



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΝΑΥΠΗΓΩΝ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΗ
“ΝΑΥΤΙΚΗ ΚΑΙ ΘΑΛΑΣΣΙΑ ΤΕΧΝΟΛΟΓΙΑ ΚΑΙ ΕΠΙΣΤΗΜΗ”

*Δυνατότητα εφαρμογής μιας προτότυπης μεθοδολογίας
στην αυτόματη αναγνώριση λειτουργικής κατάστασης /
βλάβης ναυτικού κινητήρα Diesel επί τη βάσει της
σχετικής ηχητικής πληροφορίας*

Μεταπτυχιακή Διπλωματική Εργασία του: Τερζάκου Γεώργιου

Επιβλέπων: Κ. Παπαοδυσσεύς

ΠΕΡΙΕΧΟΜΕΝΑ

1. ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΑΥΤΟΜΑΤΗ ΑΝΑΓΝΩΡΙΣΗ ΒΛΑΒΗΣ ΝΑΥΤΙΚΟΥ ΚΙΝΗΤΗΡΑ DIESEL ΕΠΙ ΤΗ ΒΑΣΕΙ ΤΗΣ ΣΧΕΤΙΚΗΣ ΗΧΗΤΙΚΗΣ ΠΛΗΡΟΦΟΡΙΑΣ.....	4
1.1 Ανάλυση του προβλήματος	4
2. ΠΡΩΤΟΥ ΕΠΙΠΕΔΟΥ ΑΝΑΛΥΣΗ ΗΧΟΥ ΚΙΝΗΤΗΡΑ DIESEL.	6
2.1 Ανάλυση στο πεδίο του χρόνου και στο πεδίο της συχνότητας	6
2.2 Αξιοποίηση του φαινομένου Masking	6
2.3 Η ιδέα των «διανυσμάτων αντιπροσώπων συχνοτικών περιοχών» .	9
3. Κριτήρια ομοιότητας ηχων ΚΙΝΗΤΗΡΑ	11
3.1 Αρχικά κριτήρια ομοιότητας ηχητικών σημάτων.....	11
3.2 Τελεσίδικο κριτήριο ομοιότητας ηχητικών σημάτων	12
4. ΥΛΟΠΟΙΗΣΗ ΤΗΣ ΠΡΟΤΕΙΝΟΜΕΝΗΣ ΜΕΘΟΔΟΛΟΓΙΑΣ.....	14
4.1 Δημιουργία ενός συνόλου διανυσμάτων αντιπρόσωπων συχνοτικών περιοχών για έναν πρότυπο ήχο κινητήρα	14
4.2 Αλγόριθμος προσαρμοστικού (adaptive) DFT	16
4.3 Μέθοδος αποδοτικής αποθήκευσης των διανυσμάτων αντιπροσώπων των πρότυπων ηχητικών σημάτων.....	18
4.4 Δημιουργία συνόλου διανυσμάτων αντιπροσώπων συχνοτικών περιοχών για ένα καταγεγραμμένο ηχητικό σήμα	25
4.5 Αλγόριθμος σύγκρισης διανυσμάτων αντιπροσώπων	26
4.6 Τελεσίδικο κριτήριο ομοιότητας	31

5. ΣΥΣΤΗΜΑ ΑΥΤΟΜΑΤΗΣ ΑΝΑΓΝΩΡΙΣΗΣ ΛΕΙΤΟΥΡΓΙΚΗΣ ΚΑΤΑΣΤΑΣΗΣ / ΒΛΑΒΗΣ ΚΙΝΗΤΗΡΑ DIESEL ΚΑΙ ΕΠΙΔΟΣΕΙΣ.....	32
5.1 Διαδικασία δημιουργίας βάσης προτύπων ηχητικών σημάτων για ένα συγκεκριμένο τύπο ναυτικού κινητήρα Diesel	33
5.2 Διαδικασία εξαγωγής των διανυσμάτων αντιπροσώπων από τα πρότυπα ηχητικά σήματα.....	35
5.3 Διαδικασία καταγραφής ηχητικών σημάτων από ιδίου τύπου κινητήρες με άγνωστη λειτουργική κατάσταση.....	35
5.4 Διαδικασία υλοποίησης αλγορίθμων ελέγχου των κριτηρίων ταυτοποίησης των ηχητικών σημάτων	35
5.5 Επιδόσεις του συστήματος αυτόματης αναγνώρισης λειτουργικής κατάστασης / βλάβης σε έναν κινητήρα Diesel.....	37
5.6 Παρατηρούμενος υπολογιστικός φόρτος της εφαρμογής.....	38
ΠΑΡΑΡΤΗΜΑ Α - Υλοποίηση αποδοτικού αναδρομικού αλγορίθμου σύγκρισης διανυσμάτων αντιπροσώπων	40
ΒΙΒΛΙΟΓΡΑΦΙΑ	51

1. ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΑΥΤΟΜΑΤΗ ΑΝΑΓΝΩΡΙΣΗ ΒΛΑΒΗΣ ΝΑΥΤΙΚΟΥ ΚΙΝΗΤΗΡΑ DIESEL ΕΠΙ ΤΗ ΒΑΣΕΙ ΤΗΣ ΣΧΕΤΙΚΗΣ ΗΧΗΤΙΚΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

1.1 Ανάλυση του προβλήματος

Η υλοποίηση ενός συστήματος που θα εκτιμά, αξιοποιώντας τον ήχο ενός κινητήρα Diesel, τη λειτουργική κατάσταση αυτού και θα αναγνωρίζει αυτόματα βλάβες σε αυτόν, εντάσσεται στα νέα συστήματα παρακολούθησης και ελέγχου (Monitoring and Control Systems) Ναυτικών Κινητήρων Diesel. Σήμερα έχουν αναπτυχθεί συστήματα παρακολούθησης και ελέγχου που βασίζονται στην επεξεργασία και ανάλυση των ακουστικών σημάτων που λαμβάνονται κατά τη λειτουργία ενός κινητήρα Diesel, με τις ακόλουθες εφαρμογές:

- Εκτίμηση της απόδοσης των κυλίνδρων από ακουστική ανάλυση.
- Μέτρηση του χρόνου ανοίγματος και κλεισίματος των βαλβίδων εισαγωγής και εξαγωγής καυσαερίων.

Στην παρούσα διπλωματική εργασία θα εξετασθεί η δυνατότητα υλοποίησης ενός συστήματος που θα αναγνωρίζει αυτόματα την λειτουργική κατάσταση / βλάβες σε έναν κινητήρα Diesel αξιοποιώντας τον ήχο αυτού με την εφαρμογή μιας πρωτότυπης μεθοδολογίας. Η μεθοδολογία αυτή έχει ήδη εφαρμοστεί, με συνολικό ποσοστό επιτυχίας πάνω από 99,5%, στην αυτόματη αναγνώριση μουσικών συνθέσεων, ομιλίας ή και ηχητικών προτύπων τα οποία έχουν υποστεί έως και σοβαρή παραμόρφωση λόγω θορύβου και αλλαγής του ρυθμού είτε της θεμελιώδους συχνότητας.

Το πρώτο βήμα για την υλοποίηση του συστήματος είναι η δημιουργία μίας βάσης ηχητικών σημάτων στην οποία θα είναι καταγεγραμμένος με την καλύτερη δυνατή πιστότητα ο ήχος ενός συγκεκριμένου τύπου κινητήρα Diesel σε διάφορες φάσεις λειτουργίας του, με την προϋπόθεση ότι ο κινητήρας δεν παρουσιάζει κανένα λειτουργικό πρόβλημα. Η εν λόγω βάση ξεκινά με την καταγραφή του ήχου του κινητήρα σε όλο το πεδίο των στροφών λειτουργίας του. Η καταγραφή γίνεται όταν διενεργούνται τελικές δοκιμές στον κινητήρα από την κατασκευάστρια εταιρεία πριν την απόδοσή του στην αγορά, ή μετά από τη διενέργεια γενικής επισκευής στον κινητήρα, όπου η λειτουργική του κατάσταση θεωρείται άριστη. Στη συνέχεια η

ηχητική βάση δεδομένων συμπληρώνεται με την καταγραφή του ήχου του κινητήρα σε όλο το πεδίο των στροφών λειτουργίας μετά από συγκεκριμένες ώρες λειτουργίας του που καθορίζονται από τις ενδιάμεσες μερικές επισκευές του κινητήρα, ώστε να ληφθεί υπόψη η φυσιολογική φθορά που έχει υποστεί ο κινητήρας κατά την λειτουργία του.

Το δεύτερο βήμα για την υλοποίηση του συστήματος είναι η συμπλήρωση της ηχητικής βάσης δεδομένων με την καταγραφή του ήχου του κινητήρα σε συγκεκριμένες πιστοποιημένες από τον επισκευαστικό φορέα βλάβες του κινητήρα κατά τις οποίες ο ήχος του διαφοροποιείται. Σε αυτό το στάδιο επιλέγονται και καταγράφονται κατ' αρχάς οι συνήθεις βλάβες που διαφοροποιούν τον ήχο όλων των Ναυτικών Κινητήρων Diesel όταν παρουσιάζονται σε αυτούς, με γνώμονα την υφιστάμενη εμπειρία χειριστών καθώς και των επισκευαστικών φορέων. Συγκεκριμένα είναι βλάβες που συνδέονται με την λειτουργία του συστήματος εισαγωγής αέρος καύσεως και εξαγωγής καυσαερίων του κυλίνδρου καθώς και με την λειτουργία του συστήματος καύσεως του κυλίνδρου.

Η ανωτέρω βάση σημάτων θα είναι η βάση προτύπων ηχητικών σημάτων του συστήματος μας. Από τον τρόπο που δημιουργήθηκε η εν λόγω βάση γίνεται αμέσως αντιληπτό ότι μπορεί να εμπλουτίζεται συνεχώς με οποιαδήποτε πιστοποιημένη από τον επισκευαστικό φορέα αλλαγή της λειτουργικής κατάστασης του συγκεκριμένου τύπου κινητήρα που διαφοροποιεί τον ήχο του.

Το τρίτο και τελευταίο βήμα για την υλοποίησή του συστήματός μας είναι η αναγνώριση από τον ήχο του ενός ίδιου τύπου κινητήρα κατά την λειτουργία του στο πλοίο, της λειτουργικής κατάστασης / βλάβης αυτού, αξιοποιώντας την υφιστάμενη βάση προτύπων σημάτων. Έτσι το πρόβλημα που καλούμεθα πλέον να αντιμετωπίσουμε, με βάση την πρωτότυπη μεθοδολογία που θα εφαρμοσθεί, μπορεί να διατυπωθεί ως εξής: Καταγράφουμε με την ίδια πιστότητα τον ήχο ενός ίδιου τύπου κινητήρα του οποίου η λειτουργική του κατάσταση είναι άγνωστη, έστω “X” το ηχητικό αυτό σήμα. Ζητείται να εντοπιστεί με την μεγαλύτερη δυνατή ακρίβεια εκείνο το πρότυπο ηχητικό σήμα που βέλτιστα αντιστοιχεί στο “X”. Εάν η αντιστοιχία αυτή πραγματοποιηθεί επιτυχώς τότε από τον καταγεγραμμένο ήχο γνωρίζουμε την λειτουργική κατάσταση του κινητήρα.

2. ΠΡΩΤΟΥ ΕΠΙΠΕΔΟΥ ΑΝΑΛΥΣΗ ΗΧΟΥ ΚΙΝΗΤΗΡΑ DIESEL

2.1 Ανάλυση στο πεδίο του χρόνου και στο πεδίο της συχνότητας

Ας θεωρήσουμε ένα καταγεγραμμένο πρότυπο ηχητικό σήμα που αφορά έναν συγκεκριμένο τύπο κινητήρα, και ένα ηχητικό σήμα που έχει καταγραφεί κατά τη λειτουργία ενός κινητήρα ιδίου τύπου. Σ' αυτή την περίπτωση επειδή η καταγραφή γίνεται με την ίδια πιστότητα, η ποσοτική πληροφορία που λαμβάνεται από την ανάλυση των δύο σημάτων στο πεδίο του χρόνου και στο πεδίο της συχνότητας δεν διαφέρει σημαντικά. Επίσης υπάρχουν κρίσιμες ομοιότητες μεταξύ των απολύτων τιμών των φασμάτων του πρότυπου ηχητικού σήματος και του προς αναγνώριση καταγεγραμμένου ήχου. Περαιτέρω ανάλυση έδειξε ότι χρησιμοποιώντας αυτές τις ομοιότητες προκύπτει μία πολύ ικανοποιητική και αποδοτική διαδικασία αναγνώρισης των δύο καταγεγραμμένων ηχητικών σημάτων, που βασίζεται στη σύγκριση δύο ιδίου μήκους, έστω N δειγμάτων, τμημάτων ήχου αυτών. Εάν τα δύο τμήματα είναι σε πλήρη αντιστοιχία, αφορούν δηλαδή κινητήρες που αφενός λειτουργούν σε ίδιες στροφές και αφετέρου παρουσιάζουν την ίδια φυσιολογική φθορά λόγω των συνολικών ωρών λειτουργίας των, τότε οι απόλυτες τιμές του Διακριτού Μετασχηματισμού Fourier (DFT) αυτών των δύο τμημάτων παρέχουν αρκετή πληροφορία για την εξακρίβωση της ομοιότητας («ταιριάσματος») των δύο ηχητικών σημάτων. Η βασική παρατήρηση που οδήγησε στην υλοποίηση της προτεινόμενης μεθοδολογίας ήταν ότι όταν το ηχητικό αποτέλεσμα των υπό σύγκριση ήχων κινητήρα είναι σχεδόν όμοιο, τότε οι κορυφές του σήματος στο μετασχηματισμό Fourier μπορεί να διαφέρουν αισθητά μεν στο πλάτος, αλλά ελαφρά ως προς τη θέση που εμφανίζονται. Για παράδειγμα, μπορεί στο πρότυπο ηχητικό σήμα μια κορυφή να εμφανίζεται στο σημείο του DFT που αντιστοιχεί στα 220 Hz ενώ στο αντίστοιχο υπό εξέταση καταγεγραμμένο σήμα να βρίσκεται ± 2 ή 3 Hz από την κανονική συχνότητα των 220 Hz.

2.2 Αξιοποίηση του φαινομένου Masking

Ένας εκτεταμένος αριθμός πειραμάτων για την εξακρίβωση της ομοιότητας των δύο ακουστικών σημάτων με επεξεργασία της πληροφορίας που λαμβάνουμε από τον υπολογισμό της απόλυτης τιμής του DFT αυτών, κατέδειξε ότι μπορούμε να μειώσουμε την ποσότητα της πληροφορίας δημιουργώντας ένα κατώφλι πλάτους για

κάθε συχνότητα f του φάσματος, χωρίς όμως να μειώνετε η ποιότητα αυτής. Αυτό επιτευχθεί αξιοποιώντας την συνάρτηση διασποράς του φαινομένου masking. Ως masking αναφέρεται στην διεθνή βιβλιογραφία το ακουστικό φαινόμενο κατά το οποίο ισχυρής εντάσεως ήχοι καλύπτουν άλλους χαμηλότερης εντάσεως, και οφείλεται στην ύπαρξη κρίσιμων ζωνών συχνοτήτων για το ανθρώπινο ακουστικό σύστημα.

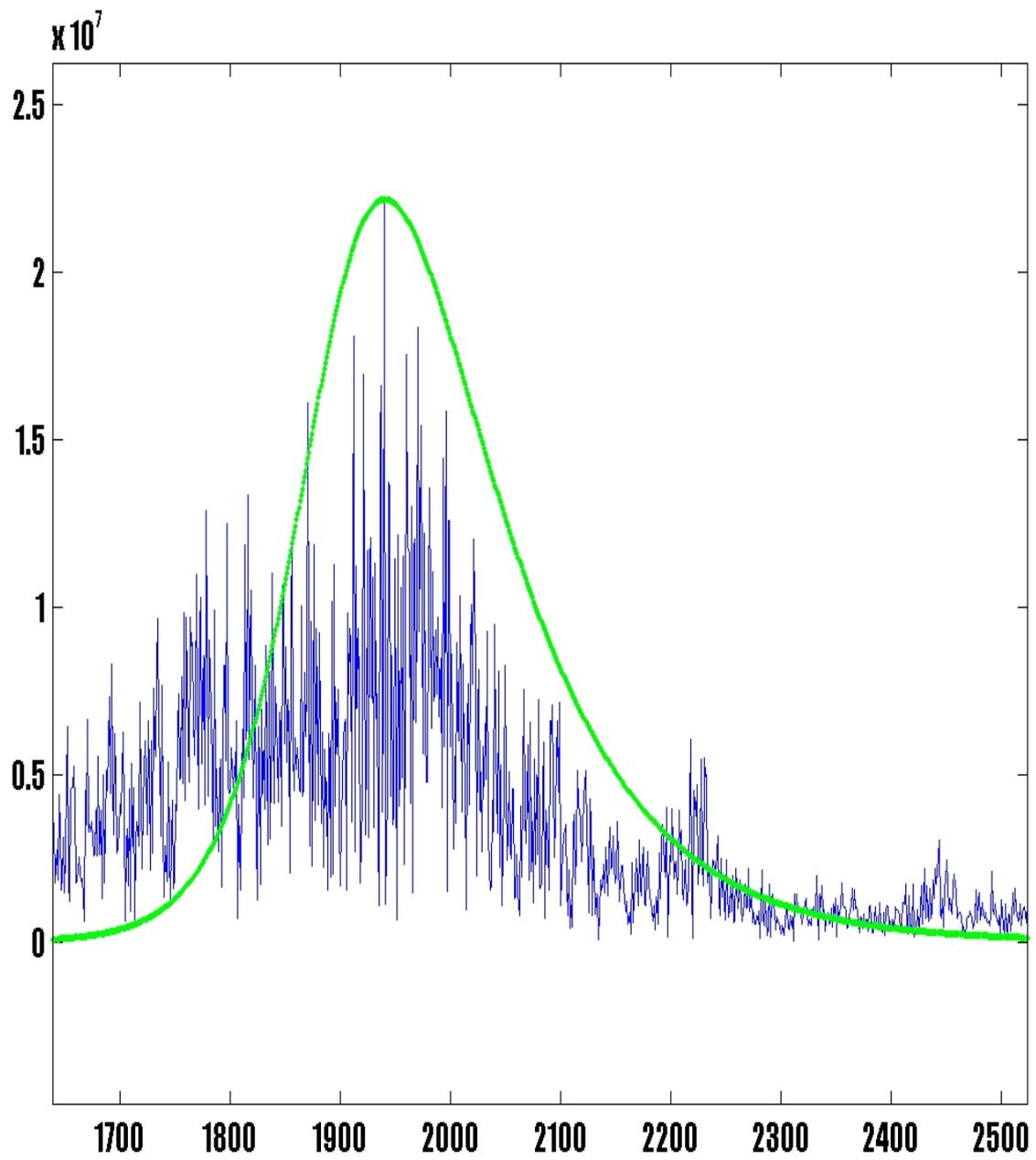
Η διαδικασία που εφαρμόστηκε, και χαρακτηρίστηκε στο πλαίσιο της παρούσης μελέτης ως διαδικασία “masking-like”, είναι η εξής: Σε κάθε επιλεγμένο τμήμα ήχου κινητήρα υπολογίζουμε το DFT. Μετά τον υπολογισμό της απόλυτης τιμής του DFT εντοπίζουμε τις κορυφές (μέγιστα) αυτού και αποθηκεύουμε τις θέσεις τους στο φάσμα καθώς και τα αντίστοιχα πλάτη τους. Στη συνέχεια ταξινομούμε αυτές τις κορυφές κατά φθίνουσα σειρά ανάλογα με το πλάτος τους. Ξεκινώντας από την συχνότητα /κορυφή f_0 με το μεγαλύτερο πλάτος εφαρμόζουμε τις ακόλουθες εξισώσεις (2.1) και (2.2) οι οποίες περιγράφουν τη συνάρτηση διασποράς του φαινομένου masking, περιορίζοντας κάθε φορά τη δράση της συνάρτησης μεταξύ των συχνοτήτων $f_0 - 27$ και $f_0 + 27$.

$$F(z) = 15.81 + 7.5 * (z + 0.474) - 17.5 * \sqrt{1 + (z + 0.474)^2} \quad (2.1)$$

$$z = 13 * \arctan(0.00076(f - f_0)) + 3.5 * \arctan\left(\left(\frac{f - f_0}{7500}\right)^2\right) \quad (2.2)$$

Εάν το πλάτος μιας κορυφής στη συχνότητα f είναι μικρότερο από την τιμή της συνάρτησης διασποράς $F(f)$ στη συγκεκριμένη θέση, τότε απομακρύνουμε την κορυφή f από τη λίστα με τις ταξινομημένες κορυφές. Επαναλαμβάνουμε την ίδια διαδικασία επιλέγοντας ως f_0 τη συχνότητα της αμέσως επόμενης κορυφής που έχει εναπομένει στη λίστα μέχρι να εξαντλήσουμε όλες τις κορυφές. Η ανανεωμένη πλέον λίστα των κορυφών για κάθε τμήμα ήχου κινητήρα φυλάσσεται προκειμένου να εφαρμοσθεί η διαδικασία που θα περιγραφεί στην επόμενη ενότητα.

Στο ακόλουθο σχήμα (Σχήμα 1) παρουσιάζεται ενδεικτικά η απεικόνιση μίας επιλεγμένης συνάρτησης διασποράς masking στον DFT ενός σήματος. Οι κορυφές που είναι «κάτω» από αυτή (καμπύλη με πράσινο χρώμα) «κόβονται».



Σχήμα 1

2.3 Η ιδέα των «διάνυσμάτων αντιπροσώπων συχνοτικών περιοχών»

Από την διαδικασία που περιγράφηκε στην ανωτέρω ενότητα δημιουργήθηκε μία ανανεωμένη λίστα κορυφών για κάθε τμήμα ήχου κινητήρα. Αυτή η λίστα περιλαμβάνει κορυφές της απόλυτης τιμής του DFT του ηχητικού σήματος που έχουν παραμείνει μετά από τη διαδικασία του “masking-like”, και αποτελεί την κρίσιμη πλέον πληροφορία στην οποία βασίζεται όλη η μεθοδολογία για την εξακρίβωση της ομοιότητας δύο ηχητικών σημάτων. Ο τρόπος συνεπώς που θα αποθηκευτεί αυτή η πληροφορία αποτελεί ένα σημαντικό παράγοντα για την ταχύτερη και αποτελεσματικότερη λειτουργία του συστήματός μας.

Η διαδικασία που εφαρμόστηκε για την αποθήκευση της εν λόγω πληροφορίας, είναι παρόμοια με αυτή που εφαρμόστηκε στη μεθοδολογία για την αυτόματη αναγνώριση μουσικών συνθέσεων, ομιλίας ή και ηχητικών προτύπων και είναι η εξής: Χωρίσαμε όλο το ηχητικό φάσμα σε 54 περιοχές συχνοτήτων όπως εμφανίζονται στον Πίνακα 1. Ο αριθμός και το μέγεθος των περιοχών συχνοτήτων καθορίστηκε μετά από πολλές διαδικασίες δοκιμής. Σε κάθε περιοχή συχνοτήτων ελέγχεται αν υπάρχουν, μετά από τη διαδικασία του “masking-like”, κορυφές της απόλυτης τιμής του DFT που να ανήκουν σε αυτή την περιοχή, και αν ναι, επιλέγεται μεταξύ αυτών, η κορυφή με το μεγαλύτερο πλάτος ως αντιπρόσωπος της περιοχής. Στη συνέχεια, οι δείκτες των περιοχών που αντιστοιχούν στις L (συνήθως $14 \leq L \leq 30$) κορυφές αντιπροσώπους με τα μεγαλύτερα πλάτη, αποθηκεύονται σε ένα ξεχωριστό πίνακα, ο οποίος θα αναφέρεται ως «διάνυσμα αντιπροσώπων περιοχών» ή «πίνακας αντιπροσώπων περιοχών», και τα στοιχεία του πίνακα θα αποκαλούνται ως «αντιπρόσωποι περιοχών».

Η παραπάνω διαδικασία εφαρμόζεται όπως είναι φυσικό σε όλα τα ηχητικά σήματα που αποτελούν την βάση προτύπων ηχητικών σημάτων για το συγκεκριμένο τύπο κινητήρα, καθώς και στο προς σύγκριση καταγεγραμμένο σήμα που έχει ληφθεί κατά τη λειτουργία ενός ιδίου τύπου κινητήρα.

Δείκτης περιοχής	Εύρος περιοχής (σε Hz)	Δείκτης περιοχής	Εύρος περιοχής (σε Hz)	Δείκτης περιοχής	Εύρος Περιοχής (σε Hz)
0	0 - 50	18	285- 306	36	1089 – 1172
1	51- 80	19	307 - 329	37	1173 – 1263
2	81- 90	20	330 - 355	38	1264 – 1360
3	91- 100	21	356 - 382	39	1361 – 1466
4	100 - 107	22	383 - 412	40	1467 – 1579
5	108 - 116	23	413 - 444	41	1580 – 1702
6	117 - 125	24	445 - 478	42	1703 – 1834
7	126 - 134	25	479 - 516	43	1835 – 1976
8	135 - 145	26	517 - 556	44	1977 – 2129
9	146 - 156	27	557 - 599	45	2130 – 2293
10	157 - 168	28	600 - 645	46	2294 – 2471
11	169 - 181	29	646 - 695	47	2472 – 2663
12	182 - 195	30	696 - 749	48	2664 – 2869
13	196 - 210	31	750 - 807	49	2870 – 3090
14	211 - 227	32	808 - 869	50	3091 – 3326
15	228 - 244	33	870 - 937	51	3327 – 3577
16	245 - 263	34	938 - 1009	52	3578 – 3835
17	264 - 284	35	1010 - 1088	53	3836 – 11025

Πίνακας 1

Διαίρεση ηχητικού φάσματος σε 54 περιοχές συχνοτήτων

3. ΚΡΙΤΗΡΙΑ ΟΜΟΙΟΤΗΤΑΣ ΗΧΩΝ ΚΙΝΗΤΗΡΑ

Λαμβάνοντας υπόψη την προαναφερθείσα ανάλυση, το ερώτημα που τίθεται είναι ποια πρέπει να είναι τα κατάλληλα κριτήρια έτσι ώστε να αποφασισθεί πότε δύο συγκρινόμενα ηχητικά σήματα αντιστοιχούν στην ίδια λειτουργική κατάσταση του κινητήρα. Στο παρόν κεφάλαιο αναλύονται διεξοδικά τα κριτήρια που επελέγησαν και οδήγησαν σε ιδιαίτερος επιτυχή ταυτοποίηση ηχητικών σημάτων.

3.1 Αρχικά κριτήρια ομοιότητας ηχητικών σημάτων

Ένας πολύ μεγάλος αριθμός πειραμάτων έδειξε ότι τουλάχιστον P από τα L στοιχεία ενός διανύσματος αντιπροσώπων συχνοτικών περιοχών συμπεριλαμβάνονται και σε δύο προς σύγκριση ηχητικά σήματα όταν αυτά αφορούν την ίδια λειτουργική κατάσταση του κινητήρα. Εξάχθηκε το συμπέρασμα ότι όταν $14 \leq L \leq 30$ τότε ο ελάχιστος αριθμός P των κοινών αντιπροσώπων περιοχών πρέπει να είναι μεγαλύτερος από $0.55 * L$. Αυτός ο αριθμός είναι ανεξάρτητος της εκάστοτε λειτουργικής κατάστασης του κινητήρα και ο περιορισμός $0.55 * L \leq P$ συνιστά ένα πρώτο κανόνα για την εύρεση πιθανού ταιριάσματος μεταξύ δύο ηχητικών τμημάτων.

Σε αυτό το σημείο όμως πρέπει να τονίσουμε ότι ο Διακριτός Μετασχηματισμός Fourier ενός μόνο πλαισίου δεν περιέχει πληροφορία για την εξέλιξη του σήματος στο χρόνο. Επιπρόσθετα, είναι αρκετά πιθανό δύο τμήματα ήχου μήκους N δειγμάτων, που προέρχονται από διαφορετικές λειτουργικές καταστάσεις κινητήρα να ικανοποιούν τον παραπάνω κανόνα των τουλάχιστον $0.55 * L$ κοινών αντιπροσώπων περιοχών. Η ποικιλία των λειτουργικών καταστάσεων του κινητήρα είναι τέτοια και ο αριθμός επίσης των παραθύρων-τμημάτων ήχων που εξάγονται από ένα και μόνο τμήμα ήχου είναι τόσο μεγάλος, που είναι πάρα πολύ πιθανό να βρεθεί ένα τμήμα καταγεγραμμένου ήχου κινητήρα και ένα τμήμα προτύπου ηχητικού σήματος να ικανοποιούν τον κανόνα, αλλά να μην αντιστοιχούν στην ίδια λειτουργική κατάσταση του κινητήρα.

Για τον ανωτέρω λόγο και προκειμένου να λαμβάνεται υπόψη και το πεδίο του χρόνου, η προηγούμενη ανάλυση επεκτάθηκε σε πολλά τμήματα (παράθυρα) των ηχητικών σημάτων, ώστε να εμπλουτιστεί ο κανόνας $0.55 * L$ κοινών αντιπροσώπων περιοχών. Έτσι, επιλέχθηκαν παράθυρα κατανεμημένα σε ίδιες αποστάσεις, σε ένα

εύρος δέκα ως δεκαπέντε δευτερολέπτων στο κάθε ηχητικό σήμα. Τότε, εάν δύο ηχητικά σήματα αντιστοιχούν στην ίδια λειτουργική κατάσταση του κινητήρα, μπορούμε να διατυπώσουμε με ασφάλεια ότι ο κανόνας των τουλάχιστον $0.55 * L$ κοινών αντιπροσώπων περιοχών ανά παράθυρο, πρέπει να ισχύει σε κάθε ένα από τα ζεύγη συγκρινόμενων παραθύρων. Με αυτό τον τρόπο υλοποιείται το πρώτο κριτήριο ομοιότητας δύο τμημάτων ηχητικών σημάτων, που περιέχει πληροφορία τόσο στο πεδίο του χρόνου όσο και στο πεδίο της συχνότητας ως εξής:

ΠΡΩΤΟ ΚΡΙΤΗΡΙΟ ΤΑΥΤΟΠΟΙΗΣΗΣ ΔΥΟ ΗΧΗΤΙΚΩΝ ΤΜΗΜΑΤΩΝ

Απαιτείται ικανοποίηση του κριτηρίου $0.55 * L \leq P$, σε M ζεύγη παραθύρων, με κάθε παράθυρο σε απόσταση ℓ δειγμάτων από το προηγούμενο, σε κάθε ηχητικό τμήμα.

Επιπρόσθετα, ένα δεύτερο κριτήριο ομοιότητας, είναι το κάτωθι:

ΔΕΥΤΕΡΟ ΚΡΙΤΗΡΙΟ ΤΑΥΤΟΠΟΙΗΣΗΣ ΔΥΟ ΗΧΗΤΙΚΩΝ ΤΜΗΜΑΤΩΝ

Απαιτείται ο μέσος όρος των κοινών αντιπροσώπων περιοχών ανά παράθυρο να είναι μεγαλύτερος από $0.71 * L$ και ταυτόχρονα το πολύ $\left\lfloor \frac{M}{3} \right\rfloor$ ζεύγη παραθύρων να έχουν κοινούς αντιπροσώπους συχνοτικών περιοχών μεταξύ $0.55 * L$ και $0.71 * L$ ($\lfloor x \rfloor$ είναι το ακέραιο μέρος του πραγματικού αριθμού x).

Η εφαρμογή των παραπάνω κριτηρίων έδωσε ένα ποσοστό επιτυχούς ταυτοποίησης καταγεγραμμένων ηχητικών σημάτων τουλάχιστον 98%.

3.2 Τελεσίδικο κριτήριο ομοιότητας ηχητικών σημάτων.

Ένα ακόμα αποτέλεσμα των πειραμάτων ήταν ότι σε μερικές σπάνιες περιπτώσεις (στο 2% περίπου όλων των περιπτώσεων), υπήρχαν δύο ή περισσότερα πρότυπα ηχητικά σήματα που ικανοποιούσαν τα προαναφερθέντα κριτήρια συγκρινόμενα με την ίδιο προς αναγνώριση καταγεγραμμένο ήχο κινητήρα. Για αυτό το λόγο έπρεπε να τεθεί ένα τελεσίδικο κριτήριο έτσι ώστε να γίνει δυνατό να διακριθεί σε ποιο ακριβώς πρότυπο ηχητικό σήμα αντιστοιχεί το καταγεγραμμένο. Το τελεσίδικο αυτό κριτήριο χρησιμοποιεί ένα μεγαλύτερο παράθυρο στο πεδίο της συχνότητας και έναν αλγόριθμο ταιριάσματος περιβαλλουσών (envelope matching) με σκοπό να αποφασισθεί οριστικά η ταυτοποίηση του καταγεγραμμένου ήχου.

Χρησιμοποιείται ένα παράθυρο μήκους WL δειγμάτων, έστω $WL = 2^{17}$, που αρχίζει στο σημείο που τα προηγούμενα κριτήρια αναφέρουν πιθανό ταίριασμα ανάμεσα στο πρότυπο και στο καταγεγραμμένο ηχητικό σήμα. Σε αυτά τα παράθυρα εφαρμόζεται ο DFT, υπολογίζεται η απόλυτη τιμή του και δημιουργούνται οι περιβάλλουσες αυτών των παραθύρων στο πεδίο της συχνότητας ακολουθώντας τα παρακάτω βήματα:

Βήμα 1: Υπολογίζονται όλα τα τοπικά ακρότατα και στα δύο τμήματα των σημάτων και εφαρμόζεται γραμμική παρεμβολή σε αυτά. Με αυτό τον τρόπο λαμβάνεται μία πρώτη εκτίμηση της περιβάλλουσας και των δύο τμημάτων. Παρ' όλα αυτά, οι δύο υπολογισμένες περιβάλλουσες ακολουθούν με μεγάλη λεπτομέρεια τα δύο σήματα, κάτι το οποίο δεν είναι ικανοποιητικό για τις ανάγκες της παρούσας εφαρμογής. Για αυτό το λόγο, επαναλαμβάνουμε την προηγούμενη διαδικασία με την έννοια ότι:

Βήμα 2: Υπολογίζονται όλα τα τοπικά ακρότατα στη πρώτη περιβάλλουσα των σημάτων και εφαρμόζουμε γραμμική παρεμβολή σε αυτά. Με αυτό τον τρόπο λαμβάνεται μία δεύτερη εκτίμηση της περιβάλλουσας για κάθε τμήμα η οποία όμως είναι ακόμα πολύ «κοντινή» στα δύο σήματα για την παρούσα εφαρμογή.

Βήμα 3: Υπολογίζονται όλα τα τοπικά ακρότατα στη δεύτερη περιβάλλουσα των σημάτων και εφαρμόζουμε γραμμική παρεμβολή σε αυτά λαμβάνοντας με αυτό τον τρόπο μία τρίτη εκτίμηση της περιβάλλουσας για κάθε τμήμα. Αυτή θεωρείται η τελική περιβάλλουσα του Διακριτού Μετασχηματισμού Fourier (βλέπε Σχήμα 2)

Βήμα 4: Υπολογίζουμε κατ' αρχήν τα ολοκληρώματα των δύο περιβαλλουσών. Ακολουθώντας τις κανονικοποιούμε διαιρώντας τις τιμές της κάθε περιβάλλουσας με το αντίστοιχο ολοκλήρωμα.

Βήμα 5: Τελικά υπολογίζεται το άθροισμα των απόλυτων τιμών των διαφορών των τιμών των δύο κανονικοποιημένων περιβαλλουσών σε όλα τα αντίστοιχα σημεία των δύο φασμάτων.

Παρατηρήσαμε ότι αν τα δύο τμήματα ήχου (πρότυπου ηχητικού σήματος και καταγεγραμμένου ήχου) αντιστοιχούν στο ίδιο τμήμα ήχου μιας συγκεκριμένης λειτουργικής κατάστασης του κινητήρα, τότε το προηγούμενο άθροισμα των απόλυτων τιμών έχει τιμή μικρότερη από ένα συγκεκριμένο κατώφλι. Η ακριβής τιμή αυτού του κατωφλιού εξαρτάται από το μήκος του παράθυρου που χρησιμοποιείται για τον υπολογισμό του Διακριτού Μετασχηματισμού Fourier.

Συμπερασματικά, μπορούμε να πούμε ότι η τιμή του αθροίσματος των απόλυτων τιμών των διαφορών των περιβαλλουσών των DFT των δύο σημάτων, είναι αναμφίβολα ένα πολύ αποτελεσματικό κριτήριο για να ληφθεί η απόφαση αν τα δύο σήματα αντιστοιχούν στην λειτουργική κατάσταση του κινητήρα. Η απευθείας όμως εφαρμογή αυτού του κριτηρίου θα απαιτούσε πολύ μεγάλο όγκο υπολογισμών, γι' αυτό χρησιμοποιείται μόνο στο τελικό στάδιο της αναγνώρισης για να πιστοποιήσει ορθή αναγνώριση, δηλαδή να ξεχωρίσει, μέσα από ένα πολύ μικρό σύνολο πρότυπων ηχητικών σημάτων, ποιο αντιστοιχεί στον άγνωστο καταγεγραμμένο ήχο κινητήρα.

4. ΥΛΟΠΟΙΗΣΗ ΤΗΣ ΠΡΟΤΕΙΝΟΜΕΝΗΣ ΜΕΘΟΔΟΛΟΓΙΑΣ

4.1 Δημιουργία ενός συνόλου διανυσμάτων αντιπρόσωπων συχνοτικών περιοχών για έναν πρότυπο ήχο κινητήρα

Προκειμένου να εφαρμοσθούν τα παραπάνω κριτήρια ομοιότητας, εκτελούνται τα παρακάτω βήματα για την εξαγωγή των διανυσμάτων αντιπρόσωπων συχνοτικών περιοχών του τμήματος του πρότυπου ήχου κινητήρα:

A1: Ηχογραφείται για διάρκεια 5 λεπτών περίπου ο κινητήρας του οποίου η λειτουργική κατάσταση θεωρείται ως πρότυπο και μετατρέπεται με κάποιο από τα γνωστά εργαλεία σε μορφή "wav", έτσι ώστε να μπορεί να επεξεργαστεί κατάλληλα από τον υπολογιστή, με συχνότητα δειγματοληψίας $F_s = 22050$. Στην ακόλουθη ανάλυση θα αναφερόμαστε σε αυτό ως το «πρότυπο ηχητικό σήμα», στη δε διάρκεια του πρότυπου ηχητικού σήματος με τον όρο ML .

A2: Στην αρχή αυτού του σήματος επιλέγεται το πρώτο παράθυρο N δειγμάτων (για το υλοποιημένο σύστημα έχει επιλεγθεί $N = 2^{13} = 8192$) και εφαρμόστηκε Διακριτός Μετασχηματισμός Fourier σε αυτό.

A3: Υπολογίζεται η απόλυτη τιμή του μετασχηματισμού και εφαρμόζεται η "masking-like" διαδικασία που περιγράφεται στην Ενότητα 2.2. Κατόπιν διαιρούνται τα N δείγματα σε 54 συχνοτικές περιοχές (βλέπε Πίνακα 1) και σε κάθε περιοχή κρατείται η κορυφή με το μεγαλύτερο πλάτος, αν βέβαια υπάρχει μέγιστο σε αυτή την περιοχή. Σε αυτή την περίπτωση το πλάτος της κορυφής θεωρείται από εδώ και στο

εξής ως το πλάτος ολόκληρης της συχνοτικής περιοχής, αλλιώς θεωρείται ότι η περιοχή έχει μέγιστη κορυφή πλάτους μηδέν.

A4: Καταχωρούνται οι δείκτες των L συχνοτικών περιοχών με το μεγαλύτερο πλάτος σε έναν πίνακα. Με αυτό τον τρόπο λαμβάνεται ένα διάνυσμα αντιπροσώπων συχνοτικών περιοχών που αντιστοιχεί στο πρώτο παράθυρο του πρότυπου ηχητικού σήματος, με περιεχόμενο τις τιμές των προαναφερθέντων δεικτών.

B1: Επιλέγεται το επόμενο παράθυρο N δειγμάτων (όπως στο **A2**), που αρχίζει ένα δείγμα μετά από την αρχή του προηγούμενου παραθύρου και εφαρμόζουμε τον προσαρμοστικό Διακριτό Μετασχηματισμό Fourier (βλέπε Ενότητα 4.2) σε αυτό.

B2: Ακολουθώς επαναλαμβάνουμε τα βήματα **A3**, **A4** και για αυτό το παράθυρο. Η διαδικασία αυτή επαναλαμβάνεται μέχρι το τελευταίο δείγμα του πρότυπου ήχου κινητήρα.

Με αυτό τον τρόπο λαμβάνονται $ML - N + 1$ διανύσματα αντιπροσώπων συχνοτικών περιοχών που αντιστοιχούν στα παραπάνω παράθυρα, τα οποία βεβαίως είναι συνεχόμενα και σε μεγάλο βαθμό επικαλυπτόμενα. Αν σκεφθούμε ότι ο πρότυπος ήχος κινητήρα έχει διάρκεια τουλάχιστον 5 λεπτών και δεδομένης της συχνότητας δειγματοληψίας των 22050 Hz που χρησιμοποιούμε προκύπτει ότι μια μουσική σύνθεση αποτελείται από $5 \cdot 60 \cdot 22050 = 6615000$ δείγματα. Άρα, προφανώς η απ' ευθείας εκτέλεση των υπολογισμών του DFT είναι απαγορευτική ακόμα κι αν χρησιμοποιηθεί πολύ γρήγορος υπολογιστής. Για να αποφευχθεί η εκτέλεση του κλασικού DFT αναπτύχθηκε ένας νέος αλγόριθμος «adaptive DFT» που προσφέρει δραματική βελτίωση της απαιτούμενης υπολογιστικής ισχύος και περιγράφεται στην Ενότητα 4.2. Εκτός από τον υπολογιστικό φόρτο για τις δομές που προκύπτουν από την παραπάνω διαδικασία απαιτείται τεράστιος αποθηκευτικός χώρος και για αυτό τον λόγο η αποθήκευση των παραπάνω διανυσμάτων γίνεται με έναν πολύ αποτελεσματικό τρόπο που έχει πολύ μικρές αποθηκευτικές απαιτήσεις και περιγράφεται στην Ενότητα 4.3.

4.2 Αλγόριθμος προσαρμοστικού (adaptive) DFT

Ας υποθέσουμε ότι έχουμε υπολογίσει τον Διακριτό Μετασχηματισμό Fourier ενός σήματος $x[n]$ των N δειγμάτων, ξεκινώντας από το σημείο a του πεδίου του χρόνου και τελειώνοντας στο σημείο $(a+N-1)$. Έστω τώρα ότι θέλουμε να

υπολογίσουμε τον μετασχηματισμό Fourier N δειγμάτων του ίδιου σήματος $x[n]$, αλλά τώρα ξεκινώντας από το σημείο $(a+1)$ και τελειώνοντας στο σημείο $(a+N)$. Ανεξάρτητα του τρόπου με τον οποίο έχει υπολογιστεί ο πρώτος μετασχηματισμός που αντιστοιχεί στο σημείο a , ο υπολογισμός του επόμενου μετασχηματισμού Fourier μπορεί να γίνει με πολύ λιγότερους υπολογισμούς, με τη χρήση της πληροφορίας που υπάρχει ήδη στον προηγούμενο.

Κατ' αρχήν υπενθυμίζουμε ότι ο μετασχηματισμός Fourier ενός παραθύρου N δειγμάτων, που ξεκινά από το σημείο a και τελειώνει στο σημείο $(a+N-1)$ δίνεται από την παρακάτω σχέση:

$$X[k] = \sum_{n=0}^{N-1} x[n+a]W^{kn} \quad (\text{A.5}), \text{ όπου } W = e^{-j\left(\frac{2\pi}{N}\right)} \quad (4.1)$$

ενώ ο μετασχηματισμός Fourier ενός παραθύρου N δειγμάτων, που ξεκινά από το σημείο $(a+1)$ και τελειώνει στο σημείο $(a+N)$ δίνεται από την πιο κάτω σχέση:

$$X_s[k] = \sum_{n=0}^{N-1} x[n+a+1]W^{kn} \quad (4.2)$$

Όμως ευθύγραμμα προκύπτει ότι:

$$X_s[k] = (-x[a] + x[a]) * W^{-k} + \left(\sum_{n=0}^{N-2} x[n+a+1]W^{k(n+1)} \right) * W^{-k} + x[a+N] * W^{k(N-1)} \quad (4.3)$$

Στο τελευταίο άθροισμα αντικαθιστούμε το $n+1$ με την βοηθητική μεταβλητή i

$$i = n + 1 \quad (4.4)$$

οπότε στη συνέχεια λαμβάνουμε:

$$X_s[k] = -x[a] * W^{-k} + x[a] * W^{-k} + \left(\sum_{i=1}^{N-1} x[i+a]W^{ki} \right) * W^{-k} + x[a+N] * W^{k(N-1)} \quad (4.5)$$

ή ισοδύναμα:

$$X_s[k] = -x[a] * W^{-k} + \left(\sum_{i=0}^{N-1} x[i+a]W^{ki} \right) * W^{-k} + x[a+N] * W^{k(N-1)} \quad (4.6)$$

Αλλά η ποσότητα $\sum_{i=0}^{N-1} x[i+a]W^{ki}$ είναι ο μετασχηματισμός Fourier του παραθύρου

που ξεκινά από το δείγμα a . Έτσι τελικά αποδεικνύεται ότι:

$$X_s[k] = -x[a]*W^{-k} + X[k]*W^{-k} + x[a+N]*W^{k(N-1)} \quad (4.7)$$

το οποίο και χρησιμοποιούμε ως βασικό τύπο για τον αναδρομικό υπολογισμό του Διακριτού Μετασχηματισμού Fourier.

Η υλοποίηση αυτού του αλγορίθμου είναι σχετικά εύκολη. Από τον τύπο (4.7) φαίνεται ότι κάθε δείγμα του μετασχηματισμού Fourier υπολογίζεται πολύ απλά. Οι ποσότητες W^{-k} και $W^{k(N-1)}$ υπολογίζονται μόνο μια φορά στην αρχή της διαδικασίας. Από κει και πέρα, για κάθε δείγμα απαιτούνται 6 πολλαπλασιασμοί πραγματικών αριθμών και 5 προσθαιρέσεις. Έχει δηλαδή συνολικά πολυπλοκότητα $O(N)$ ή αναλυτικά χρειάζεται $6N$ πολλαπλασιασμούς. Από την άλλη, ο απλός αλγόριθμος γρήγορου μετασχηματισμού Fourier έχει πολυπλοκότητα $O\left(\frac{N}{2} * \log_2(N)\right)$ αλλά σε μιγαδικές πράξεις, επομένως χρειάζεται $2N * \log_2(N)$ πολλαπλασιασμούς πραγματικών αριθμών. Προφανώς ο παρουσιαζόμενος αλγόριθμος είναι πολύ γρηγορότερος από τον κλασικό DFT. Για μέγεθος παραθύρου μεγαλύτερο από 16 δείγματα, ο προσαρμοστικός μετασχηματισμός Fourier είναι πάντα πιο γρήγορος, ενώ με το χρησιμοποιούμενο παράθυρο των 8192 δειγμάτων, είναι 4,5 περίπου φορές γρηγορότερος, αν κανείς λάβει υπόψη τον αριθμό των πλέον πολύπλοκων πράξεων, των πολλαπλασιασμών. Σύγκριση του απαιτούμενου αριθμού πράξεων για τον υπολογισμό του φάσματος, πρώτα μέσω του απλού DFT και εν συνεχεία με τη χρήση του προσαρμοστικού (Adaptive) DFT, για διάφορα μεγέθη παραθύρων, με τα οποία η εφαρμογή μπορεί να λειτουργήσει με πολύ καλά αποτελέσματα, φαίνεται στον πιο κάτω πίνακα.

	Μέγεθος Παραθύρου		
	N=8*1024	N=16*1024	N=32*1024
Αριθμός Πράξεων για Απλό DFT	212992	458752	983040
Αριθμός Πράξεων για Adaptive DFT	49152	98304	196608

Πίνακας 2

Στην πράξη παρατηρήθηκε ότι ο αλγόριθμος έδινε πραγματικά αποτελέσματα 4,5 με 5 φορές ταχύτερα από τον απλό DFT. Σε αυτόν το χρόνο παίζουν ρόλο τόσο οι προσθήσεις και οι αφαιρέσεις, όσο και προβλήματα κατανομής και εκχώρησης μνήμης.

Η τελική εφαρμογή του προσαρμοστικού αλγορίθμου ξεκινά με την εκτέλεση ενός μετασχηματισμού Fourier. Στη συνέχεια εφαρμόζεται ο προσαρμοστικός μετασχηματισμός Fourier, ενώ κάθε 30000 περίπου δείγματα, το σύστημα επαναρχικοποιείται, εφαρμόζοντας ξανά ένα κλασικό DFT, για να αποφευχθεί η περίπτωση του αριθμητικού λάθους. Να σημειωθεί πάντως, ότι ο προσαρμοστικός DFT δοκιμάστηκε και σε πρότυπα ηχητικά σήματα, χωρίς καμία επανεκκίνηση, χωρίς να παρουσιάσει κανένα πρόβλημα αριθμητικής ευστάθειας και δίνοντας ακριβώς τα ίδια αποτελέσματα με τον κλασικό DFT, όταν αυτός λειτουργεί χωρίς λάθος πεπερασμένης ακρίβειας. Διότι αξίζει να σημειωθεί ότι ο κλασικός DFT μπορεί να αποτύχει τελείως για διάφορα σήματα εισόδου, λόγω του γενόμενου λάθους πεπερασμένης ακρίβειας.

4.3 Μέθοδος αποδοτικής αποθήκευσης των διανυσμάτων αντιπροσώπων των πρότυπων ηχητικών σημάτων

Όπως προαναφέρθηκε, για κάθε πρότυπο ηχητικό σήμα, λαμβάνεται ένα σύνολο από $ML - N + 1$ διανύσματα που το καθένα αποτελείται από L στοιχεία. Είναι πολύ συνηθισμένο δύο ή περισσότερα συνεχόμενα δείγματα να αντιστοιχούν σε πανομοιότυπα διανύσματα αντιπροσώπων. Για αυτό το λόγο σε κάθε τέτοιο διάνυσμα αποδίδεται ο αριθμός των δειγμάτων για τον οποίο παραμένει αναλλοίωτο, ο οποίος αποκαλείται «αριθμός επαναλήψεων του διανύσματος». Με αυτό τον τρόπο αποθηκεύεται το κάθε διαφορετικό διάνυσμα μόνο μία φορά, μαζί με τον αντίστοιχο αριθμό επαναλήψεων.

Παρ' όλα αυτά, τα διανύσματα που προκύπτουν είναι πολύ μεγάλα για να αποθηκευθούν σε μορφή ASCII ή ακόμα και σε τυπική δυαδική μορφή. Γι' αυτό το λόγο, δημιουργήσαμε ένα πρωτόκολλο για την αποθήκευση των δεδομένων σε κωδικοποιημένα αρχεία, σύμφωνα με το οποίο, σε κάθε διάνυσμα του πρότυπου ηχητικού σήματος ανατίθεται ένας δυαδικός πίνακας με πενήντα τέσσερα (54) στοιχεία. Κάθε στοιχείο αυτού του δυαδικού πίνακα αντιπροσωπεύει μία από τις πενήντα τέσσερις συχνοτικές περιοχές στις οποίες αποφασίσαμε να διαιρέσουμε

ολόκληρο το πεδίο της συχνότητας. Η τιμή “1” αποδίδεται σε ένα στοιχείο του πίνακα όταν η αντίστοιχη περιοχή ανήκει στις L περιοχές με το μεγαλύτερο μη μηδενικό πλάτος, ενώ αποδίδεται η τιμή “0” στην αντίθετη περίπτωση. Όπως είναι φανερό το πολύ L στοιχεία μπορεί να πάρουν τιμή “1”.

Ένα παράδειγμα θα διευκρινίσει περισσότερο την προτεινόμενη κωδικοποίηση. Πράγματι, ας υποθεθεί ότι για το παράθυρο που εκκινά στο δείγμα n στο πεδίο του χρόνου, έχουν υπολογισθεί οι $L = 15$ μεγαλύτεροι αντιπρόσωποι περιοχών οι οποίοι, ταξινομημένοι σε φθίνουσα σειρά, σχηματίζουν το παρακάτω διάνυσμα

[3128 3002 1523 1294 912 731 677 392 340 271 212 174 150 111 89]

με τις τιμές των στοιχείων σε Hertz. Προκειμένου να μειωθεί ο αποθηκευτικός χώρος, αντικαθίσταται η τιμή της συχνότητας (σε Hertz) αντιπροσώπου κάθε περιοχής με τον αντίστοιχο δείκτη της συχνοτικής περιοχής (σύμφωνα με τον Πίνακα 1). Έτσι λαμβάνεται το διάνυσμα:

[50 49 40 38 33 30 29 22 20 17 14 11 9 5 2]

Προκειμένου να μειωθεί περαιτέρω ο αποθηκευτικός χώρος, δημιουργήσαμε ένα δυαδικό πίνακα από 54 δυαδικά ψηφία, όλα τα στοιχεία του οποίου είναι μηδέν, εκτός από αυτά με θέση που αντιστοιχεί σε δείκτη συχνοτικής περιοχής στο παραπάνω διάνυσμα. Έτσι λαμβάνεται ο πίνακας

[00100100010100100100101000000110010000101000000011000]

ο οποίος αποθηκεύεται ως δυαδική ακολουθία. Τέλος, αποθηκεύεται μαζί με αυτή την ακολουθία bits ένας δυαδικός αριθμός που αντιστοιχεί στον αριθμό των επαναλήψεων του διανύσματος δηλ. τον αριθμό των συνεχόμενων δειγμάτων στο πεδίο του χρόνου για τα οποία η συγκεκριμένη ακολουθία εμφανίζεται. Κατ’ αυτόν τον τρόπο, για κάθε ένα διάνυσμα χρειάζονται 7 bytes για τα στοιχεία του και 4 για τον αριθμό των επαναλήψεων. Με αυτή τη μέθοδο δυαδικής αποθήκευσης μειώνεται σημαντικά ο αριθμός των bytes που απαιτούνται για την αποθήκευση του συνόλου των διανυσμάτων αντιπροσώπων ενός πρότυπου ηχητικού σήματος.

Η αποκωδικοποίηση των δεδομένων είναι σχετικά απλή και ευθύγραμμη διαδικασία: Από τη δυαδική λέξη των 7 bytes μπορούμε εύκολα να ανακτήσουμε το αρχικό διάνυσμα αντιπρόσωπο. Ξεκινώντας από το 1ο bit αν βρεθεί η τιμή ‘1’

σημειώνουμε το δείκτη συχνοτικής περιοχής που αντιστοιχεί στο bit αυτό στην πρώτη θέση του διανύσματος αντιπροσώπου. Ο δείκτης που αντιστοιχεί στο δεύτερο μη μηδενικό bit σημειώνεται στη δεύτερη θέση του αποκωδικοποιημένου διανύσματος αντιπροσώπου κ.ο.κ. Στην περίπτωση που δεν συμπληρωθούν οι L συνολικές θέσεις του διανύσματος, οι θέσεις του διανύσματος που απομένουν παίρνουν την τιμή '-1', αφού αυτό σημαίνει ότι δεν είχαμε ανιχνεύσει L συχνοτικές περιοχές με μη μηδενικό μέτρο στη διαδικασία κατασκευής του διανύσματος. Με τη διαδικασία αυτή, που απαιτεί μόνο ένα γραμμικό πέρασμα σε κάθε bit του κωδικοποιημένου διανύσματος και καμία διαδικασία ταξινόμησης, το διάνυσμα που ανακατασκευάζεται έχει ακριβώς τη μορφή που είχε και πριν, δηλαδή τα στοιχεία του είναι ταξινομημένα με φθίνουσα σειρά.

Παρότι η κωδικοποίηση αυτή προσφέρει σαφώς καλύτερη αποθήκευση των διανυσμάτων, έγινε ακόμα μια προσπάθεια μείωσης του όγκου που καταλαμβάνουν τα αποθηκευμένα στοιχεία. Η ιδέα είναι να εφαρμόσουμε μια τεχνική που θα επιτρέπει την αποθήκευση του αριθμού των επαναλήψεων με ένα, δύο ή τρία bytes, ανάλογα με το μέγεθος του αριθμού των επαναλήψεων κάθε φορά, ώστε όταν αυτός είναι μικρότερος από 256, να χρησιμοποιούμε ένα byte, όταν είναι μεταξύ 256 και 65535, να χρησιμοποιούμε δύο bytes κι όταν είναι μεγαλύτερος από 65536 να χρησιμοποιούμε τρία bytes. Το πάνω όριο του αριθμού των επαναλήψεων με τη χρήση 3 bytes είναι 16777315, δηλαδή αντιστοιχεί σε πάνω από 12 λεπτά ηχογράφησης πρότυπου ηχητικού σήματος κινητήρα γεγονός που καλύπτει πρακτικά την εφαρμογή μας.

Λαμβάνοντας υπ' όψιν την προηγούμενη ανάλυση, μπορούμε τώρα να ορίσουμε την έννοια «διάνυσμα προς αποθήκευση», ως το διάνυσμα, το οποίο συνίσταται από ένα διάνυσμα αντιπρόσωπο και τον αριθμό επαναλήψεων αυτού.

Μια εναλλακτική μορφή συμπίεσης των αποθηκευμένων διανυσμάτων αντιπροσώπων, μπορεί να επιτευχθεί αν εκμεταλλευτούμε το γεγονός ότι, στην πράξη, ο αριθμός των μη κοινών στοιχείων ανάμεσα σε ένα τυχόν διάνυσμα προς αποθήκευση και στο αμέσως επόμενο του, είναι ελάχιστος. Πραγματικά, τις περισσότερες φορές το ένα διάνυσμα από το επόμενο του διαφέρει κατά ένα ή δύο στοιχεία. Η σκέψη λοιπόν είναι, να μην αποθηκεύουμε όλο το διάνυσμα αν αυτό δεν έχει πάρα πολλές αλλαγές σε σχέση με το προηγούμενό του, αλλά να αποθηκεύουμε,

με κατάλληλο τρόπο, μόνο τις αλλαγές που έχουν συμβεί σε σχέση με το προηγούμενο διάνυσμα.

Υπενθυμίζουμε ότι ο στόχος είναι να πετύχουμε μικρότερη κωδικοποίηση από τα 7 bytes η οποία επετεύχθη με τη χρήση της προαναφερθείσας δυαδικής κωδικοποίησης. Άρα, αν χρησιμοποιούμε δύο bytes για κάθε αλλαγή που συμβαίνει μεταξύ του υπό εξέταση διανύσματος προς αποθήκευση και του προηγούμενου διανύσματος και με δεδομένο ότι θα χρησιμοποιήσουμε κάποια bits για ένδειξη της μεθόδου αποθήκευσης, συμφέρει να αποθηκεύουμε τις αλλαγές όταν αυτές είναι μέχρι και τρεις.

Η τεχνική που χρησιμοποιούμε είναι η εξής: Έστω ότι μεταξύ του προηγούμενου και του υπό εξέταση διανύσματος έχει συμβεί μια αλλαγή. Αυτό σημαίνει ότι στο υπό εξέταση διάνυσμα αντιπροσώπων, θα εμφανίζεται ένας δείκτης συχνοτικής περιοχής που δεν υπήρχε στο προηγούμενο διάνυσμα, ενώ θα λείπει ένας δείκτης, ο οποίος υπήρχε στο προηγούμενο και όχι στο υπό εξέταση διάνυσμα. Στην αποθήκευση θα χρησιμοποιήσουμε δύο bytes. Στο πρώτο αποθηκεύεται ο δείκτης της συχνοτικής περιοχής που «εισέρχεται», δηλαδή δεν υπήρχε στο προηγούμενο αλλά υπάρχει στο υπό εξέταση διάνυσμα. Στο δεύτερο αποθηκεύεται η θέση στο προηγούμενο διάνυσμα αντιπρόσωπο, όπου βρισκόταν ο δείκτης συχνοτικής περιοχής που δεν υπάρχει στο υπό εξέταση διάνυσμα. Με τον τρόπο αυτό, όπως θα φανεί αργότερα, ελαττώνουμε τον αριθμό των πράξεων που απαιτούνται για να βρεθεί η θέση στην οποία βρίσκεται ο δείκτης περιοχής που «εξέρχεται», αργότερα, στη διαδικασία αναζήτησης. Όταν οι αλλαγές είναι παραπάνω από μία, αποθηκεύονται ζευγάρια αλλαγών, όπως αυτό που περιγράφηκε, με πάντα πρώτο ζευγάρι αυτό με τις συχνοτικές περιοχές με τους μεγαλύτερους δείκτες.

Στο ακόλουθο παράδειγμα φαίνεται πως αποθηκεύεται μια σειρά από τρία διανύσματα για $L = 20$, χωρίς να λάβουμε υπόψη αρχικά τον αριθμό των επαναλήψεων του κάθε διανύσματος.

Η αρχική τριάδα διανυσμάτων προς αποθήκευση είναι:

$$A=[53\ 52\ 51\ 50\ 48\ 47\ 46\ 42\ 40\ 33\ 30\ 28\ 22\ 16\ 11\ 9\ 8\ 4\ 1\ 0]$$

$$B=[53\ 52\ 51\ 50\ 48\ 47\ 46\ 42\ 40\ 33\ 30\ 28\ 22\ 16\ 11\ 9\ 8\ 3\ 1\ 0]$$

$$\Gamma=[53\ 52\ 51\ 50\ 48\ 47\ 46\ 42\ 40\ 33\ 30\ 28\ 22\ 16\ 11\ 8\ 7\ 3\ 2\ 0]$$

Το πρώτο διάνυσμα αποθηκεύεται με τον απλό δυαδικό τρόπο, όπως εξηγήθηκε ωρύτερα, θέτοντας '1' σε κάθε θέση που αντιστοιχεί σε συχνοτική περιοχή η οποία εμφανίζεται στο διάνυσμα αντιπρόσωπο. Επομένως θα έχουμε:

$$A=[110010001101000010000010000010100100000010100011101111]$$

όπου κάθε '1' ή '0' αντιστοιχεί σε ένα απλό δυαδικό ψηφίο.

Το δεύτερο διάνυσμα προς αποθήκευση παρατηρούμε ότι έχει μια αλλαγή με την έννοια ότι στο διάνυσμα A εμφανίζεται ο δείκτης '4', ενώ στο διάνυσμα B, ο δείκτης αυτός δεν υπάρχει και αντ' αυτού υπάρχει ο '3'. Κατά συνέπεια, αντί για το διάνυσμα B θα αποθηκευθεί πρώτα ο δείκτης συχνοτικής περιοχής που «εισέρχεται» σ' αυτό και κατόπιν η θέση στο διάνυσμα A που βρισκόταν ο δείκτης που δεν εμφανίζεται στο B, δηλαδή η θέση που βρισκόταν ο '4' άρα το '18'. Άρα, το διάνυσμα B αποθηκεύεται ως εξής:

$$B=[3,18]$$

Το τρίτο διάνυσμα (το Γ) παρατηρούμε ότι έχει δύο αλλαγές σε σχέση με το B. Δηλαδή, εμφανίζονται οι δείκτες '7' και '2' που δεν βρίσκονται στο B, ενώ λείπουν οι δείκτες '9' και '1' οι οποίοι βρίσκονται στις θέσεις '16' και '19'. Άρα, αντί για το διάνυσμα Γ αποθηκεύουμε το:

$$\Gamma=[7,16,2,19]$$

Σύμφωνα με τα παραπάνω, έχουν προκύψει ως τώρα τρεις διαφορετικοί τρόποι αποθήκευσης του αριθμού των επαναλήψεων, με ένα, δύο ή τρία bytes και τέσσερις διαφορετικοί τρόποι αποθήκευσης του διανύσματος αντιπροσώπων, ανάλογα με τον αριθμό των «αλλαγών» που έχουν συμβεί από το ένα προς αποθήκευση διάνυσμα στο αμέσως επόμενο του.

Το τελευταίο βήμα είναι να καταγράψουμε τον τρόπο κωδικοποίησης που έχει επιλεγεί για κάθε διάνυσμα προς αποθήκευση. Για αυτό το λόγο χρησιμοποιούμε έναν επιπλέον χαρακτήρα, τον οποίο ονομάζουμε *χαρακτήρα κωδικοποίησης*. Ο χαρακτήρας αυτός δηλώνει τη μέθοδο που έχει χρησιμοποιηθεί για την κωδικοποίηση του διανύσματος αντιπροσώπων και του μεγέθους του αριθμού των επαναλήψεων. Οι τιμές που λαμβάνει ο χαρακτήρας αυτός φαίνονται στον ακόλουθο πίνακα 3:

Αριθμός αλλαγών\μέγεθος αριθμού επαναλήψεων	επαναλ.<256	255<επαναλήψεις <64536	Επαναλ.>65535
Μία Αλλαγή	0	3	6
Δύο αλλαγές	1	4	7
Τρεις αλλαγές	2	5	8
Πάνω από τρεις αλλαγές	9	10	11

Πίνακας 3

Τιμή του χαρακτήρα κωδικοποίησης ανάλογα με το μέγεθος του αριθμού των επαναλήψεων και του αριθμού των αλλαγών.

Ο αριθμός των bytes που χρησιμοποιούνται για κάθε περίπτωση είναι φυσικά διαφορετικός και αυτό δηλώνεται στον χαρακτήρα κωδικοποίησης. Άρα, τελικά, η μορφή ενός διανύσματος αντιπροσώπων θα είναι όπως στον ακόλουθο πίνακα 4:

Χαρακτήρας Κωδικοποίησης	Πληροφορία διανύσματος αντιπροσώπων	Αριθμός επαναλήψεων
1 byte	2-7 bytes	1-3 bytes

Πίνακας 4

Στον πίνακα 5 φαίνεται πόσα bytes χρησιμοποιούνται για την αποθήκευση ανάλογα με την τιμή του κωδικού, και πόσα αφιερώνονται στην αποθήκευση κάθε τμήματος του διανύσματος αντιπροσώπων.

Χαρακτήρας Κωδικοποίησης	0	1	2	3	4	5	6	7	8	9	10	11
Χρησιμοποιούμενα bytes διανύσματος	2	4	6	2	4	6	2	4	6	7	7	7
Χρησιμοποιούμενα bytes επαναλήψεων	1	1	1	2	2	2	3	3	3	1	2	3
Σύνολο Χρησιμοποιούμενων bytes	4	6	8	5	7	9	6	8	10	9	10	11

Πίνακας 5

Πειραματικά διαπιστώθηκε ότι με την ανωτέρω διαδικασία, κατά την αποθήκευση όλων των διανυσμάτων αντιπροσώπων ενός πρότυπου ηχητικού σήματος, το μέγεθος του αρχείου που καταλαμβάνουν είναι το 1/5 περίπου από το αντίστοιχο μέγεθος του αρχείου κατά την απλή αποθήκευση των διανυσμάτων. Ο χρόνος ανάγνωσης από τον σκληρό δίσκο των αρχείων με τα διανύσματα αντιπροσώπων κάθε πρότυπου ηχητικού σήματος, είναι ένας από τους κύριους παράγοντες καθυστέρησης της διαδικασίας αναγνώρισης ηχητικών σημάτων. Επομένως, μειώνοντας το μέγεθος των αρχείων, επιταχύνεται σημαντικά η διαδικασία αυτή.

Η διαδικασία αποκωδικοποίησης των διανυσμάτων αντιπροσώπων είναι σχετικά ευθύγραμμη. Διαβάζεται πρώτα ο χαρακτήρας κωδικοποίησης και ανάλογα με την τιμή του, εφαρμόζεται μια από τις τεχνικές που έχουν περιγραφεί νωρίτερα για την εξαγωγή του διανύσματος αντιπροσώπου και του αριθμού των επαναλήψεων.

4.4 Δημιουργία συνόλου διανυσμάτων αντιπροσώπων συχνοτικών περιοχών σε έναν καταγεγραμμένο ηχητικό σήμα

Ας υποθεθεί ότι έχουμε ηχογραφήσει για διάρκεια T δευτερολέπτων, με $T \geq 4$, τον ήχο από έναν ιδίου τύπου εγκατεστημένο κινητήρα κατά τη λειτουργία του. Επειδή η ηχογράφηση του ήχου γίνεται κατά τον ίδιο τρόπο με αυτή των

προτύπων ηχητικών σημάτων το καταγεγραμμένο τμήμα ήχου μετατρέπεται με κάποιο από τα γνωστά εργαλεία σε μορφή "wav", έτσι ώστε να μπορεί να επεξεργαστεί κατάλληλα από τον υπολογιστή, με συχνότητα δειγματοληψίας $F_s = 22050$. Στο εξής θα αναφερόμαστε σε αυτό με το όνομα «καταγεγραμμένο ηχητικό σήμα».

Για να προχωρήσουμε στη διαδικασία ταυτοποίησης με τον αντίστοιχο πρότυπο ηχητικό σήμα, είναι απαραίτητο στο συγκεκριμένο ηχητικό σήμα, να προηγηθεί η εξαγωγή των διανυσμάτων αντιπροσώπων συχνοτικών περιοχών. Εφαρμόζεται λοιπόν και σε αυτό, η διαδικασία που εφαρμόστηκε στο πρότυπο ηχητικό σήμα. (βλέπε Ενότητα 4.1). Στην αρχή του σήματος αυτού (στο υλοποιημένο σύστημα επιλέχθηκε χρονικό διάστημα T για το καταγεγραμμένο ηχητικό σήμα που αντιστοιχεί σε $BL = 2^{18}$ δείγματα) επιλέγεται ένα παράθυρο από N ($N = 8192$) δείγματα και εφαρμόζεται ο Διακριτός Μετασχηματισμός Fourier στο παράθυρο αυτό. Ακολούθως, επαναλαμβάνονται τα βήματα **A3** και **A4** και με αυτό τον τρόπο δημιουργείται το πρώτο διάνυσμα που αποτελείται από L στοιχεία.

Στη συνέχεια επιλέγεται ένα δεύτερο παράθυρο από N δείγματα (το N ίδιο ακριβώς όπως στο βήμα **A2**) το οποίο αρχίζει σε σταθερή απόσταση ℓ (στο υλοποιημένο σύστημα επιλέξαμε $\ell = 22050$) δειγμάτων από το πρώτο δείγμα του τμήματος του καταγεγραμμένου ηχητικού σήματος και εφαρμόζεται ακριβώς η ίδια διαδικασία όπως στο πρώτο παράθυρο κατασκευάζοντας με αυτό τον τρόπο ένα δεύτερο διάνυσμα από L στοιχεία.

Η παραπάνω διαδικασία επαναλαμβάνεται για M (στο υλοποιημένο σύστημα επιλέξαμε $M = 11$) παράθυρα, με κάθε παράθυρο να λαμβάνεται σε απόσταση ℓ δειγμάτων στο πεδίο του χρόνου από το προηγούμενο, με αποτέλεσμα τη δημιουργία ενός συνόλου M διανυσμάτων που το καθένα αποτελείται από L στοιχεία.

4.5 Αλγόριθμος σύγκρισης διανυσμάτων αντιπροσώπων

Ας υποθεθεί ότι έχει ολοκληρωθεί η διαδικασία που περιγράφηκε στην προηγούμενη Ενότητα 4.4, οπότε και έχει υπολογισθεί ένα σύνολο από διανύσματα αντιπροσώπων που αντιστοιχούν στο τμήμα του καταγεγραμμένου ηχητικού σήματος. Την ίδια στιγμή υπάρχουν αποθηκευμένα όλα τα σύνολα διανυσμάτων που αντιστοιχούν στα πρότυπα ηχητικά σήματα της βάσης δεδομένων. Για τη διαδικασία ταυτοποίησης χρησιμοποιείται ένας αλγόριθμος σύγκρισης μεταξύ του συνόλου

διανυσμάτων αντιπροσώπων του καταγεγραμμένου ηχητικού σήματος αφ' ενός, και του συνόλου διανυσμάτων αντιπροσώπων κάθε πρότυπου ηχητικού σήματος αφ' εταίρου, αλγόριθμος, ο οποίος υλοποιεί τα κριτήρια ταιριάσματος που παρουσιάζονται στην Ενότητα 3.1. Πιο συγκεκριμένα:

Συγκρίνουμε το πρώτο διάνυσμα αντιπρόσωπο, έστω V_1 , του καταγεγραμμένου ηχητικού σήματος με το πρώτο διάνυσμα αντιπρόσωπο, έστω $U_{1,n}$, ενός πρότυπου ηχητικού σήματος (ο δείκτης n δείχνει το δείγμα στο πεδίο του χρόνου που αρχίζει το παράθυρο στο οποίο αναφέρεται το τρέχων διάνυσμα αντιπροσώπων), και εν συνεχεία:

A₁) Εάν ο αριθμός των κοινών στοιχείων είναι μικρότερος από $0.55 * L$ θεωρούμε ότι η σύγκριση είναι ανεπιτυχής, με την έννοια ότι δεν επετεύχθη ταυτοποίηση, και ξαναρχίζει η διαδικασία με τη σύγκριση του διανύσματος V_1 με το επόμενο διάνυσμα αντιπροσώπων συχνοτικών περιοχών του πρότυπου ηχητικού σήματος, έστω $U_{1,(n+1)}$.

B₁) Εάν ο αριθμός των κοινών στοιχείων είναι μεγαλύτερος ή ίσος με $0.55 * L$, τότε και μόνο τότε προχωρούμε στη σύγκριση μεταξύ του δεύτερου διανύσματος αντιπροσώπων, έστω V_2 , του καταγεγραμμένου ηχητικού σήματος και του διανύσματος $U_{2,(n+l)}$ του πρότυπου ηχητικού σήματος που αντιστοιχεί στο δείγμα $n + l$. Στη συνέχεια,

A₂) Εάν ο αριθμός των κοινών στοιχείων μεταξύ του V_2 και του $U_{2,(n+l)}$ είναι μικρότερος από $0.55 * L$ η σύγκριση δεν συνεχίζεται περαιτέρω. Θεωρούμε ότι η ως τώρα σύγκριση ήταν ανεπιτυχής και η διαδικασία ξαναρχίζει με την σύγκριση του διανύσματος V_1 , με το επόμενο στο $U_{1,n}$ διάνυσμα αντιπροσώπων περιοχών του πρότυπου ηχητικού σήματος $U_{1,(n+1)}$, όπως στην **A₁**.

B₂) Εάν ο αριθμός των κοινών στοιχείων των V_2 και $U_{2,(n+l)}$ είναι μεγαλύτερος ή ίσος με $0.55 * L$, τότε και μόνο τότε προχωρούμε στη σύγκριση μεταξύ του τρίτου διανύσματος αντιπροσώπου V_3 του καταγεγραμμένου ηχητικού σήματος και του διανύσματος $U_{3,(n+2*l)}$ του συνόλου διανυσμάτων αντιπροσώπων του πρότυπου ηχητικού σήματος που αντιστοιχεί στο δείγμα $n + 2 * l$.

•
•
•

A_M) Εάν ο αριθμός των κοινών στοιχείων μεταξύ του V_M και του $U_{M,(n+(M-1)*\ell)}$ είναι μικρότερος από $0.55 * L$ η σύγκριση δε συνεχίζεται περαιτέρω. Θεωρούμε ότι η ως τώρα σύγκριση ήταν ανεπιτυχής και η διαδικασία ξαναρχίζει με τη σύγκριση του διανύσματος V_1 , με το επόμενο διάνυσμα αντιπρόσωπο περιοχής του πρότυπου ηχητικού σήματος $U_{1,(n+1)}$, όπως στην **A₁**.

B_M) Εάν ο αριθμός των κοινών στοιχείων των V_M και $U_{M,(n+(M-1)*\ell)}$ είναι μεγαλύτερος ή ίσος με $0.55 * L$ τότε και μόνο τότε προχωρούμε στη εύρεση της μέσης τιμής των κοινών στοιχείων μεταξύ όλων των προηγούμενων ζευγών διανυσμάτων αντιπρόσωπων συχνοτικών περιοχών:

$$(V_1, U_{1,n}), (V_2, U_{2,(n+\ell)}), \dots, (V_M, U_{M,n+(M-1)*\ell}).$$

C) Εάν αυτή η μέση τιμή είναι μεγαλύτερη ή ίση από $0.71 * L$, προχωρούμε στο βήμα **D**. Στην άλλη περίπτωση, εάν η μέση τιμή είναι μικρότερη από $0.71 * L$ τότε το κριτήριο ομοιότητας δεν ικανοποιείται. Θεωρούμε την ως τώρα διαδικασία ανεπιτυχή, οπότε και επαναρχίζει η σύγκριση του διανύσματος V_1 , με το επόμενο διάνυσμα αντιπρόσωπο του πρότυπου ηχητικού σήματος $U_{1,(n+1)}$, όπως ακριβώς στο **A₁**.

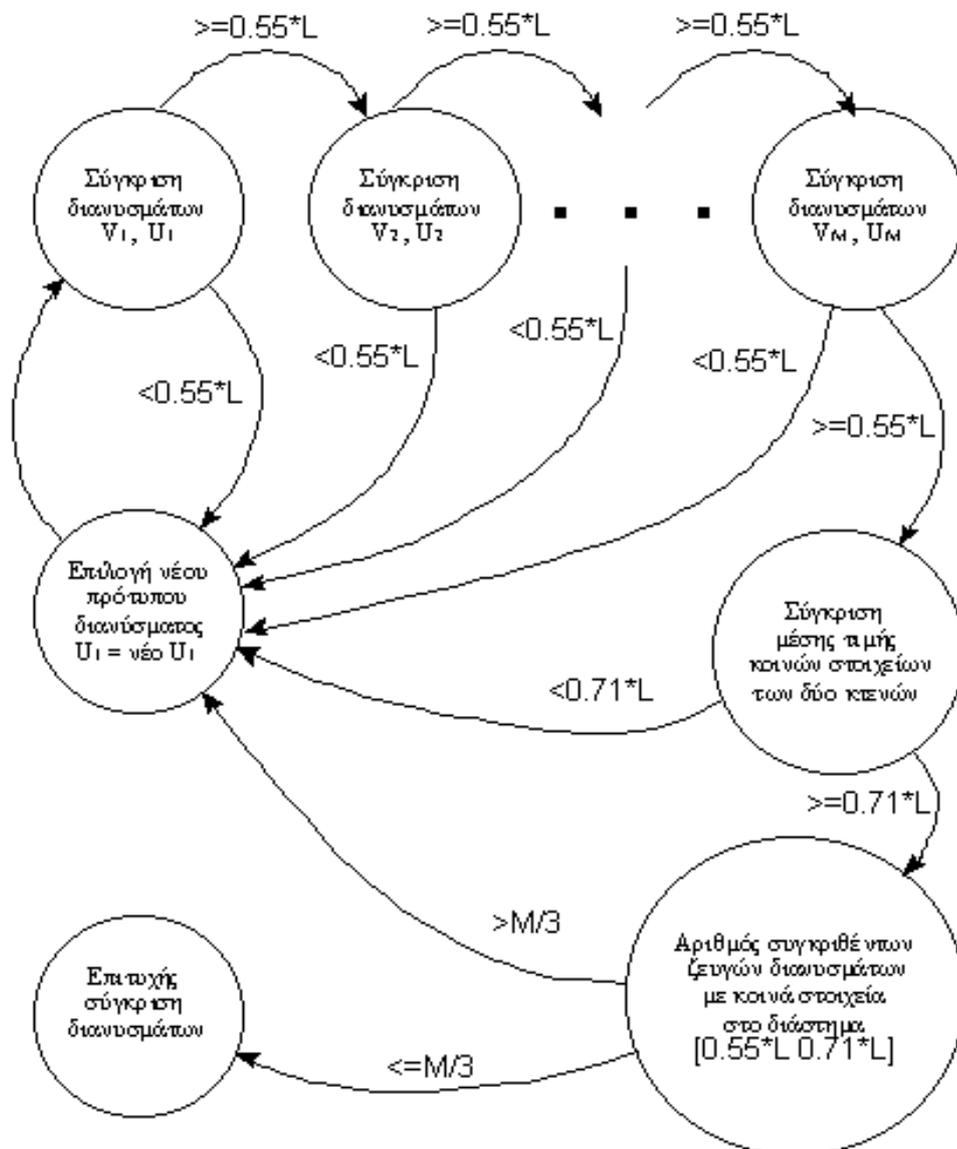
D) Εάν περισσότερα από $\left\lfloor \frac{M}{3} \right\rfloor$ ζεύγη διανυσμάτων των παραθύρων που συγκρίνονται έχουν κοινούς αντιπροσώπους μεταξύ $0.55 * L$ και $0.71 * L$, τότε το κριτήριο ομοιότητας ικανοποιείται. Αλλιώς, όπως και προηγουμένως, ξαναρχίζει η σύγκριση του διανύσματος V_1 , με το επόμενο διάνυσμα αντιπρόσωπο του πρότυπου ηχητικού σήματος $U_{1,(n+1)}$, όπως ακριβώς στο **A₁**.

Πρέπει να σημειωθεί ότι το διάνυσμα $U_{i,n}$ έχει αυξημένη πιθανότητα να είναι ακριβώς ίδιο με το $U_{i,(n+1)}$, $i=1,2,\dots,M$, εφ' όσον τα δύο αυτά διανύσματα αντιστοιχούν σε γειτονικά δείγματα στο πεδίο του χρόνου. Σε αυτή την περίπτωση δεν πραγματοποιείται σύγκριση στο συγκεκριμένο στάδιο, αλλά χρησιμοποιείται το

αποτέλεσμα της προηγούμενης σύγκρισης προκειμένου να γίνει ταχύτερη η διαδικασία.

Μια αποδοτική υλοποίηση του παραπάνω αλγόριθμου σύγκρισης παρουσιάζεται στο Παράρτημα Α.

Στις περιπτώσεις που ολοκληρώνεται επιτυχώς η προηγούμενη διαδικασία προχωρούμε στο τελεσίδικο κριτήριο. Αν όλα τα σύνολα διανυσμάτων αντιπροσώπων των πρότυπων ηχητικών σημάτων εξεταστούν χωρίς επιτυχία, το σύστημα αποφασίζει ότι το υπό καταγεγραμμένο ηχητικό σήμα δεν αντιστοιχεί σε κανένα από τα πρότυπα ηχητικά σήματα που υπάρχουν στη βάση δεδομένων. Η προηγούμενη διαδικασία φαίνεται σχηματικά στο ακόλουθο διάγραμμα.



4.6 Τελεσίδικο κριτήριο ομοιότητας

Ας υποθεθεί ότι με βάση τα προαναφερθέντα κριτήρια και τη διαδικασία σύγκρισης που περιγράφηκε στην προηγούμενη ενότητα, προτείνεται από το σύστημα ένα σύνολο από πρότυπα ηχητικά σήματα, τα οποία έχουν πιθανότητα ταιριάσματος με το καταγεγραμμένο ηχητικό σήμα, σε συγκεκριμένα δείγματα για κάθε πρότυπο ηχητικό σήμα. Για να αποφασίσουμε αν όντως κάποιο από τα πρότυπα ηχητικά σήματα αντιστοιχεί στο καταγεγραμμένο ηχητικό σήμα, το σύστημα συνεχίζει όπως περιγράφεται παρακάτω, αξιοποιώντας την ανάλυση της Ενότητας 3.2:

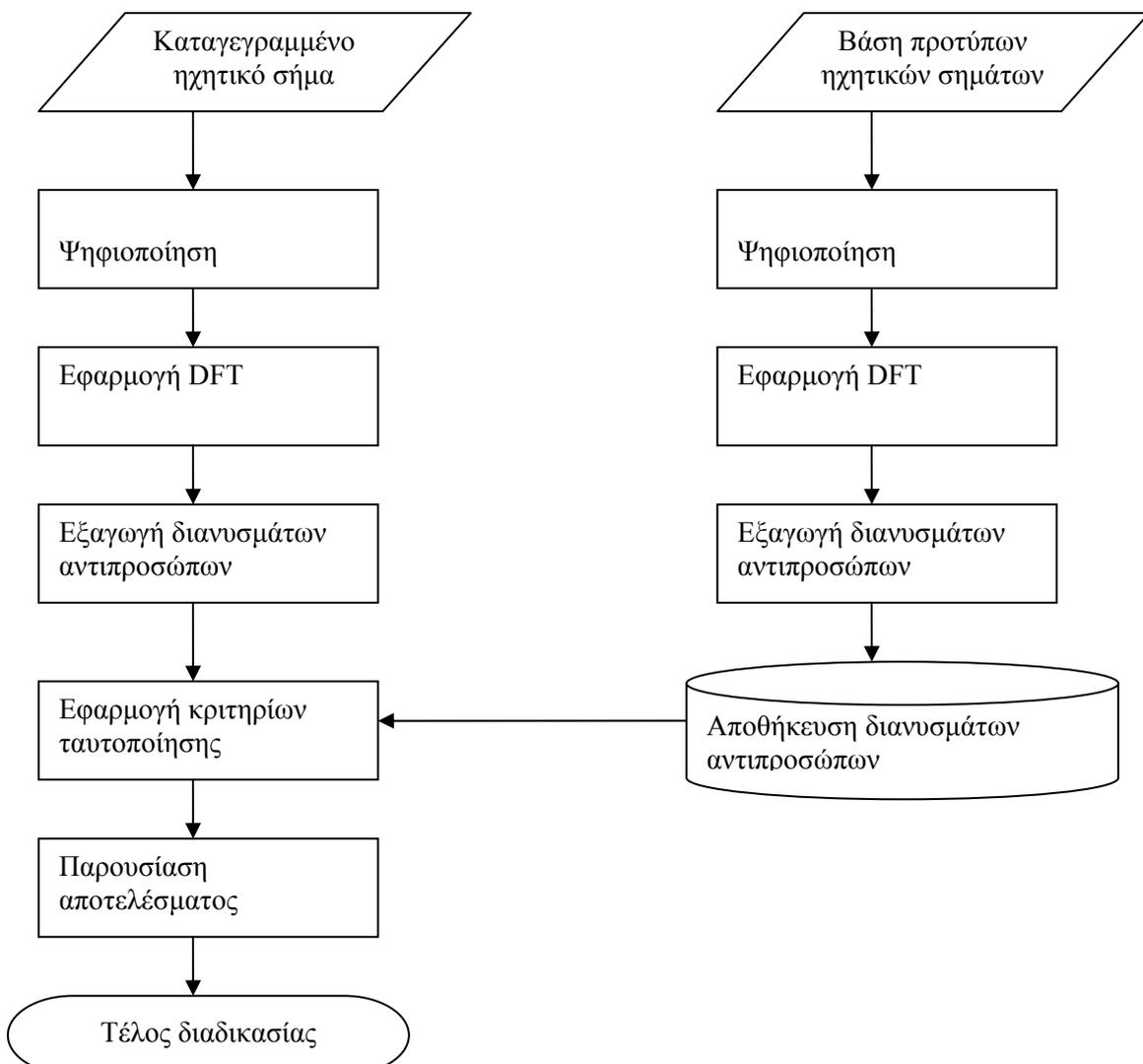
Αρχικά, υπολογίζεται ο DFT σε 2^{17} δείγματα του τμήματος του καταγεγραμμένου ηχητικού σήματος, αρχίζοντας από το πρώτο της δείγμα στο πεδίο του χρόνου και υπολογίζεται η κανονικοποιημένη περιβάλλουσα του προκύπτοντος αποτελέσματος με τον τρόπο που περιγράφεται στην Ενότητα 3.2.

Στη συνέχεια, σε κάθε πρότυπο ηχητικό σήμα που ικανοποιεί τα κριτήρια της προηγούμενης παραγράφου, το σύστημα επιλέγει ένα παράθυρο μήκους 2^{17} δειγμάτων που αρχίζει στο δείγμα όπου αναφέρεται το πιθανό ταίριασμα. Ο DFT 2^{17} δειγμάτων εκτελείται σε κάθε τέτοιο παράθυρο και υπολογίζονται οι κανονικοποιημένες περιβάλλουσες των παραπάνω DFT.

Τελικά, υπολογίζεται το άθροισμα των απόλυτων τιμών των διαφορών των τιμών δύο κανονικοποιημένων περιβαλλουσών, για κάθε πιθανό πρότυπο ηχητικό σήμα ταιριάσματος αφ' ενός, και για το καταγεγραμμένο ηχητικό σήμα αφ' εταίρου. Αν αυτό το άθροισμα είναι μικρότερο από 0,85, αναφέρεται τελεσίδικο ταίριασμα. Διαφορετικά, το σύστημα αποφασίζει ότι το καταγεγραμμένο ηχητικό σήμα δεν αντιστοιχεί στο συγκεκριμένο πρότυπο ηχητικό σήμα. Πρέπει να τονίσουμε ότι, συνήθως, δύο διαφορετικά πρότυπα ηχητικά σήματα έχουν μέσο άθροισμα απόλυτων διαφορών μεγαλύτερο από 1,45. Το σύστημα εξετάζει με αυτό τον τρόπο όλα τα πρότυπα ηχητικά σήματα που ικανοποίησαν τα αρχικά κριτήρια. Αν όλα τα πρότυπα ηχητικά σήματα αποτύχουν στο τελεσίδικο κριτήριο το σύστημα αναφέρει πως δεν υπάρχει κανένα ταίριασμα με το προς αναγνώριση καταγεγραμμένο ηχητικό σήμα.

5. ΣΥΣΤΗΜΑ ΑΥΤΟΜΑΤΗΣ ΑΝΑΓΝΩΡΙΣΗΣ ΛΕΙΤΟΥΡΓΙΚΗΣ ΚΑΤΑΣΤΑΣΗΣ / ΒΛΑΒΗΣ ΚΙΝΗΤΗΡΑ DIESEL ΚΑΙ ΕΠΙΔΟΣΕΙΣ

Με βάση τους αλγορίθμους που περιγράφηκαν στις προηγούμενες ενότητες υλοποιήθηκε ένα ολοκληρωμένο σύστημα λογισμικού αυτόματης αναγνώρισης λειτουργικής κατάστασης / βλάβης σε έναν κινητήρα Diesel με αξιοποίηση τον ήχου αυτού. Το σύστημα αποτελείται από 4 ανεξάρτητα τμήματα που αποσκοπούν: α) στη δημιουργία μιας βάσης προτύπων ηχητικών σημάτων για ένα συγκεκριμένο τύπο ναυτικό κινητήρα Diesel στην οποία είναι καταγεγραμμένος ο ήχος που αντιστοιχεί σε διαφορετικές λειτουργικές καταστάσεις αυτού συμπεριλαμβανομένων των βλαβών β) στην εξαγωγή των διανυσμάτων αντιπροσώπων από την εν λόγω βάση προτύπων ηχητικών σημάτων, γ) στη καταγραφή ήχου από έναν ίδιου τύπου κινητήρα για τον οποίο δεν γνωρίζουμε την λειτουργική του κατάσταση και δ) στην υλοποίηση αλγορίθμων ελέγχου των κριτηρίων ταύτισης ενός πρότυπου ηχητικού σήματος με το καταγεγραμμένο.



5.1 Διαδικασία δημιουργίας βάσης προτύπων ηχητικών σημάτων για ένα συγκεκριμένο τύπο ναυτικού κινητήρα Diesel.

Η συγκεκριμένη διαδικασία έχει σαν σκοπό τη δημιουργία μιας βάσης η οποία θα περιλαμβάνει ηχογραφήσεις ενός συγκεκριμένου τύπου κινητήρα για τον οποίον η λειτουργική του κατάσταση κατά την διάρκεια της ηχογράφησης θα θεωρείται ως πρότυπο. Για το σύστημα μας, με τον όρο ‘πρότυπη λειτουργική κατάσταση’ ενός συγκεκριμένου τύπου κινητήρα, θεωρούμε την πιστοποιημένη από τους επισκευαστικούς φορείς κατάσταση που κινητήρα η οποία μπορεί να αναφέρεται σε έναν κινητήρα που δεν παρουσιάζει καμία απολύτως βλάβη, αλλά και σε έναν κινητήρα ο οποίος παρουσιάζει συγκεκριμένη βλάβη η οποία επηρεάζει τον ήχο του κατά τη λειτουργία. Στο υλοποιημένο σύστημα επιλέξαμε έναν 20κύλινδρο σε διάταξη V υπερτροφοδοτούμενο με δύο Turbo ναυτικό κινητήρα Diesel, ο οποίος χρησιμοποιείται για πρόωση. Η διάταξη που χρησιμοποιήθηκε για την καταγραφή του ήχου του κινητήρα περιλαμβάνει τα εξής στοιχεία:

- Έναν φορητό υπολογιστή.
- Μία εξωτερική κάρτα ήχου με δυνατότητα αναλογικής εισόδου ήχου.
- Ένα ‘δυναμικό’ στερεοφωνικό μικρόφωνο.

Η διαδικασία καταγραφής των πρότυπων ηχητικών ήχων που αποτέλεσαν και την βάση του συστήματος μας έχει ως εξής: Κάθε συγκεκριμένη λειτουργική κατάσταση του κινητήρα ηχογραφείται για διάρκεια 5 λεπτών περίπου με το στερεοφωνικό μικρόφωνο. Ταυτόχρονα η ηχογράφηση καταγράφεται ψηφιακά και αποθηκεύεται σε ξεχωριστό αρχείο με τη μορφή PCM 16-byte, 2 κανάλια, με συχνότητα δειγματοληψίας 44100Hz στον φορητό υπολογιστή. Αυτό γίνεται με κατάλληλο πρόγραμμα ανάγνωσης και αποθήκευσης σε μορφή αρχείου «.wav».

Η διαδικασία αυτή προϋποθέτει τον ανθρώπινο παράγοντα καθ’όλη τη διάρκεια της ηχογράφησης / αποθήκευσης. Προκειμένου να επιτευχθεί το καλύτερο αποτέλεσμα, η ηχογράφηση διενεργήθηκε περιφέροντας το μικρόφωνο περιμετρικά του κινητήρα.

Με αυτόν τον τρόπο δημιουργήθηκε η βάση προτύπων ηχητικών σημάτων του συστήματός μας καταγράφοντας από ίδιου τύπου κινητήρες διαφορετικές 'πρότυπες λειτουργικές καταστάσεις' αυτών, η οποία παρουσιάζεται στον ακόλουθο πίνακα 6:

ΠΡΟΤΥΠΗ ΛΕΙΤΟΥΡΓΙΚΗ ΚΑΤΑΣΤΑΣΗ ΚΙΝΗΤΗΡΑ	ΠΛΗΘΟΣ ΠΡΟΤΥΠΩΝ ΗΧΗΤΙΚΩΝ ΣΗΜΑΤΩΝ
Ηχος νεοτοποθετηθέντος καινούργιου κινητήρα σε όλο το φάσμα στροφών (RPM) λειτουργίας του: $H_0^{RPM(i)}$ RPM(i): 800,900,1000,1100,1200,1300,1400,1500	8
Ηχος κινητήρα που δεν παρουσιάζει βλάβη, μετά από 1500 ώρες λειτουργίας από την τοποθέτησή του στο πλοίο, σε όλο το φάσμα στροφών λειτουργίας του: $H_1^{RPM(i)}$	8
Ηχος κινητήρα που δεν παρουσιάζει βλάβη, μετά από 3000 ώρες λειτουργίας από την τοποθέτησή του στο πλοίο, σε όλο το φάσμα στροφών λειτουργίας του: $H_2^{RPM(i)}$	8
Ηχος κινητήρα που δεν παρουσιάζει βλάβη, μετά από 4500 ώρες λειτουργίας από την τοποθέτησή του στο πλοίο, σε όλο το φάσμα στροφών λειτουργίας του: $H_3^{RPM(i)}$	8
Ηχος κινητήρα που δεν παρουσιάζει βλάβη, μετά από 6000 ώρες λειτουργίας από την τοποθέτησή του στο πλοίο, σε όλο το φάσμα στροφών λειτουργίας του: $H_3^{RPM(i)}$	8
Ηχος κινητήρα που παρουσιάζει πλημμελή λειτουργία βαλβίδος εξαγωγής καυσαερίων στον Νο1 κύλινδρο μετά από 2500 ώρες λειτουργίας από την τοποθέτησή του στο πλοίο, σε όλο το φάσμα στροφών λειτουργίας του: $H_{B1}^{RPM(i)}$	8
Ηχος κινητήρα που παρουσιάζει πλημμελή λειτουργία βαλβίδος εξαγωγής καυσαερίων στον Νο5 κύλινδρο μετά από 5000 ώρες λειτουργίας από την τοποθέτησή του στο πλοίο, σε όλο το φάσμα στροφών λειτουργίας του: $H_{B2}^{RPM(i)}$	8
Ηχος κινητήρα που παρουσιάζει πλημμελή λειτουργία καυστήρα στον Νο8 κύλινδρο μετά από 1300 ώρες λειτουργίας από την τοποθέτησή του στο πλοίο, σε όλο το φάσμα στροφών λειτουργίας του: $H_{B3}^{RPM(i)}$	8
Ηχος κινητήρα που παρουσιάζει πλημμελή λειτουργία Πρωραίου Turbo μετά από 4000 ώρες λειτουργίας από την τοποθέτησή του στο πλοίο, σε όλο το φάσμα στροφών λειτουργίας του: $H_{B4}^{RPM(i)}$	8

Πίνακας 6

5.2 Διαδικασία εξαγωγής των διανυσμάτων αντιπροσώπων από τα πρότυπα ηχητικά σήματα

Η συγκεκριμένη διαδικασία έχει σαν σκοπό να εξάγει και να αποθηκεύσει όλα τα διανύσματα αντιπροσώπων για κάθε πρότυπο ηχητικό σήμα, από την υφιστάμενη βάση προτύπων ηχητικών σημάτων ενός συγκεκριμένου τύπου κινητήρα Diesel.

Τα αποθηκευμένα, με την διαδικασία που παρουσιάζεται στην προηγούμενη Ενότητα, στον φορητό υπολογιστή αρχεία δίνονται ως είσοδος στο λογισμικό εξαγωγής των διανυσμάτων αντιπροσώπων. Το λογισμικό είναι υπεύθυνο για την παραγωγή ενός αρχείου για κάθε πρότυπο ηχητικό σήμα, το οποίο περιέχει τα συμπίεσμένα, με την τεχνική που παρουσιάζεται στην Ενότητα 4.3, διανύσματα αντιπροσώπων αυτού. Τα αρχεία αυτά των διανυσμάτων αντιπροσώπων των προτύπων ηχητικών σημάτων, αποτελούν τα δεδομένα που χρησιμοποιεί το υποσύστημα αναγνώρισης λειτουργικής κατάστασης κινητήρα.

5.3 Διαδικασία καταγραφής ηχητικών σημάτων από ιδίου τύπου κινητήρες με άγνωστη λειτουργική κατάσταση

Η συγκεκριμένη διαδικασία έχει σαν σκοπό την καταγραφή τμημάτων ήχου κατά την λειτουργία ιδίου τύπου κινητήρων, των οποίων η λειτουργική κατάσταση θεωρείται άγνωστη. Τα εν λόγω ηχητικά τμήματα αποτελούν το πεδίο δοκιμής του συστήματος. Η ηχογράφηση, ψηφιακή καταγραφή και αποθήκευση έγινε με τον τρόπο που παρουσιάζεται στην Ενότητα 5.1.

5.4 Διαδικασία υλοποίησης αλγορίθμων ελέγχου των κριτηρίων ταυτοποίησης των ηχητικών σημάτων

Στο υλοποιημένο σύστημα συμπεριλαμβάνονται εξήντα τέσσερα (64) πρότυπα ηχητικά σήματα που αντιστοιχούν σε διαφορετικές λειτουργικές καταστάσεις ενός συγκεκριμένου τύπου κινητήρα Diesel όπως αναφέρονται στο Πίνακα Α.6. Για να ελεγχθεί η απόδοση του συστήματος, χρησιμοποιείται ένα σύνολο οκτώ (8) καταγεγραμμένων ηχητικών τμημάτων από ισάριθμους ιδίου τύπου κινητήρες των οποίων η λειτουργική κατάσταση είναι άγνωστη, και οι ώρες λειτουργίας τους από την τοποθέτησή του στο πλοίο ήταν περίπου ισοκατανεμημένες μεταξύ των 1000 και των 5500 ωρών. Από τα καταγεγραμμένα

ηχητικά τμήματα τα επτά (7) αντιστοιχούν σε πρότυπα ηχητικά σήματα που περιλαμβάνονται στο σύστημα ενώ το ένα (1) αντιστοιχεί σε λειτουργική κατάσταση κινητήρα που δεν περιλαμβάνεται στη βάση δεδομένων του συστήματος.

Και στα 8 καταγεγραμμένα ηχητικά σήματα εκτελέστηκε ένας μεγάλος αριθμός ψευδοδειγματοληψιών, όπου με τον όρο ψευδοδειγματοληψία εννοούμε την διαδικασία σύμφωνα με την οποία, κάθε φορά, επιλέγεται ένα τμήμα $BL = 250000$ δειγμάτων από το συγκεκριμένο σήμα, η αρχή του οποίου διαλέγεται με τη βοήθεια μιας γεννήτριας τυχαίων αριθμών. Σε αυτό το τμήμα των BL δειγμάτων του καταγεγραμμένου ηχητικού σήματος εφαρμόζεται όλη η παρακάτω διαδικασία αναγνώρισης και καταγράφεται το αποτέλεσμα της.

1. Το σύστημα εξάγει, από το κάθε τμήμα των BL δειγμάτων του καταγεγραμμένου ηχητικού σήματος, ένα σύνολο M διανυσμάτων αντιπροσώπων.
2. Το σύνολο των M διανυσμάτων συγκρίνεται με κάθε αποθηκευμένο αρχείο διανυσμάτων αντιπροσώπων των πρότυπων ηχητικών σημάτων.
3. Κάθε φορά που πληρείται το πρώτο κριτήριο, το σύστημα ελέγχει αν ισχύει και το δεύτερο, τελεσίδικο, κριτήριο.
4. Αν ισχύει και το δεύτερο κριτήριο, τότε το σύστημα αποφαινεται ότι το συγκεκριμένο καταγεγραμμένο ηχητικό σήμα ταυτοποιείται με το συγκρινόμενο πρότυπο ηχητικό σήμα. Στην συγκεκριμένη φάση των πειραμάτων, συνεχίζεται ο έλεγχος και με τα υπόλοιπα αποθηκευμένα σύνολα διανυσμάτων, ώστε να διαπιστωθεί αν κάποιο καταγεγραμμένο ηχητικό σήμα ταυτοποιείται λανθασμένα, με περισσότερα από ένα πρότυπα ηχητικά σήματα.

Το τελικό αποτέλεσμα είναι ένα αρχείο αναφοράς, για στατιστική επεξεργασία, με στόχο τη βελτίωση των παραμέτρων του συστήματος και την ανίχνευση κάθε πιθανού προβλήματος.

Στο αρχείο αναφοράς καταγράφονται:

- Κάθε πρότυπο ηχητικό σήμα για το οποίο διαπιστώθηκε εκπλήρωση του πρώτου κριτηρίου.

- Το ακριβές σημείο (δείγμα) του πρότυπου ηχητικού σήματος στο οποίο έχει συμβεί η εκπλήρωση του πρώτου κριτηρίου.
- Ο αριθμός των κοινών στοιχείων των διανυσμάτων αντιπροσώπων μεταξύ του καταγεγραμμένου και του πρότυπου ηχητικού σήματος που οδηγεί στην εκπλήρωση του πρώτου κριτηρίου.
- Το άθροισμα των απόλυτων τιμών των διαφορών των τιμών των δύο κανονικοποιημένων περιβαλλουσών του DFT, του καταγεγραμμένου ηχητικού σήματος αφ' ενός και κάθε πρότυπου ηχητικού σήματος αφ' εταίρου, για τα οποία ικανοποιείται το πρώτο κριτήριο. Καταγράφεται, τέλος, το αν η διαδικασία σύγκρισης με το κριτήριο αυτό οδηγεί σε επιτυχή ταυτοποίηση των ήχων / λειτουργικών καταστάσεων του κινητήρα ή όχι.

Οι ψευδοδειγματοληψίες και η σχετική αυτόματη διαδικασία αναγνώρισης σε κάθε άγνωστο καταγεγραμμένο ηχητικό σήμα επαναλαμβάνονται μέχρι

$$\frac{(\text{Αριθμός δειγμάτων καταγεγραμμένου ήχου})}{(\text{Αριθμός τυχαίων ψευδοδειγματοληψιών})} < 1000.$$

Για παράδειγμα, σε έναν καταγεγραμμένο ηχητικό σήμα διάρκειας τριών λεπτών, η ψευδοδειγματοληψία και η σχετική διαδικασία αυτόματης αναγνώρισης εκτελούνται περισσότερο από τρεις χιλιάδες φορές.

5.5 Επιδόσεις του συστήματος αυτόματης αναγνώρισης λειτουργικής κατάστασης / βλάβης σε έναν κινητήρα Diesel

Το σύστημα αναγνώρισε επιτυχώς επτά (7) άγνωστες καταγεγραμμένες λειτουργικές καταστάσεις του συγκεκριμένου τύπου κινητήρα Diesel με την ταυτοποίηση του εκάστοτε καταγεγραμμένου ηχητικού σήματος με το αντίστοιχο πρότυπο ηχητικό σήμα για όλες τις χιλιάδες ψευδοδειγματοληψίες που εφαρμόστηκαν στο καταγεγραμμένο ηχητικό σήμα. Με άλλα λόγια, το σύστημα είχε εκατό τοις εκατό (100%) ποσοστό επιτυχούς ταυτοποίησης για τις επτά (7) άγνωστες λειτουργικές καταστάσεις, ανεξάρτητα από το που λήφθηκε το τμήμα των *BL* δειγμάτων από το αντίστοιχο καταγεγραμμένο ηχητικό σήμα τους.

Επίσης το σύστημα απέρριψε επιτυχώς ένα (1) καταγεγραμμένο ηχητικό σήμα καθώςον επαληθεύτηκε εκ των υστέρων ότι αντιστοιχούσε σε βλάβη κινητήρα η οποία δεν συμπεριλαμβάνονταν στη βάση του συστήματος.

Βασισμένοι στα προαναφερθέντα αποτελέσματα, μπορεί να διατυπωθεί με ασφάλεια ότι το σύστημα που περιγράφεται στην παρούσα εργασία υπόσχεται πολύ μεγάλο ποσοστό επιτυχούς αναγνώρισης λειτουργικής κατάστασης / βλάβης Ναυτικών κινητήρων Diesel αξιοποιώντας την ηχητική πληροφορία.

Επιπρόσθετα ένα σημαντικό πλεονέκτημα του εν λόγω συστήματος, που φάνηκε από τα αρχικά στάδια υλοποίησης του είναι ότι δεν προϋποθέτει ουδεμία μετατροπή στον κινητήρα, καθώς επίσης και ότι λειτουργεί ανεξάρτητα από κάθε ήδη εγκατεστημένο σύστημα παρακολούθησης και ελέγχου.

5.6 Παρατηρούμενος υπολογιστικός φόρτος της εφαρμογής

Ένα ακόμα σημαντικό τμήμα των επιδόσεων του συστήματος, αφορά την ταχύτητα με την οποία εκτελείται το σύνολο της διαδικασίας αναγνώρισης καταγεγραμμένων ηχητικών σημάτων. Πρέπει να επισημανθεί ότι δεν μας απασχολεί ο χρόνος που απαιτείται για τη διαδικασία εξαγωγής των διανυσμάτων αντιπροσώπων των πρότυπων ηχητικών σημάτων, διότι αυτή γίνεται άπαξ και εντελώς εκτός της κυρίως διαδικασίας αναγνώρισης. Η μελέτη σε αυτή την ενότητα αφορά μόνο το υποσύστημα που εκτελεί την κύρια διαδικασία αναγνώρισης καταγεγραμμένων ηχητικών σημάτων.

Μελετώντας διεξοδικά τη διαδικασία αυτή, γίνεται φανερό ότι το τμήμα που απαιτεί το μεγαλύτερο υπολογιστικό φόρτο είναι αυτό της σύγκρισης των M διανυσμάτων αντιπροσώπων του καταγεγραμμένου ηχητικού σήματος με το σύνολο των διανυσμάτων των πρότυπων ηχητικών σημάτων. Επίσης, δεν είναι ευκαταφρόνητος και ο χρόνος που απαιτείται για την ανάγνωση των κωδικοποιημένων διανυσμάτων από τους μαγνητικούς σκληρούς δίσκους στους οποίους έχουν αποθηκευθεί, ακόμα και αν αυτοί είναι υψηλής ταχύτητας.

Παρατηρήθηκε ότι η χρήση της μη κωδικοποιημένης μορφής των διανυσμάτων, που οδηγεί σε μεγαλύτερα δυαδικά αρχεία, δεν ενδείκνυται, όχι μόνο γιατί το μέγεθος των δεδομένων είναι μεγάλο και άρα θα απαιτείται πολύς αποθηκευτικός χώρος, αλλά και γιατί ο χρόνος ανάγνωσης από τα μαγνητικά μέσα αποθήκευσης είναι πολύ μεγαλύτερος του χρόνου αποκωδικοποίησης και ανάγνωσης των δυαδικών συμπιεσμένων διανυσμάτων.

Επίσης ο χρόνος για την αναγνώριση ενός καταγεγραμμένου ηχητικού σήματος είναι ανάλογος του αριθμού και της διάρκειας των προτύπων ηχητικών

σημάτων. Μεγάλης διάρκειας πρότυπα ηχητικά σήματα, αποτελούνται από πολλά δείγματα, άρα, το προς εξέταση σύνολο διανυσμάτων αντιπροσώπων είναι κατά κανόνα μεγαλύτερο. Ένας επιπλέον, μη προφανής παράγοντας που επηρεάζει την ταχύτητα με την οποία ένα πρότυπο ηχητικό σήμα θα εξετασθεί, είναι ο βαθμός ομοιότητας του πρότυπου ηχητικού σήματος με το υπό εξέταση άγνωστο σύνολο διανυσμάτων. Δηλαδή, όσα περισσότερα από τα M διανύσματα αντιπροσώπων έχουν πάνω από $0.55 * L$ κοινά στοιχεία κατά τη διαδικασία αναζήτησης, τόσο περισσότερο καθυστερεί η εξέταση του συγκεκριμένου πρότυπου ηχητικού σήματος. Αυτό συμβαίνει, γιατί μετά το πρώτο διάνυσμα, ο έλεγχος καθυστερεί αρκετά παραπάνω, όπως προκύπτει και από το Παράρτημα Α.

Κατά συνέπεια, είναι δυνατό, τα διανύσματα αντιπροσώπων ενός καταγεγραμμένου ηχητικού σήματος να συγκριθούν με αυτά του πρότυπου ηχητικού σήματος σε χρόνο που κυμαίνεται από 15 έως 30 εκατοστά του δευτερολέπτου. Για την υπάρχουσα βάση των 64 προτύπων ηχητικών σημάτων ο μέσος χρόνος αναγνώρισης ενός καταγεγραμμένου ηχητικού σήματος είναι περίπου 0.5 λεπτό. Ο χρόνος αυτός μπορεί να δώσει ένα σύστημα που θα λειτουργεί σε πραγματικό χρόνο, ακόμα και για πολύ μεγαλύτερο αριθμό προτύπων ηχητικών σημάτων.

ΠΑΡΑΡΤΗΜΑ Α - Υλοποίηση αποδοτικού αναδρομικού αλγορίθμου σύγκρισης διανυσμάτων αντιπροσώπων

Ο αλγόριθμος σύγκρισης διανυσμάτων αντιπροσώπων που περιγράφηκε στην Ενότητα 4.5 μπορεί να υλοποιηθεί έτσι ώστε να είναι σχετικά γρήγορος και αποδοτικός με χρήση ευφών αλγορίθμων για τον υπολογισμό του αριθμού των κοινών στοιχείων των διανυσμάτων αντιπροσώπων. Προς την κατεύθυνση αυτή βοηθά αρκετά και η μέθοδος κωδικοποίησης που έχουμε χρησιμοποιήσει και έχει αναφερθεί νωρίτερα στην Ενότητα 4.3. Αναλυτικότερα, τα δεδομένα μας σε κάθε σύγκριση είναι τα εξής:

- Το υπό εξέταση σύνολο διανυσμάτων αντιπροσώπων V του καταγεγραμμένου ηχητικού σήματος.
- Το σύνολο των διανυσμάτων αντιπροσώπων U^* του πρότυπου ηχητικού σήματος. Για την αποθήκευσή τους χρησιμοποιείται ένας πίνακας, έστω F_M , μεγέθους $Y \times L$, όπου Y ο αριθμός των διαφορετικών συνεχόμενων διανυσμάτων ενός πρότυπου ηχητικού σήματος. Σε κάθε γραμμή του πίνακα αυτού, αποθηκεύεται κι ένα διαφορετικό διάνυσμα αντιπροσώπων. Υπενθυμίζεται ότι τα διανύσματα αποθηκεύονται με τα στοιχεία τους ταξινομημένα κατά φθίνουσα σειρά.
- Ένας μονοδιάστατος πίνακας, έστω EP , μεγέθους $Y \times I$. Σε κάθε στοιχείο, i , του πίνακα αυτού αποθηκεύεται ο αριθμός του δείγματος του πρότυπου ηχητικού σήματος στο οποίο αντιστοιχεί το διάνυσμα αντιπροσώπων L στοιχείων που βρίσκεται στη γραμμή i , του πίνακα F_M .
- Ένας μονοδιάστατος πίνακας, έστω L_I , μεγέθους τόσο όσο το μήκος των δειγμάτων του πρότυπου ηχητικού σήματος. Σε κάθε στοιχείο j , του πίνακα αυτού αποθηκεύεται ο αριθμός της γραμμής του πίνακα F_M στην οποία βρίσκεται αποθηκευμένο το διάνυσμα αντιπροσώπων που αντιστοιχεί στο στοιχείο αυτό.

Οι τρεις πίνακες αυτοί προκύπτουν πολύ εύκολα αφού ο F_M είναι ο πίνακας των διανυσμάτων, ο EP προκύπτει με απλή άθροιση των μέχρι τότε αριθμών επαναλήψεων, ενώ στον L_I διατηρείται η τιμή της τρέχουσας γραμμής του F_M για τόσα στοιχεία όσα καθορίζει ο αριθμός των επαναλήψεων. Χρησιμοποιώντας το συνδυασμό των πινάκων F_M , EP και L_I μπορούμε ανά πάσα στιγμή και με δύο (2)

το πολύ βήματα να γνωρίζουμε ποιό διάνυσμα αντιπροσώπων αντιστοιχεί σε ποιό χρονικό δείγμα.

Μια πρώτη σημαντική παρατήρηση είναι ότι ο μεγαλύτερος αριθμός των συγκρίσεων που πραγματοποιείται αφορά το πρώτο διάνυσμα V_1 του καταγεγραμμένου ηχητικού σήματος. Κάθε φορά αυτό συγκρίνεται με ένα διάνυσμα αντιπροσώπων από το πρότυπο ηχητικό σήμα και αν η σύγκριση αποτύχει, τότε το V_1 πρέπει να συγκριθεί με το διάνυσμα αντιπροσώπων που αντιστοιχεί στο επόμενο δείγμα του πρότυπου ηχητικού σήματος. Αφού όμως τις περισσότερες φορές τα διανύσματα που αντιστοιχούν σε συνεχόμενα χρονικά δείγματα είναι όμοια, δεν είναι ανάγκη να συγκρίνουμε το V_1 παρά μόνο μια φορά με κάθε διάνυσμα του πίνακα F_M . Πρακτικά αυτό σημαίνει ότι για κάθε πρότυπο ηχητικό σήμα θα γίνουν οι μισές συγκρίσεις διανυσμάτων.

Άρα, επειδή ο μεγαλύτερος υπολογιστικός φόρτος είναι πραγματικά οι συγκρίσεις του V_1 , αναπτύχθηκε ένας αλγόριθμος συγκρίσεων που βασίζεται στην πληροφορία των αλλαγών από το ένα διάνυσμα του πίνακα F_M στο επόμενο. Έτσι εισάγουμε τρεις απλούς πίνακες που επιταχύνουν σημαντικά τη διαδικασία σύγκρισης.

- Ο μονοδιάστατος πίνακας C_C , μεγέθους $Y \times I$. Σε αυτόν περιέχεται η πληροφορία του αριθμού των αλλαγών που έχουν συμβεί από ένα διάνυσμα αντιπροσώπων στο επόμενο. Η τιμή αυτή μπορεί να είναι 1, 2 ή 3, ώστε να προκύπτει άμεσα από την κωδικοποίηση που έχει πραγματοποιηθεί. Αν ένα διάνυσμα αντιπροσώπων έχει παραπάνω από 3 αλλαγές, αποθηκεύουμε τον ενδεικτικό αριθμό '0'.
- Ο πίνακας WIS , μεγέθους $Y \times 3$. Σε κάθε στοιχείο n του πίνακα αυτού περιέχονται οι τιμές των δεικτών που εμφανίζονται στο διάνυσμα U_n^* , ενώ δεν υπήρχαν στο διάνυσμα U_{n-1}^* .
- Ο πίνακας EKS , μεγέθους $Y \times 3$. Σε κάθε στοιχείο n του πίνακα αυτού περιέχονται οι θέσεις των δεικτών που υπήρχαν στο διάνυσμα U_{n-1}^* , ενώ δεν εμφανίζονται στο διάνυσμα U_n^* .

Οι πίνακες αυτοί υλοποιούνται μια και μόνη φορά για κάθε πρότυπο ηχητικό σήμα, κατά τη διαδικασία της αποκωδικοποίησης των διανυσμάτων. Στην περίπτωση που ο αριθμός των αλλαγών που έχουν πραγματοποιηθεί είναι μικρότερος από 3, οι

τελευταίες τιμές των πινάκων *EKS* και *EIS* μας είναι αδιάφορες, όπως κι όταν ο αριθμός των αλλαγών είναι μεγαλύτερος από 3, οπότε και οι πίνακες δεν χρησιμοποιούνται.

Με χρήση των ανωτέρω, μπορεί να διατυπωθεί ο κάτωθι γρήγορος αναδρομικός αλγόριθμος εξέτασης του πρώτου διανύσματος αντιπροσώπων:

Υπόθεση: Έστω ότι έχει ελεγχθεί η γραμμή $i-1$, του πίνακα F_M , κι έχει ήδη υπολογισθεί ο αριθμός των κοινών στοιχείων της με το διάνυσμα V_1 . Αν βρισκόμαστε στο πρώτο στοιχείο του πίνακα F_M , μεταβαίνουμε στο Βήμα 2A.

Βήμα 1: Ελέγχεται αρχικά, αν ο αριθμός $C_C[n]$ είναι μηδέν. Αν είναι μηδέν πηγαίνουμε στο Βήμα 2A. Αν ο αριθμός δεν είναι 0 (επομένως είναι 1,2 ή 3), μεταβαίνουμε στο Βήμα 2B.

Βήμα 2A: Το διάνυσμα αντιπροσώπων U_n^* έχει πάνω από 3 διαφορές από το U_{n-1}^* .

Σε αυτή την περίπτωση προχωρούμε στη σύγκριση των στοιχείων του U_n^* με τα στοιχεία του V_1 , για να βρεθεί ο αριθμός των κοινών τους στοιχείων. Για τη σύγκριση αυτή, εκμεταλλευόμαστε το γεγονός ότι και τα δύο διανύσματα είναι ταξινομημένα με φθίνουσα σειρά. Έτσι αντί να πραγματοποιήσουμε L^2 συγκρίσεις, κάτι που θα ήταν αναπόφευκτο αν τα στοιχεία ήταν με τυχαία σειρά, πραγματοποιούμε $2*L$ συγκρίσεις. Η μέση περίπτωση είναι περίπου ίδια με τη χειρότερη, δηλαδή κοντά στις 40 συγκρίσεις όταν $L=20$. Τέλος, να σημειωθεί ότι κατά τη σύγκριση υπολογίζεται ο αριθμός των μη κοινών στοιχείων, από τον οποίο με μια απλή αφαίρεση από το L υπολογίζεται ο ζητούμενος αριθμός. Η σύγκριση βασίζεται στο γεγονός ότι τα στοιχεία των διανυσμάτων είναι καταταγμένα σε φθίνουσα σειρά και πραγματοποιείται ως εξής:

Αρχικές συνθήκες: Χρησιμοποιούνται δύο μετρητές για τη μέτρηση των μη κοινών στοιχείων, ο ένας μετρά χρησιμοποιώντας πληροφορία από το διάνυσμα V_1 (μετρητής i_v) και ο άλλος από το U_n^* (μετρητής i_u). Επίσης χρησιμοποιούνται δύο μετρητές (έστω c_v και c_u), για να διατρέξουμε το κάθε ένα διάνυσμα, από το πρώτο του στοιχείο (το μεγαλύτερο δείκτη συχνοτικής περιοχής) ως το τελευταίο του, μετρητές που αρχικοποιούνται ώστε να δείχνουν το πρώτο στοιχείο κάθε διανύσματος.

Έλεγχος 2A.1: Ελέγχουμε αν το στοιχείο $V_1[c_v]$ είναι το ίδιο με το $U_n^*[c_u]$ και διακρίνουμε περιπτώσεις

Βήμα 2A.1A, $V_1[c_v]=U_n^[c_u]$:* Αν είναι ίσα, αυξάνουμε και τους δύο δείκτες (c_v και c_u) και επαναλαμβάνουμε τον έλεγχο 2A.1.

Βήμα 2A.1B, $V_1[c_v]>U_n^[c_u]$:* Αυξάνουμε το δείκτη c_v και το μετρητή i_v , αφού λόγω του γεγονότος ότι τα διανύσματα έχουν καταταγεί σε φθίνουσα σειρά, είμαστε σίγουροι ότι το στοιχείο $V_1[c_v]$ δεν μπορεί να υπάρχει στο διάνυσμα U_n^* .

Βήμα 2A.1Γ, $V_1[c_v]<U_n^[c_u]$:* Αυξάνουμε το δείκτη c_u και το μετρητή i_u , αφού, πάλι λόγω του γεγονότος ότι τα διανύσματα έχουν καταταγεί σε φθίνουσα σειρά, είμαστε βέβαιοι ότι το υπό εξέταση στοιχείο του U_n^* δεν υπάρχει στο V_1 .

Τελικό βήμα: Όταν ένας από τους δύο δείκτες c_v ή c_u φθάσει στο L η διαδικασία σταματά. Το αποτέλεσμα που ψάχνουμε (ο αριθμός των μη κοινών στοιχείων των δύο διανυσμάτων) είναι ο μεγαλύτερος από τους δύο μετρητές, i_u ή i_v .

Το Βήμα 2A ολοκληρώνεται με την εύρεση του αριθμού των κοινών (ή αντίστοιχα μη κοινών στοιχείων) των διανυσμάτων. Αμέσως μετά επιστρέφουμε στο Βήμα 1, θέτοντας νέα τιμή στο n , αλλά ταυτόχρονα αποθηκεύουμε την τιμή που έχουμε υπολογίσει.

Βήμα 2B: Το διάνυσμα αντιπροσώπων U_n^* έχει μέχρι 3 διαφορετικά στοιχεία σε σχέση με το U_{n-1}^* . Άρα προχωρούμε στον υπολογισμό της τιμής των κοινών στοιχείων, βασιζόμενοι στην πληροφορία των αλλαγών.

Αρχικές συνθήκες: Ξεκινούμε χρησιμοποιώντας πάλι δύο μετρητές, όπου ο ένας, έστω c_n , διατρέχει το υπό εξέταση διάνυσμα U_n^* , και ο άλλος, έστω c_{n-1} , διατρέχει το προηγούμενό του διάνυσμα U_{n-1}^* . Οι δύο μετρητές θα διατρέξουν τα διανύσματα αυτά ανεξάρτητα ο ένας από τον άλλο, οπότε είναι δυνατό σε μια υλοποίηση να χρησιμοποιηθεί και η ίδια μεταβλητή γι' αυτούς. Έχουμε επίσης δύο μετρητές, c_{EKS} και c_{EIS} αντίστοιχα, οι οποίοι διατρέχουν τη γραμμή n των πινάκων **EKS** και **EIS** που κρατούν τις αλλαγές. Όλοι αυτοί οι μετρητές αρχικοποιούνται ώστε να δείχνουν στο πρώτο στοιχείο της αντίστοιχης γραμμής.

Βήμα 2B1: Ελέγχεται αν το στοιχείο **EIS**[c_{EIS}] που εμφανίζεται στο διάνυσμα U_n^* ενώ δεν υπήρχε στο U_{n-1}^* , υπάρχει στο διάνυσμα V_1 . Αν το στοιχείο αυτό υπάρχει στο V_1 , τότε αυξάνουμε τον αριθμό των κοινών στοιχείων κατά ένα, σε σχέση με τον αριθμό που είχε υπολογιστεί μεταξύ του V_1 και του U_{n-1}^* . Αν όχι, προχωρούμε χωρίς να αλλάξουμε τον αριθμό αυτό. Αμέσως μετά συνεχίζουμε αυξάνοντας τον δείκτη **EIS** κατά ένα. Αν τότε έχουμε περάσει το όριο $C_C[n]$ προχωρούμε στο επόμενο βήμα.

Υπάρχουν δύο, σχεδόν ισοδύναμες μέθοδοι με τις οποίες γίνεται ο έλεγχος για το αν τα στοιχεία του **EIS** εμφανίζονται στο διάνυσμα V_1 . Αρχικά, επειδή και τα δύο διανύσματα είναι ταξινομημένα, μπορεί να εφαρμοστεί μια μέθοδος ακριβώς όμοια με αυτή που χρησιμοποιήθηκε για το Βήμα 2A. Μια τέτοια μέθοδος έχει πολυπλοκότητα $L+3$ συγκρίσεις. Εναλλακτικά μπορεί να χρησιμοποιηθεί και η πολύ γνωστή μέθοδος δυαδικής αναζήτησης, εφαρμοζόμενη μία φορά για κάθε ένα στοιχείο του πίνακα **EIS**, αλλά κάθε φορά στο μη ελεγμένο τμήμα του διανύσματος V_1 . Αυτή η μέθοδος δίνει πολυπλοκότητα $(2 * \log_2 L + 2 * \log_2 (L - 1) + 2 * \log_2 (L - 2))$, η οποία στην περίπτωση $L=20$ είναι χειρότερο από το $L+3$. Παρ' όλα αυτά, επειδή στην πράξη ο μέσος αριθμός των στοιχείων που εναλλάσσονται είναι μικρότερος του 3, οι δύο μέθοδοι αποδεικνύονται σχεδόν ισοδύναμες. Να τονιστεί κι εδώ η σπουδαιότητα της ταξινομημένης αποθήκευσης των διανυσμάτων, που οδηγεί σε ταχύτερες μεθόδους αναζήτησης και σύγκρισης.

Παρακάτω δίνεται ένα παράδειγμα των πράξεων που απαιτούνται με τις δύο μεθόδους για μια τυπική τριάδα στοιχείων του πίνακα **EIS**.

Έστω το διάνυσμα $V_1=[60\ 57\ 54\ 50\ 48\ 47\ 46\ 42\ 40\ 33\ 30\ 28\ 22\ 16\ 11\ 9\ 8\ 4\ 1\ 0]$
και ο πίνακας **EIS**[n]=[45,28,16]

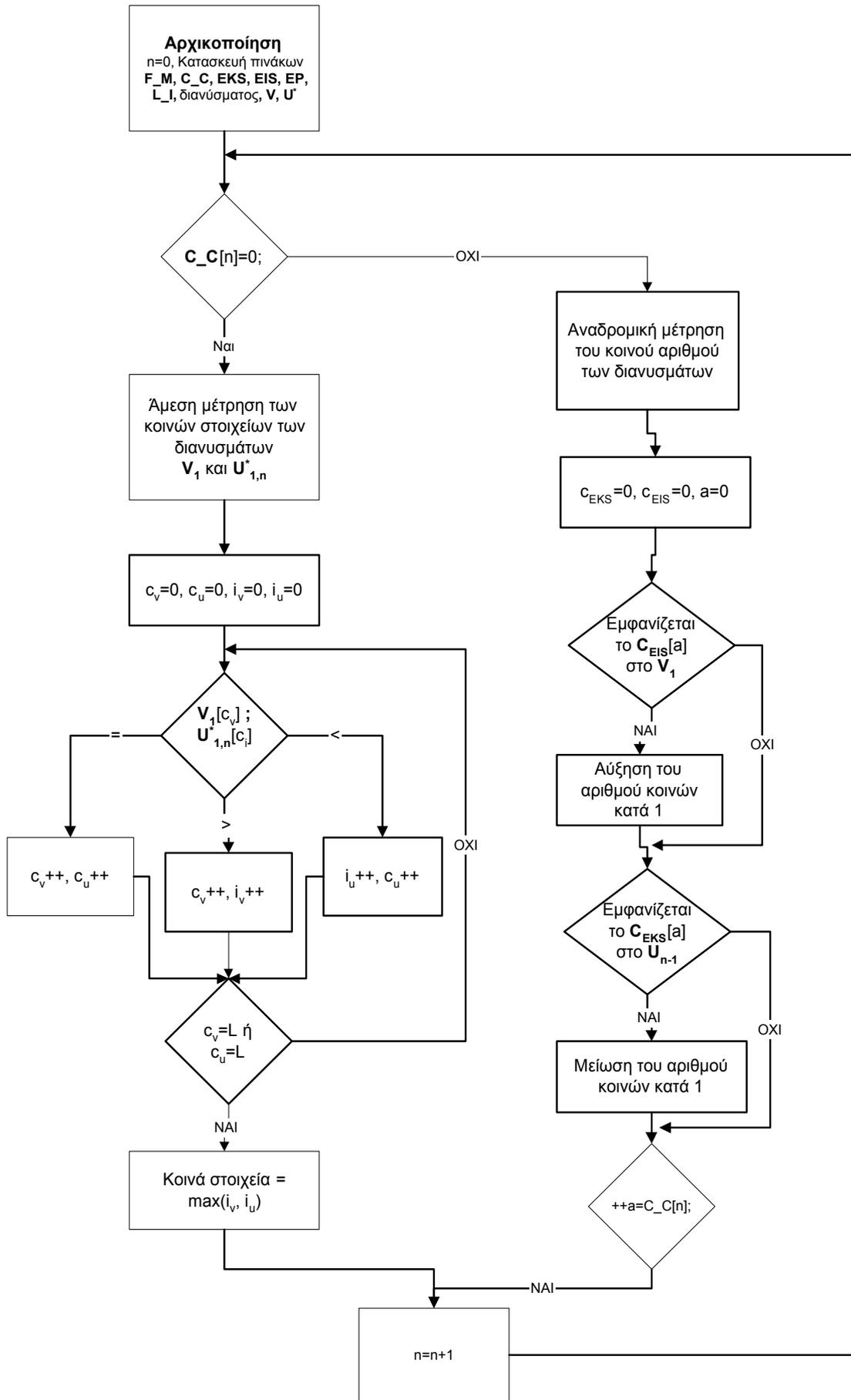
Τότε ο γραμμικός αλγόριθμος (2A) θα χρησιμοποιήσει οχτώ συγκρίσεις, για το πρώτο στοιχείο, πέντε για το δεύτερο και δύο για το τρίτο, σύνολο 15.

Ο αλγόριθμος δυαδικής αναζήτησης θα χρησιμοποιήσει 8 για το πρώτο στοιχείο, 5 για το δεύτερο και 3 για το τρίτο σύνολο 16 συγκρίσεις.

Βήμα 2B2: Ελέγχεται αν το στοιχείο $EKS[c_{EKS}]$ που υπήρχε στο διάνυσμα U_{n-1}^* ενώ δεν εμφανίζεται στο U_n^* , υπάρχει στο διάνυσμα V_1 . Αν το στοιχείο αυτό υπάρχει στο V_1 , τότε μειώνουμε τον αριθμό των κοινών στοιχείων κατά ένα, σε σχέση με τον αριθμό που είχε υπολογιστεί από το προηγούμενο βήμα. Αν όχι, προχωρούμε χωρίς να αλλάξουμε τον αριθμό αυτό. Αμέσως μετά συνεχίζουμε αυξάνοντας τον δείκτη EIS κατά ένα. Αν τότε έχουμε περάσει το όριο $C_C[n]$ προχωρούμε στο επόμενο βήμα.

Ο συνολικός αριθμός προκύπτει αμέσως μετά το βήμα αυτό. Ιδιαίτερη προσοχή χρειάζεται όταν οι πίνακες EKS και EIS περιέχουν κάποια τιμή '-1'. Τότε πρέπει να μετρηθεί η διαφορά του αριθμού των '-1' ανάμεσα στα διανύσματα U_n^* στο U_{n-1}^* ώστε οι να εξαχθούν τα ορθά συμπεράσματα.

Με τον αλγόριθμο που περιγράφηκε προηγουμένως, η διαδικασία σύγκρισης του πρώτου διανύσματος αντιπροσώπου, υλοποιείται ταχύτατα. Η διαδικασία αυτή φαίνεται σχηματικά στο παρακάτω διάγραμμα.



Με γρήγορο τρόπο γίνεται όμως και η διαδικασία της σύγκρισης των υπόλοιπων διανυσμάτων, όπως περιγράφεται πιο κάτω. Στη διαδικασία αυτή εισερχόμεθα μόνον εφ' όσον ισχύει το κριτήριο των $0.55 * L$ κοινών στοιχείων για το πρώτο συγκρινόμενο διάνυσμα ενός συνόλου. Για μια φορά ακόμα υπεισέρχεται μια τουλάχιστον διαδικασία σύγκρισης. Η διαδικασία αυτή γίνεται χρησιμοποιώντας την πληροφορία που υπάρχει αποθηκευμένη στους πίνακες F_M , EP και L_I .

Ο αλγόριθμος για τη σύγκριση αυτή έχει ως εξής

Βήμα 3: Αρχικά υπολογίζεται σε ποιο χρονικό δείγμα στο καταγεγραμμένο ηχητικό σήμα αντιστοιχεί το επόμενο από τα M διανύσματα αντιπροσώπους. Συγκεκριμένα, αν σε κάποιο σημείο έχει γίνει ήδη, επιτυχώς, η σύγκριση του k -οστού (από τα M) διανύσματος, τότε ο τύπος που δίνει τη θέση \underline{n}_{k+1} του επόμενου διανύσματος αντιπροσώπου V_{k+1} , είναι $\underline{n}_{k+1} = n + k * l$, όπου n η θέση στην οποία είχε υπολογιστεί ταίριασμα για το πρώτο διάνυσμα αντιπρόσωπος. Τονίζεται για μια φορά ακόμα ότι στο πρότυπο ηχητικό σήμα ορίζουμε ως το διάνυσμα αντιπροσώπων αυτού που θα πρέπει να συγκριθεί με το διάνυσμα V_{k+1} , το διάνυσμα που αντιστοιχεί στο χρονικό δείγμα \underline{n}_{k+1} , δηλαδή το $F_M[L_I[n+k*l]]$, δεδομένου ότι ο πίνακας L_I μας οδηγεί στο διάνυσμα που αντιστοιχεί στο τυχόν δείγμα n_1 . Δεν αρκεί όμως να εξετάσουμε αυτό το διάνυσμα και μόνο, εφ' όσον, όπως έχει αναφερθεί, το διάνυσμα $F_M[L_I[n+k*l]]$ μπορεί να αντιστοιχεί σε περισσότερα του ενός χρονικά δείγματα, το πλήθος των οποίων φανερώνει ο αριθμός των επαναλήψεων $EP[n]$. Επομένως, για πλήρη ασφάλεια στην αναγνώριση, θα πρέπει να εξεταστεί το διάστημα των χρονικών δειγμάτων που αρχίζει από το $F_M[L_I[n+k*l]]$ και τελειώνει στο $F_M[L_I[n+k*l + EP[n]]]$. Για ν' αντιμετωπίσουμε δε πιθανά προβλήματα που μπορούν να προκύψουν από αλλαγές ρυθμού, μπορούμε να εισάγουμε κι ένα μικρό παράγοντα ανοχής κατά μερικά δείγματα επιπλέον ($\pm b$). Επομένως, τελικά, το διάστημα των δειγμάτων των οποίων τα διανύσματα αντιπροσώπων πρέπει να εξεταστούν είναι το $[L_I[n+k*l-b], L_I[n+k*l+EP[n]+b]]$. Το διάστημα αυτό δεν αντιστοιχεί συνήθως παρά σε ένα μικρό αριθμό διανυσμάτων. Ορίζουμε τους αριθμούς $Z_{low} = L_I[n+k*l-b]$ και $Z_{high} = n+k*l + EP[n]+b$ και υπολογίζουμε το μέσο του προαναφερθέντος διαστήματος ως $Z_{start} = (Z_{low} + Z_{high})/2$ και θέτουμε ως αρχικό διάνυσμα που αναζητούμε το $F_M[Z_{start}]$.

Βήμα 4: Χρησιμοποιώντας τον αλγόριθμο που περιγράφεται στο Βήμα 2Α, υπολογίζουμε τον αριθμό των κοινών στοιχείων των διανυσμάτων $F_M[Z_{start}]$ και V_k . Ανάλογα με την τιμή του αριθμού των κοινών στοιχείων των διανυσμάτων αυτών μεταβαίνουμε στο επόμενο βήμα ως εξής:

Συνθήκη 4α: Ο αριθμός είναι μεγαλύτερος από $0.55 * L$. Τότε έχουμε επιτυχώς ελέγξει το ταίριασμα του διανύσματος V_k , οπότε αυξάνουμε το k κατά 1. Αν το k δεν έχει ξεπεράσει το M , επιστρέφουμε στο Βήμα 3, διαφορετικά θα πρέπει να μεταβούμε στο βήμα 5, για να γίνει ο έλεγχος του μέσου όρου των κοινών στοιχείων των διανυσμάτων.

Συνθήκη 4β: Ο αριθμός είναι μικρότερος από $0.55 * L$. Σ' αυτήν την περίπτωση συνεχίζουμε να ψάχνουμε για κάποιο διάνυσμα, που μπορεί να καλύψει την προϋπόθεση των $0.55 * L$ κοινών στοιχείων μέσα στο διάστημα $[Z_{low}, Z_{high}]$. Χρησιμοποιούμε ένα περιστροφικό αλγόριθμο, όπου το επόμενο διάνυσμα που θα εξετάσουμε είναι αυτό που αντιστοιχεί στο δείγμα $Z_{start+i}$, με το i να παίρνει τιμές ως εξής $(1, -1, 2, -2, 3, -3, \dots)$, εφ' όσον θεωρείται πιο πιθανό το ταίριασμα να πραγματοποιηθεί πιο κοντά στο κεντρικό δείγμα Z_{start} . Αν ο έλεγχος αποτύχει για όλα τα σημεία του διαστήματος $[Z_{low}, Z_{high}]$, τότε εγκαταλείπουμε τον έλεγχο του συνόλου διανυσμάτων αντιπροσώπων που αντιστοιχεί στο χρονικό δείγμα n , θέτουμε $n=n+1$ και επιστρέφουμε στο βήμα 1.

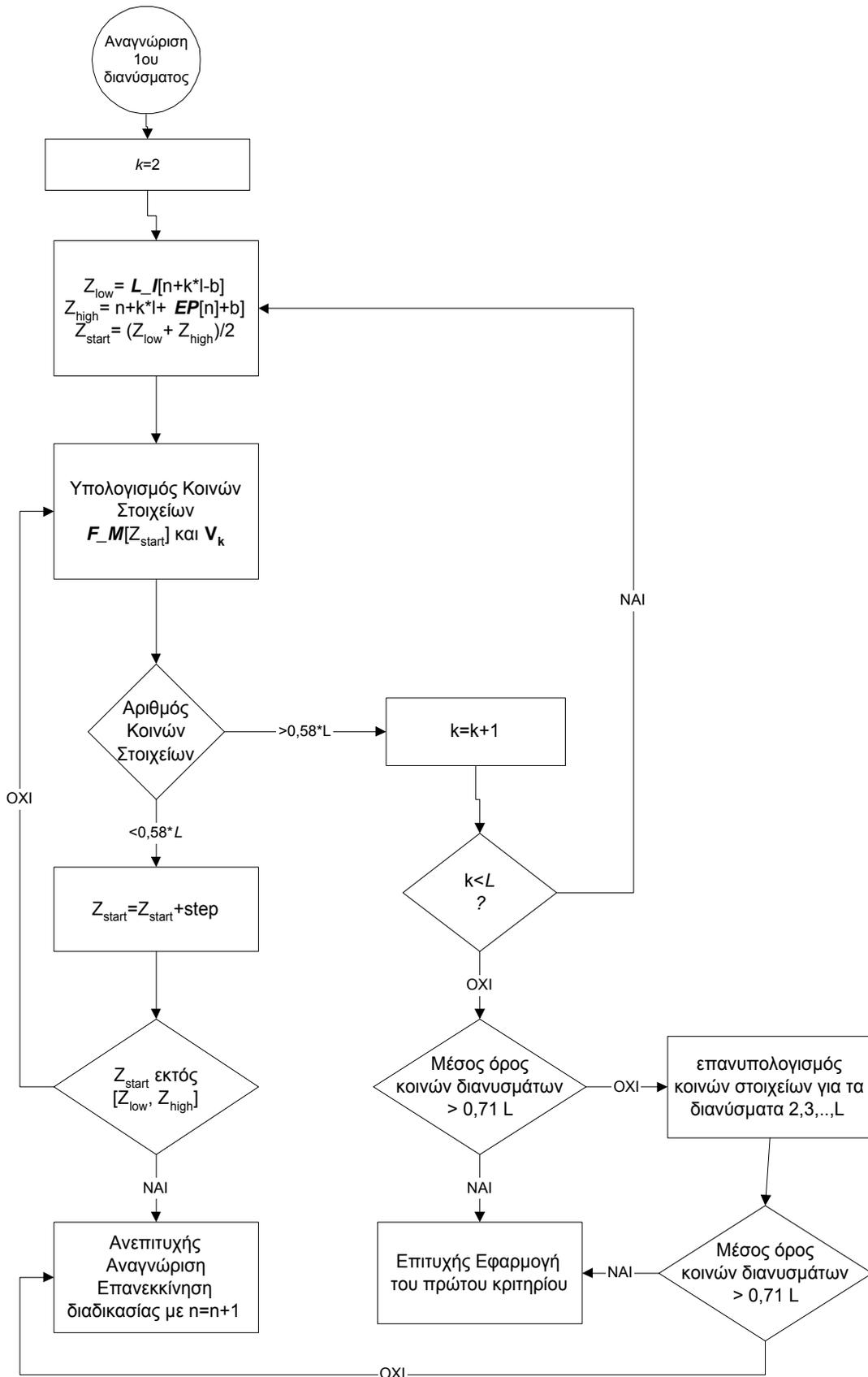
Βήμα 5: Φτάνοντας στο βήμα αυτό, όλα τα M διανύσματα αντιπροσώπων που έχουν υπολογιστεί πληρούν το κριτήριο των $0.55 * L$ κοινών στοιχείων με τ' αντίστοιχα διανύσματα αντιπροσώπων του υπό εξέταση πρότυπου ηχητικού σήματος. Τότε, αθροίζονται όλες οι τιμές που έχουν υπολογιστεί για τα κοινά στοιχεία των M διανυσμάτων και το αποτέλεσμα διαιρείται με το M . Αν ο προκύπτων αριθμός είναι μεγαλύτερος από το $0.71 * L$, δηλώνουμε επιτυχία ως προς το πρώτο κριτήριο αναγνώρισης. Αν όμως ο αριθμός αυτός είναι μικρότερος από $0.71 * L$, πρέπει να πραγματοποιηθεί ένας ακόμα έλεγχος, πριν αποφασίσουμε ότι το κριτήριο δεν πληρείται, όπως περιγράφεται στο βήμα 6.

Όπως παρατηρήθηκε στη συνθήκη 4α, ο στόχος ήταν να υπολογιστεί ένα μόνο διάνυσμα, το πρώτο που θα βρισκόταν, το οποίο να ικανοποιεί το κριτήριο των $0.55 * L$ κοινών στοιχείων. Υπάρχει όμως περίπτωση, μέσα στο διάστημα $[Z_{low}, Z_{high}]$ να μπορεί να βρεθεί κι ένα άλλο διάνυσμα αντιπρόσωπος, που να πληρεί

με το κριτήριο των $0.55 * L$. Επειδή δεν μπορούμε να αποκλείσουμε τη μικρή πιθανότητα, η αύξηση της τιμής των κοινών στοιχείων για το διάνυσμα αυτό, να οδηγήσει σε εκπλήρωση και του κριτηρίου των $0.71 * L$, πρέπει σε κάθε τέτοιο διάστημα να υπολογίζεται η καλύτερη τιμή ταιριάσματος. Παρατηρήθηκε, ότι συμφέρει από άποψη υπολογιστικού χρόνου, να μην γίνεται μια τέτοια αναζήτηση από την αρχή, όταν δηλαδή αναζητείται διάνυσμα που να έχει $0.55 * L$ κοινά στοιχεία. Αυτό θα σήμαινε πολλούς υπολογισμούς σε κάθε βήμα για να βρεθεί η καλύτερη τιμή ταιριάσματος. Στην πράξη, η πλειονότητα των συνόλων διανυσμάτων που φτάνουν στο Βήμα 5, πληρούν το κριτήριο των $0.71 * L$, ενώ το ποσοστό αυτών που δεν το πληρούν είναι τόσο μικρό, που συμφέρει να πραγματοποιηθεί σε αυτό το σημείο ο έλεγχος για το αν μπορεί να βρεθεί καλύτερο ταιρίασμα σε κάθε διάνυσμα, παρά κατά τη διάρκεια των βημάτων 3 και 4.

Βήμα 6: Επαναλαμβάνουμε τη διαδικασία των βημάτων 3 και 4, για κάθε ένα από τα διανύσματα V_k , $k=2,3,\dots,L$, και μόνο για όσα διανύσματα που αντιστοιχούν σε κάθε διάστημα $[Z_{low}, Z_{high}]$ δεν έχουμε ήδη εξετάσει. Αυτό βέβαια προϋποθέτει ότι θα πρέπει να έχουμε αποθηκεύσει ενδιάμεσα τα σημεία του $[Z_{low}, Z_{high}]$ όπου σταμάτησε ο έλεγχος για κάθε k -διάνυσμα. Αφού ολοκληρωθεί η διαδικασία αυτή, ελέγχουμε πάλι το μέσο όρο των κοινών στοιχείων με τον αριθμό $0.71 * L$. Αν ο μέσος όρος είναι ξανά μικρότερος του $0.71 * L$, τότε το κριτήριο δεν πληρείται, και θα πρέπει να εγκαταλείψουμε τον έλεγχο του συνόλου διανυσμάτων αντιπροσώπων που αντιστοιχεί στο χρονικό δείγμα n , να θέσουμε $n=n+1$ και να επιστρέψουμε στο βήμα 1. Αν είναι μεγαλύτερος δηλώνουμε την εκπλήρωση του κριτηρίου αυτού και ότι οδηγούμαστε σε επιτυχή ταυτοποίηση.

Η διαδικασία ταιριάσματος για τα $L-1$ μη πρώτα διανύσματα δίνεται γραφικά στο πιο κάτω σχήμα:



BIBΛΙΟΓΡΑΦΙΑ

- [1] M. Mongeau, D. Sankoff, "Comparison of musical sequences", *Computers and the Humanities*, 24, 161-175, 1990.
- [2] R.J. McNab, L.A. Smith, I.H. Witten, C.L. Henderson, S.J. Cunningham, "Towards the digital music library: Tune retrieval from acoustic input", *Proceedings of ACM Digital Libraries*, 11-18, 1996.
- [3] R.J. McNab, L.A. Smith, D. Bainbridge, I.H. Witten, " The New Zealand Digital Library MELody inDEX", *D-Lib Magazine*, May 1997.
- [4] A. Pikrakis, S. Theodoridis, D. Kamarotos, "Recognition of Isolated Musical Patterns in the Context of Greek Traditional Music", *Proceedings of the International Conference on Electronics Circuits and Systems (ICECS, organized by IEEE)*.
- [5] A. Pikrakis, S. Theodoridis, D. Kamarotos, "Recognition of Isolated Musical Patterns using Discrete observation Hidden Markov Models", *Proceedings of the European Signal Processing Conference (EUSIPCO) 1998*.
- [6] C.J. Stevens, C.R. Latimer, "Music recognition: An illustrative application of a connectionist model", *Psychology of Music*, 25, 161-185, 1997.
- [7] C.J. Stevens, C.R. Latimer, "A comparison of connectionist models of music recognition and human performance", *Minds and machines*, 2, 279-400, 1992.
- [8] C.J. Stevens, C.R. Latimer, "Recognition of short tonal compositions by connectionist models and listeners: Effects of feature manipulation and training", *Musikometrika*, 5, 197-224, 1993.
- [9] K. Brandenburg, G. Stoll, "ISO-MPEG-1 Audio: a generic standard for coding of high quality digital audio", *Journal of the Audio Engineering Society* 42 (10) 780-792, October 1994.
- [10] Davis Pan, "A tutorial on MPEG/Audio Compression", *IEEE Multimedia Journal*, Summer 1995 issue.
- [11] J. H. Rothweiler, "Polyphase Quadrature Filters - a new subband coding technique", *Proc. of the Int. Conf. IEEE ASSP*, 27.2, pp. 1280-1283, Boston 1983.

- [12] M. Bosi et al, "ISO/IEC MPEG-2 Advanced Audio Coding", Journal of the Audio Engineering Society, No 10, Oct. 1997, pp. 789-813.
- [13] E. Zwicker, H. Fastl, "Psychoacoustics", Springer - Verlag, Berlin, Germany, 1990.
- [14] Painter T., Spanias A., A Review of Algorithms for Perceptual Coding of Digital Audio Signals
- [15] McNab R.J, Smith L.A., Witten I.H., Henderson C.L., Cunningham S.J., Towards the digital music library: Tune retrieval from acoustic input, Proceedings of ACM Digital Libraries, (1996) pp. 11-18.
- [16] Moore B.C.J., An Introduction to the Psychology of Hearing, Academic Press (1997).
- [17] Portnoff M.R., Implementation of the Digital Phase Vocoder Using the Fast Fourier Transform, IEEE Transactions on Acoustic, Speech and Signal Processing ASSP-24 (1976) 243-248.
- [18] Allen J.B. and Rabiner L.R., A Unified Approach to Short-time Fourier Analysis and Synthesis, Proc. IEEE 65 (1977) 1558-1564.
- [19] A.V. Aho, J.E. Hopcroft, and J.D. Ulmann, "Data structures and Algorithms", Addison-Wesley Pub. Co., 1983.
- [20] R.M. Karp and M.O. Rabin, "Efficient Randomize Pattern-Matching Algorithms", IBM Journal of Research and Development, 31, 249-260, 1987.
- [21] Y. Matias and U. Vishkin, "On Parallel Hashing and Integer Sorting", Journal of Algorithms, 12(4), 573-606, 1991.
- [22] D.E. Knuth, J.H. Morris and V.R. Pratt, "Fast Pattern Matching in Strings", SIAM J. Comp., 6, 323-350, 1977.
- [23] D. Harel and R.E. Tarjan, "Fast Algorithms for Finding Nearest Common Ancestor", Computer and System Science, 13, 338-355, 1984.
- [24] D.E. Knuth "The Art of Computing Programming", Vol. 3: Sorting and Searching", Addison-Wesley, Reading, MA, 1973.
- [25] P. Weiner "Linear Pattern Matching Algorithm", Proc. 14 IEEE Symposium on Switching and Automata Theory, 1-11, 1973.

- [26] B. Baker “A Theory of Parametrized Pattern Matching: Algorithms and Applications” Proc. of the 25th Ann. ACM Symp. On Theory of Computing, 71-80, 1993.
- [27] E. Zwicker, “Subdivision of the audible frequency range into critical bands (Frequenzgruppen)”, Journal of the Acoustical Society of America, 33, 248, 1961.
- [28] B.C.J. Moore, B.R. Glasberg, “Suggested formulae for calculating auditory-filter bandwidths and excitation patterns”, Journal of the Acoustical Society of America, 74, 750-753, 1983b.
- [29] L.R Rabiner and Gold, “Theory and Application of Digital Signal Processing”, Prentice Hall, 1975.
- [30] Frerking, “Digital Signal Processing in Communication Systems”, Kluwer, 1994.