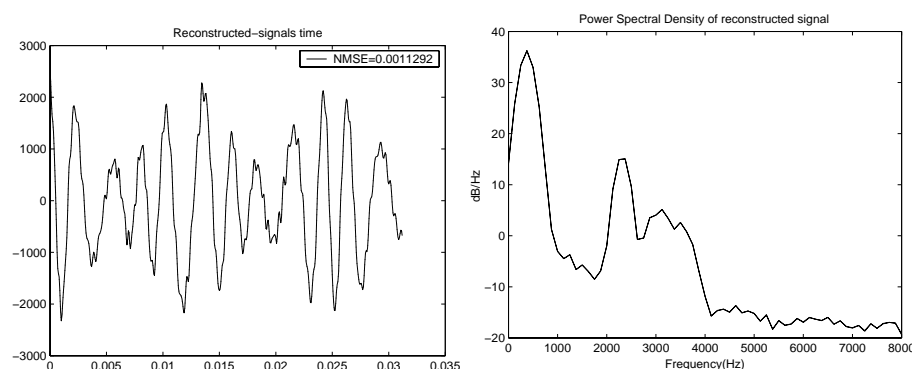
Συγκρίνοντας κατ' αρχάς τις εκτιμήσεις που δίνουν τα *bKES* και *mKES* μεταξύ τους διαπιστώνεται ότι οι εκτιμήσεις των περιβαλλουσών των συνιστωσών του σήματος είναι παρόμοιες, με αυτές από το δεύτερο μοντέλο να είναι ίσως λίγο πιο θορυβώδεις. Παρατηρούνται έντονες διαμορφώσεις στο πλάτος των δύο πιο υψίσυχων συνιστωσών με συχνότητες αρκετά μεγαλύτερες της θεμελιώδους, κάτι που ίσως αποτελεί ένδειξη της ορθότητας του μη γραμμικού μοντέλου αφού δεν είναι εύκολο να εξηγηθούν διαφορετικά τα σύνθετα σήματα που προκύπτουν. Όσο για τις στιγμιαίες συχνότητες, στις εκτιμήσεις του πρώτου μοντέλου για όλες τις συνιστώσες εκτός της πρώτης παρατηρούνται εκρήξεις παρόμοιες με αυτές που ανιχνεύτηκαν και σχολιάστηκαν στην περίπτωση των συνθετικών σημάτων φωνής. Οφείλονται κατά κύριο λόγο σε μικρές τιμές του εκτιμώμενου στιγμιαίου πλάτους στις θέσεις των εκρήξεων. Όπως και στην περίπτωση των συνθετικών σημάτων, τα σήματα των στιγμιαίων συχνοτήτων που προκύπτουν από το *mKES* είναι περισσότερο ομαλά, με τις εκρήξεις να είναι περιορισμένες.

Από την άλλη, τα αποτελέσματα του MBDA είναι πολύ πιο θορυβώδη και μια παρατήρηση που θα μπορούσε να γίνει είναι το γεγονός ότι τα στιγμιαία πλάτη παίρ-





Σχήμα 5.10: Ανάλυση με βάση το μοντέλο  $mKES$  του συνθετικού σήματος που προσομοιάζει σήμα φωνής και περιγράφεται στην εξίσωση 5.2. (α), (γ), (ε) οι εκτιμήσεις των στιγμιαίων πλατών των τριών συνιστωσών του σήματος (β), (δ), (στ) οι εκτιμήσεις των αντίστοιχων στιγμιαίων συχνοτήτων. Στα υπομνήματα φαίνονται τα αντίστοιχα μέσα τετραγωνικά λάθη ενώ με διακεκομμένες γραμμές είναι τα πραγματικά σήματα.



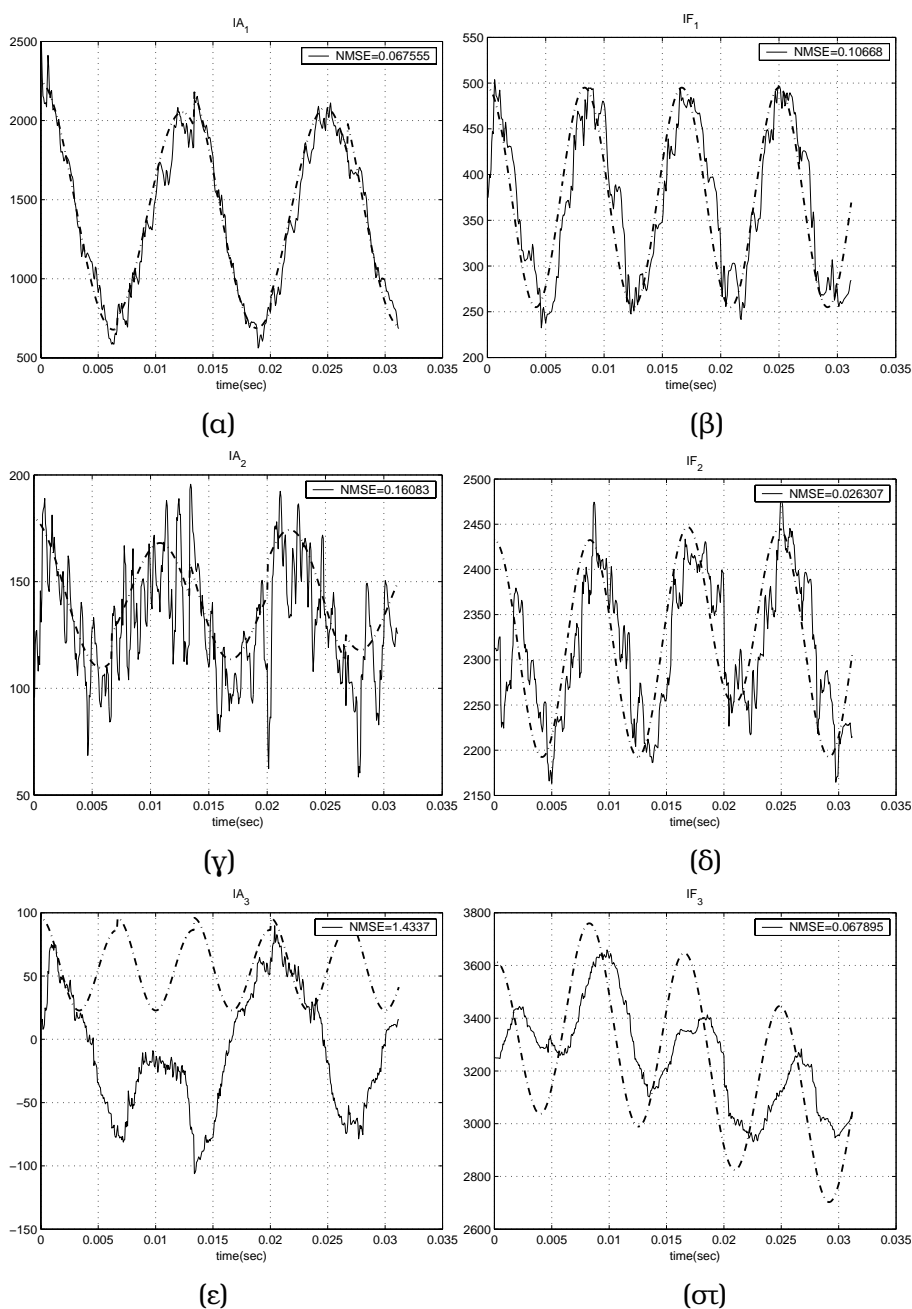
Σχήμα 5.11: Ανακατασκευή του συνθετικού σήματος φωνής της εξίσωσης 5.2 με βάση τις εκτιμήσεις με χρήση του μοντέλου *mKES* (α) Το ανακατασκευασμένο σήμα μετά την εκτίμηση με το *mKES*, (β) Η φασματική πυκνότητα ισχύος του ανακατασκευασμένου σήματος υπολογισμένη με τη μέθοδο Welch. Σημειώνεται ότι στις γραφικές παραστάσεις υπερτίθενται τα πραγματικά σήματα. Οι διαφορές όμως είναι τόσο μικρές που δεν διακρίνονται ξεχωριστές γραμμές.

νουν και αρνητικές τιμές, κάτι το οποίο έχει ξανασχολιαστεί και το οποίο αποφεύγεται στο *mKES* με τη χρήση πληροφορίας από τον ESA. Το πρόβλημα με τις αρνητικές τιμές στην περιβάλλουσα είναι ότι δεν συμφωνούν με το μη γραμμικό μοντέλο όπως έχει προταθεί, όπου είναι γενικώς αποδεκτό ότι τα σήματα εκτός από διαμορφωμένα στη συχνότητα είναι και διαμορφωμένου πλάτους αλλά με φέρον (AMWC), δηλαδή το στιγμιαίο πλάτος τους είναι θετικό.

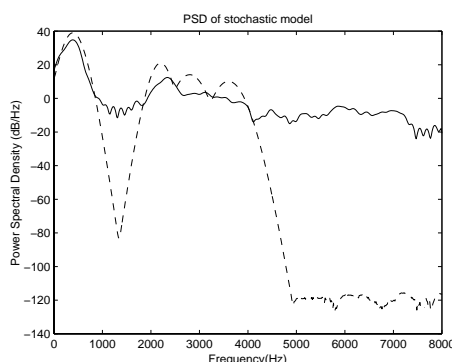
### 5.3.2 Παρακολούθηση κεντρικών συχνοτήτων συντονισμών φωνής

Στη συνέχεια παρουσιάζεται η χρήση της παραλλαγής *mKES* για παρακολούθηση των κεντρικών συχνοτήτων των συντονισμών (formants). Χρησιμοποιήθηκε και το *bKES* στην αντίστοιχη εφαρμογή χωρίς όμως να έχει ικανοποιητικά αποτελέσματα. Η αρχικοποίηση των  $q_{f_i}$  γίνεται και πάλι με τη χρήση του πρόχειρου αλγορίθμου παρακολούθησης κεντρικών συχνοτήτων, όπως έχει περιγραφεί προηγουμένως. Το σήμα που χρησιμοποιείται είναι το ίδιο με αυτό που χρησιμοποιούν οι Lu & Doerschuk (υποενότητα 3.6.2).

Όπως φαίνεται και στο Σχήμα 5.18, η παρακολούθηση έχει πρόβλημα στο σημείο όπου διασταυρώνονται οι δύο ενδιαμέσοι συντονισμοί. Μετά τη διασταύρωση βέβαια, όπως και στην αρχή, η παρακολούθηση γίνεται χωρίς προβλήματα. Για το σπεκτρογράφημα αναφέρεται ότι σε κάθε πλαίσιο και για κάθε συνιστώσα θεωρήθηκε η αντίστοιχη μέση τιμή της εκτίμησης της κεντρικής συχνότητας που δίνει το μοντέλο σε όλο το μήκος του πλαισίου. Η εξέλιξη αυτής της μέσης τιμής για κάθε



Σχήμα 5.12: Ανάλυση με βάση τον MBDA του συνθετικού σήματος που προσομοιάζει σήμα φωνής και περιγράφεται στην εξίσωση 5.2. (α), (γ), (ε) εκτιμήσεις των στιγμιαίων πλατών των τριών συνιστωσών του σήματος (β), (δ), (σ) οι εκτιμήσεις των αντίστοιχων στιγμιαίων συχνοτήτων. Στα υπομνήματα φαίνονται τα αντίστοιχα μέσα τετραγωνικά λάθη, ενώ με διακεκομμένες γραμμές είναι τα πραγματικά σήματα.



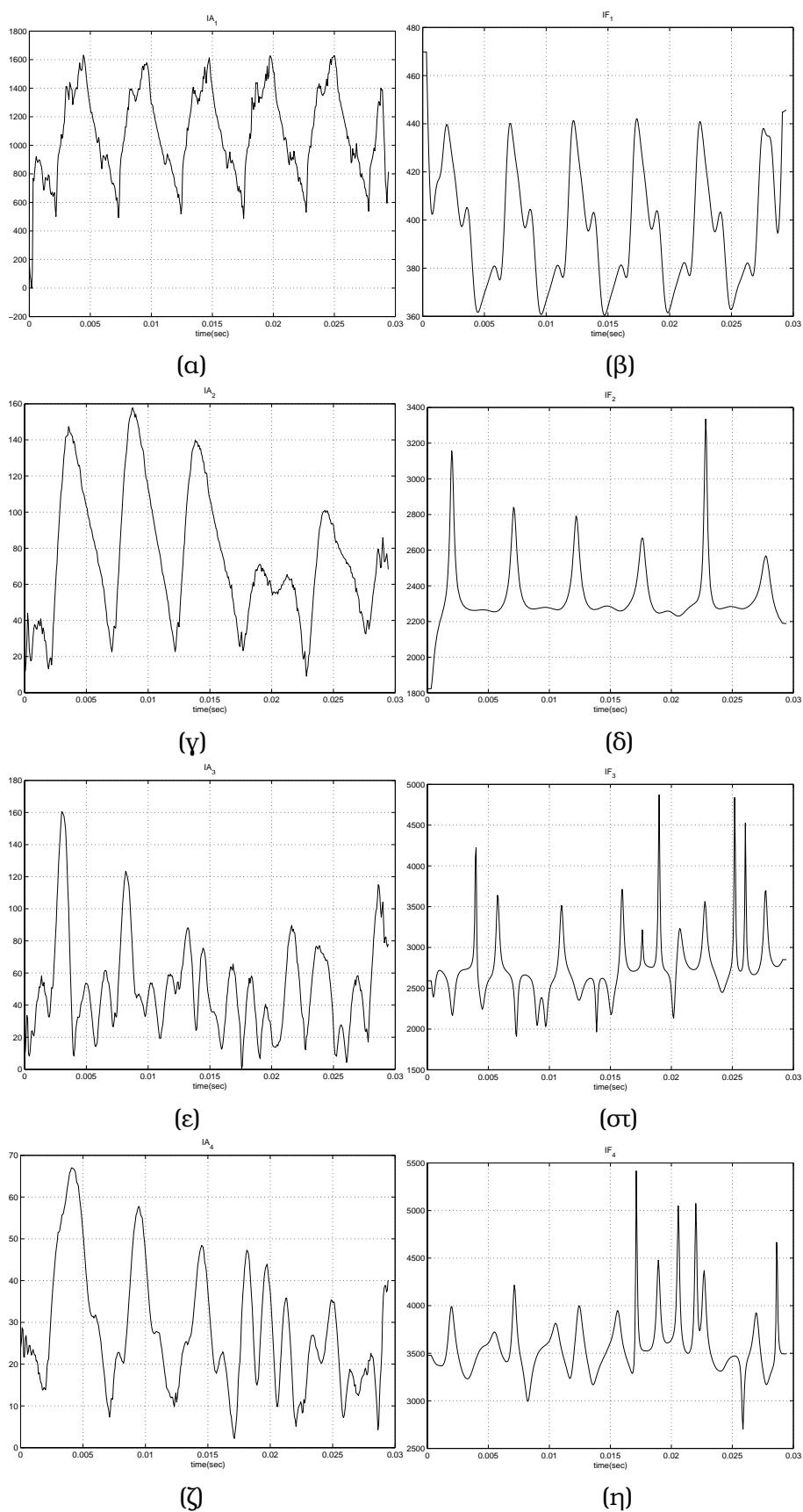
Σχήμα 5.13: Με διακεκομμένες γραμμές εμφανίζεται η φασματική πυκνότητα ισχύος όπως υπολογίζεται με βάση το στοχαστικό μοντέλο του *MBDA*, όπως αρχικοποιείται αυτό για την ανάλυση του φωνήματος /ee/, ενώ με συνεχή γραμμή είναι η πραγματική φασματική πυκνότητα ισχύος του φωνήματος, όπως εκτιμάται με χρήση της μεθόδου *Welch*

συνιστώσα είναι ουσιαστικά αυτή που έχει υπερτεθεί στο σπεκτρογράφημα. Σημειώνεται πάντως ότι δεν φαίνεται να υπάρχει πραγματική βελτίωση στα αποτελέσματα σε σύγκριση με τον *MBDA*, βελτίωση τουλάχιστον τέτοια που να δικαιολογεί την αυξημένη πολυπλοκότητα που έχει η εκτίμηση με βάση το μοντέλο *mKES*.

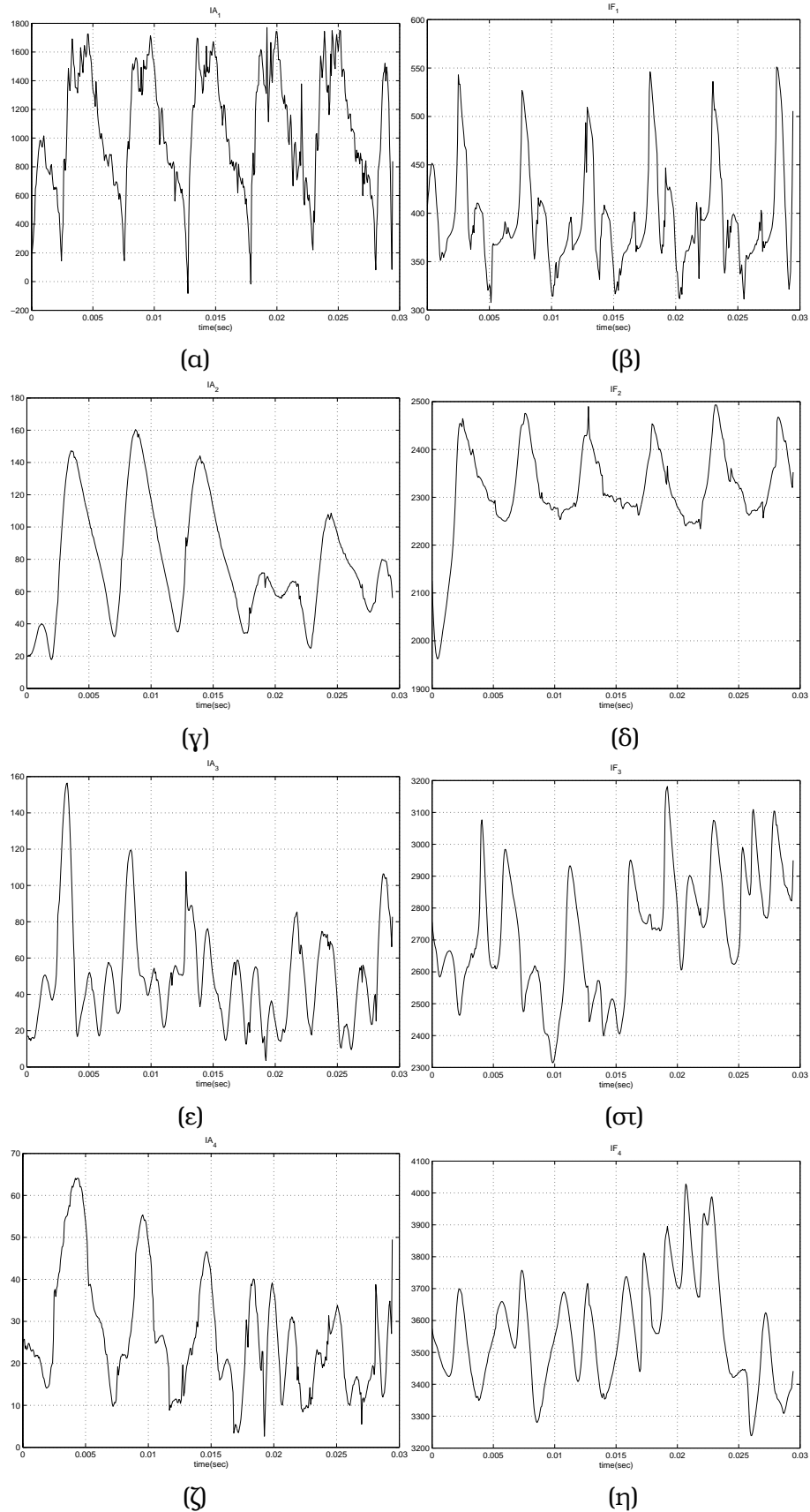
## 5.4 Συμπεράσματα

Κρίνοντας από τα αποτελέσματα που παρουσιάστηκαν και με βάση διάφορα άλλα σχετικά πειράματα που πραγματοποιήθηκαν γίνεται φανερό πως η χρησιμοποίηση του *ESA* από κάποιο δυναμικό μοντέλο με το οποίο εφαρμόζεται στη συνέχεια εκτίμηση *Kalman*, έχει ιδιαίτερα καλές προοπτικές στην ανάλυση φωνής. Επιτρέπει με αξιοπιστία την εφαρμογή του *ESA* σε όλες τις συνιστώσες ενός σήματος φωνής και σε πολλές περιπτώσεις μπορεί να έχει και βελτιωμένα αποτελέσματα σε σχέση με αυτόν, λόγω ύπαρξης κάποιας στοχαστικής δυναμικής. Οι εκτιμήσεις πάντως που προκύπτουν είναι ιδιαίτερα ομαλές, διευκολύνοντας την αναζήτηση επιπλέον πληροφορίας από τα σήματα των στιγμιαίων πλατών και των στιγμιαίων συνιστωσών καθώς και την εύρεση διαφορών συσχετίσεων μεταξύ τους.

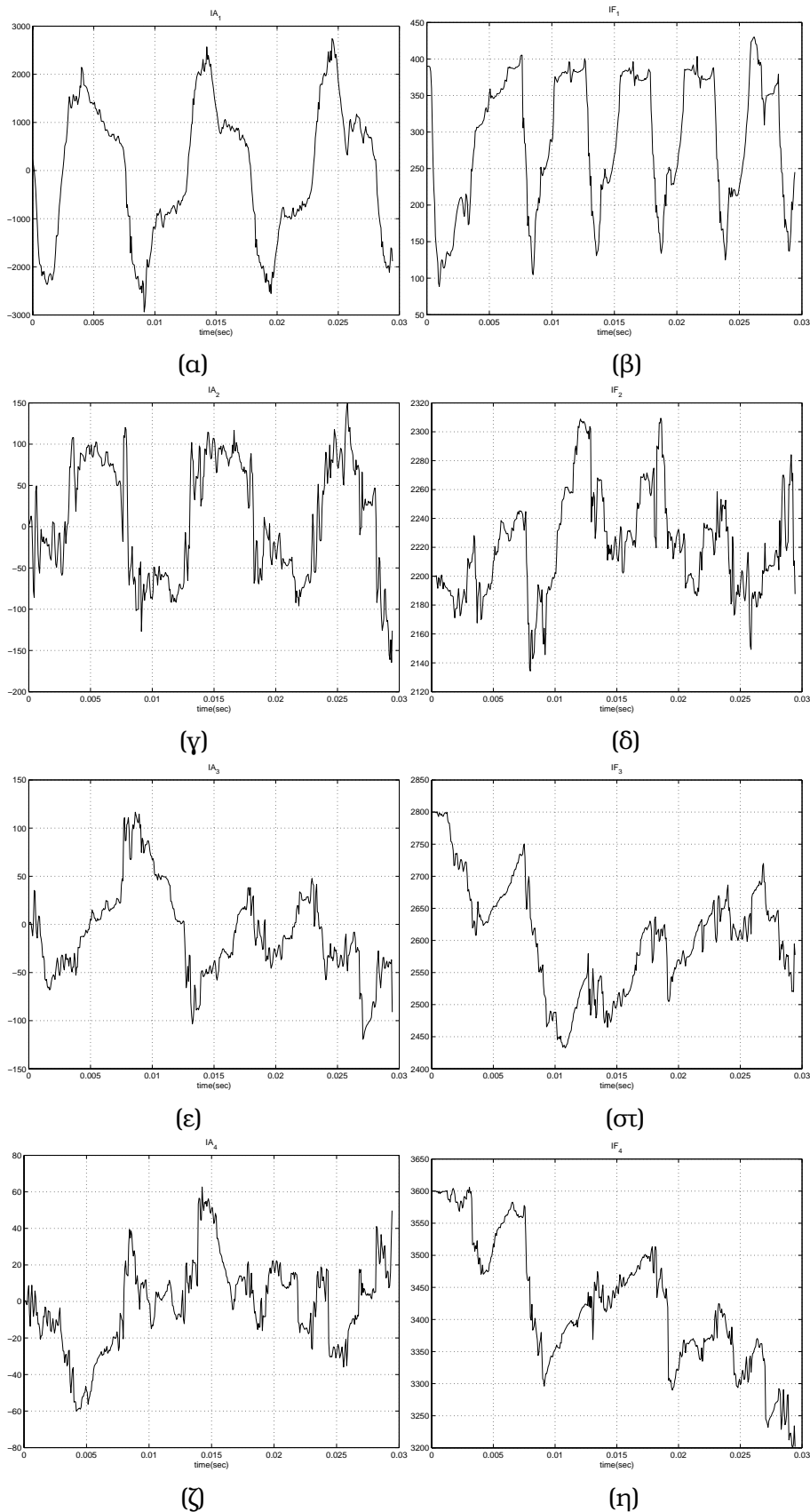
Από την άλλη ελπιδοφόρα κρίνεται και η εφαρμογή της νέας προσέγγισης στην παρακολούθηση των κεντρικών συχνοτήτων των συντονισμών της φωνής. Ήδη τα αποτελέσματα του *mKES* είναι ανάλογα αυτών του *MBDA*, ενώ παραμένει ανοιχτό ακόμα το θέμα κατάλληλου μοντέλου εξέλιξης των κεντρικών συχνοτήτων, αφού, όπως προκύπτει από πολλά σχετικά πειράματα το μοντέλο του τυχαίου περιπάτου είναι μάλλον ανεπαρκές για μακράς διάρκειας σήματα φωνής, στα οποία μπορεί να υπάρχουν πολλές εναλλαγές. Όσον αφορά στο *bKES* από την άλλη, δεν μπορεί να ελέγξει ακριβώς αυτήν την εξέλιξη των κεντρικών συχνοτήτων με αξιοπιστία αφού



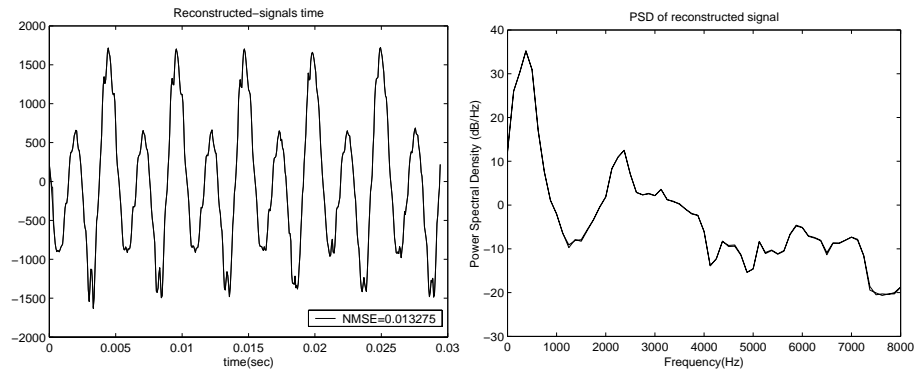
Σχήμα 5.14: Ανάλυση του φωνήματος /ee/ με τη χρήση του *bKES* (α), (γ), (ε), (ζ) Τα στιγμιαία πλάτη των τεσσάρων συνιστωσών όπως υπολογίζονται με το *bKES* (β), (δ), (σ), (η) Οι εκτιμήσεις των στιγμιαίων συχνοτήτων των αντίστοιχων συνιστωσών



Σχήμα 5.15: Ανάλυση του φωνήματος /ee/ με τη χρήση του μοντέλου *mKES* (α), (γ), (ε), (ζ) Τα στιγμιαία πλάτη των τεσσάρων συνιστωσών όπως υπολογίζονται με το *mKES* (β), (δ), (στ), (η) Οι εκτιμήσεις των στιγμιαίων συχνοτήτων των αντίστοιχων συνιστωσών



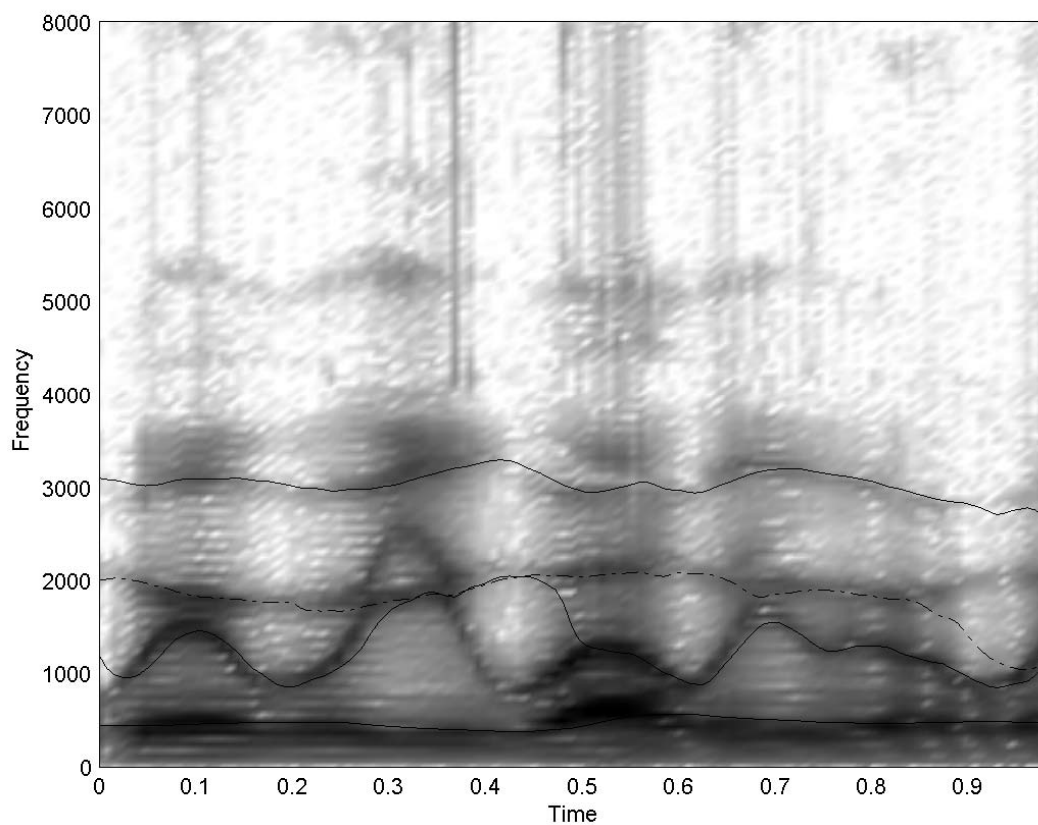
Σχήμα 5.16: Ανάλυση του φωνήματος /ee/ με τον MBDA (α), (γ), (ε), (ζ) Τα στιγμιαία πλάτη των τεσσάρων συνιστωσών όπως υπολογίζονται με τον MBDA, (β), (δ), (στ), (η) Οι εκτιμήσεις των στιγμιαίων συχνοτήτων των αντίστοιχων συνιστωσών που προκύπτουν ως  $\hat{IF}_i = \hat{f}_i + \hat{v}_i$



Σχήμα 5.17: Ανακατασκευή του φωνήματος /ee/ από τις εκτιμήσεις του *bKES* (α) Το ανακατασκευασμένο σήμα με τις εκτιμήσεις από το μοντέλο *bKES*, όπου στο υπόμνημα δίνεται το κανονικοποιημένο μέσο τετραγωνικό λάθος της ανακατασκευής. (β) Η φασματική πυκνότητα ισχύος του ανακατασκευασμένου σήματος υπολογισμένη με τη μέθοδο Welch

προκύπτει πως υπάρχει πρόβλημα ακαταλληλότητας των αρχικών παραμέτρων. Δε γίνεται κατάλληλη ταυτοποίηση του συστήματος ώστε να μπορεί να ανταπεξέλθει σε έντονα μεταβαλλόμενα και ανομοιογενή σήματα με μεγάλο μήκος. Χρειάζεται οπωσδήποτε εκτενέστερη διερεύνηση του θέματος αλλά και γενικότερα της προσέγγισης KESDA ώστε να μπορέσει να θεμελιωθεί και θεωρητικά και να προσδιοριστούν με σαφήνεια οι δυνατότητες και οι περιορισμοί της στον πολυσύνθετο χώρο της φωνής.





Σχήμα 5.18: Παρακολούθηση συντονισμών του σήματος φωνής "Where were you while we were away". Στο σπεκτρογράφημα υπερτίθενται οι τροχιές των κεντρικών συχνοτήτων των συντονισμών της φωνής, όπως εκτιμώνται από το *mKES*

## Κεφάλαιο 6

# Συνεισφορές-Συμπεράσματα της Εργασίας και Κατευθύνσεις για μελλοντική Έρευνα

Το κεφάλαιο αυτό αποτελεί την κατακλείδα της όλης εργασίας. Ανακεφαλαιώνονται τα βασικά σημεία της παρουσίασης που προηγήθηκε, παρουσιάζονται τα συμπεράσματα που προκύπτουν και γίνεται κάποια διερεύνηση ποικίλων θεμάτων σχετικών με τις προοπτικές που αναδεικνύονται και τις κατευθύνσεις προς τις οποίες θα μπορούσε να κινηθεί η μελλοντική σχετική έρευνα.

### 6.1 Ανακεφαλαίωση

Με την παρούσα εργασία προτάθηκε και μελετήθηκε μια νέα προσέγγιση στο πρόβλημα της αποδιαμόρφωσης AM-FM σημάτων με πολλές συνιστώσες. Η προσέγγιση αυτή είναι εμπνευσμένη από τη διαίσθηση ότι ο συνδυασμός στατιστικής εκτίμησης Kalman και ενεργειακών αλγορίθμων θα μπορούσε να βελτιώσει τις δυνατότητες των απλών στατιστικών αλγορίθμων που έχουν ως τώρα παρουσιαστεί, ως εκπρόσωπος των οποίων περιγράφεται ο αλγόριθμος Kalman-MBDA. Συγκεκριμένα, προτάθηκαν τέσσερις διαφορετικές παραλλαγές δυναμικών συστημάτων για τη μοντελοποίηση σημάτων. Τα συστήματα αυτά αξιοποιούν, με τον έναν ή τον άλλο τρόπο, πληροφορία που προέρχεται από τον αλγόριθμο διαχωρισμού με υπολογισμό της ενέργειας (ESA). Η εφαρμογή του αλγορίθμου γίνεται στις εξόδους προσαρμοζόμενης, ως προς την κεντρική συχνότητα, συστοιχίας φίλτρων Gabor, ένα φίλτρο για κάθε συνιστώσα του σήματος. Η κεντρική συχνότητα κάθε φίλτρου εκτιμάται μέσω του αλγορίθμου Kalman ενώ παράλληλα λαμβάνονται και βελτιωμένες εκ των υστέρων (a posteriori) εκτιμήσεις των στιγμιαίων πλατών και των στιγμιαίων συχνοτήτων των συνιστωσών του σήματος.

Η παρουσίαση κάθε τέτοιου συστήματος συνοδεύτηκε από παραδείγματα εφαρμογών του σε απλές αλλά ενδεικτικές περιπτώσεις συνθετικών σημάτων. Με τέτοιες

εφαρμογές έγινε και μια πρώτη προσπάθεια συνολικής συγκριτικής αξιολόγησης των επιμέρους συστημάτων αλλά και του Kalman-MBDA. Εξετάστηκαν διάφορες περιπτώσεις διαμόρφωσης πλάτους και συχνότητας ενώ παρουσιάστηκαν και τα αποτελέσματα πειραμάτων για διαφορετικούς σηματοθορυβικούς λόγους (SNR) ώστε να αναδειχτούν οι δυνατότητες των συστημάτων στην περίπτωση ύπαρξης θορύβου.

Ιδιαίτερη βαρύτητα δόθηκε στην εφαρμογή της νέας προσέγγισης σε σήματα φωνής. Η ανάλυση σημάτων φωνής εξάλλου απετέλεσε και το πρωταρχικό κίνητρο αναζήτησης μιας νέας μεθόδου πολυζωνικής αποδιαμόρφωσης, μιας μεθόδου η οποία θα είχε την προοπτική να αποκαλύψει νέες πληροφορίες με βάση το μη γραμμικό μοντέλο φωνής του οποίου δόθηκε η βασική ιδέα. Αναλύθηκαν συνθετικά σήματα που προσομοιάζουν σήματα φωνής είτε με βάση το γραμμικό είτε το μη γραμμικό μοντέλο με στόχο την εξαγωγή χρήσιμων συμπερασμάτων για την συμπεριφορά της νέας προσέγγισης σε πολυσύνθετα σήματα. Τέλος, παρουσιάστηκαν τα αποτελέσματα κάποιων πρώτων πειραμάτων στο χώρο της φωνής, που αφορούν κυρίως στην ανάλυσή της ως σήμα AM-FM με τόσες συνιστώσες όσες και συντονισμούς αλλά και στην παρακολούθηση των κεντρικών συχνοτήτων των συντονισμών της (formants).

## 6.2 Συμπεράσματα

Στη συνέχεια δίνονται τα συμπεράσματα της εργασίας:

- Το βασικότερο συμπέρασμα που προκύπτει είναι η επιβεβαίωση, σε κάποιο βαθμό, της διαίσθησης, ότι με τη συμβολή ενεργειακών μεθόδων είναι δυνατή η αναβάθμιση των απλών στατιστικών μεθόδων για την αποδιαμόρφωση AM-FM σημάτων με πολλές συνιστώσες. Αυτό γίνεται φανερό από τα αποτελέσματα της σχετικής σύγκρισης που παρουσιάστηκε σε συνθετικά σήματα, με την οποία φάνηκε γενικότερη βελτίωση των εκτιμήσεων, με κριτήριο τα αντίστοιχα μέσα τετραγωνικά σφάλματα. Γενικότερα, η νέα προσέγγιση παρουσιάζει ιδιαίτερα ενδιαφέροντα χαρακτηριστικά γνωρίσματα, τα οποία συνοψίζονται στα εξής:

- { Επεξεργάζεται σήματα ευρέος συχνοτικού περιεχομένου, χωρίς να χρειάζεται κάποια προεπεξεργασία, όπως θα ήταν για παράδειγμα, η απομόνωση κάποιας συγκεκριμένης ζώνης.
- { Η παρουσία θορύβου περιλαμβάνεται με πολύ « φυσικό » τρόπο, αρκεί να γίνει κατάλληλη επιλογή των παραμέτρων που ρυθμίζουν την ισχύ του θορύβου στις εξισώσεις κατάστασης και μέτρησης των συστημάτων.
- { Η εκτίμηση των στιγμιαίων πλατών και συχνοτήτων στηρίζεται στα αποτελέσματα ενός ιδιαίτερα αξιόπιστου αλγορίθμου, ο οποίος έχει και ισχυρό φυσικό υπόβαθρο και δεν είναι απλά το αποτέλεσμα στοχαστικής διαδικασίας.
- { Ιδιαίτερα στην περίπτωση της φωνής, λόγω αυτού του φυσικού υποβάθρου, ενδεχόμενα διευκολύνεται πιθανή σύνδεση του σήματος της φωνής με το φυσικό σύστημα που το παράγει.

- Είναι εμφανές ότι η νέα προσέγγιση συνδυάζει πλεονεκτήματα τόσο των στατιστικών όσο και των ενεργειακών μεθόδων. Αυτό βέβαια γίνεται με κόστος κάποια επιβάρυνση της υπολογιστικής πολυπλοκότητας. Η εφαρμογή του φιλτραρίσματος Gabor και του αλγορίθμου διαχωρισμού με υπολογισμό της ενέργειας επαναληπτικά αυξάνει σε κάποιο βαθμό την πολυπλοκότητα του αλγορίθμου. Οι προσπάθειες που έγιναν για αντιμετώπιση αυτής της αύξησης ενώ είχαν ικανοποιητικά αποτελέσματα εκτίμησης (*fKES*, *iKES*), τελικά αυτά δεν ήταν τόσο καλά όσο αυτά των δύο άλλων παραλλαγών (*bKES*, *mKES*), οι οποίες βέβαια είναι πολυπλοκότερες.
- Πάντως, σε σύγκριση με τον αλγόριθμο Kalman-MBDA, η προσέγγιση που προτάθηκε προβάλλει σε πολλές περιπτώσεις ως περισσότερα υποσχόμενη. Φαίνεται να είναι σε μικρότερο βαθμό ευαίσθητη στις διάφορες παραμέτρους αλλά και στην αρχικοποίηση του συστήματος. Επιπλέον, έχει γενικά καλύτερη συμπεριφορά σε περιπτώσεις υψηλού θορύβου, ειδικά η παραλλαγή *bKES*. Ενώ και σε περιπτώσεις υψηλής διαμόρφωσης τα αποτελέσματα των παραλλαγών *bKES*, *mKES* τουλάχιστον, είναι πιο καλά από αυτά του Kalman-MBDA. Συγκριτικά τέλος, πολύ καλύτερη είναι και η συμπεριφορά σε περιπτώσεις ανάλυσης φωνής. Οι εκτιμήσεις των στιγμιαίων πλατών και συχνοτήτων είναι λιγότερο θορυβώδεις και η μορφή τους είναι περισσότερο σύμφωνη με το μη γραμμικό μοντέλο φωνής.
- Από την άλλη, συγκρίνοντας τις δύο επικρατέστερες παραλλαγές, δηλαδή τα μοντέλα *bKES*, *mKES* για το πρώτο αξίζει να σημειωθεί ότι γενικά μπορεί να προσαρμοστεί σε σήμα πιο εύκολα από ότι το δεύτερο, το οποίο περιλαμβάνει περισσότερες παραμέτρους προς ταυτοποίηση από αυτές του MBDA. Από την άλλη, το δεύτερο μοντέλο φαίνεται να αντιμετωπίζει καλύτερα τυχόν σφάλματα του αλγορίθμου διαχωρισμού με υπολογισμό της ενέργειας και έτσι δίνει ικανοποιητικότερα αποτελέσματα σε κάποια σήματα φωνής. Γενικά πάντως, είναι δύσκολη η ασφαλής εξαγωγή συμπερασμάτων για το ποια παραλλαγή τελικά είναι καλύτερη και θα έπρεπε να χρησιμοποιηθεί σε εφαρμογές. Για τέτοια συμπεράσματα αλλά και για βελτιστοποίηση της νέας προσέγγισης, χρειάζεται σίγουρα συστηματική μελλοντική έρευνα.

### 6.3 Μελλοντική έρευνα

Η μελλοντική έρευνα για τις δυνατότητες και τους περιορισμούς της νέας προσέγγισης θα μπορούσε να κινηθεί σε τουλάχιστον τρεις άξονες, ο ένας αφορά στην εφαρμογή της σε σήματα φωνής, ο άλλος στη θεωρητική διερεύνησή της και ο τρίτος στη μείωση της πολυπλοκότητας των διαφόρων παραλλαγών της.

### 6.3.1 Εφαρμογή στη φωνή

Όπως προαναφέρθηκε, πρωταρχικός χώρος εφαρμογών της νέας προσέγγισης είναι η φωνή. Στον χώρο αυτό πρέπει να γίνουν διεξοδικά και συστηματικά πειράματα ώστε να διαπιστωθεί αν πράγματι η ιδέα του συνδυασμού στατιστικών και ενεργειακών μεθόδων μπορεί να ικανοποιήσει τις προσδοκίες, να φανερώσει δηλαδή την όψη του σήματος φωνής από την οπτική γωνία της μη γραμμικής μοντελοποίησης. Επιπλέον, βέβαια, οι περαιτέρω σχετικοί πειραματισμοί, μπορεί να οδηγήσουν σε χρήσιμες παρατηρήσεις που ίσως αφορούν άμεσα την ιδέα της παραδοχής του μοντέλου AM-FM για τη φωνή.

Στο ίδιο γενικό πλαίσιο εντάσσεται και η βελτίωση των προτεινόμενων παραλλαγών προς την κατεύθυνση της καλύτερης μοντελοποίησης της εξέλιξης των κεντρικών συχνοτήτων. Πιο συγκεκριμένα, διάφορα πειράματα έχουν δείξει ότι η μοντελοποίηση των κεντρικών συχνοτήτων με τυχαίους περιπάτους (random walks), είναι μάλλον ανεπαρκής στην περίπτωση της φωνής. Οδηγεί σε ιδιαίτερα θορυβώδεις εκτιμήσεις των κεντρικών συχνοτήτων κάποιες φορές και μπορεί και σε αστάθεια. Μια εναλλακτική θα ήταν ίσως η χρησιμοποίηση περιορισμών που να επιβάλλουν κάποια ομαλότητα στην εξέλιξη των εν λόγω μεταβλητών κατάστασης στα προτεινόμενα μοντέλα ενώ μια άλλη θα ήταν ίσως η μοντελοποίηση τους ως δυναμικά fractal σήματα. Σε κάθε περίπτωση πάντως, το θέμα χρειάζεται συστηματική διερεύνηση.

### 6.3.2 Θεωρητική Διερεύνηση

Ενδιαφέρον επίσης παρουσιάζει και η θεωρητική θεμελίωση της νέας προσέγγισης. Είναι ιδιαίτερα σημαντικό να διερευνηθούν και θεωρητικά οι όποιες δυνατότητές της. Μια σχετική κατεύθυνση υπέδειξαν οι Lu & Doerschuk, υπολογίζοντας για τον αλγόριθμο Kalman-MBDA το όριο Cramer-Rao, το οποίο στην ουσία αποτελεί ένα άνω φράγμα της επίδοσης του αλγορίθμου ως εκτιμητή (estimator). Με τη χρήση τέτοιων ορίων θα μπορούσαν να συγκριθούν οι δυνατότητες υποβέλτιστων εκτιμητών (όπως είναι οι παραλλαγές που προτάθηκαν) με κάποια αντικειμενικά κριτήρια.

Από την άλλη, η θεωρητική ανάλυση θα μπορούσε να δώσει λύση στο πρόβλημα της συστηματικής βέλτιστης αρχικοποίησης του αλγορίθμου καθώς και της ταυτοποίησης συστήματος (system identification), ουσιαστικά δηλαδή της κατάλληλης επιλογής των διαφόρων παραμέτρων. Αυτό είναι ιδιαίτερα σημαντικό για αλγορίθμους που εφαρμόζονται σε σήματα φωνής, τα οποία χαρακτηρίζονται από μεγάλη συνθετότητα και ποικιλομορφία.

### 6.3.3 Μείωση πολυπλοκότητας

Ένα τελευταίο θέμα που αξίζει ιδιαίτερη προσοχή είναι η υπολογιστική πολυπλοκότητα της νέας προσέγγισης. Ήδη έχει γίνει κάποια προσπάθεια για καλύτερη αποδοτικότητα αλλά δεν έχει επιτευχθεί βελτιστοποίηση αυτού του παράγοντα. Είναι κάποιες πτυχές, όπως το φιλτράρισμα Gabor, που θα μπορούσαν να βελτιωθούν. Τα

φίλτρα Gabor, για παράδειγμα, έχουν το μειονέκτημα ότι δεν έχουν μεγάλη διακριτότητα στο χρόνο. Για σωστή εφαρμογή τους, απαιτείται ένα αρκετά μεγάλο παράθυρο του σήματος. Αυτό προκαλεί πρόσθετη, και ίσως τη σημαντικότερη, επιβάρυνση, δεδομένου ότι με τον αλγόριθμο Kalman γίνεται επαναληπτική εφαρμογή των φίλτρων αυτών. Η αντικατάσταση των φίλτρων Gabor με κάποια πιο κατάλληλα ζωνοπερατά φίλτρα για αναδρομικούς αλγορίθμους προβάλλει ως ιδιαίτερα σημαντική προοπτική βελτίωσης.

Τέλος, στην ίδια κατεύθυνση, ενδιαφέρον παρουσιάζει η επανεξέταση των παραλλαγών που χρησιμοποιούν πλαίσια του σήματος (*fKES*, *iKES*). Η ιδέα στην οποία στηρίζονται είναι ιδιαίτερα ελκυστική αλλά μάλλον απαιτείται περαιτέρω διερεύνησή τους ώστε να τροποποιηθούν κατάλληλα και να μπορέσουν να δώσουν καλύτερα αποτελέσματα.

Παρουσιάστηκαν συνοπτικά κάποια θέματα που αφορούν στην νέα προσέγγιση που προτάθηκε και αναφέρθηκαν σχετικές παρατηρήσεις και ιδέες. Είναι εμφανές πως ενώ από τη μία τα αποτελέσματα της προσέγγισης είναι ελπιδοφόρα, από την άλλη χρειάζεται διεξοδική μελλοντική έρευνα προς διάφορες κατευθύνσεις ώστε να γίνει δυνατή η συστηματική χρησιμοποίησή της ως εργαλείο πολυζωνικής αποδιαμόρφωσης AM-FM σημάτων, με άμεση εφαρμογή στην ανάλυση σημάτων φωνής.

# Βιβλιογραφία

- [1] Δ.Καλλίνικος Γ.Καραγιάννης, *Σήματα και Συστήματα*, ΣΥΜΕΩΝ.
- [2] Μ. Ραγκούση Γ. Καραγιάννης, *Ψηφιακή Επεξεργασία Σημάτων*, ΣΥΜΕΩΝ.
- [3] Τρύφων Γ. Κουσιουρής, *Θεωρία Γραμμικών Πολυμεταβλητών Συστημάτων*, Εθνικό Μετσόβιο Πολυτεχνείο, 1990.
- [4] A. Potamianos and P. Maragos, *A comparison of the energy operator and the Hilbert transform approach to signal and speech demodulation*, 1994.
- [5] Alan C. Bovik, Petros Maragos, Thomas F. Quatieri, *AM-FM Energy Detection and Separation in Noise Using Multiband Energy Operators*, IEEE Transactions on Signal Processing **41** (1993), vo. 12, 3245–3265.
- [6] Alexandros Potamianos, Petros Maragos, *Speech formant frequency and bandwidth tracking using multiband energy demodulation*, J. Acoust. Soc. Am. **99** (1996), vo. 6, 3795–3806.
- [7] Athanasios Papoulis, S. Unnikrishna Pillai, *Probability, Random Variables and Stochastic Processes*, "Mc Graw Hill", 2002.
- [8] Balasubramaniam Santhanam, *Multicomponent AM-FM Demodulation via Periodicity-Based Algebraic Separation and Energy-Based Demodulation*, IEEE Transactions on Communications **48** (2000), vo. 3, 473–490.
- [9] Balasubramaniam Santhanam, Petros Maragos, *Energy Demodulation of Two-Component AM-FM Signal Mixtures*, IEEE Signal Processing Letters **3** (1996), vo. 11, 294–298.
- [10] Barbara F. La Scala, Robert R Bitmead, *Design of an Extended Kalman Filter Frequency Tracker*, IEEE Transactions on Signal Processing **44** (1996), vo. 3, 739–742.
- [11] E. Wan and R. van der Merwe, *The Unscented Kalman Filter for Nonlinear Estimation*, 2000.

- [12] Gary E. Kopec, *Formant Tracking Using Hidden Markov Models and Vector Quantization*, IEEE Transactions on Acoustics, Speech and Signal Processing **34** (1986), vo. 4, 709-729.
- [13] Greg Welch and Gary Bishop, *An introduction to the Kalman Filter*.
- [14] Helen M. Hanson, Petros Maragos, Alexandros Potamianos, *A System for Finding Speech Formants and Modulations via Energy Separation*, IEEE Transactions on Speech and Audio Processing **2** (1994), vo. 3, 436-443.
- [15] H.M.Teager, *Some Observations on Oral Air Flow During Phonation*, IEEE Transactions on Acoustics, Speech, Signal Processing **28** (1980), vo. 5, 599-601.
- [16] H.M.Teager, S.M.Teager, *Active Fluid Dynamic Voice Production Models or There is a Unicorn in the Garden*, Appears in book Vocal Fold Physiology Biomechanics, Acoustics and Phonatory Control.
- [17] J.F.Kaiser, *On a simple algorithm to calculate the energy of a signal*, Appears in book Vocal Fold Physiology Biomechanics, Acoustics and Phonatory Control.
- [18] ———, *Some Observations on Vocal Tract Operation from a Fluid Flow Point of View*, Appears in book Vocal Fold Physiology Biomechanics, Acoustics and Phonatory Control.
- [19] Kiyoshi Nishiyama, *A Nonlinear Filter for Estimating a Sinusoidal Signal and its Parameters in White Noise: On the Case of a Single Sinusoid*, IEEE Transactions on Signal Processing **45** (1997), vo. 4, 970-981.
- [20] L.R.Rabiner, B.-H.Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.
- [21] M. Faundez-Zanuy, G. Kubin, W. B. Kleijn, P. Maragos, S. McLaughlin, A. Esposito, A. Hussain, J. Schoentgen, *Nonlinear speech processing: overview and applications*, Int. J. Control Intelligent Syst. **30** (2002), vo. 1, 1-10.
- [22] M. Niranjan, I.J. Cox, *Recursive Tracking of Formants in Speech Signals*, ολ. II, 1994, σσ. 205-208.
- [23] Mohinder.S.Grewal, Angus P. Andrews, *KALMAN FILTERING*, "Prentice Hall", 1993.
- [24] Peter S. Maybeck, *Stochastic models, estimation and control*, ολ. 1, Academic Press, 1993.
- [25] Petros Maragos, Alexandros Potamianos, *Higher Order Differential Energy Operators*, IEEE Signal Processing Letters **2** (1995), vo. 8, 152-154.



- [26] Petros Maragos, James F. Kaiser, Thomas F. Quatieri, *Energy Separation in Signal Modulations with Application to Speech Analysis*, IEEE Transactions on Signal Processing **41** (1993), vo. 10, 3024–3051.
- [27] ———, *On Amplitude and Frequency Demodulation Using Energy Operators*, IEEE Transactions on Signal Processing **41** (1993), vo. 4, 1532–1550.
- [28] Robert J. McAulay and Thomas F. Quatieri, *Speech Analysis Synthesis Based on a Sinusoidal Representation*, IEEE Transactions on Acoustics, Speech and Signal Processing **34** (1986), vo. 4, 744–754.
- [29] Robert N. McDonough, A. D. Whalen, *Detection of Signals in Noise*, Academic Press, 1995.
- [30] Rudolph van der Merwe and Eric A. Wan, *The Square-Root Unscented Kalman Filter For State And Parameter-Estimation*.
- [31] S. Julier, *The Scaled Unscented Transformation*, 1999.
- [32] S. Julier and J. Uhlmann, *A new extension of the Kalman filter to nonlinear systems*, 1997.
- [33] Shan Lu, Peter C. Doerschuk, *Nonlinear Modeling and Processing of Speech based on Sums of AM-FM Formant Models*, IEEE Transactions on Signal Processing **44** (1996), vo. 4, 773–782.
- [34] Simon Haykin, “Συστήματα Επικοινωνίας”, Παπασωτηρίου, 1993.
- [35] Wan-Chieh Pai, Peter C. Doerschuk, *Statistical AM-FM Models, Extended Kalman Filter Demodulation, Cramer-Rao Bounds, and Speech Analysis*, IEEE Transactions on Signal Processing **48** (2000), vo. 8, 2300–2313.
- [36] Wan, Eric and van der Merwe, Rudolph, *Kalman Filtering and Neural Networks*, σφ. 7, Wiley, 2001.