



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΜΟΥΣΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΣΤΗΝ ΜΟΥΣΙΚΗ ΕΚΠΑΙΔΕΥΣΗ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΒΑΓΙΑ Χ. ΟΥΡΑΝΙΑ-ΕΙΡΗΝΗ

Επιβλέπων : ΚΑΡΑΓΙΑΝΝΗΣ ΓΕΩΡΓΙΟΣ

Καθηγητής Ε.Μ.Π.

Αθήνα, ΑΥΓΟΥΣΤΟΣ 2007



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΜΟΥΣΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΣΤΗΝ ΜΟΥΣΙΚΗ ΕΚΠΑΙΔΕΥΣΗ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΒΑΓΙΑ Χ. ΟΥΡΑΝΙΑ-ΕΙΡΗΝΗ

Επιβλέπων : ΚΑΡΑΓΙΑΝΝΗΣ ΓΕΩΡΓΙΟΣ

Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 28^η ΑΥΓΟΥΣΤΟΥ 2007.

.....

Καραγιάννης Γεώργιος

.....

Καμπουράκης Γεώργιος

.....

Μαραγκός Πέτρος

Αθήνα, Αύγουστος 2007

.....

ΒΑΓΙΑ Χ. ΟΥΡΑΝΙΑ-ΕΙΡΗΝΗ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

.....

ΒΑΓΙΑ Χ. ΟΥΡΑΝΙΑ-ΕΙΡΗΝΗ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright ©ΟΥΡΑΝΙΑ-ΕΙΡΗΝΗ ΒΑΓΙΑ, 2007.

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

Στην παρούσα διπλωματική διατριβή ασχολούμαστε με εξαγωγή χαρακτηριστικών από μονοφωνικά μουσικά σήματα. Συγκεκριμένα, σε πρώτο στάδιο μελετείται το πρόβλημα της αναγνώρισης του μουσικού σήματος, με κύριο στάδιο την εξαγωγή του ύψους των νοτών γνωστό στη διεθνή βιβλιογραφία ως monophonic pitch tracking/ estimation. Σε δεύτερο στάδιο χειριζόμαστε το πρόβλημα της αντιστοίχισης της ακολουθίας εισόδου, δηλαδή της ακολουθίας των αναγνωρισμένων νοτών από το πρώτο στάδιο, με την ακολουθία αναφοράς, την παρτιτούρα του μουσικού κομματιού (score following/ matching). Καθώς οι δύο μονάδες δρουν αλυσιδωτά και η έξοδος της πρώτης αποτελεί είσοδος της δεύτερης, είναι απαραίτητη η αξιολόγηση της επιτυχίας της μονάδας αναγνώρισης. Η αξιολόγηση αυτή γίνεται με ένα σύνολο εκτελέσεων αρχαρίων, ηχογραφημένων σε πραγματικές συνθήκες και προτείνονται βελτιώσεις στις περιπτώσεις που η μονάδα παρουσιάζει αδυναμίες. Τέλος, ασχολούμαστε με τη μονάδα αξιολόγησης της μουσικής εκτέλεσης (performance evaluation), η οποία παίρνοντας ως είσοδο την αντιστοιχισμένη ακολουθία, υπολογίζει τα λάθη στην εκτέλεση του μουσικού κομματιού . Αναζητούνται επίσης τρόποι αυτόματου προσδιορισμού των λαθών στην ποιότητα του τόνου, με βάση τα φασματογράμματα των προβληματικών ήχων.

Λέξεις κλειδιά

Εκτίμηση , ανίχνευση, ύψος, νότα, τόνος, μονοφωνική ,μουσική, αντιστοίχιση, ακολουθία, παρτιτούρα, αξιολόγηση, εκτέλεση, φασματογράφημα, σήμα

Abstract

In this diploma thesis we handle the feature extraction problem from monophonic music signals. More specifically, firstly, we study the music recognition problem, which first and main step is pitch estimation /tracking of each note. Secondly, we handle the following and matching problem of the recognized input sequence with the reference sequence, i.e. the score of the musical piece. Because of the two modules performing in chain and the output of the first module is input in the second module, we evaluate the success of the Recognizer module. The evaluation is held with a set of test-cases performed by beginner students under real-world conditions and we propose improvements in cases, where the module fails to recognize correctly some notes. Last, we study a performance evaluation module, which takes as input the matched sequence and calculates the mistakes of the performer. We also search ways of automatic identification of bad quality tones, based on the spectrograms of these tones.

Key words

Estimation, tracking, pitch, note, tone, monophonic, music, score, following, matching, performance, evaluation, spectrogram, signal

Ευχαριστίες

Ένα μεγάλο ευχαριστώ στον κ. Γιώργο Καραγιάννη για την εμπιστοσύνη που μου έδειξε και την ευκαιρία που μου έδωσε να ασχοληθώ με τη διπλωματική μου εργασία.

Επίσης, θα ήθελα να ευχαριστήσω ιδιαίτερα τον επιβλέποντά μου διδάκτορα κ. Σπύρο Ράπτη για την πολύτιμη καθοδήγηση και βοήθειά του καθ' όλη τη διάρκεια εκπόνησης της παρούσας εργασίας.

Επιπλέον, ευχαριστώ την ομάδα του Ι.Ε.Λ. και ιδιαίτερα τον Κώστα Περήφανο και τον Άγγελο Γκικόκα για την πολύτιμη συμβολή τους.

Τέλος, τους φίλους και την οικογένειά μου για την υπομονή και τη στήριξη τους σε όλη τη διάρκεια της εργασίας.

Εισαγωγή

Η παρούσα διπλωματική εκπονήθηκε στον Τομέα ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ υπό την εποπτεία του Καθ. κ. Γ. Καραγιάννη. Μέρος της πραγματοποιήθηκε στο Ινστιτούτο Επεξεργασίας του Λόγου / Ερευνητικό Κέντρο «Αθηνά» όπου προσφέρθηκε η δυνατότητα ενασχόλησης με τις εφαρμογές που περιγράφονται στην παρούσα διατριβή.

Σκοπός της διπλωματικής είναι η διερεύνηση των σταδίων επεξεργασίας που απαιτούνται για την ανάλυση, την αναγνώριση, την εξαγωγή χαρακτηριστικών και, τελικά, την αξιολόγηση μουσικών εκτελέσεων αρχαρίων μαθητών σε μονοφωνικά όργανα.

Πιο συγκεκριμένα, αρχικά αναλύεται το θέμα της μουσικής αναγνώρισης με την ανίχνευση του ύψους ενός μονοφωνικού μουσικού ήχου ως συνάρτηση του χρόνου. Το ζήτημα αυτό είναι γνωστό στη διεθνή βιβλιογραφία ως *μονοφωνική εκτίμηση/παρακολούθηση μουσικού ύψους* (monophonic pitch tracking/estimation). Με απλά λόγια, η εξαγωγή του μουσικού ύψους ενός ήχου μας δίνει τη μουσική νότα.

Σε δεύτερο επίπεδο χρησιμοποιούνται οι προηγούμενες πληροφορίες ως είσοδος σε μία βαθμίδα η οποία αντιστοιχίζει την ακολουθία εισόδου με την ακολουθία αναφοράς (την παρτιτούρα) και παράγει ένα σύνολο υποθέσεων για τα λάθη των μαθητών τα οποία αποτελούν την βάση για την αξιολόγηση της εκτέλεσης. Γνωστό στη διεθνή βιβλιογραφία ως *αντιστοίχιση της εκτέλεση με την παρτιτούρα* (score matching), το ζήτημα αυτό έχει απασχολήσει πλήθος ερευνητών από το 1984 και συνεχίζει να είναι ένα πεδίο ενεργούς έρευνας.

Τέλος, περιγράφεται η βαθμίδα αξιολόγησης της εκτέλεσης. Με βάση φασματικά χαρακτηριστικά του ήχου, ακολουθεί η *αξιολόγηση της ποιότητας της νότας* (note quality assessment). Καταγράφεται ένα σύνολο λαθών που γίνονται συχνά σε εκτελέσεις αρχαρίων, γίνεται αναζήτηση των παραγόντων που οδηγούν στα συγκεκριμένα λάθη και περιγράφονται φασματικά οι τύποι αυτοί λαθών.

Η μέθοδος της εργασίας βασίζεται στην αναλυτική μελέτη της απόδοσης των αντίστοιχων υπολογιστικών βαθμίδων επεξεργασίας που έχουν αναπτυχθεί στο πλαίσιο του ερευνητικού έργου VEMUS/IMUTUS. Με βάση τις βαθμίδες αυτές, αναλύεται ένα σύνολο εκτελέσεων-αναφοράς, ηχογραφημένων υπό ρεαλιστικές συνθήκες από εκτελέσεις σπουδαστών, και αξιολογείται η επιτυχία περάτωσης του στόχου της κάθε βαθμίδας. Σε περιπτώσεις αποτυχίας, μελετάται η φύση του προβλήματος, οι πιθανές αιτίες καθώς και πιθανές λύσεις.

Στο πλαίσιο της παρούσας εργασίας, για την αξιολόγηση των βαθμίδων του VEMUS χρησιμοποιήθηκαν 37 εκτελέσεις κλαρινέτου και 52 εκτελέσεις φλογέρας σε 35 διαφορετικά κομμάτια από 80 διαφορετικούς μαθητές, από αρχαρίους έως μέσου επιπέδου καλύπτοντας ένα μεγάλο εύρος δυνατών περιπτώσεων χρήσης του προγράμματος.

Η σχέση μεταξύ των θεμάτων που ερευνώνται στην παρούσα διπλωματική εργασία είναι αλληλεξάρτησης. Συγκεκριμένα για να γίνει σωστή εκτίμηση του ύψους μιας νότας, θα πρέπει αυτή να εκτελεστεί επαρκώς καλά από τον μουσικό. Με την σειρά της, η σωστή εκτίμηση του ύψους είναι απαραίτητη προϋπόθεση για να γίνει σωστά η αντιστοίχιση της εκτέλεσης με την παρτιτούρα. Σε ένα πραγματικό σύστημα μουσικής τεχνολογίας, οι βαθμίδες δουλεύουν αλυσιδωτά και οι έξοδοι της μιας είναι είσοδοι της άλλης. Επομένως, η ακρίβεια των αποτελεσμάτων που παράγει κάθε βαθμίδα παίζει καθοριστικό ρόλο στην επίτευξη ικανοποιητικών συνολικών αποτελεσμάτων. Επιπρόσθετα, κάθε

βαθμίδα θα πρέπει να έχει την δυνατότητα να χειριστεί κατάλληλα τα πιθανά λάθη ή/και την αβεβαιότητα στην πληροφορία που λαμβάνει από την προηγούμενη, ώστε να επιτευχθεί συνολικά μεγαλύτερος βαθμός ευρωστίας.

Η εργασία επικεντρώνεται σε *μουσικές εκτελέσεις αρχαρίων*. Αυτές αποτελούν μια ειδική (και ιδιαίτερα απαιτητική) κατηγορία μουσικών όπου περιλαμβάνονται ποικιλία λαθών, εν γένει χαμηλή ποιότητα εκτέλεσης αλλά και συγκεκριμένες ανάγκες εντοπισμού λαθών και κατάλληλης αναφοράς τους στον χρήστη. Τα δεδομένα δοκιμής και οι ηχογραφήσεις που χρησιμοποιήθηκαν ως βάση της παρούσας μελέτης προέρχονται από εκτελέσεις αρχαρίων.

Καθένα από τα παραπάνω θέματα παρουσιάζεται ξεχωριστά στα αντίστοιχα κεφάλαια της διπλωματικής.

Θα πρέπει να τονιστεί ότι η διπλωματική εργασία πραγματεύεται μονοφωνική μουσική, που αποτελεί και το ζητούμενο του έργου VEMUS. Για μονοφωνική μουσική, πολλά από τα παραπάνω θέματα παρουσιάζουν σημαντικά μειωμένο βαθμό πολυπλοκότητας σε σχέση με την αναγνώριση πολυφωνικής μουσικής ή την αναγνώριση πιο σύνθετων ηχητικών σημάτων, όπως η φωνή. Εντούτοις, τα προβλήματα που εμφανίζονται εξακολουθούν να διατηρούν το ενδιαφέρον τους και συνεχίζουν να παρουσιάζουν σημαντικές προκλήσεις για την αποδοτική αντιμετώπισή τους, γεγονός που αντανακλάται και από το έντονο ερευνητικό ενδιαφέρον που ελκύουν οι περιοχές αυτές. Επιπρόσθετα, στο πεδίο της μουσικής εκπαίδευσης, η μονοφωνική μουσική παρουσιάζει ιδιαίτερο ενδιαφέρον καθώς παγκοσμίως η ευρεία πλειοψηφία των μαθημάτων μουσικής, ξεκινά με μονοφωνικά όργανα όπως η φλογέρα.

Η διπλωματική αυτή πραγματοποιήθηκε στα πλαίσια του ερευνητικού έργου VEMUS. Το VEMUS ("Virtual European Music School") είναι ένα έργο έρευνας και ανάπτυξης που συγχρηματοδοτείται από την Ευρωπαϊκή Επιτροπή στο πλαίσιο του Προγράμματος των Τεχνολογιών της Κοινωνίας της Πληροφορίας (IST - Information Society Technologies) του Έκτου Προγράμματος-Πλαισίου (FP6 - Sixth Framework Program). Περισσότερες πληροφορίες για το έργο VEMUS είναι διαθέσιμες στο Παράρτημα I.

“Wonderful things would come out of that box if only we knew how to evoke them”

John R.Pierce,1965: Portrait of the Computer as a Young Artist

Πίνακας Περιεχομένων

ΠΕΡΙΛΗΨΗ	1
ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ	1
ABSTRACT	2
KEY WORDS	2
ΕΥΧΑΡΙΣΤΙΕΣ	3
ΕΙΣΑΓΩΓΗ	4
1 ΒΑΣΙΚΕΣ ΈΝΝΟΙΕΣ ΜΟΥΣΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΑΚΟΥΣΤΙΚΗΣ	9
1.1 ΜΟΥΣΙΚΗ ΤΕΧΝΟΛΟΓΙΑ	9
1.1.1 Ορισμός	9
1.1.2 Τομείς Σύγχρονης Έρευνας στη Μουσική Τεχνολογία	10
1.1.3 Εφαρμογές της Μουσικής Τεχνολογίας	12
1.1.4 Η Μουσική Τεχνολογία στην Εκπαίδευση	13
1.2 ΒΑΣΙΚΕΣ ΈΝΝΟΙΕΣ ΑΚΟΥΣΤΙΚΗΣ	16
1.2.1 Ήχος	16
1.2.2 Ηχηρότητα (Loudness)	21
1.2.3 Μουσικό ύψος του ήχου (Pitch)	24
1.2.4 Χροιά ή Ηχόχρωμα (Timbre or Tone colour)	24
1.2.5 Συμπέρασμα	25
1.3 Η ΜΟΥΣΙΚΗ ΚΛΙΜΑΚΑ	26
1.4 ΜΟΥΣΙΚΗ ΚΑΤΑΓΡΑΦΗ ΗΧΗΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ	26
2 ΑΝΑΓΝΩΡΙΣΗ ΜΟΥΣΙΚΗΣ	28
2.1 ΣΥΝΤΟΜΗ ΕΠΙΣΚΟΠΗΣΗ	29
2.2 ΔΙΑΧΩΡΙΣΜΟΣ ΜΟΥΣΙΚΗΣ/ΘΟΡΥΒΟΥ	30
2.3 ΑΝΙΧΝΕΥΣΗ ΣΙΩΠΗΣ	31
2.4 ΕΚΤΙΜΗΣΗ ΘΕΜΕΛΙΩΔΟΥΣ ΣΥΧΝΟΤΗΤΑΣ	35
2.5 ΚΑΤΑΤΜΗΣΗ ΗΧΗΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ	41
2.6 ΑΝΙΧΝΕΥΣΗ ΤΗΣ ΑΡΧΗΣ (ONSET) ΤΗΣ ΝΟΤΑΣ	43
2.7 ΜΗΧΑΝΗ ΑΠΟΦΑΣΗΣ 3 ΕΠΙΠΕΔΩΝ	47
2.8 ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΤΟΝΟΥ ΟΡΓΑΝΟΥ	48
2.9 ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΔΟΚΙΜΗΣ	50
2.10 ΑΞΙΟΛΟΓΗΣΗ ΒΑΘΜΙΔΑΣ/ ΣΥΜΠΕΡΑΣΜΑΤΑ	54
2.10.1 Διάσπαση νότας σε δύο με διαφορά ενός ημιτονίου	54
2.10.2 Διάσπαση νότας σε δύο ίδιες νότες	57
2.10.3 Συγχώνευση Νοτών με Διαφορά ενός Ημιτονίου σε $1/8^{\circ}$	58
2.10.4 Συγχώνευση νοτών με διαφορά ενός ημιτονίου σε $1/16^{\circ}$	59
2.10.5 Συγχώνευση νοτών με διαφορά ενός τόνου	60
2.11 ΣΥΜΠΕΡΑΣΜΑΤΑ/ ΠΡΟΤΑΣΕΙΣ ΒΕΛΤΙΩΣΗΣ	61
3 ΑΚΟΛΟΥΘΙΑ /ΑΝΤΙΣΤΟΙΧΙΣΗ ΠΑΡΤΙΤΟΥΡΑΣ	63
3.1 ΓΕΝΙΚΗ ΕΠΙΣΚΟΠΗΣΗ	63
3.2 ΥΠΑΡΧΟΥΣΑ ΈΡΕΥΝΑ	63
3.3 Η ΒΑΘΜΙΔΑ ΑΝΤΙΣΤΟΙΧΙΣΗΣ ΤΟΥ VEMUS	65
3.3.1 Εισαγωγή	65

3.3.2	<i>Ρόλος και Τεχνική Περιγραφή Βαθμίδας</i>	67
3.3.3	<i>Αλγοριθμική Περιγραφή</i>	70
3.3.4	<i>Η Μηχανή Αντιστοίχισης</i>	74
3.3.5	<i>Περιβάλλον του Εργαλείου Αντιστοίχισης</i>	75
3.3.6	<i>Αποθήκευση των MTX αρχείων των αποτελεσμάτων</i>	80
3.3.7	<i>Διεπαφή γραμμής εντολών</i>	80
4	ΑΞΙΟΛΟΓΗΣΗ ΕΚΤΕΛΕΣΗΣ	81
4.1	ΒΑΘΜΙΔΑ ΑΞΙΟΛΟΓΗΣΗΣ ΕΚΤΕΛΕΣΗΣ ΤΟΥ VEMUS	81
4.2	ΕΞΑΡΤΗΣΗ ΑΞΙΟΛΟΓΗΣΗΣ ΑΠΟΔΟΣΗΣ ΑΠΟ ΤΟ ΕΠΙΠΕΔΟ ΤΟΥ ΜΑΘΗΤΗ	83
4.3	ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΒΑΘΜΙΔΑΣ ΑΞΙΟΛΟΓΗΣΗΣ ΤΗΣ ΕΚΤΕΛΕΣΗΣ	85
4.4	ΠΟΙΟΤΗΤΑ ΤΟΥ ΤΟΝΟΥ	85
4.5	ΚΑΤΗΓΟΡΙΕΣ ΠΡΟΒΛΗΜΑΤΙΚΩΝ ΤΟΝΩΝ	86
4.6	ΦΑΣΜΑΤΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΠΡΟΒΛΗΜΑΤΙΚΩΝ ΝΟΤΩΝ ΣΕ ΕΚΤΕΛΕΣΕΙΣ ΑΡΧΑΡΙΩΝ	87
4.6.1	<i>Τριζήματα</i>	88
4.6.2	<i>Κενές Νότες</i>	92
4.6.3	<i>Ασταθείς τόνοι</i>	96
4.6.4	<i>Διπλοί Τόνοι</i>	99
4.7	ΣΥΜΠΕΡΑΣΜΑΤΑ.....	102
5	ΣΥΜΠΕΡΑΣΜΑΤΑ	103
	ΠΑΡΑΡΤΗΜΑ – ΤΟ ΕΡΕΥΝΗΤΙΚΟ ΈΡΓΟ VEMUS	104
	ΠΑΡΑΡΤΗΜΑ – ΠΙΝΑΚΑΣ ΜΕΤΑΦΡΑΣΗΣ ΕΠΙΛΕΓΜΕΝΩΝ ΌΡΩΝ	108
	ΕΥΡΕΤΗΡΙΟ ΕΙΚΟΝΩΝ ΚΑΙ ΠΙΝΑΚΩΝ	109
	ΒΙΒΛΙΟΓΡΑΦΙΑ	112

1

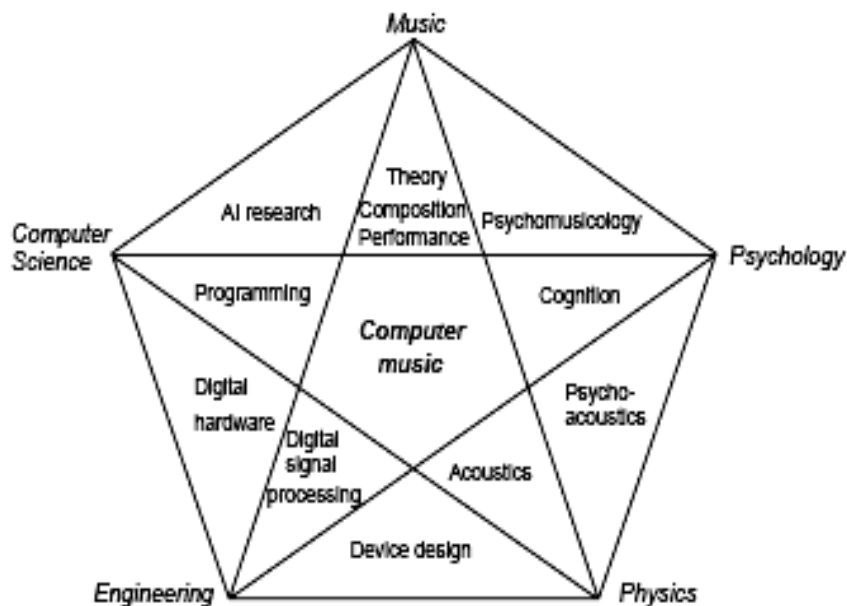
Βασικές Έννοιες Μουσικής Τεχνολογίας και Ακουστικής

1.1 Μουσική Τεχνολογία

1.1.1 Ορισμός

Η **Μουσική Τεχνολογία (Music Technology)** ορίζεται ως η ανάπτυξη τεχνολογιών και διαδικασιών, σχετικά με μεθοδολογίες, εκτέλεση, αντίληψη, επεξεργασία, ανάλυση και σύνθεση μουσικής.

Η Μουσική Τεχνολογία ή η πιο περιορισμένη έννοια της **Υπολογιστικής Μουσικής (Computer Music)** είναι ένα χαρακτηριστικό παράδειγμα διεπιστημονικού ερευνητικού πεδίου, που βρίσκεται στο σταυροδρόμι πολλών επιστημονικών κλάδων, όπως η Μουσικολογία, η Ακουστική, η Επεξεργασία Σήματος, η Ψυχοακουστική, η Τεχνητή Νοημοσύνη και η Επικοινωνία Ανθρώπου-Υπολογιστή (Human-Computer Interaction, HCI).



Εικόνα 1: Απεικόνιση διεπιστημονικότητας της Μουσικής Τεχνολογίας

Μια σημαντική επίπτωση αυτής της "διεπιστημονικότητας" είναι ότι επηρεάζει τη μεθοδολογία που ακολουθείται, αφού συνυπάρχουν ταυτόχρονα μεθοδολογίες που προέρχονται από διαφορετικές ερευνητικές παραδόσεις της Μουσικής, της Τεχνολογίας και των Κοινωνικών Επιστημών. Υπάρχει ένας μεγάλος αριθμός ερευνητικών πεδίων που είναι αποτέλεσμα αυτής της διεπιστημονικής έρευνας. Η ενασχόληση των ερευνητών καλύπτει τόσο καλλιτεχνικά όσο και επιστημονικά/ μηχανικά θέματα.

1.1.2 Τομείς Σύγχρονης Έρευνας στη Μουσική Τεχνολογία

Η Μουσική Τεχνολογία πάντα παρασύρεται από τεχνολογικά επιτεύγματα και ερευνητικές εξελίξεις που προέρχονται από άλλα πεδία. Επομένως, παρακολουθώντας τα τεχνολογικά ρεύματα είναι δυνατόν να αποκομίσουμε μία σαφή εικόνα των εφαρμογών που θα επιτευχθούν στο εγγύς μέλλον. Η Μουσική Τεχνολογία πάντα υπήρξε ένα εφαρμοσμένο πεδίο πολύ κοντά στη μουσική βιομηχανία.

Υπάρχουν πολλά ενεργά ερευνητικά πεδία στη Μουσική Τεχνολογία. Από τα πρώτα πειράματα μουσικής σε υπολογιστές στα τέλη της δεκαετίας του '50 ως τη σύγχρονη έρευνα, το πεδίο της έχει διευρυνθεί απρόσμενα. Αυτή την εποχή μία ουσιαστικής κατηγοριοποίηση της ενεργούς έρευνας είναι η παρακάτω:

1. Η **Μουσική Παραγωγή / Μοντελοποίηση (Music Generation/Modeling)** είναι κυρίως συνυφασμένη με αυτό που είναι γνωστό ως Αλγοριθμική Σύνθεση (Algorithmic Composition) και εμπεριέχει κυρίως το καλλιτεχνικό στοιχείο. Εξαιτίας αυτής της καλλιτεχνικής της φύσης, είναι δύσκολο να αναλυθεί ποια είναι τα κύρια επιτεύγματα της έρευνας στη Μουσική Παραγωγή. Τα επιτεύγματα στις δεκαετίες '70-'80 ήταν σημαντικά αλλά τίποτα ιδιαίτερα καινοτομικό δεν έχει επακολουθήσει την τελευταία δεκαετία. Παρόλα ταύτα, αυτή η ερευνητική περιοχή έχει εγείρει πολύ ενδιαφέροντα ζητήματα τόσο στη μουσικολογία, όσο στα μαθηματικά και το σχεδιασμό λογισμικού. Αυτή την εποχή υπάρχει πληθώρα προσεγγίσεων και συστημάτων που έχουν αυτοματοποιήσει τη διαδικασία Σύνθεσης Μουσικής και η κινητήρια δύναμη του συγκεκριμένου πεδίου κείται στο ίδιο το περιεχόμενο της Σύγχρονης Μουσικής.

Από πολιτιστικής άποψης, η Υπολογιστική Μουσική ήταν πρωτογενώς πολύ κοντινή στη Σύγχρονη Μουσική. Στην πραγματικότητα, για τους περισσότερους η Υπολογιστική Μουσική είναι συνώνυμη με τη Σύγχρονη Μουσική. Τα τελευταία χρόνια αυτό έχει αλλάξει αρκετά και η αισθητική αντιμετώπιση είναι ανεξάρτητη από την έρευνα της Υπολογιστικής Μουσικής. Ουσιαστικά, αυτός είναι ο λόγος που συνηθίζουμε να χρησιμοποιούμε τον όρο Μουσική Τεχνολογία αντί του όρου Υπολογιστική Μουσική, ώστε να αποφύγουμε τη συνειρμική σημασία της Σύγχρονης Μουσικής.

2. Η **Παραγωγή / Μοντελοποίηση Ήχων (Sound Generation/Modeling)** υπήρξε ένας από τους ιδιαίτερα παραγωγικούς τομείς έρευνας για τα τελευταία 40 χρόνια. Η μουσική που παράγεται ηλεκτρονικά στη σημερινή εποχή, στηρίζεται στην επεξεργασία δειγμάτων, την πιο άμεση ψηφιακή αναπαράσταση ενός δεδομένου ήχου, αν και υπάρχουν άλλες περισσότερο ισχυρές αναπαραστάσεις και μοντέλα. Αυτή η προτίμηση δεν σχετίζεται μόνο με την ποιότητα του ήχου που παράγεται από έναν συγκεκριμένο αλγόριθμο σύνθεσης, αλλά περισσότερο με τον έλεγχο και την εκφραστικότητά του.

Για παράδειγμα μία πολλά υποσχόμενη κατεύθυνση στηρίζεται σε σύνθεση βασισμένη στα γνωρίσματα του ήχου (feature-driven), γνωστή και ως Audio Mosaicing. Τέτοια συστήματα στηρίζονται σε φιλτράρισμα με τη βοήθεια περιφερειακών, κατάτμηση και εξαγωγή χαρακτηριστικών, αναγνώριση προτύπων, συσσώρευση και ταξινόμηση και περνούν από άκουσμα σε μάθηση μέσω μοντελοποίησης των νοητικών διαστάσεων της ανθρώπινης ακοής.

3. **Ανάλυση / Σύνθεση Εκτέλεσης (Music Performance Analysis/Synthesis)**. Αν και ο εκτελεστής ενός μουσικού κομματιού είναι καθοριστικό στοιχείο στην αλυσίδα παραγωγής της μουσικής, έχει μελετηθεί επιστημονικά τι προσδίδει ο εκτελεστής σε αυτή την αλυσίδα. Η

δεξιοτεχνία στην εκτέλεση καλλιεργείται με ένα περισσότερο διαισθητικό τρόπο και για το λόγο αυτό είναι αρκετά αόριστη στην επιστημονική ανάλυση. Παρόλα ταύτα, τα τελευταία χρόνια έχει μετατραπεί σε ιδιαίτερα ενεργό και καρποφόρο τομέα της έρευνας. Διάφορες προσεγγίσεις χρησιμοποιούνται τώρα για την ανάλυση και τη μοντελοποίηση της απόδοσης στην εκτέλεση ενός κομματιού. Το μεγαλύτερο μέρος της τρέχουσας έρευνας είναι βασισμένο είτε στη μέτρηση πραγματικών εκτελέσεων, εφαρμόζοντας Analysis-by-Synthesis (AbS) μεθόδους είτε αναπτύσσοντας πρότυπα εκτέλεσης με τις μεθόδους εκμάθησης μηχανών.

Αρχίζουμε να είμαστε σε θέση να συλλάβουμε τις διαφορές μεταξύ των διαφορετικών εκτελεστών, έχουμε τώρα τα πρότυπα για τη σύλληψη της διάθεσης σε ένα δεδομένο μουσικό ύφος, είμαστε σε θέση να μετρήσουμε μερικά εκφραστικά χαρακτηριστικά γνωρίσματα από μια ακουστική καταγραφή μιας εκτέλεσης, και είμαστε επίσης ικανοί να παραγάγουμε μερικές πτυχές μιας εκφραστικής εκτέλεσης. Οι πρακτικές εφαρμογές όλης αυτής της έρευνας δεν είναι μακρινές.

4. Μουσικές Διεπαφές (Music Interfaces). Στα πλαίσια των ηλεκτρονικών οργάνων είναι εφικτό να χωριστεί η πτυχή του ελέγχου ενός μουσικού οργάνου από τις δυνατότητές του στην παραγωγή ήχων. Μπορούμε να χτίσουμε τους ελεγκτές και τις διεπαφές ώστε να συλλάβουμε τις «χειρονομίες» της εκτέλεσης και τις μονάδες ήχου για να παραγάγουμε τους ήχους. Με αυτό το διαχωρισμό οι δυνατότητες που δίνονται είναι εξαιρετικές. Από την εφεύρεση των πρώτων ηλεκτρονικών οργάνων έχει υπάρξει ιδιαίτερη έρευνα για την ανάπτυξη των νέων ελεγκτών με τους οποίους θα διερευνηθούν οι νέες δημιουργικές δυνατότητες και τα πρωτόκολλα επικοινωνίας για να διασυνδεθούν οι ελεγκτές με τις συσκευές παραγωγής ήχου. Κατά συνέπεια, η έννοια της εκτέλεσης παίρνει μια νέα έννοια και κατά συνέπεια και η ίδια η έννοια του οργάνου. Υπάρχει ανοικτό έδαφος για την έρευνα στην εξέταση της διεπαφής εκτελεστή-οργάνου στο γενικό πλαίσιο της αλληλεπίδρασης ανθρώπου-υπολογιστή. Αυτό είναι ένας ερευνητικός τομέας που έχει επίσης ένα πολύ ισχυρό μουσικό συστατικό και για τον οποίο δεν είναι εύκολο να καθιερωθούν κριτήρια καθορισμού της προόδου. Μερικά από τα ενεργά ερευνητικά θέματα είναι η μελέτη των μουσικών αλγορίθμων χαρτογράφησης, η ανάπτυξη των ευφυών ελεγκτών και νέων ελεγκτών για ομαδικές εκτελέσεις. Ένα σαφές παράδειγμα της μετατόπισης που λαμβάνει χώρα συσχετίζεται με Διαδίκτυο. Το Internet γίνεται όλο και πιο πολύ ένα μέσο στο οποίο παράγεται μουσική και κατά συνέπεια αποτελεί και μια νέα μουσική διεπαφή προς εξερεύνηση.

5. Κατανόηση/ Ανάκτηση μουσικής και ήχου (Music and Audio Understanding /Retrieval). Αυτήν την περίοδο αυτός είναι ο πιο ενεργός ερευνητικός τομέας στην Τεχνολογία της Μουσικής. Είμαστε μακριά από την κατανόηση της σχέσης μεταξύ της ακουστικής, της δομής μουσικής, της αναγνώρισης, και της συγκίνησης, αλλά αυτές οι σχέσεις προσφέρουν ένα μεγάλο έδαφος για την επιστημονική και τεχνολογική έρευνα.

Υπάρχουν πολλά ζητήματα που περιλαμβάνονται στη διαδικασία που ξεκινά όταν ένα ηχητικό κύμα έρχεται στο αυτί μας και τελειώνει με την μουσική αίσθηση που λαμβάνουμε από αυτόν. Καταλαβαίνουμε μερικά από τα χαμηλού επιπέδου (σύστημα ακοής) ζητήματα αλλά είμαστε μακριά από την κατανόηση ζητημάτων αναγνώρισης. Αυτήν την περίοδο υπάρχουν καλές ιδέες στις πτυχές της ακουστικής αντίληψης, μερικές από τις οποίες είναι: βασικές ακουστικές διαδικασίες, μηχανισμοί ομαδοποίησης χαμηλού-υψηλού τόνου, αντίληψη χροιάς, συχνότητας, χρόνου και ρυθμού. Πέρα από αυτό το χαμηλό επίπεδο υπάρχουν πολλά ανοικτά ερευνητικά θέματα σχετικά με θέματα όπως η ψυχοακουστική, η

ψυχολογία, η αντίληψη μουσικής, η αναγνώριση και η νευρολογία. Με τη χρησιμοποίηση των τεχνικών επεξεργασίας σήματος και εκμάθησης μηχανών ήμαστε σε θέση να εξαγάγουμε και να προσδιορίσουμε τις σημαντικές και χρήσιμες πληροφορίες από τα μουσικά σήματα. Αυτές οι αυτόματα εξαγόμενες πληροφορίες μπορούν να χρησιμοποιηθούν σε πολλά από τα προβλήματα που είναι υπό μελέτη αυτήν την περίοδο μέσα στην περιοχή ανάκτησης πληροφοριών μουσικής. Αρχίζουμε να έχουμε τα αποτελέσματα που μπορούν να χρησιμοποιηθούν στις πρακτικές εφαρμογές και η πρόοδος είναι τόσο γρήγορη που είναι δύσκολο να προβλέψουμε τι θα συμβεί στα επόμενα έτη.

1.1.3 Εφαρμογές της Μουσικής Τεχνολογίας

Η Μουσική Τεχνολογία βρίσκει εφαρμογή κυρίως στον τομέα της εκπαίδευσης και της σύνθεσης.

Η περιγραφή προηγμένων τεχνικών σύνθεσης για μοντελοποίηση οργάνων και ο έλεγχος των παραμέτρων των μοντέλων έχει ανοίξει το δρόμο σε νέες δημιουργικές εφαρμογές. Επίσης έχει προταθεί ρεαλιστική μοντελοποίηση χορδών (plucked strings) [KVJ93] και εγχόρδων οργάνων [SH93].

Η συμβολική παράσταση της μουσικής επέτρεψε την ανάπτυξη συστημάτων αυτόματης συνοδείας μουσικής (Automatic Music Accompaniment), βασισμένων σε συμβολικές αναπαραστάσεις [Dan85, Ver85, DM87] ενώ η δειγματοληπτική σύνθεση άνοιξε το δρόμο για την ανάπτυξη πιο σύνθετων συστημάτων παραγωγής [Pac02]. Η συμβολική αναπαράσταση μουσικής έχει επιτρέψει επίσης την αυτοματοποίηση της συμβολικής αναγνώρισης προτύπων [Hur95]. Συστήματα αντιστοίχισης μιας μουσικής εκτέλεσης με την αντίστοιχη παρτιτούρα (score alignment και score following) ξεκίνησαν με συμβολικά δεδομένα και τώρα συνδυάζουν την επεξεργασία σήματος με μουσικολογικούς κανόνες [Rap01b, OD01].

Οι παράμετροι των ηχητικών «εφέ» μπορούν να ελεγχθούν αυτόματα με προσαρμοστικό τρόπο [Ver04]. Η εξαγωγή χαρακτηριστικών από διαφορετικά αντιπροσωπευτικά σήματα αρχίζει να χρησιμοποιείται για τέτοια ηχητικά «εφέ», έτσι ώστε η ηχητική πηγή να μπορεί να χρησιμοποιηθεί ως έλεγχος πάνω στον εκάστοτε αλγόριθμο δημιουργίας τους.

Ο συμβολισμός για το ύψος και την αρχή της νότας φαίνεται να βελτιώνουν σημαντικά τους αλγόριθμους επεξεργασίας ήχου, όπως χρονική κλιμάκωση (time scaling) [RSB05], διατηρώντας τα αντιληπτά χαρακτηριστικά των attack (δηλαδή, των εισαγωγικών μερών των νοτών). Νέες διεπαφές χρησιμοποιούνται για το realignment των drum loops με τις οποίες ένα ηχητικό αντικείμενο μπορεί να μετατοπιστεί μέσα σε μία ηχογραφημένη δομή [AP05]. Η λεπτομερής τροποποίηση υπαρχόντων μελωδιών απαιτεί συμβολισμό τόσο του μουσικού ύψους του ήχου όσο και του χρόνου του onset [GPA03b].

Τα τελευταία συστήματα σύνθεσης από δείγματα χρησιμοποιούν βάσεις δεδομένων με τμήματα ήχων, σημειωμένων με χαρακτηριστικά ύψους και δομής [Cas94]. Η μορφή αυτής της τράπεζας ήχων επιτρέπει την δημιουργία, αποθήκευση, ανταλλαγή και τροποποίηση αυτών των δειγμάτων. Η σύνθεση από δείγματα μπορεί να χρησιμοποιηθεί με πολλούς δημιουργικούς τρόπους στη δημιουργία νέων μουσικών οργάνων με βάση αυτόματα κατατετημένους ήχους [ADH04].

Παρόμοια, η αυτόματη κατηγοριοποίηση τμημάτων ήχων εφαρμόζεται σε δημιουργικές εφαρμογές όπως το micro-montage προς δημιουργία νέας υφής ήχων[Cai04]. Ο ταξινομητής συσσωρεύει σύντομες ηχητικές «λωρίδες» σε κατηγορίες υφής, που μπορούν να συνδεθούν και να ελεγχθούν από

έναν φάκελο πλάτους. Το Music Mosaicing [ZP01, Cas05] είναι μια νέα εφαρμογή που διαλέγει από ένα σχολιασμένο σώμα ηχητικών δεδομένων υπάρχοντα δείγματα προς χρησιμοποίηση και μίξη. Άλλη δειγματοληπτική σύνθεση βασισμένη σε κατάτμηση ήχων περιέχεται στο [Jeh04, Col04].

Τα συστήματα Προτεινόμενης Μουσικής (Music recommendation) και ταξινόμησης είδους πρέπει να πάρουν ένα μικρό αριθμό από αποφάσεις χρησιμοποιώντας ένα μεγάλο σώμα από μουσικές ηχογραφήσεις, και άρα να επεξεργαστούν μεγάλο όγκο δεδομένων σε περιορισμένο χρόνο. Η αναγνώριση της μουσικής σε πραγματικό χρόνο έχει καταστεί εφικτή καθώς τα αποτελέσματα μπορούν να ληφθούν σε κάποιες δεκάδες δευτερόλεπτα [Wan03]. Κάποιες εφαρμογές όπως η «Query by Music» απαιτεί συγκεκριμένα την εξαγωγή σημασιολογικών δεδομένων [PPW05, Tza02]. Σύνθετα συστήματα συχνά βασίζονται σε μια τεχνική κατηγοριοποίησης για τη συσσώρευση στοιχείων στη βάση δεδομένων. Οι ταξινομητές πολύ συχνά χρησιμοποιούν την εξαγωγή χαρακτηριστικών των σημάτων σε περιορισμένα χρονικά τμήματα. Η ταχύτητα είναι η πιο σημαντική απαίτηση για τις μηχανές αναζήτησης.

1.1.4 Η Μουσική Τεχνολογία στην Εκπαίδευση

Έχουν γίνει κατά καιρούς αρκετές προσπάθειες χρήσης των υπολογιστών στην εκπαίδευση της μουσικής. Συχνά οι εφαρμογές που υλοποιούνται προς αυτή την κατεύθυνση έχουν ως στόχο:

1. **Τη διδασκαλία των βασικών εννοιών της μουσικής:** Τα περισσότερα προγράμματα Μουσικής Εκπαίδευσης έχουν επικεντρωθεί σε δραστηριότητες όπως διδασκαλία της σημειογραφίας της μουσικής και την εκτέλεση ακουστικών τεστ, σχετικών με αναγνώριση και ορθογραφία ακολουθιών νοτών, συγχορδιών, μελωδιών και αρμονίας [Hof88]. Η εξάσκηση στον υπολογιστή προσφέρει στους μαθητές ένα περιβάλλον με λιγότερο άγχος σε σχέση με τις παραδοσιακές αίθουσες διδασκαλίας μουσικής, αφού δεν υπάρχει το ανθρώπινο ακροατήριο [LeB97]. Στην πραγματικότητα, αυτή ήταν μια από τις πρώτες χρήσεις των υπολογιστών στην εκπαίδευση γενικά [OS83]. Τα σχετικά προγράμματα εξέταζαν κάθε πιθανή διαδρομή παραδειγμάτων που παρουσιάζονται στο σπουδαστή. Καθώς ο αριθμός των πιθανών διαδρομών μπορεί να γίνει πολύ μεγάλος, η προετοιμασία αυτού του είδους υλικού απαιτεί τεράστια προσπάθεια. Για να ελαχιστοποιηθεί αυτήν την προσπάθεια θα μπορούσε να χρησιμοποιηθεί ένα πρότυπο, αντί της προ-αποθήκευσης των ερωταποκρίσεων. Η τεχνική αυτή, που ονομάζεται αυτή η παραγωγική υπολογιστή εκμάθηση, θα μπορούσε να ελέγξει - υπό μια περιορισμένη έννοια - το θέμα και το επίπεδο δυσκολίας του επόμενου παραδείγματος με βάση κάποια εν των προτέρων διευκρινισμένη στρατηγική. Συστήματα όπως το GUIDO [Hof75, Hof81], το οποίο χρησιμοποιήθηκε για την εξάσκηση ακουστικών δεξιοτήτων σχετικά με τα μουσικά διαστήματα, τη μελωδία, την αρμονία και το ρυθμό στηρίζονται στις απαντήσεις του σπουδαστή για την επιλογή του επόμενου υλικού που θα παρουσιαστεί, την ταχύτητα της υπαγόρευσης ή ακόμα και το χρόνο που επιτρέπεται για τις απαντήσεις που δίνονται. Άλλα τέτοια προγράμματα για τη διδασκαλία βασικών αρχών της μουσικής είναι το MiBAC, το Music Ace και το Practica Musica. Οι περισσότερες από αυτές τις εφαρμογές χρησιμοποιούν τις τεχνικές παρουσίασης πολυμέσων και συσκευές MIDI εκτενώς και οι υπολογιστές παίζουν το ρόλο του ιδιαίτερα ειδικευμένου χειριστή ερωτηματολογίων πολλαπλής επιλογής.
2. **Τη διδασκαλία δεξιοτήτων στην εκτέλεση της μουσικής:** Οι εφαρμογές διδασκαλίας βασικών αρχών της μουσικής μπορούν να αντιμετωπισθούν ως βοηθητικές στη διδασκαλία

δεξιοτήτων στην εκτέλεση της μουσικής. Οι εφαρμογές αυτές δεν βελτιώνουν σημαντικά την απόδοση των σπουδαστών [Swa79] από μόνες τους, αφού είναι απαραίτητο να αναπτυχθούν κι άλλες ακουστικές ικανότητες σχετικές με την μουσική εκτέλεση. Σε αυτό το σημείο, θα περιγράψουμε μερικές προσπάθειες βελτίωσης ικανοτήτων όπως την εκτέλεση «με το αυτί» («playing by ear») και τη χρησιμοποίηση ακουστικής ανατροφοδότησης για τη διόρθωση της εκτέλεσης.

Το πρόγραμμα Tunemaster [Kir86] εξετάζει την ικανότητα εκτέλεσης «με το αυτί», καλώντας το μαθητή να παίξει μια μελωδία που παράγεται από το σύστημα χρησιμοποιώντας ένα πίνακα αφής. Εδώ, δεν υπάρχει καμία ανάγκη προηγούμενης γνώσης της συμβατικής μουσικής σημειολογίας και ο σπουδαστής παρακινείται μέσω ενός μουσικού παιχνιδιού. Οι εφαρμογές αυτές χρησιμοποιούν *οπτική ανατροφοδότηση* με τη μορφή μιας διεπαφής πιάνου. Μια παρόμοια προσέγγιση χρησιμοποιήθηκε στο Piano Tutor Project [Dan90]. Η εφαρμογή αυτή στηρίζεται στην ανατροφοδότηση της ακρίβειας της εκτέλεσης του αρχαρίου, που υποστηρίζεται από διαλογικά βίντεο ενός πραγματικού δασκάλου και μιας εφαρμογής αντιστοίχισης για τη σύγκριση της εκτέλεσης του σπουδαστή με τις προ-αποθηκευμένες πρότυπες εκτελέσεις. Τεχνικές ακόλουθου παρτιτούρας χρησιμοποιούνται ως βάση για την ανίχνευση των λαθών των σπουδαστών, ενώ έχει επίσης τη δυνατότητα προσαρμογής στις ανάγκες του κάθε σπουδαστή.

Η ανάπτυξη και η βελτίωση των δεξιοτήτων εκτέλεσης μουσικής στηρίζονται σε εργαλεία με ακουστική και οπτική ανατροφοδότηση ως κεντρικά στοιχεία. Οι προσεγγίσεις που στηρίζονται σε εκτελέσεις ειδικών χρησιμοποιούν την ικανότητα ερμηνείας (INTERPRET) [Bak92] των μαθητών για τη βελτίωση των ικανοτήτων τους, αλλά περιορίζεται από το μικρό πλήθος των πρότυπων εκτελέσεων.

Το μειονέκτημα αυτό των προ-αποθηκευμένων κομματιών ξεπεράστηκε μερικώς στο σύστημα pianoFORTE [SWK95], όπου υλοποιήθηκε ένα πρότυπο για την εκφραστικότητα της εκτέλεσης με τη βοήθεια δασκάλων στο πιάνο, που κωδικοποίησαν αυτή τη γνώση στο σύστημα. Οι εκτελέσεις των μαθητών σε MIDI πλήκτρα συλλαμβάνονται και ανατροφοδοτούνται οπτικά εκφραστικές πτυχές απόδοσης όπως ο ρυθμός, ο συγχρονισμός, η δυναμική και η άρθρωση που παρουσιάζονται στο σπουδαστή πάνω στην παρτιτούρα του κομματιού.

3. **Την ανάλυση της μουσικής:** Η ανάλυση μουσικής εξετάζει τον προσδιορισμό των θεμελιωδών στοιχείων μιας μουσικής δομής και την έρευνα των λειτουργιών αυτών των στοιχείων μέσα σ' αυτή τη δομή [Ben87].

Σε αυτό το τμήμα δίνουμε μια περίληψη της χρήσης των υπολογιστών στην ανάλυση μουσικής ως εργαλείο διδασκαλίας ή ως διαδικασία για έρευνα. Οι εφαρμογές που εξετάζονται έχουν χρησιμοποιηθεί για την εξέταση θεωριών μουσικής [Bak89a,b; Rob94], για τον έλεγχο της «πατρότητας» μουσικών κομματιών [Gro75], ή ακόμα και για τον προσδιορισμό θεσπισμένων κανόνων σε μουσικά κομμάτια [Blo81].

Οι υπολογιστές στην ανάλυση μουσικής χρησιμοποιούνται χαρακτηριστικά για υπολογισμό γεγονότων, ταξινόμηση, αναγνώριση προτύπων και στατιστική ανάλυση [Alp80]. Όλα αυτά τα προγράμματα αναγνωρίζουν τα περιστατικά τόνων, τιμών νοτών, διαστημάτων καθώς και συνδυασμών των προηγούμενων μουσικών στοιχείων.

Μια από τις πρώτες προσπάθειες χρήσης υπολογιστών στην ανάλυση μουσικής έγινε στο [Gro75]. Αναπτύχθηκε ένα σύνολο ρουτινών για τη μελωδική και κάθετη ανίχνευση σχεδίων, την ανίχνευση θεμάτων, την αρμονική ανάλυση, την καθορισμένη θεωρία και την αποθήκευση μιας συσσωρευτικής απαρίθμησης των αποτελεσμάτων. Αναλύθηκαν αντιπροσωπευτικά κομμάτια από διαφορετικές μουσικές μορφές που συντέθηκαν από τους

Bach, Haydn, Chopin και Dallapiccola και τα αποτελέσματα ήταν, στο μεγαλύτερο μέρος τους, ακριβή και παρείχαν τα χρήσιμα ποσοτικά στοιχεία.

Πιθανοτικές τεχνικές και τεχνικές βασισμένες στην αναπαράσταση της γνώσης, που υποστηρίζονται από τις καθιερωμένες θεωρίες μουσικής είναι οι κυρίαρχες προσεγγίσεις στην ανάλυση μουσικής. Το επόμενο τμήμα παρουσιάζει μια μεγαλύτερη ποικιλομορφία προσεγγίσεων με στόχο τη σύνθεση μουσικής, ως αποτέλεσμα την ατέρμονης φύσης αυτού του τομέα.

4. **Τη διδασκαλία δεξιοτήτων στη σύνθεση μουσικής:** Σε αυτό το τμήμα, εξετάζουμε εφαρμογές των υπολογιστών στη σύνθεση μουσικής που κυμαίνονται από διαλογικά εκπαιδευτικά παιχνίδια ως εξειδικευμένα ευφυή συστήματα διδασκαλίας. Οι στρατηγικές διδασκαλίας κυμαίνονται από απλές παρουσιάσεις εννοιών ως περισσότερες διερευνητικές προσεγγίσεις, και οι πιθανοί χρήστες από αρχαρίους ως πεπειραμένους συνθέτες.

Το Music Logo [Bam74] είναι ένα αντιπροσωπευτικό παράδειγμα χρήσης ενός διαλογικού εκπαιδευτικού παιχνιδιού στη μουσική σύνθεση. Στόχος του είναι να εφαρμόσει τις ιδέες της γλώσσας προγραμματισμού Logo στη μουσική και ο σπουδαστής μαθαίνει χτίζοντας και εξετάζοντας πρότυπα. Πειράματα χειρισμού των μουσικά σημαντικών στοιχείων υποστηρίζουν την άποψη του Bamberger για τα οφέλη του παιχνιδιού στην βελτίωση της μουσικής γνώσης του μαθητή [Bam91].

Άλλοι ερευνητές παρουσίασαν προσεγγίσεις βασισμένες στις διεπαφές όπως μουσικά παιχνίδια που περιλαμβάνουν μετασχηματισμούς σε σκιαγραφημένες καμπύλες πάνω σε πεντάγραμμα [Lam82]. Διαδικασίες όπως παραμόρφωση του χρόνου ή του πλάτους ή και μετάθεση που μπορούσε να εφαρμοστεί στα σκιαγραφημένα αποσπάσματα μελωδιών, παρήγαγε ενδιαφέροντα αποτελέσματα με αμφισβητήσιμη όμως εκπαιδευτική αξία.

Άλλες εργασίες βασίζονται σε διάφορες εκπαιδευτικές στρατηγικές για τη διδασκαλία βασικών θεωρητικών εννοιών και τη χρήση τους στην αναγνώριση, την εκτέλεση και τη σύνθεση αρμονιών [Sor87 Tob88].

Οι προσεγγίσεις των Έξυπνων Συστημάτων Διδασκαλίας (Intelligent Tuition Systems, ITS), βασισμένες στις θεωρίες τονικής αναγνώρισης για τη μελωδία [Nar90] ή την αρμονία [Bal80] μπορούν επίσης να βρεθούν στο MOTIVE [SH94; Smi95] και στο Harmony Space [Hol89, Hol93]. Το MOTIVE είναι ένα εργαλείο εκμάθησης για τους αρχαρίους προορισμένο να ερευνήσει τη σύνθεση από τις μελωδίες μιας εικονικής διεπαφής που βασίζεται στην παραδοσιακή σημειολογία της μουσικής. Το Harmony Space είναι ένα ιδιαίτερα διαλογικό εργαλείο για την τονική αρμονία που βασίζεται σε μια αντιπροσώπευση των αρμονικών σχέσεων σε μια δισδιάστατη μήτρα. Παρόλο που η διεπαφή δεν είναι βασισμένη στην παραδοσιακή σημειολογία της μουσικής, η αξιολόγηση του συστήματος δείχνει ότι με κάποια αρχική καθοδήγηση οι αρχάριοι θα μπορούσαν εύκολα να παραγάγουν μουσικά ενδιαφέροντα ακομπανιαμένα.

Ο [Coo93] ενθαρρύνει υψηλού επιπέδου συνθετικές δεξιότητες μέσω της μοντελοποίησης του δάσκαλου και του αρχαρίου σε δύο διαφορετικούς ρόλους. Παρουσιάζει ένα αληθοφανές γνωστικό μοντέλο για το πώς οι συνθέτες αντιλαμβάνονται την τονικότητα συνθέτοντας.

Σε πιο πρόσφατες μελέτες, χρησιμοποιήθηκε ένα πλαίσιο συμβουλευτικής γνώσης για να ερευνήσει τις αλληλεπιδράσεις δασκάλων-αρχαρίων στην μουσική σύνθεση, παρέχοντας μια ταξινόμια των παιδαγωγικών στόχων που περιλαμβάνονται.

Η μουσική σύνθεση δεν είναι ένας καθορισμένος με σαφήνεια στόχος. Ως στόχος της θα μπορούσε να καθοριστεί «το ενδιαφέρον της σύνθεσης» [Lev85]. Ως συνέπεια της απέραντης αυτής περιοχής, χρησιμοποιούνται τεχνικές που κυμαίνονται από διαλογικά παιχνίδια χωρίς οποιοδήποτε είδος καθοδήγησης σε ιδιαίτερα έξυπνα συστήματα διδασκαλίας με πολλαπλές στρατηγικές διδασκαλίας.

1.2 Βασικές Έννοιες Ακουστικής

Στο σημείο αυτό, πριν ασχοληθούμε με το κύριο αντικείμενο έρευνας της συγκεκριμένης διατριβής, παραθέτουμε μια εισαγωγή στις βασικές έννοιες που θα χρησιμοποιηθούν στη συνέχεια.

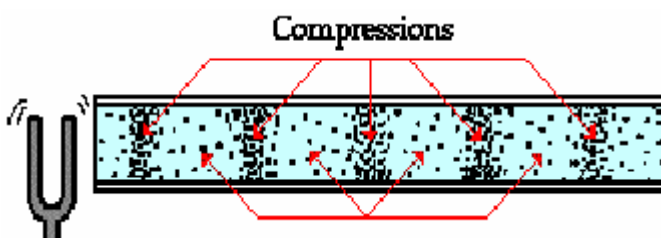
1.2.1 Ήχος

Ως **ήχος** μπορεί να οριστεί μια διάδοση κυμάτων στον αέρα ή άλλο ελαστικό μέσο(ερέθισμα) ή ως εκείνη η διέγερση του μηχανισμού ακρόασης που οδηγεί στην αντίληψη ενός ήχου (αίσθηση). Το ποιος ορισμός ισχύει εξαρτάται από το εάν η προσέγγιση είναι φυσική ή ψυχοφυσική. Ο τύπος του προβλήματος υπαγορεύει και την προσέγγιση στον ήχο. Εάν το ενδιαφέρον είναι στη διαταραχή στον αέρα που δημιουργείται από ένα μεγάφωνο, τότε η μελέτη του ήχου ανάγεται σε ένα φυσικό πρόβλημα. Εάν το ενδιαφέρον είναι πώς ηχεί σε ένα πρόσωπο πλησίον στο μεγάφωνο, τότε θα πρέπει να χρησιμοποιηθούν ψυχοφυσικές μέθοδοι.

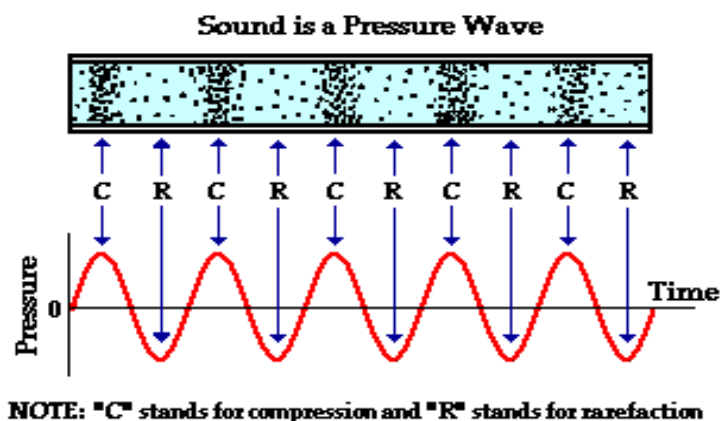
Οποιοσδήποτε ήχος, ό,τι είδους κι αν είναι, προκαλείται από μια δόνηση. Χωρίς δόνηση δεν μπορεί να υπάρξει κανένας ήχος. Το δονούμενο σώμα αναγκάζει τα μόρια αέρα γύρω του να δονηθούν και εκείνα τα μόρια αέρα, με τη σειρά τους, αναγκάζουν τα μόρια δίπλα τους να δονηθούν. Κατά αυτόν τον τρόπο μια διαταραχή του αέρα κινείται από την πηγή του ήχου και μπορεί τελικά να φθάσει στα αυτιά ενός ακροατή. Όταν ακούμε έναν ήχο, ο αέρας που φτάνει στα τύμπανά μας δονείται αναγκάζοντάς τα να δονηθούν επίσης. Αυτές οι δονήσεις ανιχνεύονται και αναλύονται από τον εγκέφαλό μας. Αν και είναι συνήθως ο αέρας που ενεργεί ως διαβιβάζον μέσο, ο ήχος μπορεί να διαβιβαστεί και από άλλα μέσα, π.χ. στερεά, υγρά.

Ο ήχος δεν αποτελείται από τον αέρα που κινείται προς εμάς, αλλά ταξιδεύει μέσω του αέρα ως ένα **ηχητικό κύμα**. Ένα ηχητικό κύμα αποτελείται από μια διαταραχή που κινείται από μια πηγή προς τις περιβάλλουσες θέσεις με αποτέλεσμα η ενέργεια να μεταφέρεται από μια θέση σε άλλη.

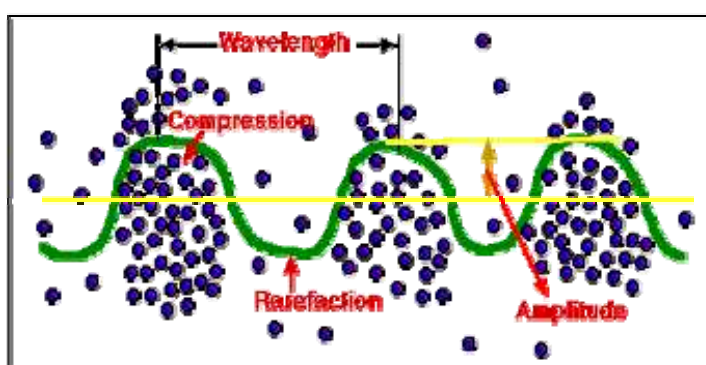
Καθώς το κύμα μεταδίδεται, η διαταραχή των μορίων είναι στην κατεύθυνση διάδοσης των κυμάτων. Η μετατόπιση των μορίων του μέσου οδηγεί σε περιοχές με εναλλασσόμενη υψηλή πυκνότητα μορίων και χαμηλή πυκνότητα μορίων. Οι περιοχές υψηλής πυκνότητας μορίων καλούνται **πυκνώσεις (compressions)**. Οι περιοχές χαμηλής πυκνότητας μορίων καλούνται **αραιώσεις (rarefactions)**.



Εικόνα 2 : Αραιώσεις και πυκνώσεις κατά τη διάδοση ενός κύματος



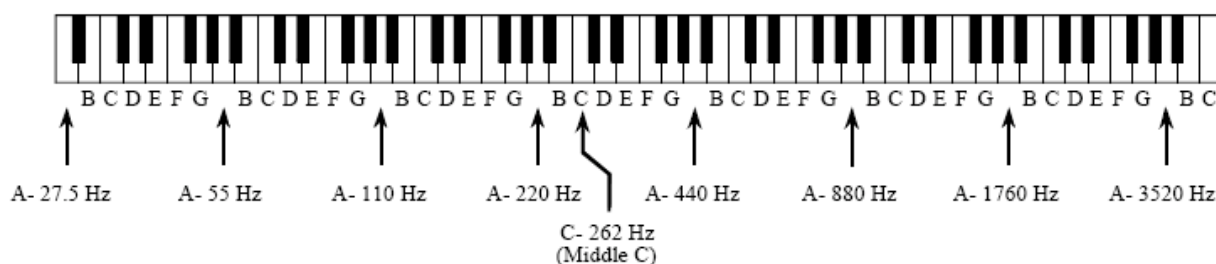
Εικόνα 3: Ο ήχος ως κύμα πίεσης



Εικόνα 4: Μήκος κύματος

Οι αραιώσεις και οι πυκνώσεις κινούνται στην κατεύθυνση διάδοσης των κυμάτων. Τα μόρια του μέσου *δεν κινούνται* ομαδικά στην κατεύθυνση της μετακίνησης κυμάτων, δονούνται γύρω από τις κανονικές θέσεις τους. Κάθε πλήρης δόνηση ενός μορίου καλείται **κύκλος** (δηλ. από την αρχική θέση του, σε μια μέγιστη απόσταση σε μια κατεύθυνση, πίσω στην αρχική θέση, έπειτα σε μια μέγιστη μετατόπιση στην αντίθετη κατεύθυνση και πίσω στην αρχική θέση).

Ο αριθμός κύκλων που ολοκληρώνονται σε ένα δευτερόλεπτο καλείται **συχνότητα** της δόνησης. Το φάσμα της ανθρώπινης ακοής εκτείνεται προσεγγιστικά από τα 16Hz έως τα 16kHz. Αυτό το φάσμα συχνοτήτων έχει εύρος ίσο με περίπου 10 οκτάβες από τον χαμηλότερο αντιληπτό τόνο ως τον υψηλότερο. Οι συχνότητες των νοτών που μπορούν να παιχτούν σε ένα πιάνο κυμαίνονται από 27.5 Hz και φτάνουν πάνω από τα 4kHz.



Εικόνα 5 : Συχνότητες νοτών στο πιάνο

Οποιαδήποτε νότα παίζεται σε ένα πιάνο θα ηχήσει διαφορετική σε μια νότα του ίδιου τόνου που παράγεται από έναν άλλο τύπο οργάνου.

Η μουσική νότα που παράγεται με έναν συντονισμό καλείται **καθαρός τόνος (pure tone)** επειδή αποτελείται από έναν τόνο μιας συχνότητας. Μια νότα που παίζεται σε ένα πιάνο, ή στα περισσότερα άλλα όργανα, αποτελείται από διάφορους τέτοιους τόνους που ηχούν μαζί σε διαφορετικές συχνότητες. Αυτές οι συχνότητες συσχετίζονται με τη συχνότητα που δίνει στη νότα το χαρακτηριστικό τόνο της, η οποία είναι συνήθως η χαμηλότερη.

Οι κυματομορφές μουσικής ή ομιλίας απέχουν πολύ από τη μορφή του απλού ημιτόνου. Είναι όμως πολύ ενδιαφέρον το γεγονός ότι όσο πολύπλοκο κι αν είναι ένα κύμα, αν είναι περιοδικό, μπορεί να περιγραφεί ως άθροισμα ημιτονικών όρων διαφορετικών συχνοτήτων, φάσεων και πλάτους.

Ο ημιτονικός όρος με τη χαμηλότερη συχνότητα καλείται **θεμελιώδης (fundamental)**. Οι άλλοι όροι καλούνται **partials** (μερικές). Εάν οι αρμονικοί ήχοι έχουν συχνότητες που είναι ακέραια πολλαπλάσια ($x_2, x_3 \dots$ μέχρι x_{14}) της θεμελιώδους συχνότητας καλούνται **αρμονικές (harmonics)**. Οι μουσικοί συχνά χρησιμοποιούν τον όρο «μερική» αντί του όρου «αρμονική», αλλά θα πρέπει σ' αυτό το σημείο να γίνει μία διάκριση μεταξύ των όρων γιατί οι μερικές πολλών οργάνων δεν είναι αρμονικά συσχετιζόμενες με την θεμελιώδη συχνότητα. Δηλαδή μπορεί να μην είναι ακέραια πολλαπλάσια της θεμελιώδους συχνότητας, όπως για παράδειγμα οι μερικές του πιάνου.

Επομένως, για την περίπτωση μονοφωνικών ήχων, το σήμα εισόδου περιλαμβάνει μία μοναδική αρμονική δομή, κάποια μη αρμονικά συστατικά και μία επιπρόσθετη ποσότητα ανεπιθύμητου θορύβου. Άρα, το σήμα μπορεί προσεγγιστικά να περιγραφεί ως:

$$x(t) = \sum kA_k(t) \sin(k\omega t + \phi_k) + e(t)$$

όπου:

$x(t)$ είναι το σήμα εισόδου στο πεδίο του χρόνου

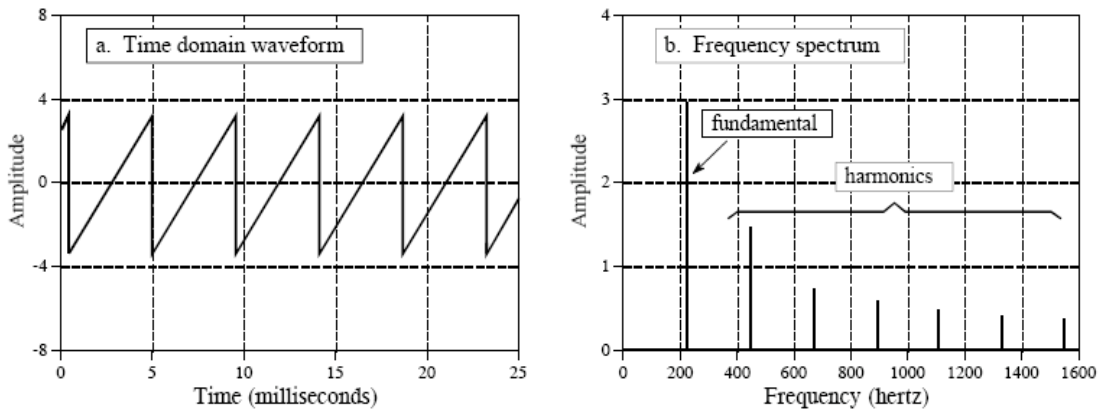
ω είναι η θεμελιώδης συχνότητα της τρέχουσας νότας

$A_k(t)$ είναι το πλάτος της k-αρμονικής τη χρονική στιγμή t

ϕ_k είναι η φάση της k-αρμονικής

$e(t)$ είναι οι λοιποί όροι (μη αρμονικά στοιχεία και θόρυβος)

Είναι η διαφορά στο αρμονικό περιεχόμενο των νοτών που δίνει σε κάθε μουσικό όργανο το χαρακτηριστικό ήχο του ή **χρoιά (timbre)**. Επομένως αν και η υψηλότερη νότα ενός πιάνου έχει μια θεμελιώδη συχνότητα μεγαλύτερη από 4kHz, ο εξοπλισμός που χρησιμοποιείται για να καταγράψει τη μουσική πρέπει να είναι σε θέση να χειριστεί πολύ υψηλότερες συχνότητες για να διατηρήσει τις αρμονικές που συνδέονται με κάθε νότα.



Εικόνα 6:Κυματομορφή στο πεδίο του χρόνου και στο πεδίο της συχνότητας

Οι ήχοι που παράγονται από κρουστικά αποτελέσματα είναι ιδιαίτερα πλούσιοι σε υψηλές αρμονικές. Αυτοί εμφανίζονται κυρίως στην έναρξη ενός ήχου, π.χ. όταν ακουμπάμε μια χορδή ή όταν χτυπάμε ένα κύμβαλο. Ο εξοπλισμός ηχογράφησης πρέπει να είναι σε θέση να αντιμετωπίσει αυτές τις υψηλές συχνότητες ειδάλλως η τονική ποιότητα των ήχων θα αλλάξει. Τα κύμβαλα, παραδείγματος χάριν, μπορούν να παραγάγουν τις συχνότητες μεταξύ 20kHz και 25kHz.

Υπάρχουν δύο τύποι ήχων: ήχοι καθορισμένου τόνου, την οποία οι μουσικοί καλούν νότα και ήχος χωρίς καθορισμένο τόνο, ο θόρυβος. «Η μουσική» περιλαμβάνει όχι μόνο τις νότες (καθορισμένου τόνου) αλλά επίσης και πολύ θόρυβο, π.χ. κρουστά. Ο θόρυβος είναι ένα αναπόσπαστο τμήμα της μουσικής. Η διαφορά μεταξύ μιας νότας και ενός θορύβου είναι προφανής στο αυτί αλλά τι προκαλεί τη διαφορά; Μια νότα αποτελείται από τις περιοδικές δονήσεις ενώ ένας θόρυβος αποτελείται από μη περιοδικές δονήσεις.

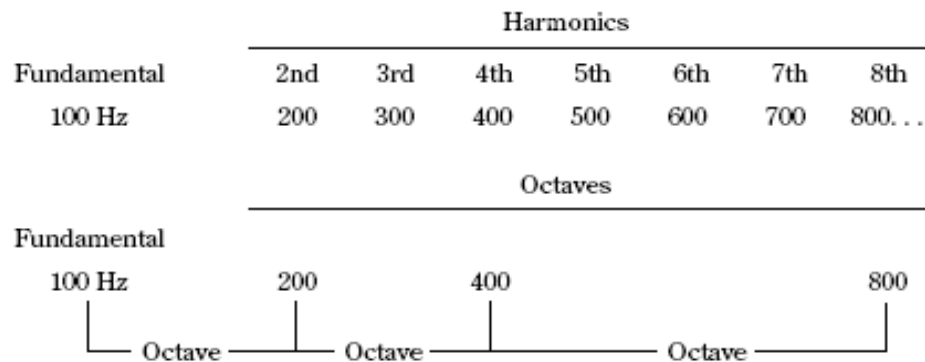
Περιοδική κίνηση είναι η μετακίνηση που είναι επαναλαμβανόμενη. Μια μουσική νότα παράγεται από την περιοδική κίνηση σε μια ευδιάκριτη συχνότητα. Η μη περιοδική κίνηση θεωρείται ως θόρυβος.

Ως **θόρυβος** μπορεί να οριστεί ένας ανεπιθύμητος ήχος. Ένας μηχανικός ήχος ενδιαφέρεται για δύο είδη του θορύβου:

1. **συστημικός θόρυβος**, εισαγόμενος από τα μικρόφωνα, τους ενισχυτές, τα όργανα καταγραφής ταινιών, κ.λπ.
2. **περιβαλλοντικός θόρυβος** σε μια αίθουσα, ένα στούντιο, ένα δωμάτιο, κ.λπ.

Ένας μηχανικός ήχος ενδιαφέρεται επίσης για την αντήχηση. Όλα τα αντικείμενα που μπορούν να δονηθούν έχουν μια ορισμένη συχνότητα στην οποία θα δονηθούν εντονότερα (δηλ. με το μέγιστο εύρος). Εάν ένα σώμα διεγείρεται σε ένα ολόκληρο φάσμα συχνοτήτων θα δονηθεί περίπου εξίσου σε όλες τις συχνότητες εκτός από εκείνες τις συχνότητες που είναι πλησιέστερα στη φυσική συχνότητά του. Σε μια συχνότητα θα δονηθεί εντονότερα. Αυτή η συχνότητα καλείται **συχνότητα συντονισμού** και ο όρος καλείται **αντήχηση**. Μια χορδή ενός μουσικού οργάνου θα δονηθεί σε μια συχνότητα, την συχνότητα συντονισμού της. Τα μόρια αέρα έχουν μάζα έτσι οποιοδήποτε εσωκλειόμενο σώμα του αέρα (π.χ. ο σωλήνας ή ο «σωλήνας» ενός κλαρινέτου) έχει μια συχνότητα συντονισμού. Ο αέρας που περιλαμβάνεται σε μια αίθουσα ή ένα στούντιο μπορεί επίσης να έχει την αντήχηση, με συνέπεια ορισμένοι ήχοι να ενισχύονται αδικαιολόγητα.

Στη μουσική, η σειρά των ήχων διαιρείται σε τμήματα γνωστά ως **οκτάβες**. Η οκτάβα είναι μία λογαριθμική σύλληψη που είναι στενά συνδεδεμένη με τις μουσικές κλίμακες και ορολογίες, εξαιτίας της σχέσης της με τα χαρακτηριστικά της ανθρώπινης ακοής.



Εικόνα 7: Σύγκριση αρμονικών και οκτάβων. Οι αρμονικές είναι γραμμικά συσχετισμένες ενώ οι οκτάβες λογαριθμικά.

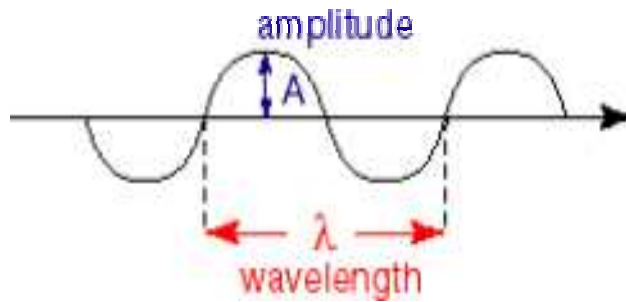
Κάθε οκτάβα αποτελείται από επτά διαφορετικές νότες και οκτώ διαστήματα μεταξύ των νοτών. Οι επτά νότες ονομάζονται A-G (Λα-Σολ) και η όγδοη νότα είναι το A της επόμενης οκτάβας. Κάθε νότα ηχεί πολύ παρόμοια με τη νότα που έχει το ίδιο όνομα σε μια άλλη οκτάβα. Ο λόγος για αυτήν την «ομοιότητα» είναι ότι μια νότα που μια οκτάβα υψηλότερη από μια άλλη με το ίδιο «όνομα» έχει δύο φορές τη συχνότητα της (π.χ. ένα Λα έχει συχνότητα 440Hz ενώ το αμέσως υψηλότερο έχει συχνότητα 880Hz). Έτσι, εάν ένα δεδομένο σύστημα ταλαντεύεται δύο φορές στη συχνότητα ενός άλλου συστήματος, οι δύο νότες που παράγονται διαφέρουν στον τόνο κατά μια οκτάβα.

Στο σημείο αυτό, στο πλαίσιο των ημιτονικών σημάτων, των αρμονικών κ.τ.λ. πρέπει να ορίσουμε την έννοια του **φάσματος (spectrum)**. Αντίστοιχα με το ορατό φάσμα του φωτός, υπάρχει και το φάσμα του ήχου, το σύνολο των συχνοτήτων που γίνονται αντιληπτά από το ανθρώπινο αυτί. Δεν μπορούμε να αντιληφθούμε την υπεριώδη ακτινοβολία, γιατί η συχνότητα της ηλεκτρομαγνητικής του ενέργειας είναι πέρα από τα όρια που μπορεί να αντιληφθεί το ανθρώπινο μάτι. Επίσης, δεν μπορούμε να δούμε την υπέρυθρη γιατί η συχνότητά του είναι πολύ χαμηλή. Αντίστοιχα, υπάρχουν ήχοι πολύ χαμηλής και πολύ υψηλής συχνότητας που το ανθρώπινο αυτί δεν μπορεί να συλλάβει. Το φάσμα μας δείχνει πώς η ενέργεια του σήματος διαμοιράζεται στις διάφορες συχνότητες.

Παρατηρώντας τα φάσματα ημιτονικών σημάτων, τετραγωνικών και τριγωνικών παλμών, παρατηρούμε συγκέντρωση της ενέργειας στις αρμονικές συχνότητες και τίποτα στο ενδιάμεσο αυτών των συχνοτήτων. Τα παραπάνω σήματα είναι χαρακτηριστικά περιοδικών σημάτων. Στην περίπτωση του τυχαίου θορύβου η ενέργεια είναι κατανεμημένη σε όλες τις συχνότητες. Ακόμα όμως και ο θόρυβος μπορεί να θεωρηθεί ως άθροισμα ημιτονικών σημάτων με συνεχώς μεταβαλλόμενη συχνότητα, πλάτος και φάση.

Τα μόρια που δονούνται λόγω της μετάβασης ενός κύματος θεωρούνται **σε φάση** εάν κινούνται προς την ίδια κατεύθυνση και έχουν την ίδια μετατόπιση (δηλ. είναι οι ίδιες αποστάσεις από τις αρχικές θέσεις τους).

Το μέτωπο του κύματος είναι μια επιφάνεια στην οποία όλα τα μόρια είναι στην ίδια φάση της δόνησης. Η απόσταση μεταξύ των διαδοχικών κυματομορφών καλείται **μήκος κύματος** του ήχου. Το μήκος κύματος αναπαριστάμεται με το ελληνικό σύμβολο *λάμδα*. Σε επίπεδο πυκνώσεων και αραιώσεων το μήκος κύματος είναι η απόσταση μεταξύ των παρακείμενων κέντρων της πύκνωσης ή των παρακείμενων κέντρων αραιώσης.



Εικόνα 8: Μήκος κύματος και πλάτος κυματομορφής.

Ο ήχος διαδίδεται με σταθερή ταχύτητα (αποκαλούμενη **ταχύτητα του ήχου**) εάν η θερμοκρασία και η πίεση του αέρα είναι σταθερές. Αυτή είναι περίπου 330 μέτρα ανά δευτερόλεπτο. Το μήκος κύματος και η συχνότητα συσχετίζονται με την ταχύτητα του ήχου με την παρακάτω σχέση:

$$v=f * \lambda$$

Υπάρχουν τρία κύρια χαρακτηριστικά του ήχου:

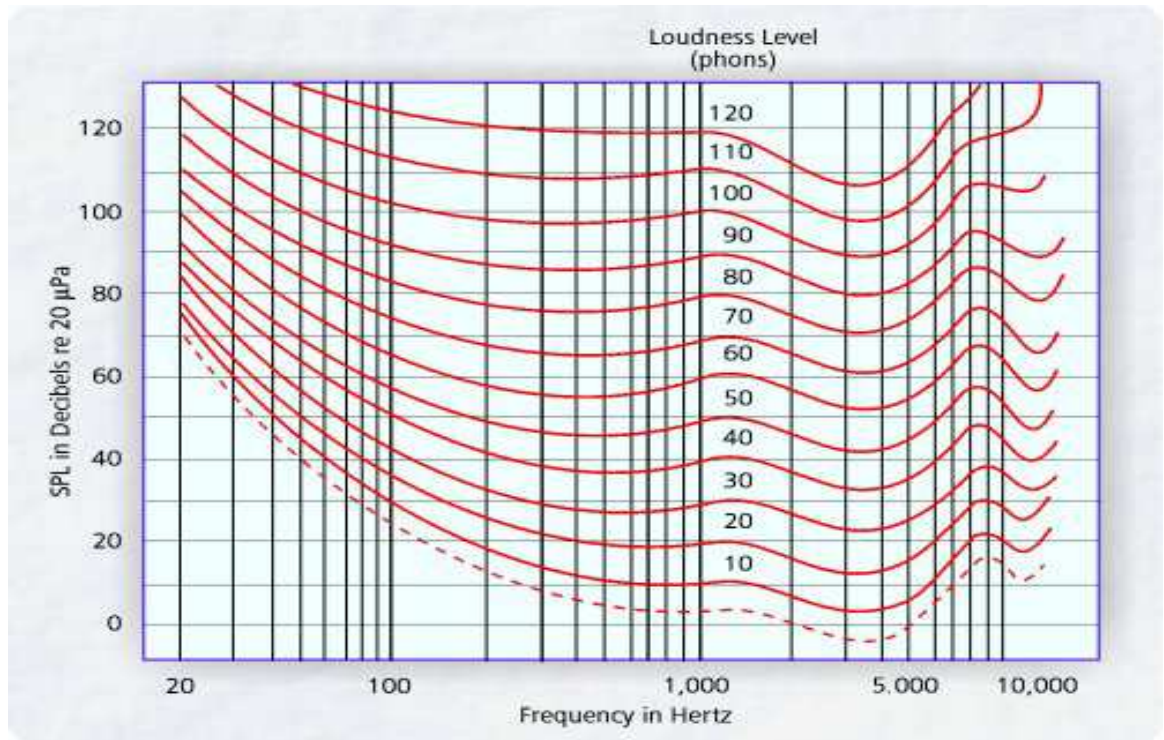
1. Ηχηρότητα που συσχετίζεται με το πλάτος
2. Τόνος που συσχετίζεται με τη συχνότητα
3. Χροιά που συσχετίζεται με το φάσμα.

1.2.2 Ηχηρότητα (Loudness)

Ένας δυνατός ήχος (νότα ή θόρυβος) παράγεται από δονήσεις ισχυρότερες από εκείνες που παράγουν έναν λιγότερο δυνατό ήχο. **Εύρος** ενός κύματος είναι η μέγιστη μετατόπιση των δονούμενων μορίων από τις θέσεις ηρεμίας τους.

Για να παραχθεί δόνηση, πρέπει να καταναλωθεί ενέργεια. Επομένως, για να παράγουμε ένα πιο δυνατό ήχο περισσότερη ενέργεια πρέπει να δαπανηθεί. Το ποσό της ενέργειας για να παραχθεί ένας ήχος ονομάζεται **ισχύς** του ήχου και μετριέται σε **Watt**. Ενώ ο ήχος εκπέμπεται από μια πηγή, η συγκέντρωση ισχύος γίνεται όλο και μικρότερη όσο η απόσταση από την πηγή μεγαλώνει αφού το ίδιο ποσό ισχύος απλώνεται σε μεγαλύτερη επιφάνεια. Το ποσό ισχύος ανά τετραγωνικό μέτρο ονομάζεται **ένταση του ήχου (intensity)**.

Οι άνθρωποι δεν αντιλαμβάνονται την ένταση του ήχου γραμμικά. Για έναν ήχο που γίνεται αντιληπτός ως δύο φορές πιο δυνατός, η ένταση πρέπει να είναι δέκα φορές μεγαλύτερη. Εξαιτίας αυτού, το αντιληπτό επίπεδο έντασης ενός ήχου μετριέται σε λογαριθμική κλίμακα χρησιμοποιώντας τη μονάδα του **decibel** (ή **dB**). Το κατώτατο όριο της ακρόασης ορίζεται ως 0 dB. Ένας ήχος που είναι τόσο δυνατός που προκαλεί πραγματικά τον πόνο - το «κατώτατο όριο του πόνου» - είναι περίπου 120 dB. Με τη λογαριθμική κλίμακα στο μυαλό, το κατώτατο όριο του πόνου (120 dB) αντιπροσωπεύει μια ένταση 1.000.000.000.000 (1 τρισεκατομμύριο) φορές μεγαλύτερες από το κατώτατο όριο της ακρόασης (0 dB)!



Εικόνα 9 : Καμπύλες ίσης ηχηρότητας

	Watts/cm ²	Decibels SPL	Example sound
	10 ⁻²	140 dB	Pain
	10 ⁻³	130 dB	
	10 ⁻⁴	120 dB	Discomfort
	10 ⁻⁵	110 dB	Jack hammers and rock concerts
	10 ⁻⁶	100 dB	
	10 ⁻⁷	90 dB	OSHA limit for industrial noise
	10 ⁻⁸	80 dB	
	10 ⁻⁹	70 dB	
	10 ⁻¹⁰	60 dB	Normal conversation
	10 ⁻¹¹	50 dB	
	10 ⁻¹²	40 dB	Weakest audible at 100 hertz
	10 ⁻¹³	30 dB	
	10 ⁻¹⁴	20 dB	Weakest audible at 10kHz
	10 ⁻¹⁵	10 dB	
	10 ⁻¹⁶	0 dB	Weakest audible at 3 kHz
	10 ⁻¹⁷	-10 dB	
	10 ⁻¹⁸	-20 dB	

Πίνακας 1: Ερμηνεύοντας τους αριθμούς

Οι ακόλουθες «εμπειροτεχνικές μέθοδοι» για τη ηχηρότητα αποδεικνύονται χρήσιμες:

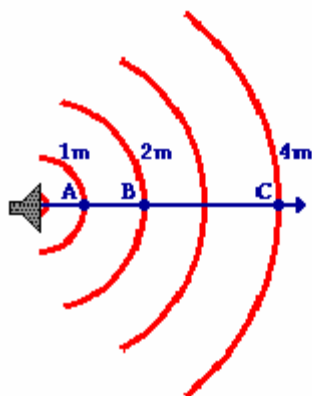
- Για να διπλασιαστεί η αντιληπτή ηχηρότητα ενός ήχου, η ισχύς (σε Watt) πρέπει να αυξηθεί δέκα φορές και το επίπεδο πίεσης του ήχου θα αυξηθεί κατά 10 dB. Παράδειγμα: Ένα επίπεδο πίεσης (SPL) 90 dB που παράγονται με ισχύ 10 Watt θα διπλασιαστεί σε αντιληπτή ηχηρότητα όταν αυξάνεται η δύναμη σε 100 Watt και η SPL αυξηθεί σε 100 dB.

- Ο διπλασιασμός της ισχύος ενός ήχου πάνει ένα κέρδος 3 dB στην αντιληπτή ηχηρότητα. Παράδειγμα: Ένα επίπεδο πίεσης (SPL) 90 dB παράγεται με ισχύ 10 Watt. Εάν η ισχύς διπλασιαστεί στα 20 Watt, το SPL θα αυξηθεί σε 93 dB.
- Κινούμενοι από μία θέση ως προς μια δεδομένη ηχητική πηγή προς μια άλλη θέση που βρίσκεται σε διπλάσια απόσταση από την πηγή, η ηχηρότητα θα μειωθεί στο $\frac{1}{4}$ (δηλ., μια απώλεια 6 dB). Παράδειγμα: Σε μια απόσταση 20 ποδιών από μια πηγή, ένα επίπεδο πίεσης (SPL) 90 dB παράγεται. Ο διπλασιασμός της απόστασης από την πηγή σε 40 πόδια θα οδηγήσει σε ένα SPL 84 dB.

Όσο μεγαλύτερο το εύρος ενός κύματος, τόσο μεγαλύτερη θα είναι η ενέργεια των δονούμενων μορίων και ο ήχος θα είναι περισσότεροι **έντονος (intensity)**. Καθώς ένα ηχητικό κύμα ταξιδεύει έξω από την πηγή, η ενέργεια μεταφέρεται από ένα δονούμενο μόριο στο επόμενο. Η ενέργεια απελευθερώνεται βαθμιαία υπό μορφή θερμότητας έτσι η ένταση του ήχου μειώνεται όσο η απόσταση από την πηγή αυξάνεται.

Η ένταση του ήχου λέγεται ότι είναι αντιστρόφως ανάλογη προς την απόσταση από την πηγή. Η μείωση υπακούει έναν αντίστροφο νόμο (ο αντίστροφος τετραγωνικός νόμος).

Όταν ένας ακροατής απομακρύνεται από μια ηχητική πηγή το ηχητικό επίπεδο δεν εμφανίζεται να μειώνεται υπό τέτοιες αναλογίες. Αυτό συμβαίνει επειδή το **αυτί** διαθέτει ένα μηχανισμό αντιστάθμισης. Το τύμπανο του αυτιού συνδέεται με το εσωτερικό αυτί με ένα σύστημα τριών ακουστικών οσταρίων που παίζουν τον ρόλο μοχλών. Οι θέσεις των αξόνων αυτών των μοχλών μπορούν να αλλάξουν για να παρέχουν μεγαλύτερη ή μικρότερη δύναμη. Η ρύθμιση αυτή επιτρέπει στους αδύνατους ήχους να φθάσουν στο εσωτερικό αυτί με τη μέγιστη δύναμη ενώ οι δυνατοί ήχοι μειώνονται για να αποτρέψουν πιθανή ζημία στο εσωτερικό αυτί. Ένας ακροατής που κινείται δύο φορές μακρύτερα από ένα μεγάφωνο δεν θα αντιλαμβανόταν μια μείωση στην ηχηρότητα στο ένα τέταρτο. Εντούτοις η (μετρήσιμη) ένταση του ήχου θα είχε μειωθεί σε εκείνη την έκταση (Εικόνα 10).



Εικόνα 10

Ηχηρότητα (loudness) κατόπιν, είναι το μέγεθος της *αίσθησης* από κάποιον που ακούει έναν ήχο. Η ένταση είναι μια μετρήσιμη, φυσική ποσότητα. Η ηχηρότητα εξαρτάται όχι μόνο από την ένταση ενός ήχου αλλά και από την ευαισθησία των αυτιών του ακροατή.

Η συσσώρευση των μορίων προκαλεί την πίεση του αέρα να είναι μεγαλύτερη από ότι κανονικά. Η πίεση είναι επομένως μέγιστη στις περιοχές της πύκνωσης και ελάχιστη στις περιοχές αραιώσης.

Εάν δύο ίδια ηχητικά κύματα φθάσουν στην ίδια θέση συγχρόνως οι περιοχές υψηλής πίεσης θα συμπέσουν και το ίδιο και οι περιοχές χαμηλής πίεσης. Αυτό θα οδηγήσει σε ένα ενιαίο κύμα

μεγαλύτερης έντασης. Τα δύο κύματα θα ενισχύσουν το ένα το άλλο και το φαινόμενο (αισθητό γεγονός) καλείται **ενίσχυση**.

Εάν ένα κύμα φθάσει μισό μήκος κύματος αργότερα, η περιοχή υψηλής πίεσης του ενός κύματος θα ακυρωθεί από την περιοχή χαμηλής πίεσης του άλλου. Σαν αποτέλεσμα δεν ακούγεται κανένας ήχος. Αυτό καλείται **ακύρωση**.

Όταν εμφανίζεται ενίσχυση, τα κύματα θεωρούνται σε φάση το ένα με το άλλο. Σε οποιαδήποτε άλλη περίπτωση τα κύματα είναι εκτός φάσης και εμφανίζεται μερική ή πλήρης ακύρωση.

1.2.3 Μουσικό ύψος του ήχου (*Pitch*)

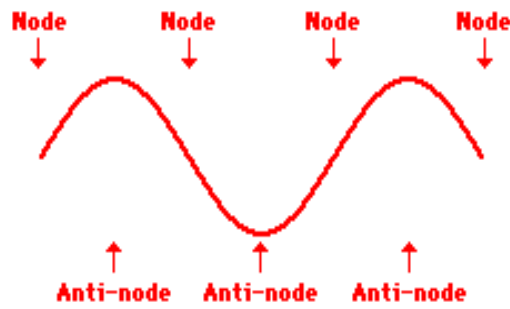
Η συχνότητα ενός ήχου γίνεται αντιληπτή ως **ύψος (pitch)**. Δεν μπορούμε να ισχυριστούμε ότι συχνότητα και το ύψος είναι το ίδιο. Συσχετίζονται, αλλά όχι με γραμμικό τρόπο. Το ύψος είναι ένας υποκειμενικός όρος, ο οποίος εξαρτάται εκτός από την συχνότητα και από το επίπεδο της πίεσης του ήχου. Η μονάδα μέτρησης του είναι το mel, ένα ύψος αναφοράς 1000 mels ορίζεται ως το ύψος ενός ήχου στα 1000 Hz με πίεση 60 dB. Οι ήχοι με υψηλή συχνότητα γίνονται αντιληπτοί ως ήχοι με μεγάλο ύψος ενώ τους ήχους με λίγους κύκλους ανά δευτερόλεπτο τους αντιλαμβανόμαστε ως ήχους με χαμηλό ύψος. Όταν η συχνότητα ενός ήχου διπλασιάζεται την αντιλαμβανόμαστε ως διπλασιασμό του ύψους.

Η συχνότητα μετριέται στα Χερτζ. Ένα **Χερτζ (Hz)** αντιστοιχεί σε ένα κύκλο ανά δευτερόλεπτο. Μια δόνηση υψηλής συχνότητας παράγει μια νότα με υψηλό τόνο, ενώ μια χαμηλής συχνότητας δόνηση δίνει μια νότα με χαμηλό τόνο.

1.2.4 Χροιά ή Ηχώχρωμα (*Timbre or Tone colour*)

Η ιδιότητα του ήχου που επιτρέπει σε μας να καθορίσουμε τη διαφορά μεταξύ ενός σαξόφωνου και ενός φλάουτου καλείται **χροιά ή ηχώχρωμα**. Οι διαφορετικές χροιές εμφανίζονται επειδή το αντιληπτό ύψος των περισσότερων ήχων περιέχει στην πραγματικότητα πολλές συχνότητες. Κατά συνέπεια, όταν ακούει κάποιος τη νότα «μέσο C» για το πιάνο, πολλές άλλες συχνότητες είναι επίσης παρούσες. Το κυρίαρχο ύψος (σε αυτήν την περίπτωση το μέσο C), καλείται **θεμελιώδης συχνότητα** του ήχου. Οι άλλες παρούσες συχνότητες εμφανίζονται σε μια μαθηματική σειρά αποκαλούμενη **αρμονική σειρά (harmonic series)** ή **σειρά αρμονικού ήχου (overtone series)**. Η συχνότητα κάθε αρμονικής είναι ακέραιο πολλαπλάσιο της θεμελιώδους συχνότητας (δηλ., 2X, 3X, 4X, κ.λπ.).

Η χροιά είναι άλλος ένας υποκειμενικός όρος. Ο ανάλογος φυσικός όρος είναι το φάσμα. Ακούγοντας μία ορχήστρα σε μία αίθουσα συναυλιών η χροιά που ακούμε είναι διαφορετική σε διαφορετικές θέσεις μέσα στο χώρο. Η μουσική αποτελείται από ένα ευρύ σύνολο συχνοτήτων, το πλάτος και η φάση των οποίων επηρεάζονται από ανακλάσεις στις διάφορες επιφάνειες του δωματίου. Ένας τρόπος ανάλυσης θα ήταν να λάβουμε το φάσμα του ήχου από διαφορετικές θέσεις. Παρόλ' αυτά, αυτές θα ήταν φυσικές μετρήσεις και η υποκειμενική χροιά ακόμα θα μας διέφευγε. Το σημαντικό είναι να συνειδητοποιήσουμε τη διαφορά μεταξύ χροιάς και φάσματος. Η χροιά καθορίζεται από το πλήθος και τη σχετική δύναμη των μερικών του οργάνου και είναι συνώνυμη με την ποιότητα του ήχου. Ενώ οι περισσότεροι μουσικοί ήχοι έχουν τους αρμονικούς ήχους που είναι ακέραια πολλαπλάσια της θεμελιώδους συχνότητας, με μερικούς μουσικούς ήχους δεν συμβαίνει το ίδιο.



Εικόνα 11 : Δεσμοί και Αντί-δεσμοί

Harmonic	# of Nodes	# of Antinodes	Pattern
1st	2	1	
2nd	3	2	
3rd	4	3	
4th	5	4	
5th	6	5	
6th	7	6	
nth	$n + 1$	n	--

Πίνακας 2 : Αρμονικές, δεσμοί και αντί-δεσμοί

Ο αριθμός των αρμονικών που είναι παρούσες και η σχετική δύναμη της καθεμιάς καθορίζουν τη χροιά του ήχου. Πολλοί ήχοι έχουν χροίες που μεταβάλλονται χρονικά. Σε τέτοιες περιπτώσεις, η σχετική ισχύς των αρμονικών ποικίλλει καθώς ο ήχος παράγεται. Για παράδειγμα, μια χαμηλή νότα που παίζεται σταθερά σε ένα πιάνο ακούγεται αρχικά «φωτεινότερη», αλλά γίνεται «πιο θαμπή» καθώς ο ήχος «σβήνει». Σε αυτή την περίπτωση, οι υψηλότερες αρμονικές χάνονται γρηγορότερα από τις χαμηλότερες αρμονικές.

Διαφορετικές χροίες έχουν σαν αποτέλεσμα διαφορετικές μορφές στο κυματοειδές. Ορισμένα κυματοειδή (και συνεπώς ορισμένες χροίες) έχουν χρησιμοποιηθεί από καιρό ως δομικές μονάδες της σύνθεσης. Αυτοί περιλαμβάνουν το κύμα ημιτόνου, το τετραγωνικό κύμα, το κύμα τριγώνων, και το πριονωτό κύμα.

1.2.5 Συμπέρασμα

Αν και ανεξάρτητα, τα τρία κύρια χαρακτηριστικά του ήχου ως σύνολο δίνουν σε κάθε ήχο που ακούμε τις μοναδικές του ιδιότητες. Μια σαφής κατανόηση του ρόλου κάθε συστατικού είναι απαραίτητη για την έρευνα στη σφαίρα της Μουσικής Τεχνολογίας.

1.3 Η Μουσική Κλίμακα

Η μουσική κλίμακα είναι μία λογαριθμική κλίμακα οργάνωσης του ύψους, βασισμένη στην οκτάβα, που είναι η αντιληπτή απόσταση μεταξύ 2 υψών, όπου το ένα είναι διπλάσιο σε συχνότητα του άλλου. Π.χ. η νότα C4 έχει συχνότητα 261.6 Hz, στην επόμενη οκτάβα η ίδια νότα έχει συχνότητα 523,2Hz κ.ο.κ.

Αν και η οκτάβα φαίνεται να είναι αντιληπτή βαθμίδα από τους ανθρώπους, η οργάνωση του ύψους μέσα στην οκτάβα διαφέρει ανάλογα με τους πολιτισμούς. Στη Δυτική μουσική, η πρωταρχική οργάνωση από τα χρόνια του Bach ήταν μια ισοσταθμισμένη κλίμακα, που χωρίζει την κλίμακα σε 12 ίσα ημιτόνια. Το διάστημα μεταξύ 2 οκτάβων αντιστοιχεί σε διπλασιασμό της συχνότητας και τα ημιτόνια είναι σε ίση απόσταση με μια πολλαπλασιαστική έννοια, έτσι ώστε ανεβαίνοντας ένα ημιτόνιο πολλαπλασιάζει την συχνότητα με τη δωδέκατη ρίζα του 2, ίσο περίπου με 1.059.

Το ημιτόνιο είναι η μικρότερη μονάδα του ύψους στη Δυτική μουσική, αλλά μικρότερες μονάδες μπορούν να γίνουν εύκολα αντιληπτές και χρησιμοποιούνται στη μουσική δομή σε αρκετούς πολιτισμούς. Φυσικοί και ψυχολόγοι έχουν συμφωνήσει στην λογαριθμική μονάδα του εκατοστού, καθορισμένου ως το ένα εκατοστό του ημιτόνιου σε μια ισοσταθμισμένη κλίμακα. Η οκτάβα είναι συνεπώς 1200 εκατοστά. Η μικρότερη διαφορά ύψους μεταξύ δύο διαδοχικών τόνων που μπορεί να γίνει αντιληπτή από το ανθρώπινο αυτί είναι 3Hz, από αυτό συνεπάγεται μία διαφορά στο ύψος περίπου 5 εκατοστών στα 1000Hz. Πάνω από τα 1000 Hz, η διάκριση σταθεροποιείται στα 4 εκατοστά.

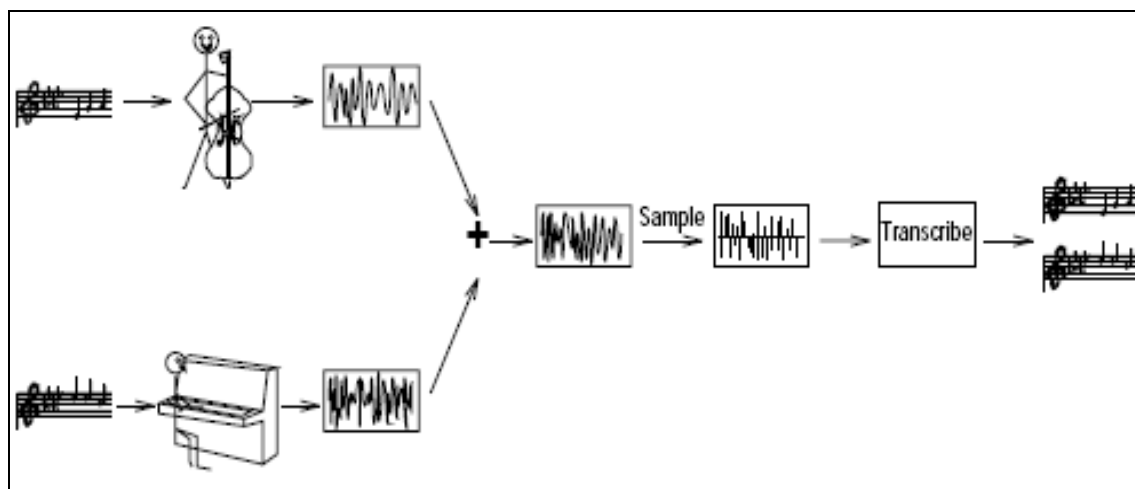
1.4 Μουσική Καταγραφή Ηχητικών Δεδομένων

Η μουσική παραδοσιακά παρουσιάζεται στο γραπτό λόγο με τη μορφή παρτιτούρας, όπου φαίνονται οι χρόνοι, οι διάρκειες και τα ύψη των ήχων που απαρτίζουν το μουσικό κομμάτι. Η μουσική καταγραφή των ηχητικών δεδομένων είναι η διαδικασία κατά την οποία από μια ακολουθία ψηφιακών δεδομένων, που αντιστοιχούν στην κυματομορφή του ήχου, εξάγονται συμβολικές πληροφορίες σχετικές με υψηλού επιπέδου μουσικές δομές που φαίνονται σε μία παρτιτούρα. Με απλά λόγια η μουσική μεταγραφή είναι η μετάφραση της μουσικής σε παρτιτούρα. Όλοι οι ήχοι που υιοθετούνται στη μουσική μπορούν να περιγραφούν από τέσσερις φυσικές παραμέτρους και τις αντίστοιχες αντιληπτές παραμέτρους τους:

1. Ρυθμός επαναληπτικότητας ή θεμελιώδης συχνότητα του ηχητικού κύματος, που συσχετίζεται με το **ύψος**.
2. Πλάτος ηχητικών κυμάτων, που συσχετίζεται με την **ηχηρότητα**.
3. Μορφή ηχητικών κυμάτων, που συσχετίζεται με τη **χροιά**.
4. **Θέση ηχητικής πηγής** όσον αφορά τον ακροατή, που συσχετίζει με τη χωρική αντίληψη του ακροατή.

Η τελευταία παράμετρος δεν θεωρείται καθοριστική για τη καταγραφή της μουσικής, και δε θα μας απασχολήσει στη συγκεκριμένη έρευνα. Οι άλλοι τρεις παράγοντες δημιουργούν τη διαφορά μεταξύ των μερών που μπορούν να καθοριστούν σε ένα μουσικό κομμάτι: η ορχήστρα και η παρτιτούρα. Η ορχήστρα είναι ο ήχος του ίδιου του οργάνου, τα συγκεκριμένα χαρακτηριστικά των οργάνων (χροιά, περιβάλλουσα), τα οποία κάνουν τον ήχο μοναδικό: η παρτιτούρα αποτελείται από τις γενικές παραμέτρους ελέγχου (ύψος, onsets, κ.λπ.), οι οποίες καθορίζουν τη μουσική που παίζεται από το όργανο. Η παρτιτούρα μας δείχνει ακριβώς ποιες νότες να παιχτούν και πότε. Ο σκοπός της παρούσας

εργασίας είναι να εξαχθούν αυτόματα τα «χαρακτηριστικά γνωρίσματα» της παρτιτούρας από μονοφωνικά μουσικά κομμάτια.



Εικόνα 12: Αναπαράσταση της Μουσικής Καταγραφής

2

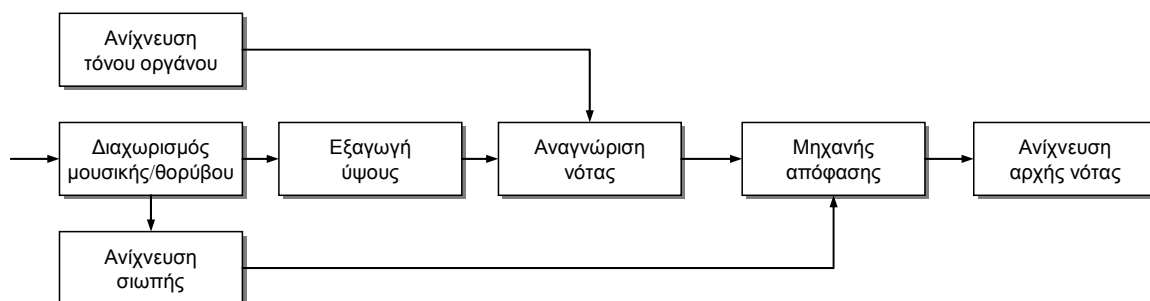
Αναγνώριση Μουσικής

Στόχος του παρόντος κεφαλαίου είναι η περιγραφή ενός συστήματος αυτόματης αναγνώρισης μονοφωνικής μουσικής. Η είσοδος ενός τέτοιου συστήματος είναι ένα ηχογραφημένο ηχητικό μονοφωνικό σήμα μουσικής και η έξοδος του είναι η ακολουθία των μουσικών νοτών που το απαρτίζουν.

Στην αναγνώριση μουσικής, περιλαμβάνεται ένα σύνολο από επιμέρους υποπροβλήματα που πρέπει να αντιμετωπιστούν, κάποια από τα σημαντικότερα είναι τα παρακάτω:

1. διαχωρισμός μουσικής/θορύβου (όπου ως θόρυβος νοείται οτιδήποτε δεν είναι μουσική)
2. ανίχνευση σιωπής (που διαφέρει από το παραπάνω καθώς η σιωπή μπορεί να αποτελεί μουσικό συμβάν –π.χ. μουσική παύση)
3. εξαγωγή ύψους
4. αναγνώριση νότας
5. ανίχνευση αρχής νότας
6. προσδιορισμός τόνου οργάνου

Το σύνολο λύσεων σε όλα αυτά τα προβλήματα μπορεί να χρησιμοποιηθεί ως βάση για τη διαμόρφωση ενός συστήματος αναγνώρισης μουσικής. Στο ακόλουθο σχήμα φαίνονται ενδεικτικά τα διάφορα στάδια επεξεργασίας που απαιτούνται για την μουσική αναγνώριση, καθώς και οι μεταξύ τους αλληλεξαρτήσεις.



Εικόνα 13 :Στάδια επεξεργασίας που απαιτούνται για την μουσική αναγνώριση και οι μεταξύ τους αλληλεξαρτήσεις.

Στις ακόλουθες παραγράφους παρουσιάζονται σε περισσότερη λεπτομέρεια καθένα από τα προβλήματα αυτά και περιγράφονται οι μέθοδοι για την επίλυσή τους, είτε όπως προκύπτουν από την βιβλιογραφία είτε όπως αντιμετωπίζονται από τις αντίστοιχες βαθμίδες στο έργο VEMUS.

Στο πλαίσιο του έργου VEMUS προκύπτει η ανάγκη για αναγνώριση τόσο σε πραγματικό όσο και σε μη πραγματικό χρόνο. Η αναγνώριση πραγματικού χρόνου αναφέρεται ως *σύγχρονη αναγνώριση* καθώς επιτελείται κατά την διάρκεια εκτέλεσης ενός μουσικού κομματιού από έναν μαθητή. Βασικός σκοπός της σύγχρονης αναγνώρισης είναι οι χαμηλές απαιτήσεις σε υπολογιστική ισχύ (λόγω της απαίτησης για λειτουργία σε πραγματικό χρόνο) καθώς και οι συνεπαγόμενες χαμηλές απαιτήσεις σε ακρίβεια και ορθότητα. Η σύγχρονη αναγνώριση επιτρέπει την υλοποίηση χρήσιμων λειτουργικοτήτων όπως, για παράδειγμα, το αυτόματο γύρισμα της σελίδας ενός κομματιού όταν ο μαθητής πλησιάζει στο τέλος της.

Αντίθετα, η *ασύγχρονη αναγνώριση*, αναφέρεται σε μια διαδικασία που εκτελείται μετά το πέρας της εκτέλεσης του μαθητή και στοχεύει στην εξαγωγή πληροφορίας σχετικά με την εκτέλεση με απώτερο σκοπό την αξιολόγησή της και την εμφάνιση κατάλληλων μηνυμάτων και προτροπών. Στην διαδικασία αυτή, η ταχύτητα επεξεργασίας λαμβάνει χαμηλότερη σημασία σε σχέση με την ποιότητα και την ακρίβεια της εξαγόμενης πληροφορίας.

Και η σύγχρονη και η ασύγχρονη αναγνώριση αποτελούνται από αντίστοιχα στάδια και χρησιμοποιούν αντίστοιχες βαθμίδες επεξεργασίας. Εντούτοις, οι σημαντικά διαφοροποιημένες απαιτήσεις τους προβάλλουν την ανάγκη για την αντιμετώπιση κάθε σταδίου με διαφορετικούς αλγορίθμους και διαφορετικά υπολογιστικά εργαλεία. Έτσι, για πολλές από τις εμπλεκόμενες βαθμίδες απαιτήθηκε η ανάπτυξη δύο διαφορετικών εκδόσεων: μία για χρήση κατά την σύγχρονη αναγνώριση και μία για χρήση κατά την ασύγχρονη. Στο υπόλοιπο κείμενο του κεφαλαίου, περιλαμβάνονται σχετικές αναφορές στις περιπτώσεις που οι σύγχρονες και ασύγχρονες εκδόσεις μιας βαθμίδας διαφοροποιούνται σημαντικά.

Και στις δύο περιπτώσεις, η ανάλυση πραγματοποιείται σε επίπεδο πλαισίου (frame-by-frame). Στην περίπτωση των βαθμίδων της σύγχρονης αναγνώρισης, το πλαίσιο ορίζεται ως ένας απομονωτής (buffer) δεδομένων ήχου όπως αυτός ηχογραφείται και προωθείται για εκτέλεση από το υλικό καταγραφής του υπολογιστή. Στην περίπτωση της ασύγχρονης αναγνώρισης, ένα πλαίσιο ορίζεται ως ένα «πακέτο» συγκεκριμένου αριθμού δειγμάτων από την εκτέλεση του μαθητή που έχει ήδη ηχογραφηθεί. Για λόγους ομοιογένειας και απλούστερης αντιμετώπισης, στο VEMUS έχει ληφθεί μέριμνα ώστε το μέγεθος των πλαισίων να είναι το ίδιο και στις δύο περιπτώσεις.

Στο σημείο αυτό θα πρέπει να τονιστεί ότι η εργασία πραγματοποιείται μονοφωνική μουσική, καθώς αυτός είναι και ο στόχος του έργου VEMUS. Για μονοφωνική μουσική, πολλά από τα παραπάνω θέματα παρουσιάζουν σημαντικά μειωμένο βαθμό πολυπλοκότητας σε σχέση με την αναγνώριση πολυφωνικής μουσικής ή την αναγνώριση πιο σύνθετων ηχητικών σημάτων, όπως η φωνή. Εντούτοις, τα προβλήματα που εμφανίζονται εξακολουθούν να διατηρούν ενδιαφέρον τους και συνεχίζουν να παρουσιάζουν σημαντικές προκλήσεις για την αποδοτική αντιμετώπισή τους, γεγονός που αντανακλάται και από το ερευνητικό ενδιαφέρον που παραμένει ενεργό στις περιοχές αυτές.

2.1 Σύντομη Επισκόπηση

Η αυτόματη αναγνώριση μουσικής είναι ένας ερευνητικός τομέας που στερείται ακόμα πλήρεις και ακριβείς λύσεις. Καμία σημαντική επένδυση στην αναγνώριση μουσικής συγκρίσιμη με τις επενδύσεις αναγνώρισης ομιλίας δεν έχει γίνει μέχρι σήμερα.

Το πρόβλημα εξαγωγής του ύψους έλαβε περισσότερη προσοχή από άλλα προβλήματα λόγω της σημασίας του για τη ανάλυση σημάτων φωνής. Η εξαγωγή του ύψους είναι προβληματική επειδή τα ηχητικά σήματα (ομιλία, μουσική) δεν είναι ποτέ από μαθηματική άποψη περιοδικά ακόμα και στα σύντομα τμήματα που εξετάζονται.

Η εξαγωγή του ύψους εκτελείται συνήθως σε δύο βήματα: (i) εξαγωγή των υποψηφίων υψών ανά τμήμα του ηχητικού σήματος (frame), (ii) μετά-επεξεργασία ύψους. Υπάρχουν πέντε σημαντικές προσεγγίσεις στην εξαγωγή υποψηφίων υψών:

- άμεσες μέθοδοι στο πεδίο του χρόνου [GR69, Kuh90]
- μέθοδοι βασισμένες στην συσχέτιση [Rab77, BHJ93, CK02]
- μέθοδοι πεδίου συχνότητας [Nol67, PG79a, Her88]
- προσαρμοστικές μέθοδοι φιλτραρίσματος [Moo78] και

- μέθοδοι που εμπνέονται από τα πρότυπα της επεξεργασίας πληροφοριών του ανθρώπινου ακουστικού συστήματος [SL90, MH91, DML+03].

Οι μέθοδοι πεδίου συχνότητας έχουν τη φήμη ως γρήγορες. Μπορούν να χρησιμοποιηθούν για την αναγνώριση της πολυφωνικής μουσικής. Οι βασισμένες στη συσχέτιση μέθοδοι είναι πιο αργές αλλά ακριβείς όσον αφορά στο θόρυβο και τα διαφορετικά ακουστικά περιβάλλοντα. Οι μέθοδοι που εμπνέονται από το ανθρώπινο ακουστικό σύστημα είναι πολύ ακριβείς αλλά αργές.

Η μετά-επεξεργασία ύψους εκμεταλλεύεται τα διαθέσιμα στοιχεία από τα γειτονικά πλαίσια. Μπορεί να γίνει μέσω αναζήτησης με δυναμικό προγραμματισμό [Boe93] ή της προσέγγισης «περιοχών εμπιστοσύνης» (islands of confidence) [Raš98]. Όσον αφορά στην αναγνώριση της πολυφωνικής μουσικής, οι τεχνικές εξαγωγής ύψους αλληλεπιδρούν έντονα με τις τεχνικές που χρησιμοποιούνται για τον διαχωρισμό πολλαπλών ηχητικών πηγών.

Οι παρούσες ακουστικές τεχνικές κατάτμησης βασίζονται συνήθως σε κάποια μορφή κατωφλίωσης (thresholding) χαρακτηριστικών γνωρισμάτων [RRS+98]. Η ενέργεια, η αρμονικότητα (harmonicity) και μερικά άλλα χαρακτηριστικά γνωρίσματα εξάγονται, τα χρονικά παράγωγά τους λαμβάνονται, ζυγίζονται και αθροίζονται. Τα χαρακτηριστικά γνωρίσματα που εξάγονται, ανιχνεύονται για τις υψηλές αιχμές που αντιστοιχούν στις χρονικές στιγμές των πιθανών onset των νοτών. Τα Hidden Markov Models (HMMs) και άλλες τεχνικές χρησιμοποιούνται επίσης για την κατάτμηση μουσικής [Ste99, SE03].

Η αναγνώριση τόνου αναφέρεται στους στόχους :

1. εκτίμηση μιας ενιαίας συνεχούς τιμής του ύψους για κάθε νότα από τα αντίστοιχα αποσπάσματα της χρονικά μεταβαλλόμενης «πορείας» του ύψους, και
2. τη σωστή αντιστοίχιση μιας συνεχούς εκτίμησης του ύψους σε έναν διακριτό αριθμό MIDI νότας.

Το πρόβλημα (1) λύνεται συνήθως με τον απλό υπολογισμό του μέσου όρου των τιμών ύψους [NI86]. Το πρόβλημα (2) λύνεται με στρογγυλοποίηση προς τα πάνω της συνεχούς τιμής του ύψους στην πλησιέστερη διακριτή συχνότητα σε μία εξίσου μετριασμένη κλίμακα μουσικών συχνοτήτων [PG79b]. Εντούτοις αυτές οι απλές λύσεις υποθέτουν ότι τα όργανα είναι καλά κουρδισμένα.

2.2 Διαχωρισμός Μουσικής/Θορύβου

Πριν από την βαθμίδα αναγνώρισης μουσικής νότας, συνήθως προηγείται μια βαθμίδα αρμόδια για την ανίχνευση μουσικής. Η εργασία που πρέπει να επιτελέσει η βαθμίδα αυτή είναι ο διαχωρισμός μεταξύ μουσικής και θορύβου. Για τις ανάγκες της παρούσας, ως «θόρυβος» εννοείται οποιοδήποτε τμήμα του ηχητικού σήματος δεν περιέχει μουσική.

Η βαθμίδα ανίχνευσης μουσικής είναι αρμόδια για την (σε πραγματικό ή μη χρόνο) διάκριση μεταξύ τμημάτων μουσικής και μη-μουσικής (θορύβου) του σήματος εισόδου. Αυτό είναι απαραίτητο για την ανίχνευση και απόρριψη ηχητικών τμημάτων που δεν περιέχουν μουσική, ούτως ώστε αφενός να αποφεύγεται η λανθασμένη αναγνώριση ψευδών νοτών και αφετέρου να μην δαπανώνται υπολογιστικοί πόροι. Ένα τέτοιο χαρακτηριστικό είναι ουσιαστικό, αφού η κύρια ομάδα-στόχος είναι νέοι μαθητές, οι οποίοι συχνά δεν κρατούν πειθαρχεία κατά τη διάρκεια του μαθήματος.

Ειδικά το πρόβλημα διάκρισης σημάτων φωνής από μουσικά σήματα, έχει γίνει όλο και πιο σημαντικό, καθώς αυτόματα σύστημα αναγνώρισης ομιλίας χρησιμοποιούνται σε εφαρμογές πολυμέσων του «πραγματικού κόσμου». Εκτενής έρευνα έχει πραγματοποιηθεί σε αυτό το θέμα και οι περισσότεροι αλγόριθμοι και τεχνικές που έχουν προταθεί βασίζονται σε εξαγωγή χαρακτηριστικών των σημάτων και σύγκριση αυτών. Στο [SS97], οι συγγραφείς εξετάζουν 13 διαφορετικές, συχνά χρησιμοποιούμενες φασματικές παραμέτρους, προκειμένου να διακρίνουν φωνητικά από ηχητικά σήματα.

Το πρόβλημα διάκρισης ενός ηχητικού σήματος σε μουσική ή μη δεν είναι τετριμμένο. Στην περίπτωση μας, ο κύριος στόχος είναι η ανάπτυξη μιας βαθμίδας αρκετά ακριβούς στη διάκριση μεταξύ μουσικής ή μη, ανεκτικής στον περιβαλλοντικό θόρυβο και επιπλέον χαμηλής υπολογιστικής πολυπλοκότητας, δεδομένου ότι καλείται να λειτουργεί σε πραγματικό χρόνο πριν τη βαθμίδα αναγνώρισης. Η ακρίβειά της βαθμίδας είναι ουσιαστική για την απόδοση των ακολούθων ενότητων.

Ο αλγόριθμος ανίχνευσης μουσικής /μη μουσικής του VEMUS έχει σχεδιαστεί με κύριο γνώμονα την ταχύτητα και τις χαμηλές υπολογιστικές απαιτήσεις. Αρχικά, για τη διάκριση μουσικής και ομιλίας ένα νευρωνικό δίκτυο είχε εφαρμοστεί, που χρησιμοποιούσε συγκεκριμένες φασματικές παραμέτρους που αποδείχτηκε πως λειτουργούν ικανοποιητικά για αυτό το σκοπό. Είναι βασισμένο σε μια frame-by-frame ανάλυση που λαμβάνει υπόψη 4 διαφορετικές παραμέτρους:

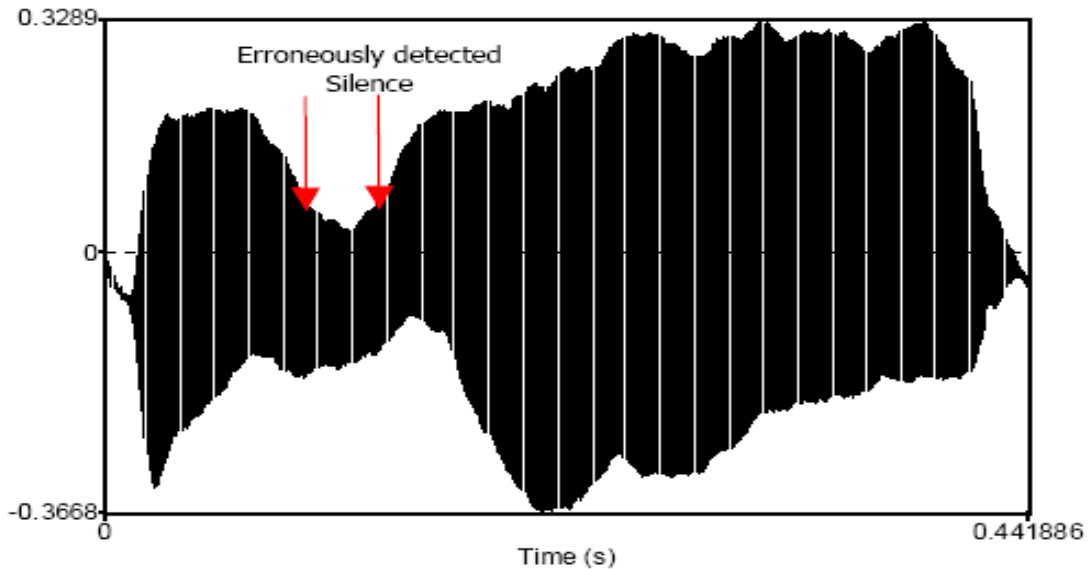
- την F_0 και το επίπεδο πλάτους,
- τη φασματική ροή,
- το επίπεδο ενέργειας και το
- φασματικό κέντρο βάρους (centroid).

Το δίκτυο εκπαιδεύτηκε με βάση επισημειωμένα ηχητικά δεδομένα ομιλίας και μουσικής και αποδείχτηκε ότι λειτουργεί ικανοποιητικά καλά.

2.3 Ανίχνευση Σιωπής

Για την βαθμίδα ανίχνευσης σιωπής στο πλαίσιο του έργου VEMUS, ερευνήθηκαν διάφορες παράμετροι, τόσο στο πεδίο του χρόνου όσο φασματικοί, κάθε μία από τις οποίες παρουσίαζε τα δικά της προτερήματα και μειονεκτήματα. Όπως προτείνεται σε διάφορες σχετικές μελέτες (όπως, για παράδειγμα, [RCG76]) ο φάκελος του σήματος είναι μια αρκετά καλή προσέγγιση, εντούτοις δεν είναι ικανοποιητική στην περίπτωση των μικροφώνων που εισάγουν τυχαίο dc offset στο σήμα, σύμφωνα με την ανομοιόμορφη απόκριση συχνότητας που μπορεί να υπακούει. Αυτό είναι προφανές στο ακόλουθο σχήμα όπου το κριτήριο φακέλου θα πρότεινε ότι υπάρχει ένα σιωπηλό μέρος μέσα στη νότα.

Άλλες απλές προσεγγίσεις για την ανίχνευση σιωπής έχουν προταθεί στο [BHJ93] και [CK01] όπου η αναλογία zero-crossing και η ενέργεια ενός πλαισίου εξετάζονται. Πάλι αυτά τα χαρακτηριστικά γνωρίσματα μόνο αποδείχτηκαν φτωχά κριτήρια για αυτόν τον στόχο αφού θα μπορούσαν να παραπλανηθούν εύκολα από τα θορυβώδη ή dc τμήματα. Στα πειράματά η αναλογία zero-crossing ήταν αρκετά φτωχή στην απόφαση της για το αν ένα πλαίσιο περιέχει φωνή ή όχι.



Εικόνα 14: Λανθασμένα ανιχνευμένα μέρη σιωπής με τη χρησιμοποίηση των απλών τιμών φακέλων

Στο VEMUS τέσσερις διαφορετικές παραμέτρους ελήφθησαν υπόψη, όλες εκ των οποίων θα πρέπει να είναι κάτω από ένα συγκεκριμένο όριο ταυτόχρονα ώστε ένα πλαίσιο να χαρακτηριστεί ως σιωπηλό. Αυτές οι παράμετροι είναι:

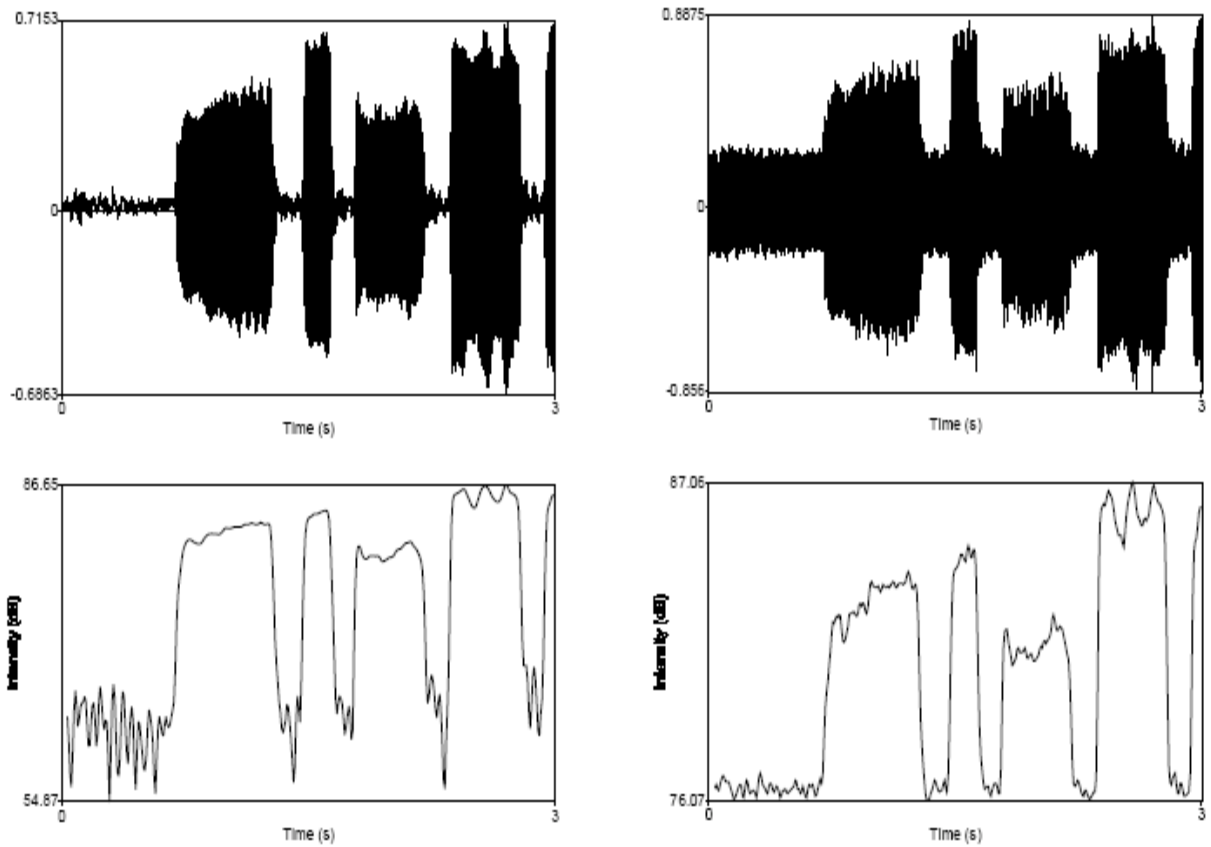
- το ανιχνευμένο επίπεδο ύψους,
- η περιοδικότητα,
- το επίπεδο εύρους και η
- ένταση ενός πλαισίου.

Τα δύο πρώτα χαρακτηριστικά γνώρισμα στοχεύουν να χαρακτηρίσουν ένα πλαίσιο για την περιοδική φύση του ή όχι, ενώ το εύρος και τα επίπεδα έντασης αποφασίζουν εάν ένα πλαίσιο σήματος είναι κάτω από το ακουστικό επίπεδο ή όχι.

Προκειμένου να υπολογιστεί η ένταση του πλαισίου, οι τιμές του σήματος υψώνονται στο τετράγωνο και γίνεται συνέλιξη με ένα παράθυρο Kaiser τάξης 20. Η χρήση της έντασης για την απόφαση εάν ένα πλαίσιο είναι σιωπή ή όχι, αποδείχτηκε αρκετά αποδοτικό. Το κύριο πλεονέκτημα αυτού του χαρακτηριστικού γνώρισματος είναι ότι μπορεί να παρέχει εξαιρετικά υψηλή ανάλυση στο πεδίο του χρόνου χωρίς να επηρεάζει την ευρωστία του αλγορίθμου. Σε αυτήν την περίπτωση το μήκος του πλαισίου καθορίζει αυτόματα το μήκος των παραθύρων της διαδικασίας. Εντούτοις, στην περίπτωση της ασύγχρονης ενότητας, το ίδιο χαρακτηριστικό γνώρισμα χρησιμοποιείται επίσης για την ανίχνευση σιωπής αλλά προκειμένου να επιτευχθεί υψηλή ανάλυση στο πεδίο του χρόνου χρησιμοποιήθηκε ένα μήκος παραθύρων 3 / ελάχιστο ύψος (δηλ. $\approx 400\text{Hz}$), το οποίο παρέχει αποτελέσματα με χρονική ανάλυση κάτω από 3msec (επίσης επικαλυπτόμενα παράθυρα). Αυτή η ακρίβεια και ανάλυση είναι υπερβολικά καλές και απαραίτητες για τη σωστή εκτίμηση του ακριβούς σημείου εκκίνησης της νότας (note onset) και της διάρκειάς της.

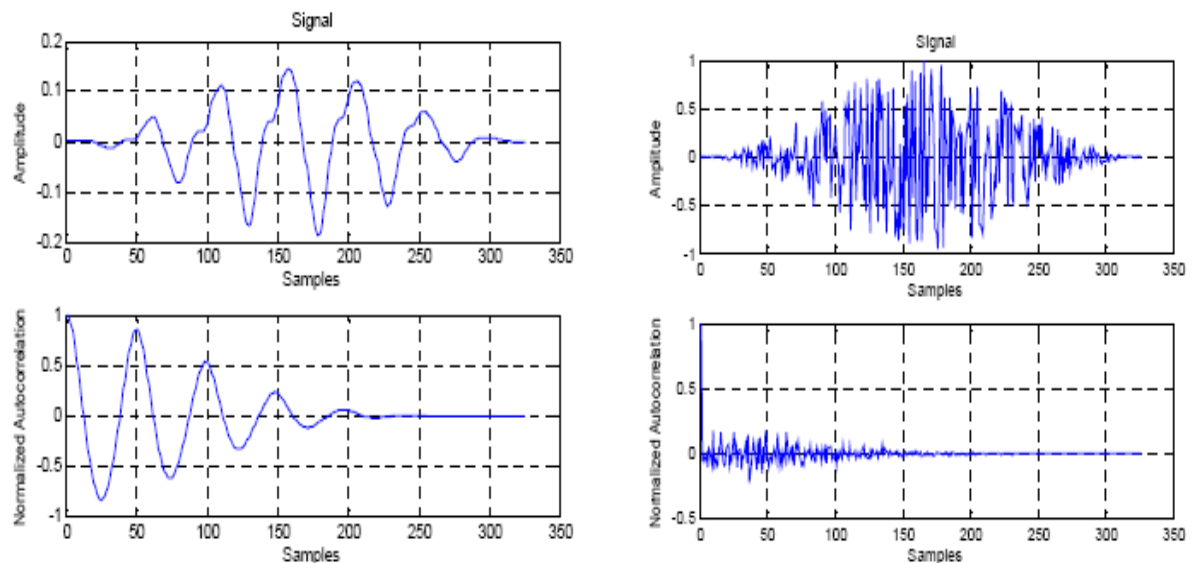
Ένα πολύ σημαντικό χαρακτηριστικό γνώρισμα της έντασης του σήματος, με τον τρόπο που αυτή υπολογίζεται, είναι η ανοχή που προσφέρει σε υψηλά επίπεδα θορύβου, όπως φαίνεται στις ακόλουθες εικόνες. Στη βαθμίδα σύγχρονης αναγνώρισης το επίπεδο κατώτατων ορίων της έντασης ώστε ένα πλαίσιο να είναι κάτω από το αντιληπτό επίπεδο είναι ορισμένο εξ αρχής, ενώ στην ασύγχρονη

βαθμίδα αναγνώρισης, το επίπεδο κατώτατων ορίων ορίζεται ως 0.01 του μέγιστου επιπέδου έντασης σε όλο το σήμα.



Εικόνα 15 : Απεικόνιση της παραμέτρου της έντασης και της αποδοτικότητάς του σε σήματα με θόρυβο

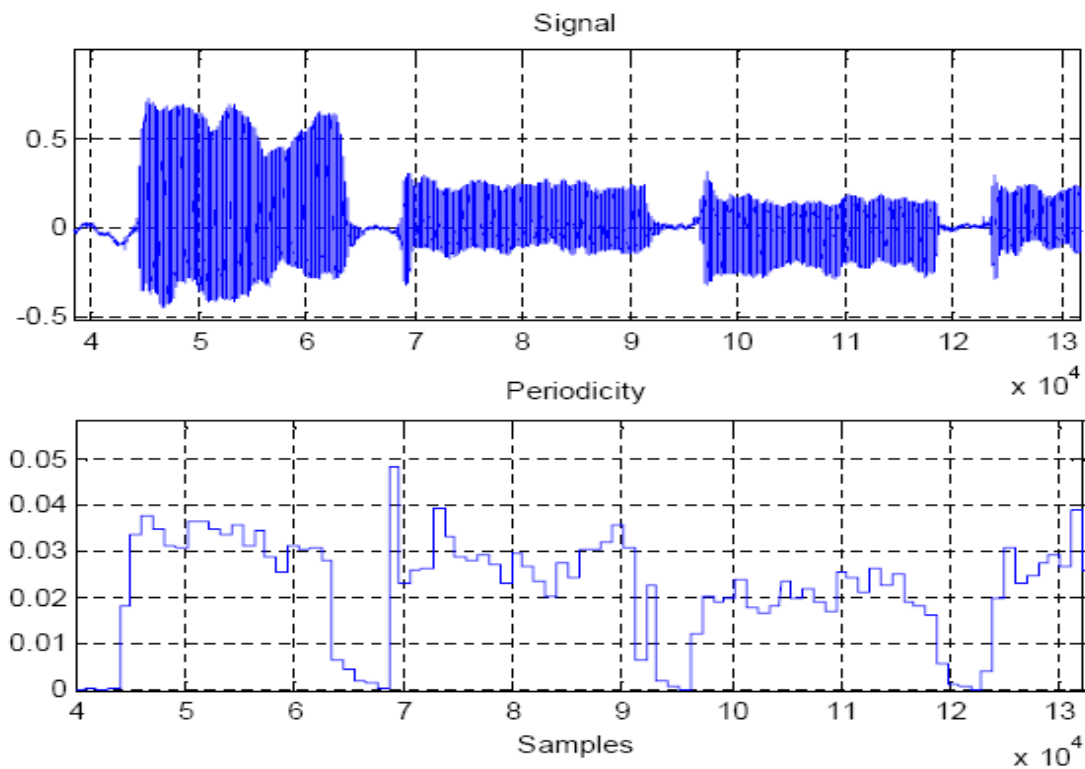
Τέλος, προκειμένου να επιβεβαιωθεί ότι οι τυχαίοι ήχοι ή θόρυβοι δεν έχουν επιπτώσεις στον αλγόριθμο, χρησιμοποιούμε επίσης την περιοδικότητα του σήματος, που μπορεί να ανιχνευθεί εύκολα από την κανονικοποιημένη ακολουθία αυτοσυσχέτισης του σήματος.



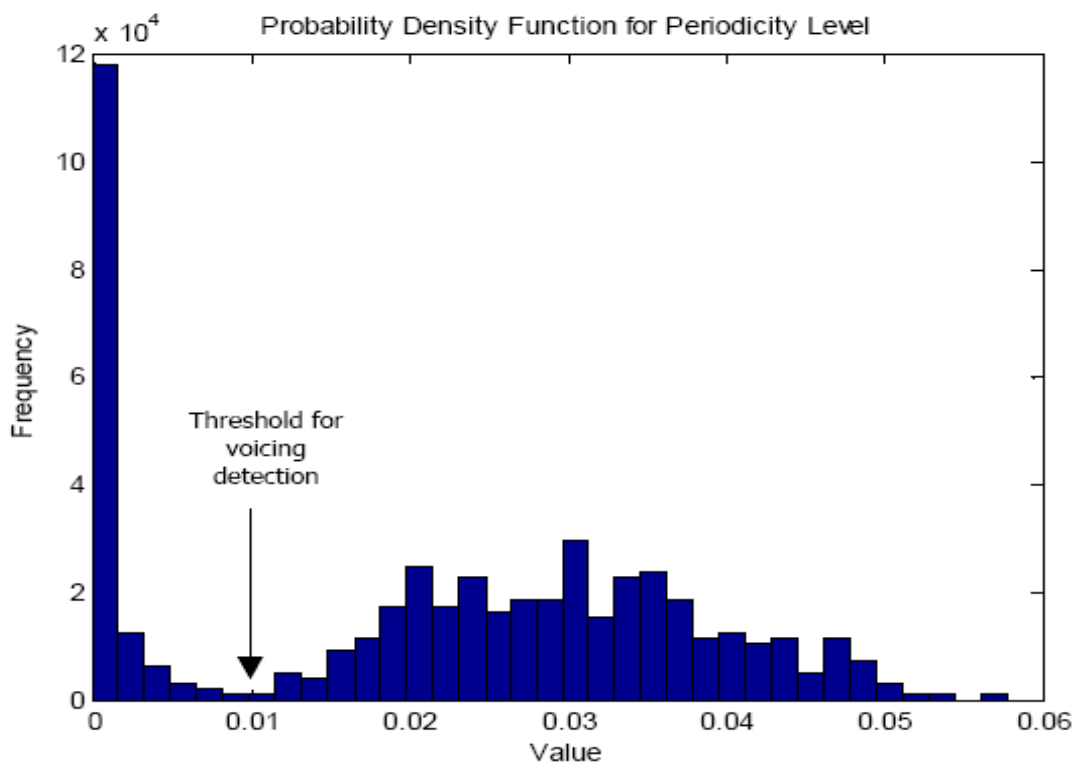
Εικόνα 16: Η ακολουθία αυτοσυσχέτισης ενός σήματος με φωνή και ενός χωρίς.

Ένας αλγόριθμος επιλογής μεγίστων (peak picking) χρησιμοποιείται στην πρώτη παράγωγο της ακολουθίας αυτοσυσχέτισης σε κάθε πλαίσιο, προκειμένου να ανιχνευθούν οι σημαντικές αιχμές (υποψήφια ύψη) μέσα στα επιτρεπτά όρια, όπως καθορίζονται από το κατώτατο και μέγιστο επίπεδο ύψους (~400-3000Hz).

Όλες οι προαναφερθείσες παράμετροι χρησιμοποιούνται και στη σύγχρονη και στην ασύγχρονη βαθμίδα, με μια μικρή διαφορά. Στη σύγχρονη βαθμίδα τα όρια για κάθε παράμετρο που χρησιμοποιούνται για την ανίχνευση ενός πλαισίου μη-μουσικής είναι προκαθορισμένα, ενώ στην ασύγχρονη βαθμίδα τα κατώτατα όρια προσαρμόζονται αυτόματα στα χαρακτηριστικά του σήματος. Παραδείγματος χάριν, στην περίπτωση της συνάρτησης αυτοσυσχέτισης, το επίπεδο κατώτατων ορίων καθορίζεται μέσω της συνάρτησης πυκνότητας πιθανότητας. Όπως φαίνεται στο παρακάτω σχήμα, το πρώτο τοπικό ελάχιστο της συνάρτησης πυκνότητας καθορίζει ένα ικανοποιητικό όριο για την ανίχνευση πλαισίων φωνής μέσα στο σήμα, εντούτοις στην περίπτωση του σύγχρονου αναλυτή που τα όρια είναι προκαθορισμένα τα αποτελέσματα δεν είναι τόσο ικανοποιητικά το οποίο όμως αντισταθμίζεται από το επόμενο επίπεδο από τη μηχανή απόφασης 3 επιπέδων.



Εικόνα 17: Απεικόνιση της περιοδικότητας για την έκφραση της ανίχνευση



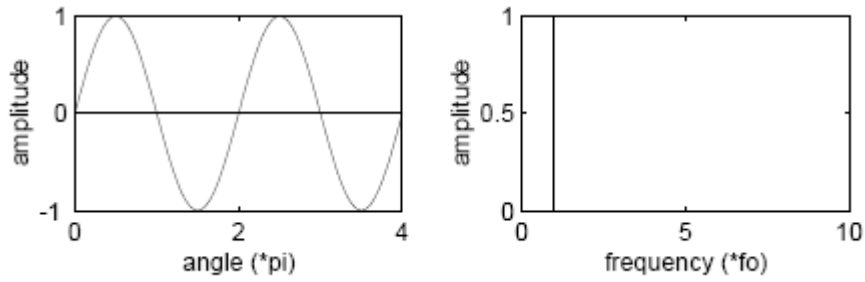
Εικόνα 18: Καθορισμός του κατώτατου ορίου έκφρασης από το ιστόγραμμα του επιπέδου περιοδικότητας

2.4 Εκτίμηση Θεμελιώδους Συχνότητας

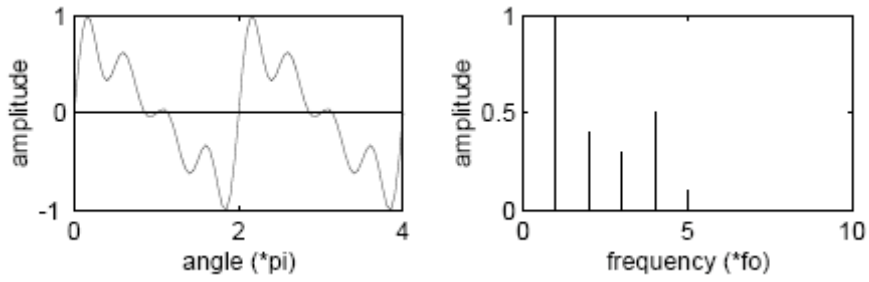
Η εκτίμηση της θεμελιώδους συχνότητας f_0 (fundamental frequency estimation) γνωστή και ως εξαγωγή του ύψους (pitch extraction) είναι το κύριο και πρώτο στάδιο που εκτελείται κατά τη διάρκεια της αναγνώρισης μουσικής. Σημαντική έρευνα έχει γίνει στο συγκεκριμένο αντικείμενο εδώ και 50 και πλέον χρόνια. Το θέμα της εξαγωγής του ύψους ανάγεται στον υπολογισμό της θεμελιώδους συχνότητας ενός μουσικού σήματος, αν και ένας εκτιμητής του ύψους θα πρέπει να λαμβάνει υπ' όψη και αντιληπτικά μοντέλα και να παράγει το αποτέλεσμα στην κλίμακα του ύψους και όχι στην κλίμακα της συχνότητας.

Υπάρχει ένα πλήθος μεθόδων για την εξαγωγή της θεμελιώδους συχνότητας, βασισμένων σε διάφορες μαθηματικές αρχές. Αφού το ύψος είναι μια αντιληπτική ποσότητα σχετιζόμενη με την θεμελιώδη συχνότητα, f_0 , μιας περιοδικής ή ψευδο-περιοδικής κυματομορφής, θα αρκούσε να προσδιορίσουμε την περίοδο της ταλάντωσης, η αντίστροφη της οποίας θα μας έδινε την συχνότητα. Το πρόβλημα προκύπτει όταν το κύμα αποτελείται από περισσότερα του ενός ημιτόνου. Καθώς αρμονικά στοιχεία προστίθενται σε ένα ημιτονοειδές κύμα, το ύψος γίνεται λιγότερο εμφανές και πρέπει να εξετάσουμε την θεμελιώδη συχνότητα ή f_0 .

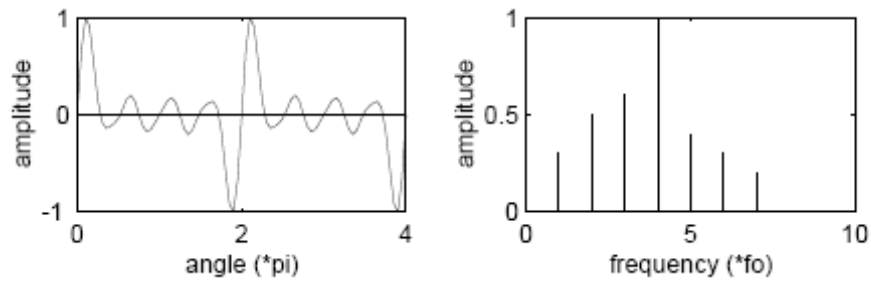
Η δυσκολία εύρεσης της f_0 μιας κυματομορφής εξαρτάται από την ίδια την κυματομορφή. Αν η κυματομορφή έχει λίγες υψηλές αρμονικές ή η ισχύς τους είναι μικρή, η f_0 είναι πιο εύκολο να ανιχνευτεί. Αν οι αρμονικές έχουν μεγαλύτερη ισχύ από την f_0 , η περίοδος είναι πιο δύσκολο να ανιχνευτεί όπως φαίνεται παρακάτω.



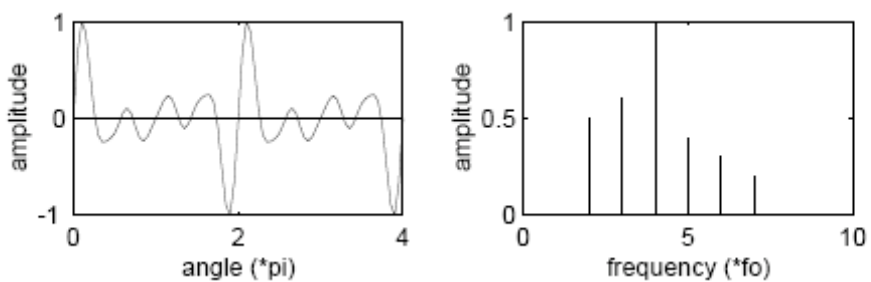
Εικόνα 19 : Κυματομορφή χωρίς υψηλότερες αρμονικές



Εικόνα 20: Κυματομορφή με υψηλότερες αρμονικές μικρότερης ισχύος



Εικόνα 21: Κυματομορφή με υψηλότερες αρμονικές μεγαλύτερης ισχύος



Εικόνα 22 :Κυματομορφή με υψηλότερες αρμονικές μικρότερης ισχύος και χωρίς θεμελιώδη συχνότητα

Εναλλακτικές προσεγγίσεις έχουν συγκριθεί βασισμένες σε ένα σύνολο μεθόδων που προσαρμόζονται στα χαρακτηριστικά του συγκεκριμένου οργάνου, συμπεριλαμβανομένων:

- τεχνικών στο πεδίο της συχνότητας (FFT, LPC και LSF, cepstrum)
- τεχνικών στο πεδίο του χρόνου (αυτοσυσχέτιση)
- αναγνώριση προτύπων
- ταξινόμηση με νευρωνικά δίκτυα.

Στο έργο VEMUS, αρχικά εφαρμόστηκαν και δοκιμάστηκαν διάφοροι αλγόριθμοι και τεχνικές εκτίμησης της θεμελιώδους συχνότητας για να διερευνηθεί κατά πόσο καλύπτουν τις ανάγκες του έργου. Πιο συγκεκριμένα δύο τεχνικές στο πεδίο του χρόνου ερευνήθηκαν, η σταθμισμένη συνάρτηση αυτοσυσχέτισης [Boe93] [Del00], και η συνάρτηση AMDF [Yin00] που αποδείχθηκε μικρής ακρίβειας στην περίπτωση θορυβωδών σημάτων. Αν και η εφαρμογή τους είναι εύκολο να επιτευχθεί, δεν παρέχουν οποιεσδήποτε πρόσθετες πληροφορίες για το φασματικό περιεχόμενο του σήματος, όπως τα partials («μερικές»).

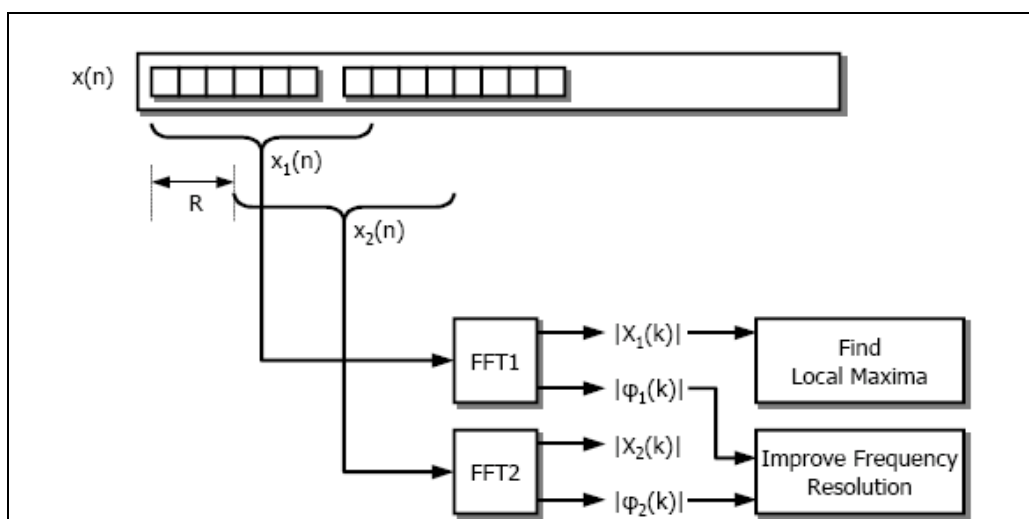
Εφαρμόστηκαν επίσης και εξετάστηκαν τεχνικές αναγνώρισης προτύπων, όπως ο αλγόριθμος μέγιστης πιθανότητας [Nol69] και η διανυσματική κβαντοποίηση με LPC και Cepstrum συντελεστές. Οι Durand και ο Gomez στο [DG00] προτείνουν μια μέθοδο ταιριάσματος αρμονικών για αυτόν τον στόχο, με απλή πρόβλεψη ενός σχεδίου αρμονικών για μια νότα και ταιριάζοντας το με το εισερχόμενο πλαίσιο, μέσω ενός μέτρου της φασματικής απόστασης. Αυτές οι μέθοδοι γενικά αποτελούν μια καλή προσέγγιση στην περίπτωση της πολυφωνικής μουσικής όπου δεν ενδιαφέρει το ακριβές επίπεδο ύψους, αλλά ο σημαντικότερος στόχος είναι η ορθή ανίχνευση πιθανών ταυτόχρονων νοτών.

Στην περίπτωσή του VEMUS, αυτές οι τεχνικές, ακόμα κι αν αποδίδουν ικανοποιητικά, αποδείχθηκαν λιγότερο κατάλληλες από άλλες αφενός εξαιτίας της ανάγκης δεδομένων εκπαίδευσης και αφετέρου λόγω του ότι η έξοδος αυτών των αλγορίθμων είναι κβαντοποιημένα επίπεδα ύψους. Με αυτόν τον τρόπο θα απαιτούταν επιπρόσθετη επεξεργασία του σήματος για την εξαγωγή των σημαντικών παραμέτρων για την βαθμίδα αξιολόγησης της απόδοσης, η οποία θα προσέθετε μεγαλύτερη υπολογιστική πολυπλοκότητα και χρόνο επεξεργασίας.

Διερευνήθηκαν επίσης αλγόριθμοι πεδίου συχνότητας, όπως οι απλοί μετασχηματισμοί Fourier [Gol64], Cepstrum [Nol67] και στιγμιαίοι αλγόριθμοι ανίχνευσης συχνότητας [Bak96; Got00], που αποδείχθηκαν πιο ακριβείς σε θορυβώδη περιβάλλοντα. Εντούτοις η ανάλυσή τους στο πεδίο χρόνου και συχνότητας εξαρτάται από το μήκος πλαισίων και τη συχνότητα δειγματοληψίας του σήματος. Ένας απλός μετασχηματισμός DFT παρέχει χαμηλή ανάλυση συχνότητας στις χαμηλές συχνότητες, ενώ η τεχνική zero-padding μπορεί να δώσει μια καλύτερη εκτίμηση τους, χωρίς τη δυνατότητα να αποφευχθούν τα φαινόμενα κβαντοποίησης συχνότητας. Αυτό μπορεί να αποφευχθεί εν τούτοις εάν οι πληροφορίες φάσης του μετασχηματισμού DFT χρησιμοποιηθούν. Αυτή η μέθοδος προέρχεται από το σύστημα Phase VOCODER, που παρουσιάζεται στο [LD99], όπου το τμήμα φάσης του μετασχηματισμού FFT εξετάζεται επίσης για την ακριβή εκτίμηση των partials του σήματος. Εναλλακτικές και πιο προηγμένες εκδόσεις αυτής της τεχνικής έχουν παρουσιαστεί στο [Dix97] και [DAFX 2000], όπου έχει αποδειχθεί ότι η ανάλυση συχνότητας αυτής της μεθόδου είναι εξαιρετικά υψηλή, ακόμη και στην περίπτωση των σημάτων χαμηλής συχνότητας δειγματοληψίας. Αυτή είναι η μέθοδος που κρίθηκε ως πιο κατάλληλη για την περίπτωσή, και που τελικά υιοθετήθηκε.

Ένας αλγόριθμος που χρησιμοποιείται συχνά για αυτό το είδος εφαρμογών είναι αυτός που έχει προταθεί από τους Gold-Rabiner στο [GR74], όπου οι συγγραφείς προτείνουν την ταυτόχρονη χρήση έξι διαφορετικών ανιχνευτών ύψους, η πλειοψηφία των οποίων αποφασίζει για τη σωστή εκτίμηση του επιπέδου του ύψους. Αν και αυτή η μέθοδος φαίνεται ελκυστική, αποδείχθηκε φτωχότερη σε ακρίβεια και ευρωστία από αυτή που χρησιμοποιήθηκε.

Η κύρια δομή του αλγορίθμου απεικονίζεται στο παρακάτω σχήμα. Ένα τμήμα μήκους N εξάγεται από κάθε R δείγματα και έπειτα εφαρμόζεται FFT.



Εικόνα 23: FFT-based δομή υπολογισμού ύψους με εκτίμηση φάσης

Η ανάλυση συχνότητας βελτιώνεται με τη χρήση των πληροφοριών φάσης, ενώ οι παράμετροι R και το μήκος πλαισίου επιτρέπουν τον έλεγχο του ρυθμού επικάλυψης, δηλαδή την ανάλυση (resolution) του αποτελέσματος στο πεδίο του χρόνου. Τέτοιες σημαντικές παράμετροι (συμπεριλαμβανομένων των ορίων για το θόρυβο, το κέρδος μικροφώνων κ.λπ.) βελτιστοποιούνται για την περίπτωση του οργάνου καταγραφής ήχου και τις συγκεκριμένες συναρτήσεις μεταφοράς των μικροφώνων.

Η κύρια ιδέα αυτού του ανιχνευτή ύψους είναι δανεισμένη από τον τυπικό VOCODER φάσης όπου η φάση του μετασχηματισμού FFT λαμβάνεται υπόψη. Εξετάζοντας τον υπολογισμό ενός FFT N -σημείων, η ανάλυση συχνότητας του FFT είναι:

$$\Delta f = \frac{f_s}{N}$$

Όπου f_s είναι η συχνότητα δειγματοληψίας του σήματος. Με την ανίχνευση της μέγιστης αιχμής του FFT, μπορούμε να πάρουμε μια εκτίμηση της θεμελιώδους συχνότητας:

$$\tilde{f}_0 = k_0 \cdot \Delta f = k_0 \frac{f_s}{N},$$

όπου k_0 είναι το FFT bin της μέγιστης αιχμής. Η αντίστοιχη κανονικοποιημένη συχνότητα μπορεί να υπολογιστεί εύκολα ως:

$$\tilde{\Omega}_0 = 2\pi \tilde{f}_0 T_s = k_0 \frac{2\pi}{N}$$

Προκειμένου να βελτιωθεί η ανάλυση συχνότητας, οι πληροφορίες φάσης χρησιμοποιούνται, αφού για ένα αρμονικό σήμα η θεμελιώδης συχνότητα μπορεί να υπολογιστεί από την παράγωγο φάσης

$$\Omega_0 = \frac{d\phi(n)}{dn}$$

Η παράγωγος φάσης μπορεί να προσεγγιστεί με τον υπολογισμό των φάσεων δύο διαδοχικών FFT, που χωρίζονται από ένα μέγεθος βήματος R δειγμάτων:

$$\bar{\Omega} = \frac{\Delta\phi}{R},$$

όπου $\Delta\phi$ η διαφορά φάσης μεταξύ των δύο FFTs που υπολογίζεται στο δείκτη FFT K_0 .

Θεωρώντας δύο διαδοχικά FFTs που χωρίζονται από R δείγματα, παίρνουμε τις αντίστοιχες φάσεις για τον ίδιο K_0 δείκτη:

$$\begin{aligned}\varphi_1 &= \angle\{X_1(k_0)\} \\ \varphi_2 &= \angle\{X_2(k_0)\}\end{aligned}$$

Λαμβάνοντας υπόψη τις προηγούμενες εξισώσεις μπορούμε να υπολογίσουμε την αναμενόμενη φάση του δεύτερου FFT μετά από ένα μέγεθος βήματος R δειγμάτων, εάν η θεμελιώδης αρμονική είναι:

$$\begin{aligned}\tilde{f}_0 &= k_0 \cdot \Delta f : \\ \varphi_{2t} &= \varphi_1 + \tilde{\Omega}_0 R = \varphi_1 + \frac{2\pi}{N} k_0 R\end{aligned}$$

Το λάθος φάσης μεταξύ της unwrapped τιμής φ_2 και της φάσης στόχου μπορεί να υπολογιστεί από:

$$\varphi_{2err} = \text{princ arg}(\varphi_2 - \varphi_{2t})$$

Η unwrapped φάση λαμβάνεται έπειτα ως $\varphi_{2u} = \varphi_{2t} - \varphi_{2err}$ και η τελική εκτίμηση της θεμελιώδους συχνότητας λαμβάνεται έπειτα ως:

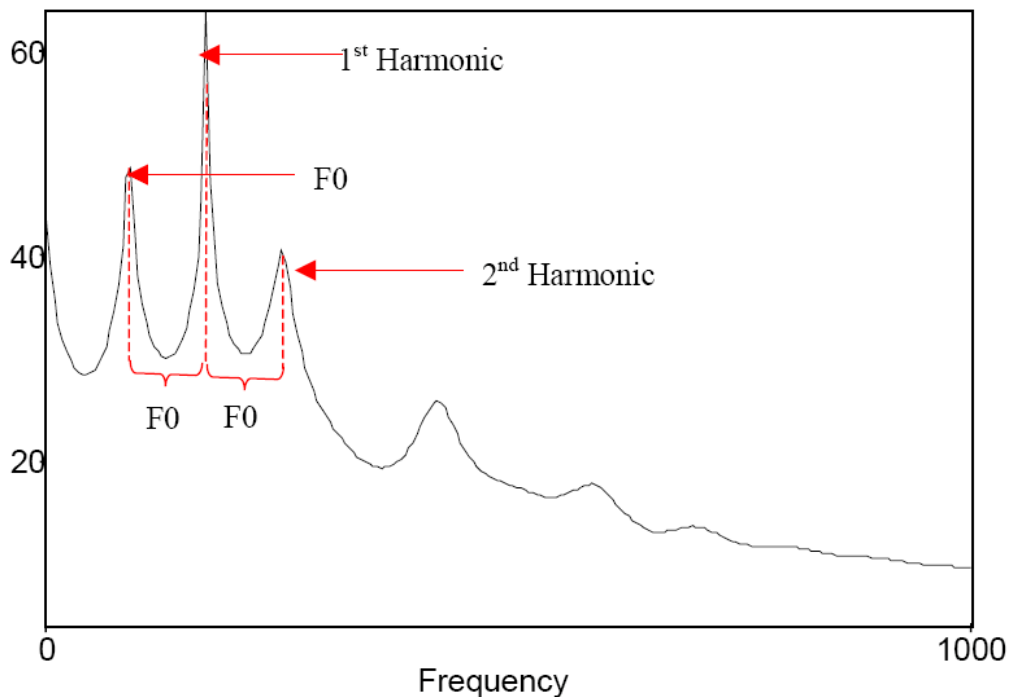
$$\bar{f}_0 = \frac{1}{2\pi} \tilde{\Omega}_0 f_s = \frac{1}{2\pi} \cdot \frac{\varphi_{2u} - \varphi_1}{R} f_s$$

Ένα απλό παράδειγμα ενός ιδανικού κύματος ημιτόνου σε μια θεμελιώδη συχνότητα 420Hz με $f_s=44.100\text{Hz}$, που αναλύεται με ένα 1024 FFT θα έδινε $\bar{f}_0=430.66\text{Hz}$ ενώ με αυτήν την μέθοδο και $R=1$ θα έδινε $\bar{f}_0=419\text{Hz}$ που είναι μία σχεδόν τέλεια εκτίμηση.

Τα κύρια πλεονεκτήματα αυτής της μεθόδου είναι η υψηλής ευκρίνειας εκτίμηση ύψους και η χαμηλή υπολογιστική πολυπλοκότητα που απαιτείται. Η εφαρμογή ενός FFT σε ένα πλαίσιο ενός σήματος είναι χαμηλού υπολογιστικού φορτίου, ενώ η χρήση αυτού του αλγορίθμου σε επίπεδο πλαισίου, μπορεί να παρέχει υψηλή ανάλυση στο πεδίο του χρόνου, εάν είναι επιθυμητό, με επικαλυπτόμενα πλαίσια. Χαρακτηριστικά, στην ενότητα ανίχνευσης ύψους σε πραγματικό χρόνο, τα πλαίσια δεν είναι επικαλυπτόμενα και ο FFT εκτελείται με 512 δείγματα με βήμα $R=1$, που παρέχει χρονική ανάλυση 32ms. Στην περίπτωση της βαθμίδας ασύγχρονης αναγνώρισης, όπου υπάρχει η ανάγκη για υψηλή ανάλυση στο πεδίο του χρόνου (προκειμένου να καθοριστούν με ακρίβεια τα σημεία εκκίνησης και τερματισμού κάθε νότας), τα πλαίσια επικαλύπτονται κατά 95%, παρέχοντας μία χρονική ανάλυση 3,2 ms. Αυτή η ανάλυση είναι απαραίτητη για την σωστή ανάλυση του attack κάθε νότας και την ακριβή εκτίμηση-ανίχνευση των ορίων των νοτών μέσα στο ακουστικό σήμα.

Στην προκειμένη περίπτωση, η θεμελιώδης συχνότητα του σήματος προέρχεται από τη χαμηλότερη (σε συχνότητα) «σημαντική» κορυφή του FFT η οποία βρίσκεται στα επιτρεπτά όρια (για παράδειγμα, για την περίπτωση της φλογέρας, τα όρια αυτά είναι μεταξύ 400 και 2500Hz). Με αυτόν τον τρόπο αποφεύγονται λανθασμένες εκτιμήσεις που οφείλονται σε πρώτες αρμονικές υψηλότερες από την θεμελιώδη συχνότητα. Ως «σημαντική» νοείται μια κορυφή υψηλού πλάτους και σχετικά ισχυρή σε σχέση με την υψηλότερη κορυφή του φάσματος.

Η καταλληλότητα του αλγόριθμου εκτίμησης θεμελιώδους συχνότητας που περιγράφηκε, ενισχύεται περαιτέρω στην περίπτωση του VEMUS όπου αντιμετωπίζονται μόνο μονοφωνικοί ήχοι, και ως εκ τούτου οι αρμονικές είναι πολλαπλάσιες της θεμελιώδους συχνότητας. Με ένα ενισχυμένο αλγόριθμο επιλογής κορυφών λαμβάνονται οι 4 πρώτες αρμονικές του σήματος, η ενδιάμεση απόσταση των οποίων παρέχει μια περισσότερο ακριβή εκτίμηση για την θεμελιώδη συχνότητα.



Εικόνα 24: Αρμονικές και εκτίμηση της F0.

Ο υπολογισμός των αποστάσεων των αρμονικών παρέχει έναν οδηγό για την κατάλληλη περιοχή φάσματος όπου θα πρέπει να αναζητηθούν υποψήφιες τιμές για την θεμελιώδη συχνότητα.

Για να αποφασιστεί κατά πόσο ένα πλαίσιο σήματος περιέχει ή όχι μουσική, τέσσερα είναι τα κύρια χαρακτηριστικά γνωρίσματα που λαμβάνονται υπόψη:

1. η ένταση του σήματος
2. το επίπεδο του πλάτους
3. το επίπεδο του ύψους
4. η περιοδικότητά του

Το σύνολο των οργάνων που υποστηρίζονται από το VEMUS καλύπτουν εύρος συχνότητας περίπου 100-4000 Hz. Η συχνότητα δειγματοληψίας των 22 kHz είναι αρκετή για να ανιχνεύσει θεμελιώδεις συχνότητες σε αυτό το εύρος χωρίς σημαντική απώλεια ακρίβειας (ανάλυση συχνότητας).

Οι βαθμίδες αναγνώρισης λειτουργούν σε επίπεδο πλαισίων (frame-based). Για κάθε πλαίσιο εξάγεται ένα σύνολο χαρακτηριστικών. Η θεμελιώδης συχνότητα και η ενέργεια είναι τα σημαντικότερα από τα χαρακτηριστικά. Επιπρόσθετα χαρακτηριστικά περιλαμβάνουν την αρμονικότητα, ένα σύνολο ενεργειακών διαδρομών που λήφθηκε από μία παραμετροποιημένη τράπεζα φίλτρων, και χρονικά παράγωγα των ανωτέρω χαρακτηριστικών γνωρισμάτων. Για την σύγχρονη αναγνώριση χρησιμοποιούνται πλαίσια ανάλυσης 10ms για να καλυφθούν οι απαιτήσεις ταχύτητας. Για την ασύγχρονη αναγνώριση χρησιμοποιούνται πλαίσια τουλάχιστον κάθε 5ms για να καλυφθούν οι απαιτήσεις χρονικής ανάλυσης. Για αυξημένη ακρίβεια, το ασύγχρονο σύστημα αναγνώρισης

χρησιμοποιεί ένα πιο αναλυτικό σχέδιο εξαγωγής χαρακτηριστικών σε σχέση με τη σύγχρονη βαθμίδα αναγνώρισης.

Η εκτίμηση της θεμελιώδους συχνότητας και της σύγχρονης και της ασύγχρονης βαθμίδας αναγνώρισης του VEMUS είναι βασισμένη στις τεχνικές συσχέτισης. Αυτή η επιλογή παρακινείται από τη σχετική ευρωστία αυτών των τεχνικών αναφορικά με τις συνθήκες ηχογράφησης, τα μικρόφωνα και τον περιβαλλοντικό θόρυβο που μπορεί να υπάρχει κατά τη διάρκεια μιας μουσικής εκτέλεσης. Οι απαιτήσεις για την υπολογιστική ταχύτητα σε πραγματικό χρόνο μπορούν να επιτευχθούν μέσω της ευρετικής βελτιστοποίησης κώδικα, π.χ. περιορισμός των χρονοβόρων υπολογισμών συσχέτισης σε ένα διάστημα κοντά στην τελευταία γνωστή θεμελιώδη συχνότητα.

Η σύγχρονη και ασύγχρονη βαθμίδα αναγνώρισης είναι βασισμένες σε διαφορετικές στρατηγικές μετά-επεξεργασίας της θεμελιώδους συχνότητας. Η πρώτη είναι σχεδιασμένη ώστε να καθυστερεί την απόφασή της για περίπου 50-100 ms (5-10 πλαίσια σημάτων) έως ότου συλλεχθούν επαρκή στοιχεία για μια σχετικά σταθερή θεμελιώδη συχνότητα (περιμένοντας μέχρι να αναγνωριστεί μια σειρά από πλαίσια στα οποία η θεμελιώδης συχνότητα παρουσιάζει απόκλιση εντός συγκεκριμένων ορίων). Η δεύτερη βαθμίδα ακολουθεί την προσέγγιση «islands of confidence» που συνίσταται στον καθορισμό προτεραιοτήτων των αποφάσεων για την θεμελιώδη συχνότητα. Οι λιγότερο διαφορούμενες αποφάσεις λαμβάνονται πρώτα, και οι εναπομένουσες πιο διαφορούμενες περιπτώσεις ύψους περιορίζονται για να είναι συνεπείς προς τις πρώτες. Η προσέγγιση «islands of confidence» επιτρέπει στις αποφάσεις να βασιστούν στο μελλοντικό πλαίσιο, βελτιώνοντας την ακρίβεια εκτίμησης της θεμελιώδους συχνότητας.

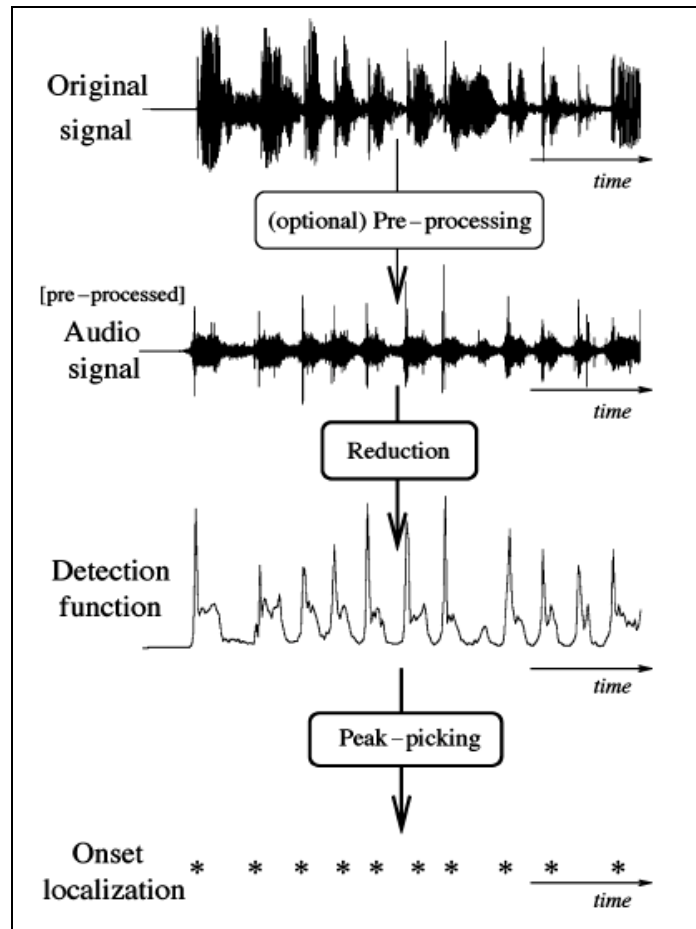
2.5 Κατάτμηση Ηχητικών Δεδομένων

Η χρονική κατάτμηση ενός ακουστικού κύματος σε πιο σύντομα τμήματα είναι ένα θεμελιώδες βήμα στο μετασχηματισμό των ήχων σε σημασιολογικά αντικείμενα (semantic objects).

Στις τελευταίες δύο δεκαετίες σημαντική έρευνα έχει αφιερωθεί σε αυτό το αντικείμενο και διαφορετικοί αλγόριθμοι έχουν αναπτυχθεί για να χωρίσουν αυτόματα τα σήματα μουσικής στα όρια των αντικειμένων του ήχου: όπου η νότα αρχίζει - onset - και τελειώνει - offset [MR97, Kla99b]. Η εξαγωγή των χρόνων των onsets είναι χρήσιμη στις εφαρμογές επεξεργασίας ήχου για την ακριβή διαμόρφωση των attacks των ήχων [Mas96, JR01], βοηθά τα συστήματα μεταγραφής στον εντοπισμό των onset των νοτών [Bel03, Kla04], και μπορεί να χρησιμοποιηθεί σε προγράμματα σύνταξης ήχων (sound editors) για τον διαχωρισμό των ηχητικών αρχείων στα λογικά τους μέρη [Smi96]. Οι μέθοδοι ανίχνευσης των onsets έχουν χρησιμοποιηθεί για την ταξινόμηση της μουσικής [GD04] και τον χαρακτηρισμό ρυθμικών προτύπων [DGW04]. Διάφορα συστήματα για ανίχνευση ρυθμού χρησιμοποιούν τα ανιχνευμένα onset για να συμπεράνουν τη θέση των beats [Sch98b, DP04]. Ένα σύστημα ικανό να ανιχνεύει αυτούς τους χρόνους των onset καθώς εμφανίζονται, ακριβώς όπως ο άνθρωπος, επιτρέπει νέες αλληλεπιδράσεις μεταξύ ακουστικών και συνθετικών οργάνων [PAZ98]. Η καθιέρωση εύρωστων μεθόδων για τη χρονική ανίχνευση των onset σε πραγματικό χρόνο είναι ένας σημαντικός στόχος για την ανάπτυξη μουσικών εγκαταστάσεων και διαδραστικών συστημάτων.

Διάφορες προσεγγίσεις για την ανίχνευση των onset σε μουσικούς ήχους παρουσιάζονται στη διεθνή βιβλιογραφία από χρονικές και filter-bank τεχνικές έως και στατιστικές μεθόδους. Αυτές οι προσεγγίσεις έχουν γενικά δύο στόχους: την κατασκευή μιας συνάρτησης ανίχνευσης των αλλαγών στο σήμα, και την επιλογή της κορυφής αυτής της συνάρτησης ώστε να εξάγουμε τους αντιλαμβανόμενους χρόνους των onset [RSB05]. Το διάγραμμα ροής που παρουσιάζεται παρακάτω

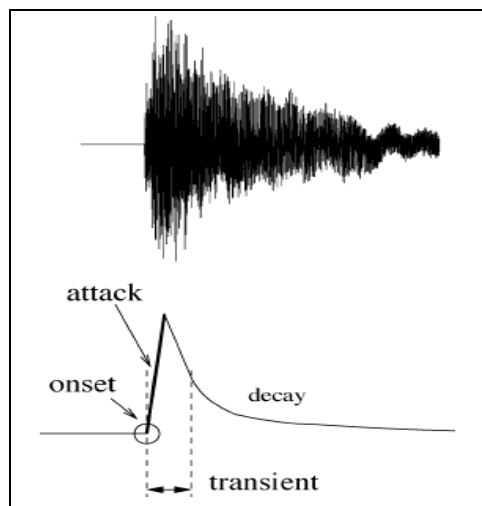
δείχνει τα βήματα που χρησιμοποιούνται για το διαχωρισμό των μουσικών σημάτων σε γεγονότα νοτών (note events). Διαφορετικές τεχνικές για κάθε στάδιο έχουν προταθεί από ερευνητές.



Εικόνα 25: Διάγραμμα ροής για το διαχωρισμό των μουσικών σημάτων σε γεγονότα νοτών

2.6 ΑΝΙΧΝΕΥΣΗ ΤΗΣ ΑΡΧΗΣ (ONSET) ΤΗΣ ΝΟΤΑΣ

Στο σημείο αυτό είναι απαραίτητο να διαχωρίσουμε σαφώς τους όρους transient, onset και attack μιας νότας. Στην παρακάτω εικόνα φαίνονται καθαρά οι τρεις όροι στην απλή περίπτωση μιας απομονωμένης νότας.



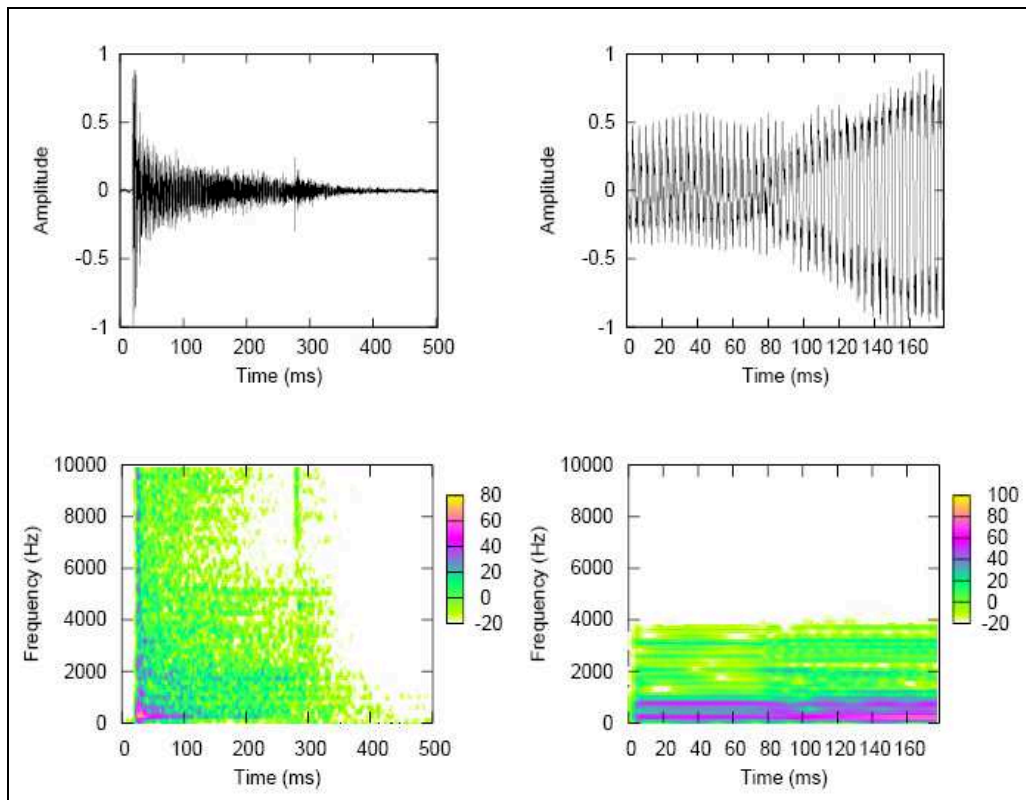
Εικόνα 26: Onset, attack, transient και decay μιας νότας

- Το attack της νότας είναι το χρονικό διάστημα κατά το οποίο η περιβάλλουσα του πλάτους αυξάνει.
- Η έννοια του transient είναι πιο δύσκολο να περιγράψει έπ' ακριβώς. Με έναν άτυπο και αρχικό ορισμό το transient είναι σύντομα χρονικά διαστήματα κατά τα οποία το σήμα μεταβάλλεται ταχέως κατά απρόβλεπτο τρόπο. Στην περίπτωση των ακουστικών οργάνων, το transient αντιστοιχεί στο χρονικό διάστημα κατά το οποίο εφαρμόζεται η διέγερση και στη συνέχεια μειώνεται αφήνοντας μια αργή εξασθένιση στις συχνότητες συντονισμού του οργάνου. Είναι σημαντικό να αναφερθούμε στην χρονική ανάλυση: το ανθρώπινο αυτί δεν μπορεί να διαχωρίσει 2 transients που απέχουν λιγότερο από 10 ms μεταξύ τους. Κατά τη διάρκεια του offset της νότας επίσης παρατηρείται μια περίοδος transient.
- Το onset μιας νότας είναι η χρονική στιγμή που ξεκινάει το transient (ή η στιγμή που μπορούμε να διακρίνουμε την αρχή του).

Οι [MR97] περιγράφουν το αντιληπτό onset σε ένα μουσικό σήμα ως την «αντιληπτή αρχή ενός διακριτού γεγονότος, που καθορίζεται από μια αξιοπρόσεκτη αύξηση στην ένταση, ή από μια ξαφνική αλλαγή στο ύψος ή στη χροιά». Ο όρος onset detection αναφέρεται εδώ στην ανίχνευση των αρχών των διακριτών γεγονότων στα ακουστικά σήματα [Kla99b].

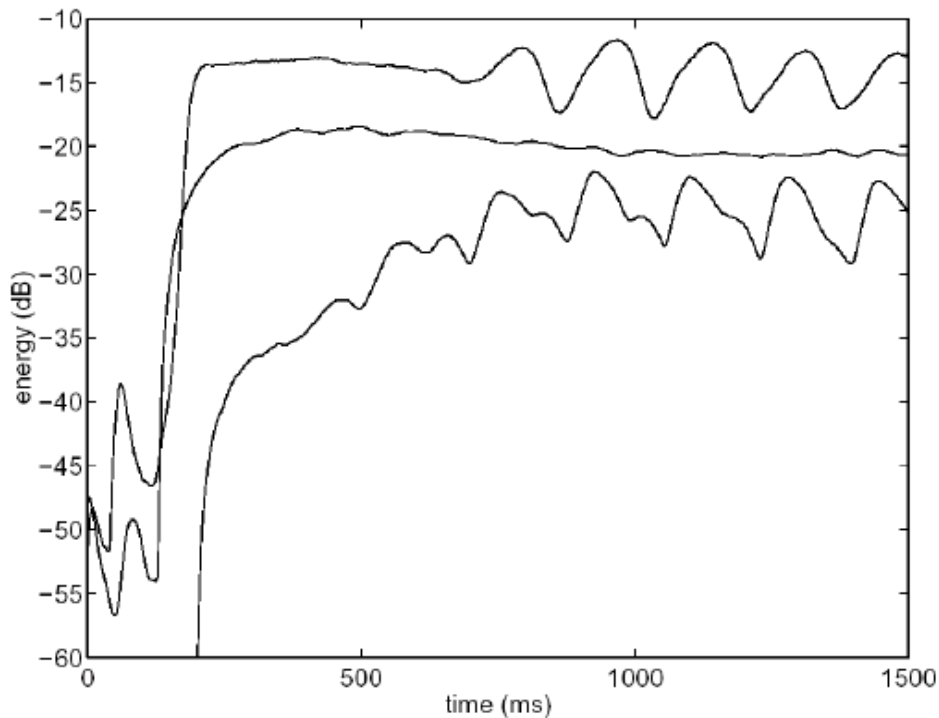
Δύο παραδείγματα ήχων που παρουσιάζουν onset παρουσιάζονται στο σχήμα 22. Ο ήχος τύμπανων στο αριστερό του σχήματος 23 παράγεται από snare τύμπανο και ξεκινά μετά από περίπου 10 ms, όπως φαίνεται στην κυματομορφή με μια ξαφνική αύξηση του πλάτους, και στο spectrogram με μια αύξηση της ενέργειας σε όλες τις συχνότητες του φάσματος. Οι κρουστικοί ήχοι όπως τα τύμπανα ή οι χορδές εμφανίζουν συχνά αιχμηρά attacks, που παρουσιάζουν μια ξαφνική αύξηση της ενέργειας στην κυματομορφή τους και μια ευρείας ζώνης «έκρηξη» στο φάσμα τους. Θα αναφερόμαστε σε αυτές τις ευρυζωνικές ενεργειακές εκρήξεις ως κρουστικές αρχές (percussive onsets). Μια δεύτερη αρχή, λιγότερο δυνατή από την πρώτη, φαίνεται στον κρουστικό ήχο του σχήματος 23, μετά από περίπου 270 ms. Αν και λιγότερο προφανής απ' ότι η πρώτη, αυτό το δεύτερο γεγονός παρουσιάζει επίσης μια ευρυζωνική αύξηση ενέργειας.

Οι χροιές άλλων οργάνων, όπως των πνευστών ή των εγχόρδων, παρουσιάζουν ομαλές μεταβάσεις από μια νότα σε άλλη, και ο εντοπισμός αυτών των αλλαγών είναι δυσδιάκριτος. Στην Εικόνα 22 παρουσιάζεται στην κυματομορφή της βιόλας (δεξιά στήλη) η μετάβαση μεταξύ δύο νοτών με διαφορετικά ύψη. Στον ήχο της βιόλας η αλλαγή από ένα ύψος σε άλλο συμβαίνει χωρίς αξιοπρόσεχτη αλλαγή στην ηχηρότητα ή τη χροιά. Τα Onset στα μη-κρουστά όργανα αναφέρονται ως τονικές αρχές (tonal onsets). Ο χαρακτηρισμός των αρχών δεν είναι εύκολος, δεδομένου ότι μπορούν να καθοριστούν από τις αλλαγές στην ηχηρότητα, το ύψος και τη χροιά. Η ενέργεια του σήματος μπορεί να μεταβληθεί από παρουσία tremolo, και η συχνότητα μπορεί να μεταβληθεί με vibrato. Αυτοί οι βαθμιαίες αλλαγές θεωρούνται ως μεταβολές στο εύρος ή τη συχνότητα, αλλά όχι ως διακριτά γεγονότα.



Εικόνα 27: Παραδείγματα ήχων όπου παρουσιάζονται τα onset

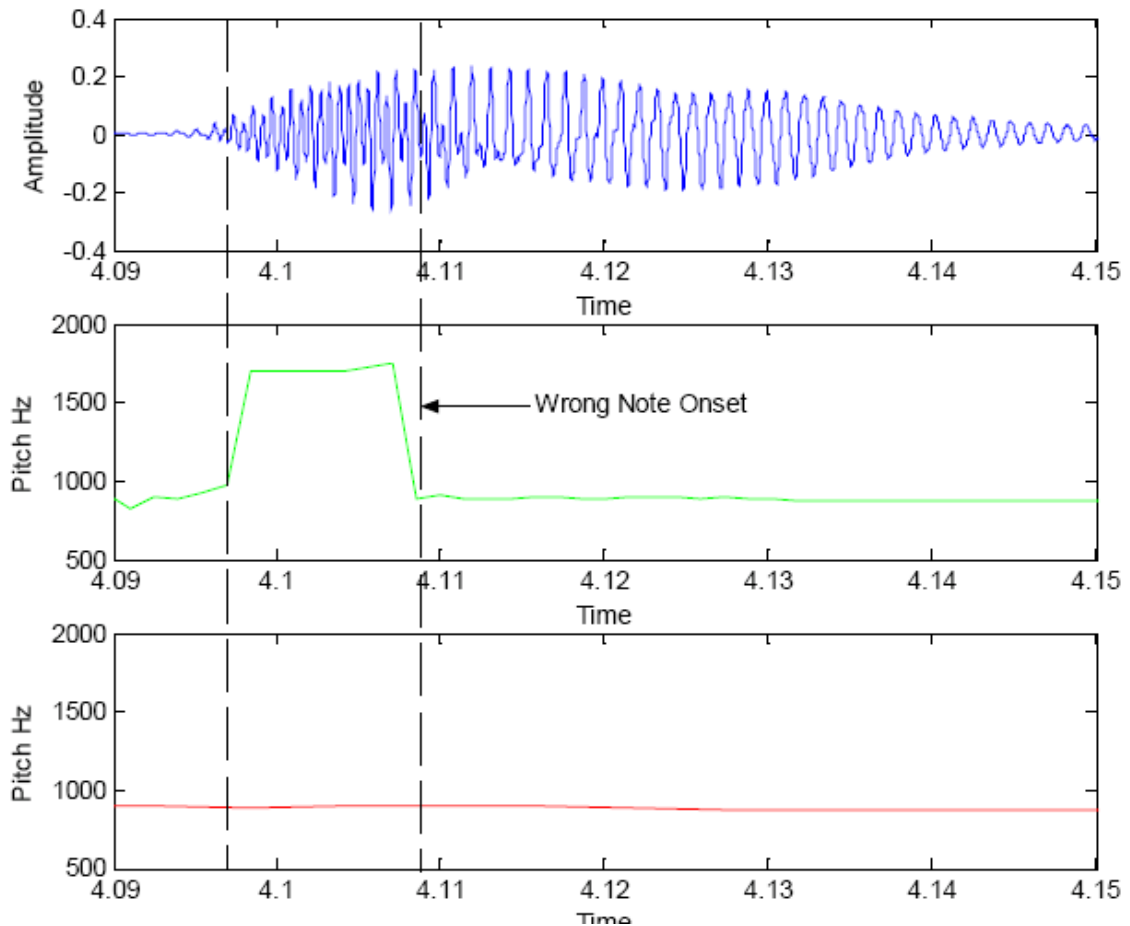
Γενικά ο στόχος ανίχνευσης της αρχής της νότας εκτελείται με την έρευνα της αλλαγής στο φασματικό περιεχόμενο της ακολουθίας των επεξεργασμένων πλαισίων του σήματος, που παρέχει μια σχετικά καλή λύση. Εντούτοις, μια πιο ακριβής λύση θα ήταν η ανίχνευση σημαντικών αλλαγών μέσα σε μια συγκεκριμένη ζώνη και όχι σε όλο το φάσμα, όπως έχει παρουσιαστεί από τον Klarinι στο [KUS01]. Αυτή η μέθοδος είναι ικανοποιητική στην περίπτωση όπου η αρχή της νότας (στιγμή του attack) και η σταθερή κατάσταση της νότας (steady state) έχουν σημαντική απόσταση. Στο [Mar98], καθώς επίσης και στα αποτελέσματά μας με τα διαφορετικά όργανα καταγραφής, έχει φανεί ότι το όργανο καταγραφής δεν πάσχει από αυτό το φαινόμενο. Αντίθετα, η αρχή του οργάνου καταγραφής και η σταθερή κατάσταση είναι σχεδόν στιγμιαίες.



Εικόνα 28:Φάκελοι εύρους για τους τόνους που παράγονται από (από κάτω προς τα πάνω) βιολί, τρομπέτα και φλάουτο. Η αρχή του φλάουτου είναι σχεδόν στιγμιαία. (Πηγή [Mar98])

Αυτό το χαρακτηριστικό μας επιτρέπει να έχουμε μεγαλύτερη βεβαιότητα στον καθορισμό των ορίων των νοτών ανιχνεύοντας τη σημαντικότερη αλλαγή για κάθε πιθανό τμήμα νοτών.

Στην περίπτωσή μας, η ανίχνευση onset και offset των νοτών εκτελείται με ευρετικούς κανόνες. Μετά τον υπολογισμό του μέσου όρου, το περίγραμμα ύψους του σήματος χωρίζεται σε « διαστήματα εμπιστοσύνης (islands of confidence)», που αποτελούνται από τουλάχιστον μια σταθερή κατάσταση ελάχιστου μήκους. Το ελάχιστο μήκος για μια σταθερή κατάσταση αντιπροσωπεύει την πιο σύντομη νότα που επιτρέπεται στην εκτέλεση, αφού πιο σύντομα τμήματα από αυτό θα απορριφθούν και θα αποδοθούν σε γειτονικές νότες που έχουν ανιχνευτεί. Μετά από αυτήν την αρχική κατάτμηση του σήματος σε νότες, ο αλγόριθμος ανίχνευσης ύψους τρέχει πάλι με μεγαλύτερα μήκη πλαισίων προκειμένου να επιτευχθεί υψηλότερη ανάλυση και ακρίβεια συχνότητας. Το μικρό μήκος πλαισίων στο πρώτο στάδιο μας παρέχει μια πολύ καλή ανάλυση, ενώ το δεύτερο πέρασμα του σήματος με πλαίσια μεγαλύτερου μήκους μας παρέχει υψηλή ανάλυση συχνότητας. Λανθασμένες εκτιμήσεις και λανθασμένα αναγνωρισμένες νότες απορρίπτονται κατά τη διάρκεια του δεύτερου πέρασματος, ενώ τα χρονικά όρια των νοτών παραμένουν ακριβή.



Εικόνα 29: Λανθασμένη νότα στο 1^ο στάδιο και σωστή στο 2^ο

Μαζί με τις αναγνωρισμένες νότες και τις αντίστοιχες διάρκειες, η βαθμίδα ακουστικής αναγνώρισης είναι αρμόδια για τον υπολογισμό των απαραίτητων παραμέτρων για την βαθμίδα αξιολόγησης της απόδοσης. Αυτοί είναι:

- Τα 4 πρώτα partials (αρμονικές, συχνότητες και επίπεδα).
- Η περιβάλλουσα του σήματος.
- Η ταχύτητα
- Τα επίπεδα εμπιστοσύνης για κάθε πλαίσιο

Επιπλέον παράμετροι, προκειμένου να ερευνηθεί περαιτέρω η ακουστική αναγνώριση, έχουμε προσθέσει μερικές πρόσθετες παραμέτρους στο μαρκάρισμα κάθε πλαισίου. Οι περισσότεροι από αυτές τις παραμέτρους χρησιμοποιούνται ως μετρικές εμπιστοσύνης για την ισχύ της αναγνώρισης σημειώσεων.

Αυτές οι τέσσερις παράμετροι είναι:

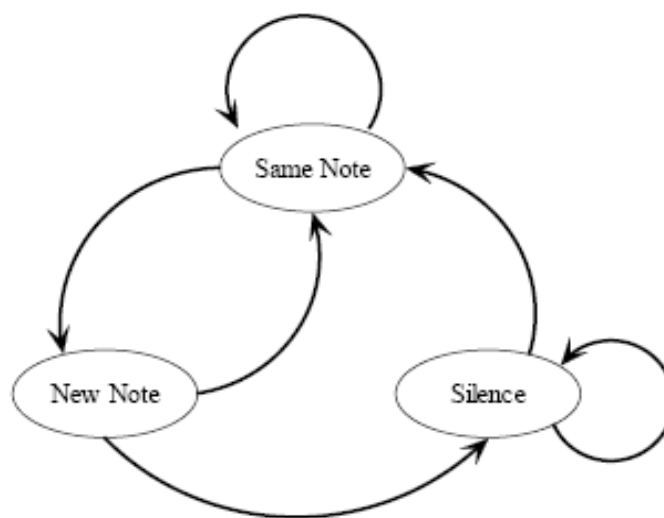
- **Το μουσικό ύψος του ήχου.** Όταν το μουσικό ύψος του ήχου (pitch), είναι αρκετά υψηλό (πέρα από ένα συγκεκριμένο κατώτατο όριο) η εκτίμηση θεωρείται σωστή.
- **Η διαφορά μεταξύ του μουσικού ύψους του ήχου και του επιπέδου της πρώτης αρμονικής.** Αυτή η παράμετρος είναι σημαντική για τη σαφήνεια του ήχου που παίζεται από τον μαθητή. Οι αρμονικές του σήματος παρέχουν επίσης μια καλή εκτίμηση της ποιότητας του ήχου, αφού όταν η F0 αιχμή είναι αρκετά υψηλότερη από τη δεύτερη f₁ το attack θεωρείται καλό.

- **Η σχετικότητα των αρμονικών και των ηχητικών επιπέδων τους.** Η απόσταση μεταξύ των διαδοχικών αρμονικών αντιπροσωπεύει μια καλή εκτίμηση της τιμής του ύψους, ενώ η διάκριση των αρμονικών αποτελεί ένα καλό μέτρο της σαφήνειας της νότας.
- **Η ένταση της νότας.** Η χαμηλή ένταση μιας απομονωμένης σύντομης νότας μπορεί να είναι σημάδι μιας πλαστής εκτίμησης.
- **Οι βασισμένες στα συμφραζόμενα πληροφορίες του περιγράμματος ύψους.** Αυτή η παράμετρος, αν και ευρετική, είναι αρκετά σημαντική προκειμένου να φιλτραριστούν οι πλαστές εκτιμήσεις, προκαλούμενες από τους γνωστούς περιορισμούς που εισάγονται στη μέθοδο από την «παραθύρωση» ενός συνεχούς σήματος. Όταν μια σύντομη νότα ανιχνεύεται και οι γειτονικές νότες εκπληρώνουν συγκεκριμένες συνθήκες, η νότα μπορεί να είναι με σιγουριά λανθασμένη εκτίμηση, που προκαλείται από συγκεκριμένους λόγους.

Αυτές οι παράμετροι ερευνώνται αυτήν την περίοδο προκειμένου να καθοριστεί ένα ακριβές μέτρο για την ποιότητα της ακουστικής αναγνώρισης εντούτοις αυτή η έρευνα διεξάγεται ακόμα.

2.7 Μηχανή Απόφασης 3 Επιπέδων

Όπως αναφέρεται ήδη, ο στόχος της βαθμίδα σύγχρονης αναγνώρισης είναι να στείλει τα μηνύματα γεγονότος νότας σε μια κεντρική δομή, ώστε να το χρησιμοποιήσουν και οι άλλες ενότητες. Προκειμένου να γίνει αυτό, η βαθμίδα χρησιμοποιεί έναν ενδιάμεσο γράφο μετάβασης 3 καταστάσεων όπου οι αντίστοιχες καταστάσεις μπορούν να είναι είτε η ίδια νότα, είτε διαφορετική, είτε σιωπή.



Εικόνα 30: Γραφική παράσταση μηχανής απόφασης 3 επιπέδων

Κατά τη διάρκεια του υπολογισμού του περιγράμματος του ύψους, προκειμένου να αποβληθούν πλαστές εκτιμήσεις, προκαλούμενος από θόρυβο ή απότομες αλλαγές των επιπέδων του ήχου, ή τα πλαίσια που χαρακτηρίζονται εσφαλμένα ως μουσική ή μη-μουσική, υπάρχει μια κοινή στρατηγική που χρησιμοποιεί πληροφορίες βασισμένες στα συμφραζόμενα γύρω από το πλαίσιο, και όχι μόνο στα πλαίσια που προηγούνται. Ως εκ τούτου, κατορθώνουμε να αποφύγουμε λανθασμένες εκτιμήσεις του επιπέδου ύψους για κάθε πλαίσιο και ως εκ τούτου λανθασμένα μηνύματα γεγονότος νότας που στέλνονται από αυτήν την ενότητα. Στην περίπτωση του VEMUS, στην σύγχρονη βαθμίδα

αναγνώρισης, χρησιμοποιούμε τρία διαδοχικά πλαίσια προκειμένου να αποφασίσουμε για το δεύτερο πλαίσιο. Πρέπει να τονίσουμε εδώ ότι αυτό εισάγει μια καθυστέρηση μήκους ενός πλαισίου (δηλ. 11msec) στο μήνυμα γεγονότος νότας, εντούτοις αυτό είναι απαραίτητο για την αποφυγή των λανθασμένων εκτιμήσεων, ενώ η καθυστέρηση δεν είναι αξιοπρόσεχτος εκτός της περίπτωσης γρήγορων κομματιών με σύντομες νότες.

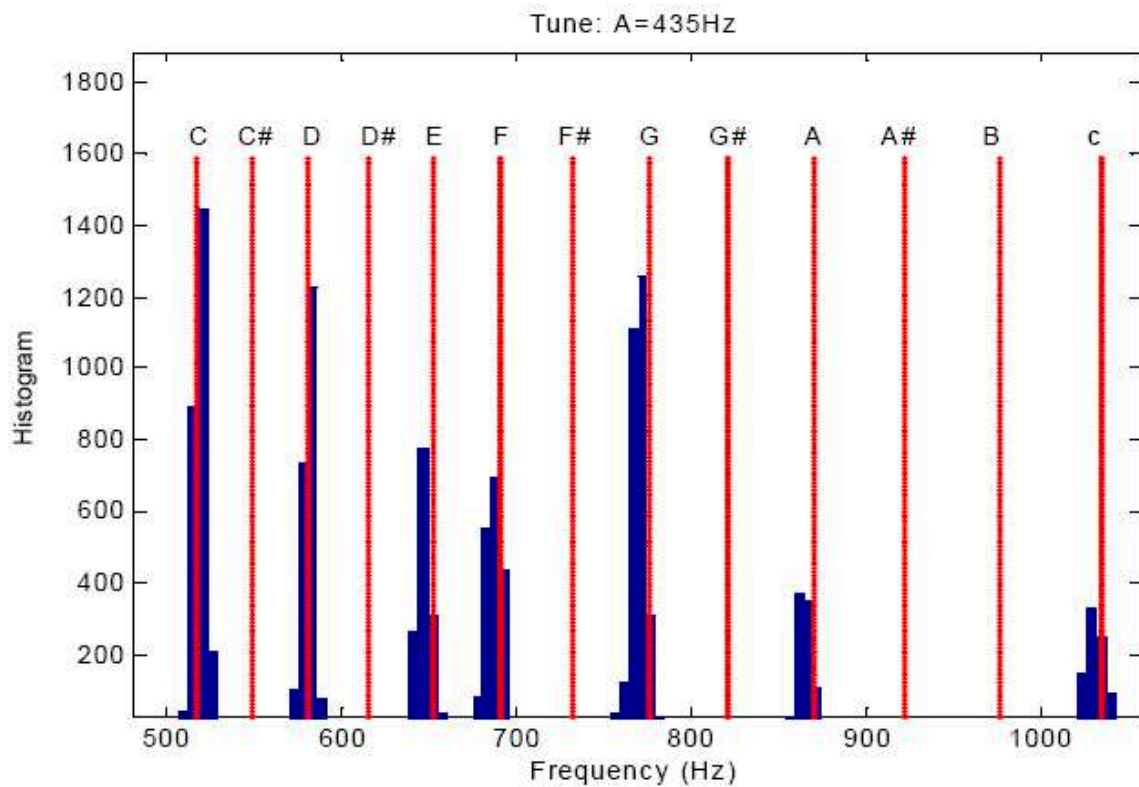
2.8 Προσδιορισμός Τόνου Οργάνου

Ο προσδιορισμός του τόνου ενός οργάνου, αναφέρεται ουσιαστικά στον προσδιορισμό του κουρδίσματός του. Αν και ο προσδιορισμός του τόνου του οργάνου είναι μια δραστηριότητα που πρέπει να επιτελείται στην αρχή της όλης διαδικασίας (στην αρχή της εκτέλεσης του προγράμματος), η συνεχής ανίχνευση του τόνου του οργάνου και κατά την διάρκεια της αναγνώρισης κάθε εκτέλεσης μπορεί να είναι απαραίτητη για διάφορους λόγους, όπως για παράδειγμα:

- Ο τόνος κάποιων οργάνων ενδέχεται να μεταβάλλεται ελαφρώς κατά την χρήση τους, μετά από κάποια ώρα. Σε αυτή την περίπτωση, το σύστημα θα πρέπει να «παρακολουθεί» τον μεταβαλλόμενο τόνο του οργάνου και να προσαρμόζει κατάλληλα την λειτουργία του.
- Δύο ή περισσότεροι μαθητές (ή ένας μαθητής και ο καθηγητής του) εκτελούν εναλλάξ, το ίδιο κομμάτι, με διαφορετικά όργανα. Σε αυτή την περίπτωση το σύστημα θα πρέπει να μπορεί να αντιλαμβάνεται και να αντιμετωπίζει κατάλληλα τους διαφορετικούς τόνους των οργάνων.

Ο προσδιορισμός του σωστού τόνου του οργάνου εκτελείται με την παραγωγή της συνάρτησης πυκνότητας πιθανότητας των κατ' εκτίμηση τιμών ύψους. Με αυτόν τον τρόπο, λαμβάνουμε τα επίπεδα νοτών γύρω από τα οποία η πλειοψηφία των κατ' εκτίμηση επιπέδων ύψους είναι διανεμημένη. Αυτή η μέθοδος αποδείχθηκε πολύ ακριβής, ενώ δεν προσθέτει μεγάλη υπολογιστική πολυπλοκότητα στο σύστημα.

Στο ακόλουθο σχήμα κάποιος μπορεί να δει τη κατανομή των επιπέδων ύψους ενός σήματος και τον τόνο του οργάνου που αναγνωρίζεται.



Εικόνα 31: Αυτόματη ανίχνευση τόνου της φλογέρας μέσω του PDF των τιμών ύψους

Η λειτουργία της αυτόματης ανίχνευσης του τόνου του οργάνου ενσωματώνεται μόνο στην ασύγχρονη βαθμίδα αναγνώρισης. Εντούτοις, θα μπορούσε να ενσωματωθεί επίσης στην σύγχρονη βαθμίδα δεδομένου ότι ίσως μπορούσε να βελτιώσει την ποιότητα των αποτελεσμάτων της σύγχρονης αναγνώρισης με σχετικά περιορισμένες επιπρόσθετες υπολογιστικές απαιτήσεις.

2.9 Ανάλυση Δεδομένων Δοκιμής

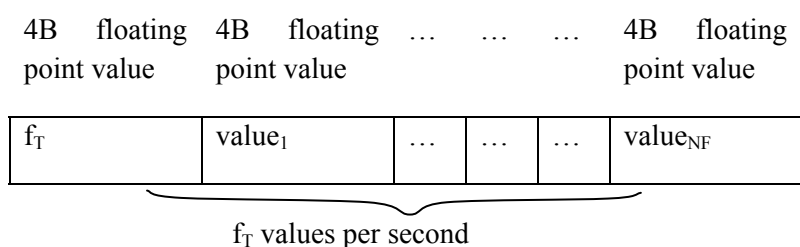
Για την αξιολόγηση των συστημάτων αναγνώρισης χρησιμοποιήθηκε ένα σύνολο στοιχείων δοκιμής. Αποτελείται από εκτελέσεις αρχαρίων και πειραμαμένων σπουδαστών που έχουν ηχογραφηθεί κάτω από ρεαλιστικές συνθήκες. Το σύνολο αυτών των δεδομένων χρησιμοποιήθηκαν επιπλέον κατά τη διάρκεια της φάσης ανάπτυξης για την εξαγωγή πληροφορίας δηλ. εξαγωγή των χαρακτηριστικών γνωρισμάτων για το κάθε όργανο και των κανόνων κατάταξης.

Παρακάτω περιγράφουμε τη διαδικασία που ακολουθήθηκε κατά την αξιολόγηση του προγράμματος αναγνώρισης μουσικής εκτέλεσης που υλοποιήθηκε.

Αρχικά βάζουμε ως είσοδο στο πρόγραμμα την ηχογραφημένη εκτέλεση σε μορφή .wav. Εκτελώντας το πρόγραμμα παίρνουμε ως έξοδο τα εξής αρχεία:

Όνομα αρχείου	Χαρακτηριστικά
f0.bin	Θεμελιώδης συχνότητα σε Hz
rms.bin	Ενεργός τιμή της ενέργειας
spectrogram.bin	Ακολουθία του φάσματος του σήματος
h1.bin	Πλάτος της 1 ^{ης} αρμονικής (θεμελιώδους)
h2.bin	Πλάτος της 2 ^{ης} αρμονικής
h3.bin	Πλάτος της 3 ^{ης} αρμονικής
h4.bin	Πλάτος της 4 ^{ης} αρμονικής
buffer.mtx	Αναγνωρισμένη εκτέλεση
buffer.Textgrid	Αναγνωρισμένη εκτέλεση προς οπτικοποίηση με το πρόγραμμα Praat

Όπου τα αρχεία **f0.bin**, **rms.bin**, **h1.bin**, **h2.bin**, **h3.bin**, και **h4.bin** έχουν την παρακάτω διαμόρφωση:

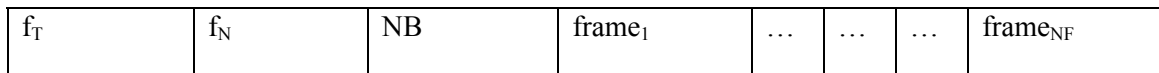


όπου

- f_T η συχνότητα δειγματοληψίας των f0/rms/h_i
- $NF = \text{διάρκεια ηχογράφησης} / f_T$;
- $\{value_1, \dots, value_N\}$ οι τιμές των χαρακτηριστικών.

Το αρχείο **spectrogram.bin** έχει την ακόλουθη διαμόρφωση:

4B floating point value 4B floating point value 4B integer value FFT data frame FFT data frame

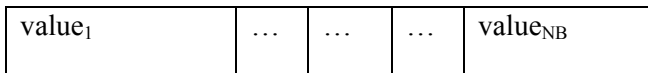


f_T frames per second

όπου

- f_T η συχνότητα δειγματοληψίας της ακολουθίας του φάσματος
- NF = διάρκεια ηχογραφήσεως / f_T
- f_N η συχνότητα Nyquist (μισή της συχνότητας δειγματοληψίας)
- NB είναι το πλήθος των bins συχνοτήτων για κάθε πλαίσιο του FFT (μισό του μεγέθους του FFT)
- Και κάθε πλαίσιο_i έχει την ακόλουθη δομή:

4B floating point value 4B floating point value



Προεπιλεγμένες τιμές (default values)

Οι υπολογισμοί του προγράμματος βασίζονται στις ακόλουθες προεπιλεγμένες τιμές των παραμέτρων:

- $f_T = 200$ Hz
- $f_N = 11025$ Hz (η αρχική κυματομορφή υπο-δειγματοληπτείται / υπέρ-δειγματοληπτείται στα 22050 Hz, αν είναι απαραίτητο)
- NB = 256 (το μέγεθος του FFT είναι 512 (23 msec))

Παρατηρήσεις στην εξαγωγή των χαρακτηριστικών

Εκτίμηση φασματογραφήματος

- Δεν υπάρχει επιλεκτική ενίσχυση συχνοτήτων
- Χρησιμοποιείται παραθύρωση Hamming
- Λαμβάνεται το ενεργειακό φάσμα.

Εκτίμηση αρμονικών

- Οι αρμονικές εκτιμώνται από το φάσμα του FFT
- Το φάσμα των FFT σαρώνεται για εύρεση ακμών γύρω από τις τιμές $NB * f_0 * i / f_N$
- Το πλάτος των ακμών και των γειτονικών δειγμάτων περνούν από τη συνάρτηση $\log_{10}()$
- Χρησιμοποιείται παραβολική παρεμβολή ανά τρεις τιμές του λογαρίθμου του φάσματος. Το μέγιστο που προκύπτει από την παραπάνω διαδικασία μετατρέπεται ξανά στη γραμμική κλίμακα και θεωρείται το πλάτος της i αρμονικής.
- Μη-αποδεκτές τιμές για τις αρμονικές είναι:
 - value = -1.0 (άφωνο πλαίσιο)
 - value = -2.0 (Η αρμονική ξεπερνά το NB (όρια του FFT)). Αυτό μπορεί να συμβαίνει για αρμονικές υψηλότερης τάξης ή ήχους μεγάλου ύψους)
 - value < 0 (Ακμή αρμονικής κοντά στο αντίστοιχο FFT bin δεν υπάρχει. Αυτό μπορεί να συμβαίνει για την 1^η αρμονική ήχων χαμηλού ύψους)

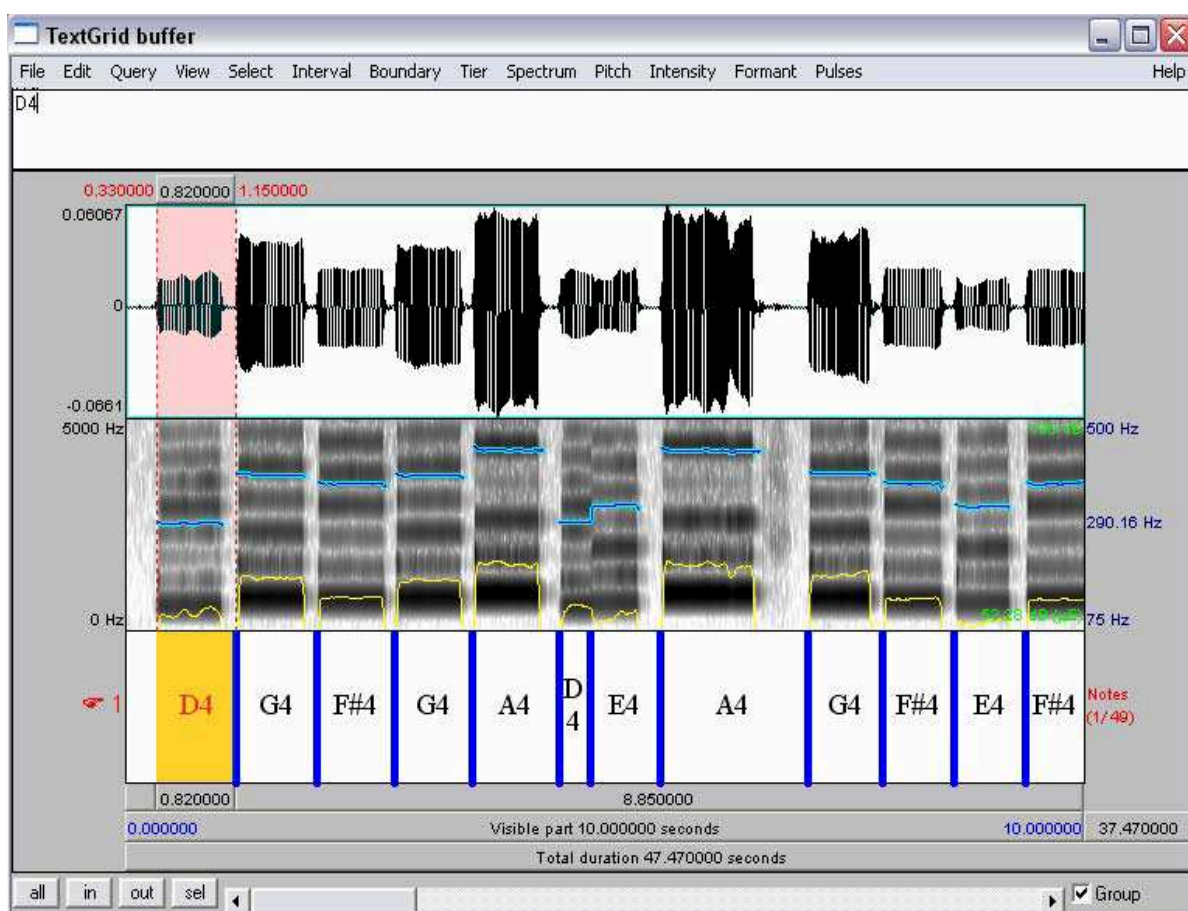
Εκτίμηση θεμελιώδους συχνότητας

- Η εκτίμηση της περιόδου γίνεται μέσω αυτοσυσχέτισης και ετεροσυσχέτισης, χρησιμοποιώντας μεταβλητό μέγεθος πλαισίου (ελάχιστο 10 msec). Χρησιμοποιούνται πλαίσια από δείγματα που λαμβάνονται ανά 5 msec;
- Η μετά-επεξεργασία της f_0 γίνεται με χρήση μιας μεταβλητής αναζήτησης δυναμικού προγραμματισμού
- Μη-αποδεκτές τιμές για την f_0 είναι:
 - value = 0.0 (άφωνο πλαίσιο)

Εκτίμηση ενεργής ισχύος

- Η εκτίμηση της ενεργής ισχύος γίνεται ταυτόχρονα με την εκτίμηση του ύψους, δηλαδή τα άφωνα πλαίσια έχουν σταθερό μήκος, ενώ για τα έμφωνα πλαίσια επιλέγεται να περιέχουν έναν ακέραιο αριθμό περιόδων της κυματομορφής.

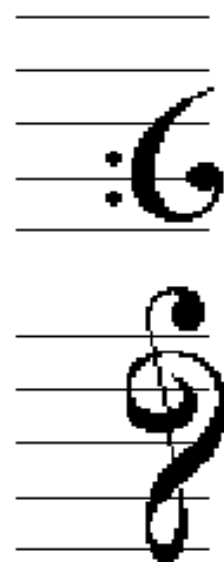
Εισάγοντας το ηχογραφημένο αρχείο (buffer.wav) και το buffer.Textgrid στο πρόγραμμα Praat έχουμε τις απαραίτητες πληροφορίες για την επεξεργασία του κομματιού. Οι νότες παρουσιάζονται χωρισμένες, σημειώνεται το onset και offset της κάθε νότας, η διάρκεια, το φάσμα, η ένταση καθώς και το ύψος της.



Εικόνα 32 : Περιβάλλον του προγράμματος Praat

Έχοντας στη διάθεσή μας τα αποτελέσματα του προγράμματος με αυτή τη μορφή μπορούσαμε να απομονώσουμε τις νότες και να εντοπίσουμε λανθασμένες εκτιμήσεις. Είχαμε επιπλέον στη διάθεσή μας τις παρτιτούρες αυτών των κομματιών έτσι ώστε να τις συγκρίνουμε με τα αποτελέσματα της αναγνώρισης καθώς και μετατροπείς υψών και ονομάτων νοτών όπως αυτός που παρουσιάζεται στην Εικόνα 33. Αντιστοιχίζοντας τα γεγονότα της εκτέλεσης και της παρτιτούρας εξάγαμε τα παρακάτω αποτελέσματα.

MIDI number	Note name	Keyboard	Frequency Hz	Period ms
21	A0		27.500	36.36
23	B0		30.868	29.135
24	C1		32.703	30.58
26	D1		36.708	34.648
28	E1		41.203	38.891
29	F1		43.654	22.91
31	G1		48.999	46.249
33	A1		55.000	51.913
35	B1		61.735	58.270
36	C2		65.406	15.29
38	D2		73.416	69.296
40	E2		82.407	77.782
41	F2		87.307	11.45
43	G2		97.999	92.499
45	A2		110.00	103.83
47	B2		123.47	116.54
48	C3		130.81	7.645
50	D3		146.83	138.59
52	E3		164.81	155.56
53	F3		174.61	5.727
55	G3		196.00	185.00
57	A3		220.00	207.65
59	B3		246.94	233.08
60	C4		261.63	3.822
62	D4		293.67	277.18
64	E4		329.63	311.13
65	F4		349.23	2.863
67	G4		392.00	369.99
69	A4		440.00	415.30
71	B4		493.88	466.16
72	C5		523.25	1.910
74	D5		587.33	554.37
76	E5		659.26	622.25
77	F5		698.46	1.432
79	G5		783.99	739.99
81	A5		880.00	830.61
83	B5		987.77	932.33
84	C6		1046.5	0.9556
86	D6		1174.7	1108.7
88	E6		1318.5	1244.5
89	F6		1396.9	0.7159
91	G6		1568.0	1480.0
93	A6		1760.0	1661.2
95	B6		1975.5	1864.7
96	C7		2093.0	0.4778
98	D7		2349.3	2217.5
100	E7		2637.0	2489.0
101	F7		2793.0	0.3580
103	G7		3136.0	2960.0
105	A7		3520.0	3322.4
107	B7		3951.1	3729.3
108	C8	J. Wolfe, UNSW	4186.0	0.2389



Εικόνα 33: Αντιστοιχία νοτών και συχνοτήτων

2.10 Αξιολόγηση Βαθμίδας/ Συμπεράσματα

Χρησιμοποιώντας ένα σύνολο από ηχογραφήσεις αρχαρίων που παρουσιάζει συγκεκριμένα χαρακτηριστικά, κάποια σημαντικά θέματα που οι βαθμίδες έπρεπε να αντιμετωπίσουν ήταν :

- Πλήθος από συνθήκες ηχογράφησης, μικρόφωνα και **περιβαλλοντικό θόρυβο** στις εκτελέσεις των μαθητών. Συχνά οι δάσκαλοι προσπαθούν να επικεντρώσουν την προσοχή του μαθητή στο tempo χτυπώντας τα δάχτυλα ή τα χέρια. Επίσης, συχνά οι εκτελέσεις διακόπτονται και οι δάσκαλοι συμβουλεύουν τους μαθητές.
- Ηχογραφήσεις δείχνουν ότι οι μαθητές αρκετές φορές εξασκούνται με όργανα με **κακό κορδίσμα**. Άρα δεν μπορεί να υποτεθεί ότι η μουσική κλίμακα έχει απόλυτη τιμή αναφοράς (A4=440Hz).
- Η φτωχή εκτέλεση νοτών από το σπουδαστή εισάγει διαστρεβλώσεις στον ήχο. Οι άνθρωποι αντιλαμβάνονται το ύψος τέτοιων νοτών με μια σημαντική **καθυστέρηση σχετικά με το onset**. Αυτό συμβαίνει και στην περίπτωση των ακουστικών συστημάτων αναγνώρισης. Οι μαθητές συχνά παίζουν νότες οι οποίες παρουσιάζουν διακυμάνσεις στο ύψος είτε η έντασή τους είναι πολύ χαμηλή.

Το πρόγραμμα μπόρεσε να αντιμετωπίσει επιτυχώς περιπτώσεις που η εκτέλεση διακόπτεται ώστε ο δάσκαλος να συμβουλεύσει τον μαθητή πάνω σε κάποια χαρακτηριστικά της εκτέλεσης. Επιπλέον, δεν παρουσίαζε καμία αδυναμία στην ανίχνευση του κορδίσματος του οργάνου στις περιπτώσεις άσχημα κορδισμένων οργάνων.

Αντίθετα σε ένα σύνολο περιπτώσεων, η βαθμίδα αναγνώρισης παρουσιάζει λάθη. Παρακάτω παρουσιάζονται ξεχωριστά οι περιπτώσεις αυτές μαζί με χαρακτηριστικά παραδείγματα, όπου παρουσιάζονται οπτικοποιημένα με το πρόγραμμα Praat τα αποτελέσματα της βαθμίδας αναγνώρισης μαζί με ηχητικά αποσπάσματα των αντίστοιχων περιπτώσεων. Επίσης, γίνεται μια έρευνα για τις αιτίες αυτής της αποτυχίας και ερευνώνται τρόποι βελτιστοποίησης.

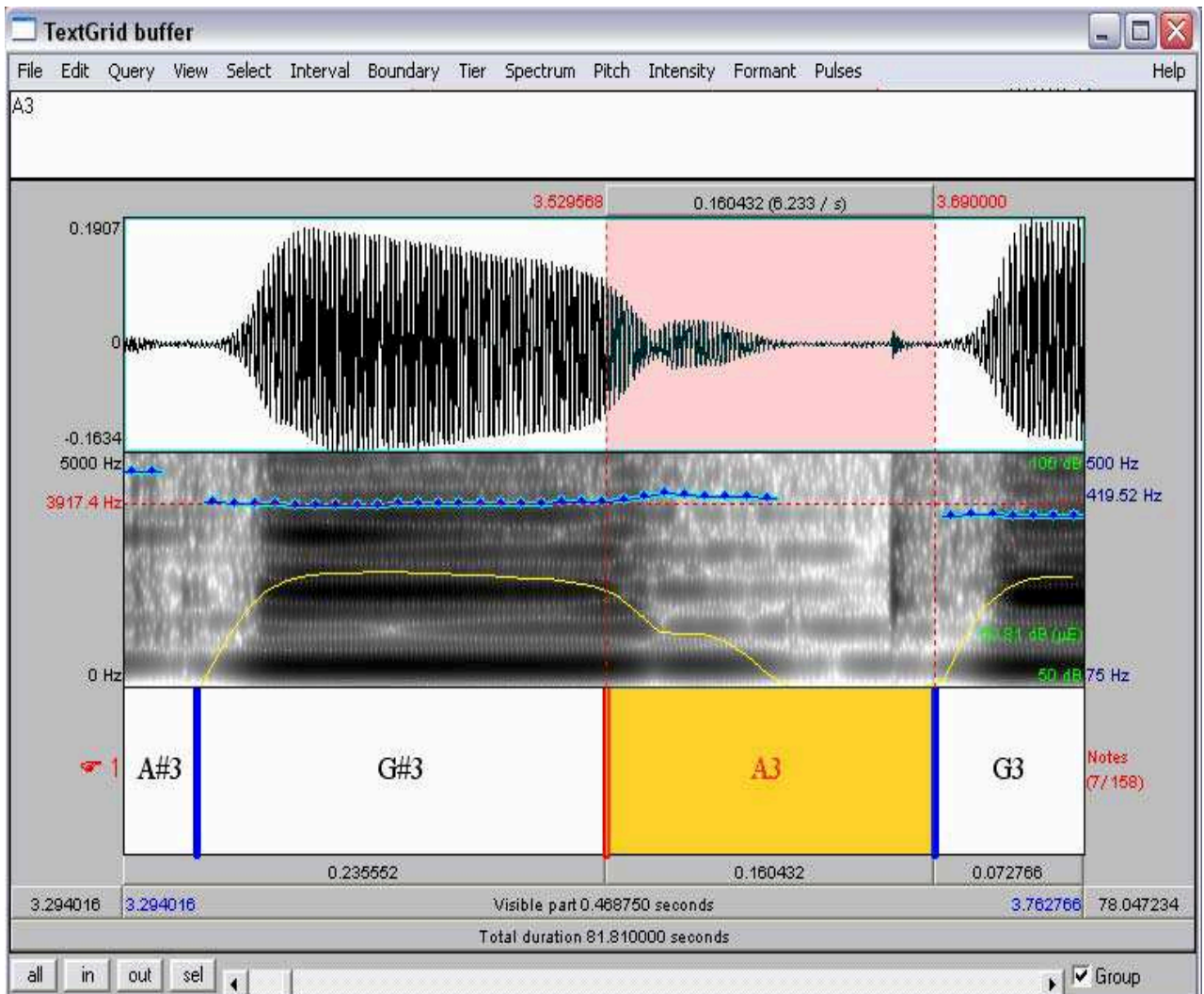
2.10.1 Διάσπαση νότας σε δύο με διαφορά ενός ημιτονίου

Ένα από τα κύρια λάθη που εμφανίζονται είναι η διάσπαση μιας νότας. Χαρακτηριστικό των νοτών οι οποίες συχνά διασπώνται είναι μικρή διακύμανση στο ύψος ταυτόχρονα με μικρή πτώση στην ένταση του ήχου. Κάποιες φορές η νότα η οποία αναγνωρίζεται δεν αποτελεί παρά άσχημο offset της νότας από τον αρχάριο μαθητή όπως φαίνεται στο παρακάτω παράδειγμα. Το ύψος το οποίο εντοπίζει η βαθμίδα αναγνώρισης στην νότα, η οποία στην πραγματικότητα αποτελεί τμήμα της προηγούμενης, παρουσιάζει μια μικρή απόκλιση από αυτή της προηγούμενης, τις περισσότερες φορές κατά ένα ημιτόνιο.

Η περίπτωση που η σπασμένη νότα δεν αναγνωρίζεται ως ίδια νότα αλλά με διαφορά ενός ημιτονίου από την συνέχειά της, το πρόβλημα δεν αντιμετωπίζεται έχοντας στην διάθεσή μας την παρτιτούρα, αφού η λανθασμένη νότα θα θεωρηθεί ως εισαγωγή νέας νότας και στην αξιολόγηση της εκτέλεσης του μαθητή θα προστεθεί ένα λάθος το οποίο θα οφείλεται κατά το μεγαλύτερο μέρος του σε λάθος της βαθμίδας αναγνώρισης (και κατά ένα μικρότερο σε κακή ποιότητα της νότας από το μαθητή).

Στην Εικόνα 33 έχουμε ένα χαρακτηριστικό παράδειγμα της παραπάνω περίπτωσης. Το τέλος της νότας G#3 (pitch=406 Hz) που εκτέλεσε ο μαθητής παρουσίαζε μια μικρή διακύμανση στο ύψος με αποτέλεσμα η βαθμίδα αναγνώρισης να εντοπίσει λανθασμένα μια νέα νότα, την A3 (pitch=419), η

οποία έχει διαφορά ενός ημιτονίου από την προηγούμενη. Παρατηρώντας το φασματογράφημα της αναγνωρισμένης ως A3 νότας αξίζει να τονίσουμε την απουσία φασματικού περιεχομένου στις υψηλές συχνότητες καθώς και την πτώση της έντασης σε όλη τη διάρκειά της.



Εικόνα 34 :Παράδειγμα διάσπασης νότας που οφείλεται σε μικρή διακύμανση του ύψους. Test-case «Martin_The Christmas song»_3.33 sec-3.699 sec.



Αρχείο ήχου : 3.33 sec -3.699 sec (G#3 και A3)

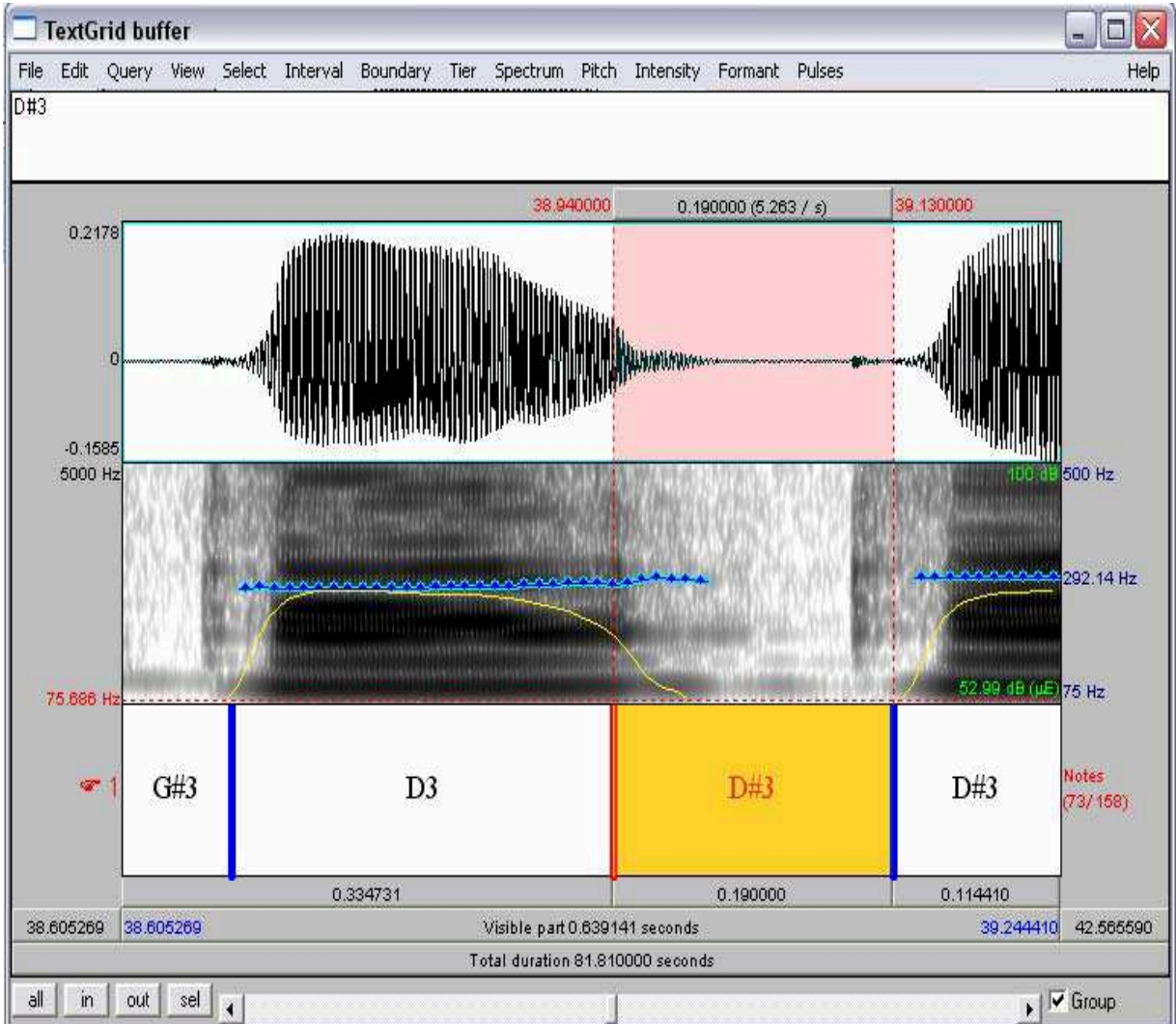


Αρχείο ήχου: 3.529 sec-3.699 sec (A3)

Από τα παραπάνω αρχεία ήχου τεκμηριώνεται και ακουστικά αυτό που παρατηρήσαμε από το φασματογράφημα των νοτών. Η περίπτωση αυτή θα μπορούσε να αντιμετωπιστεί χρησιμοποιώντας το χαρακτηριστικό της έντασης των νοτών και παρατηρώντας ότι σε όλη τη διάρκεια της A3 δεν παρουσιάστηκε παρά μόνο πτώση της σε σχέση με την προηγούμενη νότα. Το γεγονός αυτό σε συνδυασμό με τη μικρή διακύμανση στο ύψος (13 Hz) και τη μικρή διάρκεια της νότας A3 είναι αρκετά για να φτάσουμε στο συμπέρασμα ότι δεν πρόκειται παρά για μια μικρή διακύμανση στο ύψος της νότας στο τέλος της εκτέλεσής της.

Συνολικά, στον αλγόριθμο θα μπορούσε να προστεθεί ένα τμήμα κώδικα όπως:

{If $\text{pitch}(\text{note}_i) \approx \text{pitch}(\text{note}_{i+1})$ and $(\text{intensity}(\text{sample}_i) < \text{intensity}(\text{sample}_{i+1}))$ } then merge_notes



Εικόνα 35 : Παράδειγμα διάσπασης νότας που οφείλεται σε μικρή διακύμανση του ύψους. Test-case «Martin_The Christmas song»_39.68 sec-39.129 sec.



Αρχείο ήχου 39.68-39.129 sec

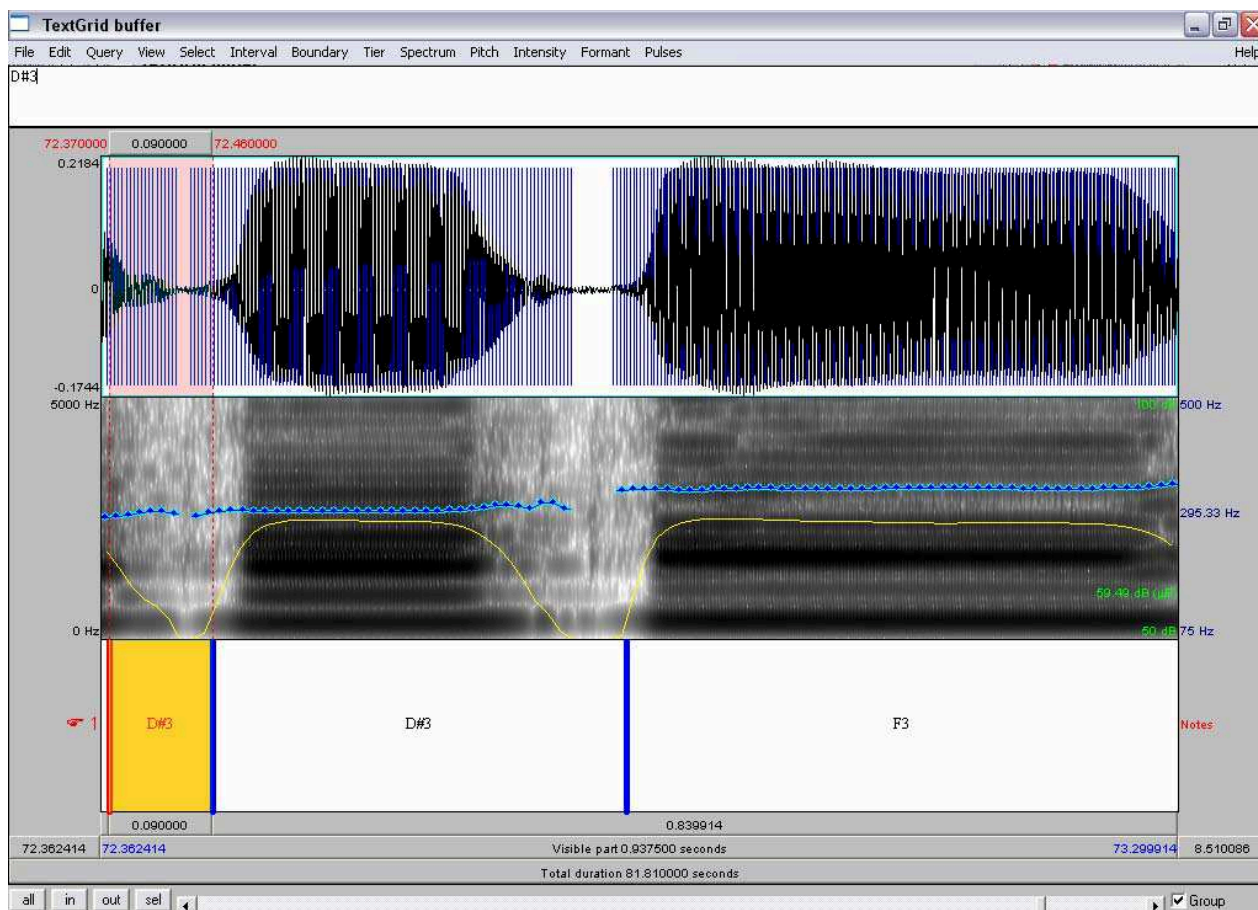


Αρχείο ήχου 38.94-39.129 sec

Στην Εικόνα 34 έχουμε το ίδιο λάθος της βαθμίδας αναγνώρισης σε μια διαφορετική περίπτωση δοκιμής. Το τέλος της νότας D3 (pitch=280.369 Hz) που εκτέλεσε ο μαθητής παρουσίαζε μια μικρή διακύμανση στο ύψος με αποτέλεσμα η βαθμίδα αναγνώρισης να εντοπίσει λανθασμένα μια νέα νότα, την D#3 (pitch=292.283), η οποία έχει διαφορά ενός ημιτονίου από την προηγούμενη. Παρατηρώντας το φασματογράφημα της αναγνωρισμένης ως D3 νότας αξίζει να τονίσουμε την απουσία φασματικού περιεχομένου στις υψηλές συχνότητες καθώς, την πτώση της έντασης σε όλη τη διάρκειά της και την απουσία υψηλών αρμονικών.

2.10.2 Διάσπαση νότας σε δύο ίδιες νότες

Χαρακτηριστικό αυτής της κατηγορίας λάθους της βαθμίδας αναγνώρισης είναι η διάσπαση μιας νότας σε δύο ίδιες νότες. Οι αιτίες σε αυτή την κατηγορία λάθους μπορεί να είναι διάφορες. Στο παρακάτω παράδειγμα η βαθμίδα αναγνώρισης σπάει τη νότα D#3 σε δύο. Η πρώτη νότα που αναγνωρίζεται διαρκεί 0.09 sec και αναγνωρίζεται ως νέα νότα εξαιτίας μιας διακοπής στο φύσημα από τον μαθητή (ασταθής ένταση). Από το φάσμα της παρατηρούμε ότι δεν έχει φασματικό περιεχόμενο στις υψηλές συχνότητες και από το ηχητικό αρχείο αποδεικνύεται ότι πρόκειται για άδεια νότα μέχρι να μπορέσει ο μαθητής να φυσήσει κατάλληλα για να εκτελέσει την επιδιωκόμενη νότα.



Εικόνα 36 : Παράδειγμα διάσπασης νότας που οφείλεται σε διακύμανση της έντασης. Test-case «Martin_The Christmas_song»_71.44 sec-71.53 sec.

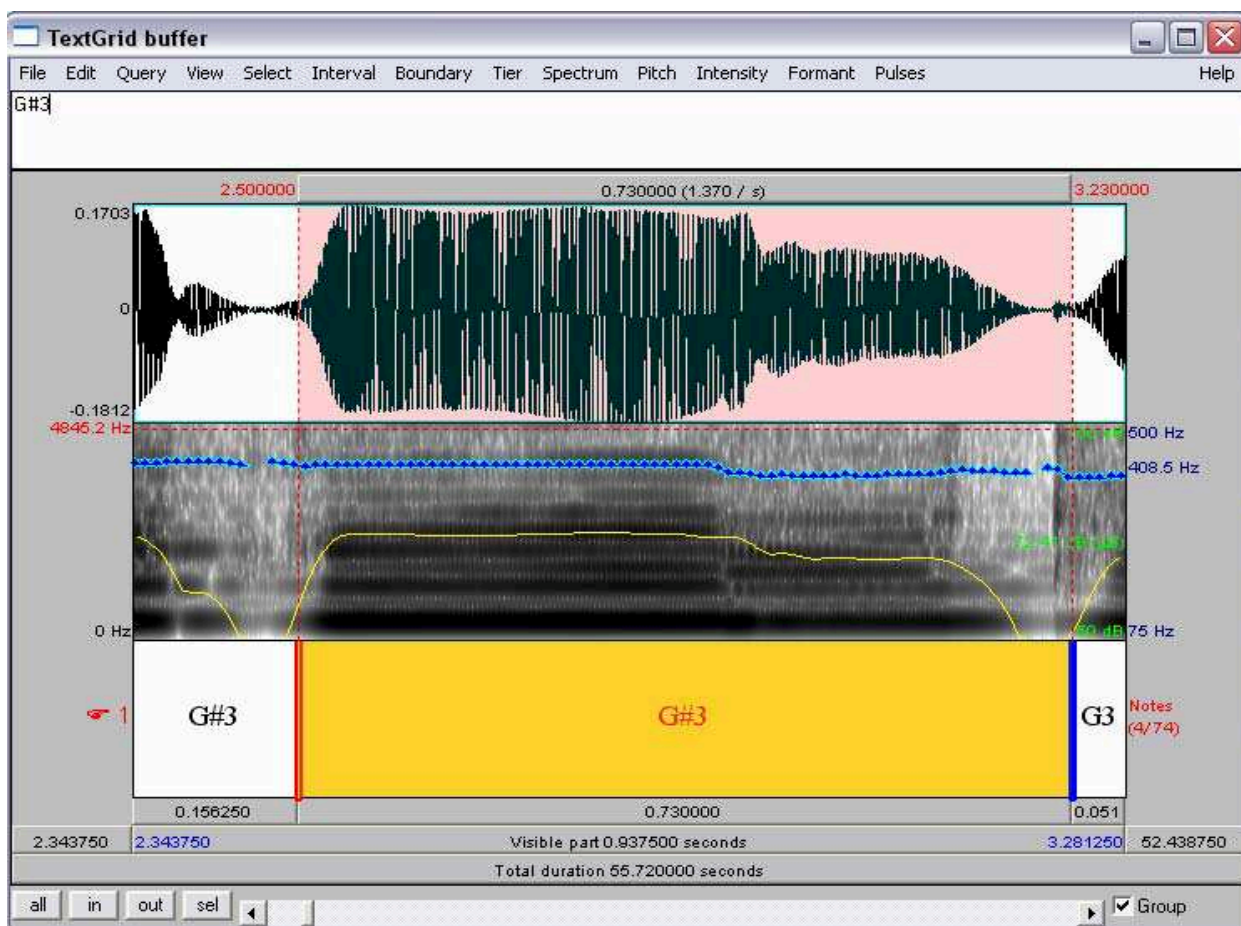


Αρχείο ήχου 71.44 sec-71.53 sec

Στην περίπτωση αυτή που μια νότα διασπάται σε δύο ίδιες, το λάθος της βαθμίδας αναγνώρισης μπορεί να διορθωθεί λαμβάνοντας υπόψη την παρτιτούρα και ενώνοντας τις δύο γειτονικές νότες που έχουν το ίδιο ύψος. Στο επόμενο κεφάλαιο αναλύεται η συμπεριφορά της βαθμίδας αντιστοίχισης της εκτέλεση με την παρτιτούρα στην περίπτωση αυτή. Επίσης αξίζει να σημειώσουμε ότι πολύ συχνά σε αυτή την κατηγορία λάθους της βαθμίδας αναγνώρισης, ένα μέρος της ευθύνης βαραίνει τους μαθητές αφού η εκτέλεσή τους δεν ήταν επαρκώς καλή. Έτσι, η βαθμίδα αξιολόγησης της εκτέλεσης, θα μπορούσε να προσθέσει ένα μικρό κόστος στο σύνολο των λαθών του μαθητή.

2.10.3 Συγχώνευση Νοτών με Διαφορά ενός Ημιτονίου σε 1/8^ο

Ένα επιπλέον σφάλμα που εμφανίζει η βαθμίδα αναγνώρισης είναι η συγχώνευση δύο ή ακόμα και τριών νοτών. Στο παρακάτω παράδειγμα οι νότες οι οποίες συγχωνεύονται έχουν διαφορά ενός ημιτονίου. Συγκεκριμένα από 2.50 sec- 2.89 sec ο μαθητής παίζει την νότα G#3 και στη συνέχεια την G3 έως τα 3,23 sec χωρίς όμως μηδενισμό της έντασης μεταξύ τους αφού έχουμε συζευγμένες νότες. Άλλο χαρακτηριστικό είναι πως οι νότες αυτές έχουν μικρή διάρκεια, εδώ 1/8^ο όπως φαίνεται και από την Εικόνα 38 όπου φαίνεται το αντίστοιχο σημείο της παρτιτούρας.



Εικόνα 37: Παράδειγμα συγχώνευσης νοτών με διαφορά ενός ημιτονίου. Test-case «Sankta_Lucia»_2.50 sec-3.23 sec.



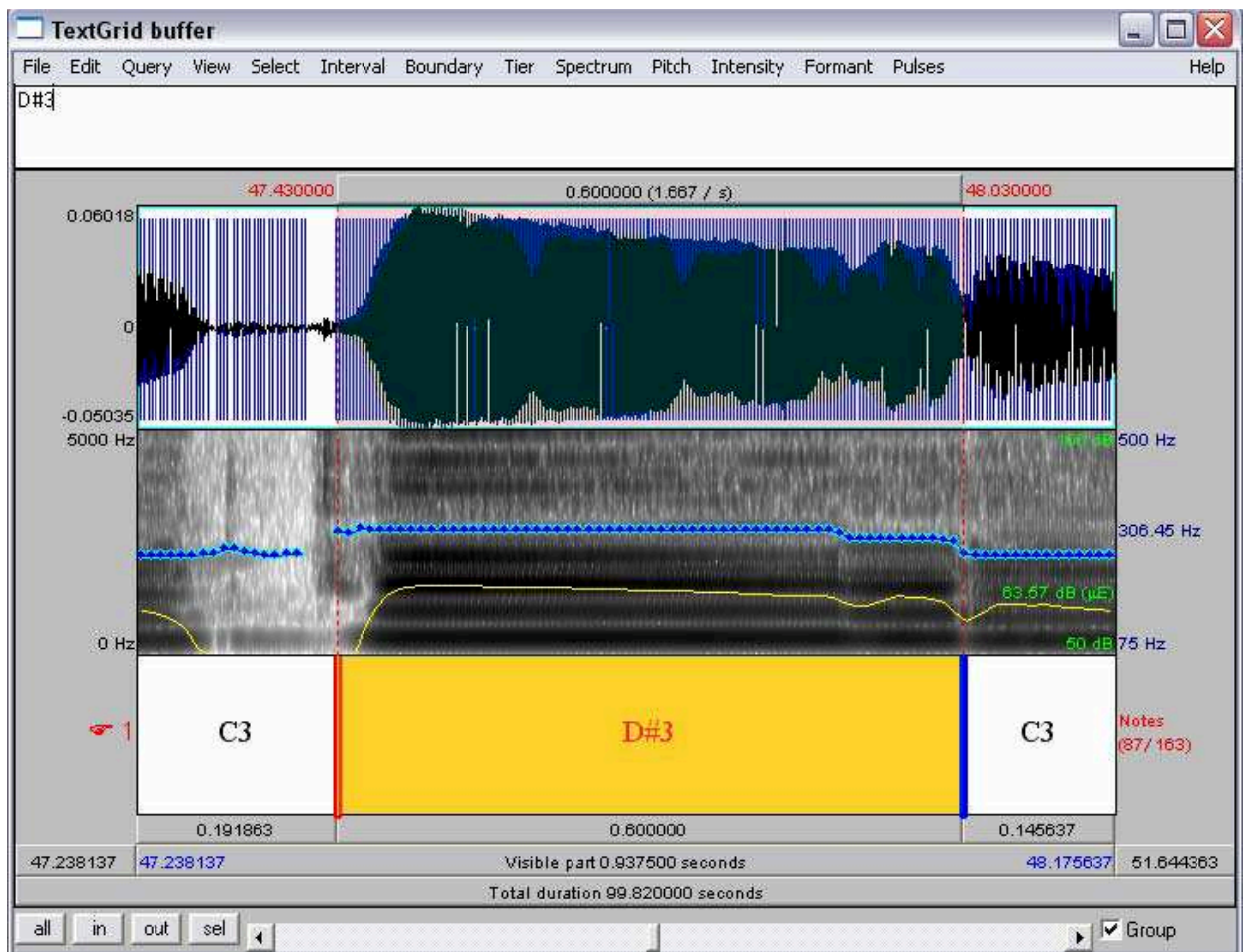
Αρχείο ήχου 2,5 sec-3,23 sec



Εικόνα 38 :Κομμάτι παρτιτούρας «Sankta_Lucia»

2.10.4 Συγχώνευση νοτών με διαφορά ενός ημιτονίου σε 1/16^ο

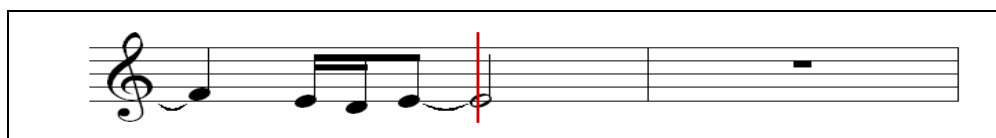
Στην Εικόνα 39 βλέπουμε την περίπτωση συγχώνευσης δύο νοτών διάρκειας 1/16^ο με διαφορά ενός ημιτονίου. Συγκεκριμένα στο χρονικό διάστημα 47.43 sec-47.90 sec το μέσο ύψος είναι 310 Hz και από 47.90 sec-48.03 sec το μέσο ύψος είναι 293 Hz. Η διαφορά ενός ημιτονίου σε συνδυασμό με τις μικρές διάρκειες των νοτών παραπλανεί την βαθμίδα αναγνώρισης. Σ' αυτό το σημείο αξίζει να σημειώσουμε ότι σε νότες με μικρές διάρκειες 1/16^ο ή και 1/8^ο η ένταση δεν προλαβαίνει να πέσει αισθητά ώστε να χρησιμοποιείται ως κριτήριο για εύρεση νέας νότας. Σε περίπτωση που έχουμε στη διάθεσή μας την παρτιτούρα του κομματιού, μπορούμε εκ των προτέρων να βρούμε τις νότες με μικρές διάρκειες και διαφορά ενός ημιτονίου, ώστε να χρησιμοποιήσουμε για το συγκεκριμένο πλαίσιο διαφορετικά βάρη στα χαρακτηριστικά που ανιχνεύουν τα onset.



Εικόνα 39 : Παράδειγμα συγχώνευσης νοτών με διαφορά ενός ημιτονίου. Test-case «Fredrik_Jag_vill_ha_en_egen» 47.43 sec-48.03 sec



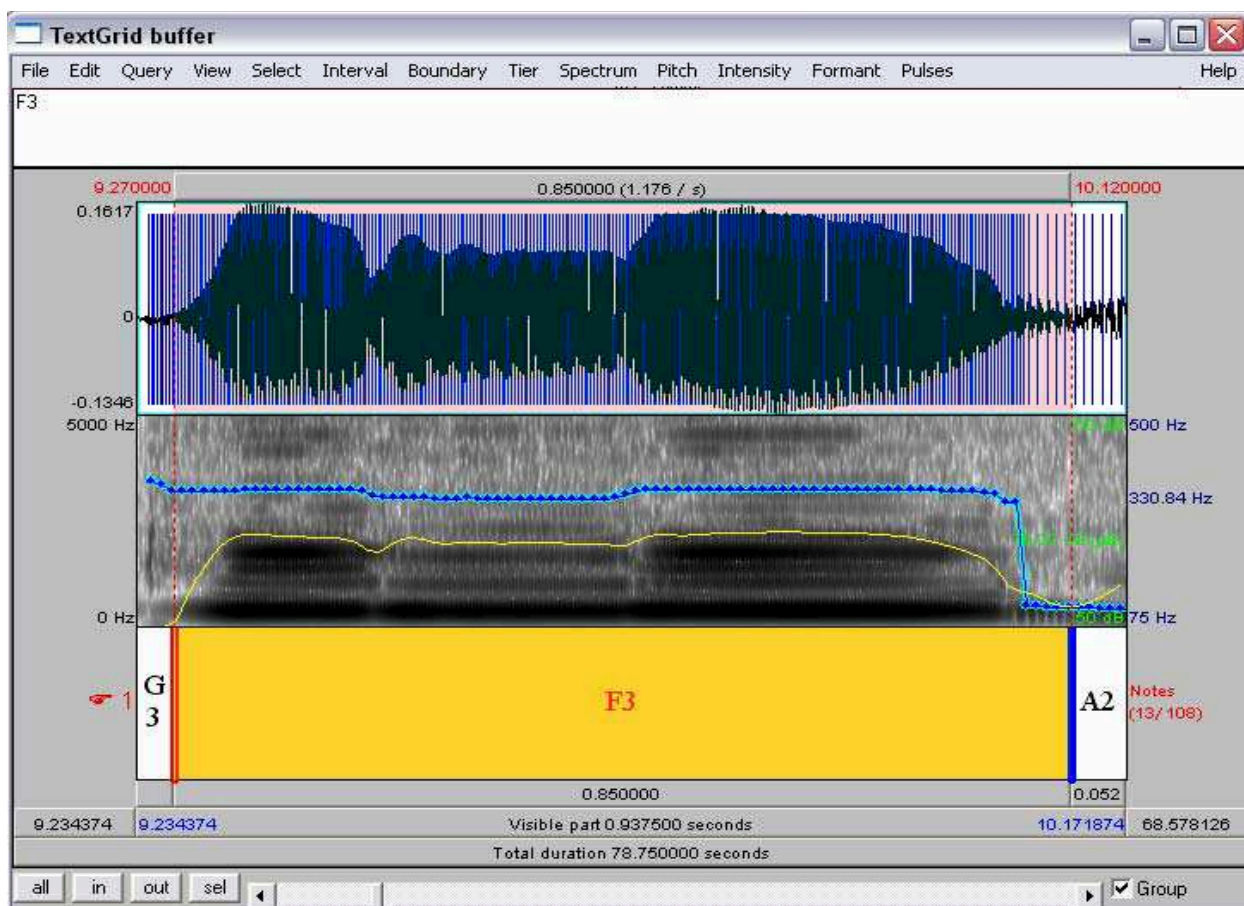
Αρχείο ήχου 47.43 sec-48.03 sec



Εικόνα 40: Κομμάτι παρτιτούρας «Jag_vill_ha_en_egen»

2.10.5 Συγχώνευση νοτών με διαφορά ενός τόνου

Στην παρακάτω εικόνα βλέπουμε την περίπτωση συγχώνευσης 3 νοτών διάρκειας 1/16, 1/16 και 1/8 αντίστοιχα, με διαφορά ενός τόνου. Συγκεκριμένα, στο χρονικό διάστημα 9.274 sec-9.464 sec το μέσο ύψος είναι 348 Hz, από 9.464 sec-9.678 sec είναι 332.27 Hz και από 9.678-10.120 sec είναι 350 Hz και οι αντίστοιχες νότες φαίνονται στην Εικόνα 42 όπου φαίνεται σημειωμένο το αντίστοιχο κομμάτι της παρτιτούρας. Η βαθμίδα αναγνώρισης βρίσκει μία νότα σε όλο το διάστημα από 9.27 sec έως 10.12 sec, την F3, για τους ίδιους λόγους που αναφέραμε στις παραπάνω περιπτώσεις συγχώνευσης.



Εικόνα 41: Παράδειγμα συγχώνευσης νοτών με διαφορά ενός τόνου Test-case «Johan_Moritz_julen er her» 47.43 sec-48.03 sec



Αρχείο ήχου 9.27 sec-10.12 sec



Εικόνα 42: Κομμάτι παρτιτούρας julen er her

2.11 Συμπεράσματα/ προτάσεις βελτίωσης

Οι εκτελέσεις των αρχαρίων παρουσιάζουν κάποια ιδιαίτερα χαρακτηριστικά που δυσκολεύουν τη σωστή εκτίμηση της νότας. Συγκεκριμένα, το ύψος των νοτών που εκτελούν οι μαθητές καθώς και η έντασή τους συχνά παρουσιάζουν αστάθειες. Αυτό το χαρακτηριστικό θα πρέπει να το λαμβάνει υπόψη της η βαθμίδα αναγνώρισης, έτσι ώστε να μην διασπάσει τις νότες. Τα κατώφλια όμως αυτά που ορίζουν την ελάχιστη χρονική διάρκεια μιας νότας θα πρέπει να εξαχθούν λαμβάνοντας υπ' όψιν τις ελάχιστες διάρκειες νοτών στην παρτιτούρα, ώστε να μην έχουμε το αντίθετο αποτέλεσμα.

Όπως είδαμε παραπάνω, αν αυτά τα όρια είναι μεγαλύτερα από τις διάρκειες κάποιων νοτών, η βαθμίδα αναγνώρισης συγχωνεύει τις νότες, χάνοντας με αυτόν τον τρόπο κάποια από τα γεγονότα νοτών. Αυτό συμβαίνει κυρίως με γειτονικές νότες που έχουν μικρή διαφορά στη θεμελιώδη συχνότητα, όπως ένα ημιτόνιο ή ακόμη και ένα τόνο. Η βαθμίδα αναγνώρισης εντοπίζει σε ένα δείγμα μία θεμελιώδη συχνότητα κοντινή με αυτή του προηγούμενου δείγματος και θεωρεί ότι πρόκειται για απλή αστάθεια, έως ότου βρει αρκετά δείγματα ώστε να ορίσει νέο γεγονός νότας. Αν η διάρκεια της νέας νότας δεν ξεπερνά το όριο που έχει οριστεί από την βαθμίδα αναγνώρισης ως ελάχιστη διάρκεια νότας, η νότες συγχωνεύονται.

Επιπλέον φτωχή εκτέλεση νότας από τον σπουδαστή εισάγει διαστρεβλώσεις στον παραγόμενο ήχο και, συνεπώς, αναγνωρίζεται μια διαφορετική νότα από αυτή που παίχτηκε. Για παράδειγμα, υπερβολικό φύσημα στη φλογέρα μπορεί να επηρεάσει σημαντικά την περιβάλλουσα που παραγόμενου ήχου, εισάγοντας μια εικονική μετατόπιση κατά μία οκτάβα.

Αναζητώντας τρόπους βελτίωσης της βαθμίδας αναγνώρισης, ψάχνουμε αρχικά στις τεχνικές που χρησιμοποιούνται για την αναγνώριση. Ελέγχοντας ένα σύνολο από διαφορετικές εκτελέσεις μπορούμε να προσαρμόσουμε τις τιμές των παραμέτρων στα επίπεδα εκείνα που θα είναι αρκετά ευαίσθητα ώστε να αναγνωρίζουν τα νέα γεγονότα νοτών, αλλά όχι τόσο ώστε να διασπά νότες με μικρές διακυμάνσεις στο ύψος τους. Επιπλέον, μπορούμε να εκμεταλλευτούμε τις παρτιτούρες των κομματιών που έχουμε στην διάθεσή μας στην συγκεκριμένη εφαρμογή. Οι πληροφορίες μιας παρτιτούρας μπορούν να ενσωματωθούν στο σύγχρονο σύστημα της αναγνώρισης, ώστε να πετυχαίνουμε μεγαλύτερη ακρίβεια στα αποτελέσματά μας σε πραγματικό χρόνο. Συγκεκριμένα, στην παρτιτούρα σημειώνονται οι τόνοι, οι διάρκειες, οι συζεύξεις και άλλες πληροφορίες που μπορεί η βαθμίδα αναγνώρισης να χρησιμοποιήσει ώστε να γνωρίζει εκ των προτέρων τα δύσκολα σημεία που μπορεί τόσο ο ίδιος όσο και ο εκτελεστής να κάνουν λάθη. Για παράδειγμα, μπορεί η βαθμίδα αναγνώρισης, όταν περιμένει ότι η επόμενη νότα που θα εκτελεστεί έχει διάρκεια $1/16^o$ μεγαλώνει την ευαισθησία του, ώστε να μην χάνει τα γεγονότα νοτών. Επίσης έχοντας στη διάθεσή του τις πληροφορίες της παρτιτούρας, μπορεί η βαθμίδα αναγνώρισης να τροποποιεί κατάλληλα τις τιμές των παραμέτρων του προκειμένου να αντιμετωπίσει με μεγαλύτερη επιτυχία τις επερχόμενες «δύσκολες» καταστάσεις.

Έχοντας μια συνολική εικόνα του τρόπου με τον οποίο η εκτέλεση έχει αντιστοιχιστεί με την παρτιτούρα, η βαθμίδα αντιστοίχισης εκτέλεσης/παρτιτούρας μπορεί να αναγνωρίσει νότες που λείπουν και κομματιασμένες νότες. Η πληροφορία αυτή μπορεί να περάσει και σε άλλο ένα επίπεδο αναγνώρισης όπου ζητείται από την ασύγχρονη βαθμίδα αναγνώρισης να εξετάσει λεπτομερέστερα κάποια τμήματα της εκτέλεσης ώστε να βρει νότες που δεν έχουν εντοπιστεί ή να επανεξετάσει τις αποφάσεις του για το αν δυο ή περισσότερες αναγνωρισμένες νότες θα πρέπει να ενωθούν σε μία νότα.

Πρέπει επίσης να σημειωθεί ότι η βαθμίδα αντιστοίχισης λαμβάνει υπόψη κάποια χαρακτηριστικά της βαθμίδα αναγνώρισης όπως, για παράδειγμα, τις διαχωρισμένες νότες που μπορεί να οφείλονται, για

παράδειγμα, στη φτωχή εκτέλεση μιας νότας από τον μαθητή και μπορεί να οδηγήσει σε διαδοχικά γεγονότα νοτών του ίδιου (ή διαφορετικού) ύψους. Αυτά τα γεγονότα νοτών ενώνονται από την βαθμίδα αντιστοίχισης σε ένα μοναδικό σύνολο και αντιστοιχίζονται σε ένα μοναδικό γεγονός, «καθαρίζοντας» με αυτόν τον τρόπο την αρχική ακολουθία νοτών συγχωνεύοντας κάθε ένα από τα σύνολα νοτών που έχουν αναγνωρισθεί σε μία μοναδική νότα (με τιμή του ύψους ισοδύναμη με το ύψος της νότας αναφοράς που έχει αντιστοιχισθεί με το σύνολο αυτό).

3

ΑΚΟΛΟΥΘΙΑ /ΑΝΤΙΣΤΟΙΧΙΣΗ ΠΑΡΤΙΤΟΥΡΑΣ

3.1 Γενική Επισκόπηση

Στο κεφάλαιο αυτό θα ασχοληθούμε με την βαθμίδα αντιστοίχισης εκτέλεσης/παρτιτούρας πραγματικού και μη πραγματικού χρόνου (score follower και score matcher αντίστοιχα). Η πρώτη πραγματοποιεί σε πραγματικό χρόνο αντιστοίχιση της ακολουθίας εισόδου με την ακολουθία αναφοράς εντοπίζοντας την τρέχουσα θέση του μαθητή στην εκτέλεση του μουσικού κομματιού. Η ασύγχρονη έκδοση της βαθμίδας δέχεται ως είσοδο την εκτέλεση ενός μουσικού κομματιού και την αντίστοιχη παρτιτούρα και αντιστοιχίζοντας τις δύο αυτές εισόδους παράγει ως έξοδο τη βέλτιστη αντιστοίχιση μεταξύ των γεγονότων εισόδου και αναφοράς.

Αρχικά, θα κάνουμε μια αναφορά στην υπάρχουσα έρευνα πάνω στο θέμα της αντιστοίχισης εκτέλεσης/παρτιτούρας. Θα μελετήσουμε την ιστορία του, δείχνοντας την εξέλιξή του στο χρόνο από τεχνολογική και μουσικολογική πλευρά. Θα συνεχίσουμε με μια λεπτομερή περιγραφή των βαθμίδων που χρησιμοποιήθηκαν στο IMUTUS και της στρατηγικής υλοποίησης των αντίστοιχων βαθμίδων του VEMUS. Κλείνοντας αυτό το κεφάλαιο, θα σημειώσουμε τις δυνατότητες και αδυναμίες των διαφόρων συστημάτων και θα προτείνουμε πιθανές βελτιώσεις των βαθμίδων αυτών.

3.2 Υπάρχουσα Έρευνα

Στο σημείο αυτό θα σχολιάσουμε τον ιστορικό ορισμό της αντιστοίχισης εκτέλεσης με την παρτιτούρα. Ωστόσο, η αντιστοίχιση ως ένα πεδίο αλληλεπίδρασης μεταξύ νέων απαιτήσεων εκτελεστών και συνθετών αφενός και επιστημονικών τεχνολογιών αφετέρου, έχει αλλάξει ορισμό και έχει εξελιχθεί τα τελευταία 20 χρόνια. Κατά συνέπεια, ένας αντικειμενικός ορισμός μπορεί εύκολα να χάσει την ισχύ του με το χρόνο. Εδώ, θα προσπαθήσουμε να δώσουμε ένα γενικό ορισμό αναφορικά με τη φύση του στην πράξη.

Η αντιστοίχιση χρησιμεύει ως μια σε πραγματικό χρόνο διεπαφή αποτύπωσης μεταξύ του μουσικού ήχου και των μουσικών συμβόλων καθώς και μεταξύ της ζωντανής εκτέλεσης ενός εκτελεστή και της εν λόγω παρτιτούρας. [14]

Η πρόκληση κείται στην επιτυχία αυτής της αποτύπωσης σε διαφορετικές συνθήκες εκτέλεσης, όπως διάφορα λάθη στις εκτελέσεις και διαφορετικά μουσικά όργανα. Μελετώντας την εξέλιξη των μεθόδων αντιστοίχισης θα δούμε πώς οι προσδοκίες μουσικών και ερευνητών βοήθησαν στην ανάπτυξη της συγκεκριμένης τεχνολογίας και οδήγησαν στις πρόσφατες ιδέες ως αποτέλεσμα 20 χρόνων εμπειρίας και αλληλεπίδρασης μεταξύ μουσικής και επιστημών.

Η ιστορία ξεκινάει το 1984 με το άρθρο του Dannenberg [Dan84] στο ICMC (International Computer Music Conference), μία από τις πρώτες δημοσιεύσεις που χειρίζεται το της αντιστοίχισης, με στόχο να δημιουργήσει ένα σύστημα αυτόματου ακομπανιαμέντο. Ο αλγόριθμος που προτείνεται από τον συγγραφέα, που μπορεί να θεωρηθεί και ως τεχνική ταιριάσματος συμβολοσειρών (string matching) βασίζεται στον δυναμικό αλγόριθμο time warping, υπό τις ακόλουθες υποθέσεις:

Η είσοδος του αλγορίθμου αποτελείται από διακριτά γεγονότα νοτών.

Ο αλγόριθμος ψάχνει για το καλύτερο ταίριασμα μεταξύ της παρτιτούρας αναφοράς και εισόδου όπου το πλήθος των αταίριαστων γεγονότων θα είναι το μικρότερο δυνατόν.

“Η εκτέλεση θα είναι κοντινή στην παρτιτούρα αναφοράς, δηλαδή ο εκτελεστής ποτέ δεν θα παραλείψει ή θα εισάγει παρά λίγα γεγονότα στην παρτιτούρα αναφοράς.”, σύμφωνα με τα λόγια του συγγραφέα.

Ο εκτελεστής αναμένεται περισσότερο να παίζει κάποιες νότες λάθος παρά να κάνει άλματα. Οι αντιστοιχίσεις σε επόμενες περιοχές τις παρτιτούρας δεν υποστηρίζονται από τους ευρεστικούς αλγορίθμους.

Αν και αναγνωρίζει τους περιορισμούς αυτού του συστήματος, κυρίως το γεγονός ότι δεν αρμόζει για πολυφωνικά όργανα, ο Dannenberg δηλώνει επιτυχή αποτελέσματα κατά τη διάρκεια των δοκιμών, όπου μία τρομπέτα χρησιμοποιήθηκε ως μουσικό όργανο. Προτείνει επίσης ότι επιπλέον ευρετικοί κανόνες και γνώση για τα κοινά λάθη και τις δυσκολίες στην εκτέλεση με το συγκεκριμένο όργανο μπορεί να βοηθήσει επιπλέον στη διαδικασία.

Μετά από την αρχική αυτή δημοσίευση, η έρευνα συνεχίστηκε σε διάφορες κατευθύνσεις. Ανάμεσα στα άλλα συστήματα που υλοποιήθηκαν, πολλοί εξακολουθούν να εστιάζουν στο σύγχρονο ακομπανιαμένο σε έναν μόνο εκτελεστή, μελετώντας επίσης και πολυφωνικά όργανα όπως το πιάνο. Κάποια συστήματα προχώρησαν ακόμα περισσότερο, προσπαθώντας να ακολουθήσουν μια φωνή σε ένα τραγούδι [GD97].

Όσο για τις μεθόδους που εφαρμόστηκαν, ποικίλουν από αλγορίθμους βασισμένους σε Dynamic Time Wrapping (DTW) ως και κρυμμένα μοντέλα Markov (HMMs) [OLS03]. Κάποιες μέθοδοι συγχωνεύουν τη διαδικασία της αναγνώρισης με αυτή της αντιστοίχισης, παίρνοντας διάφορα χαμηλού επιπέδου χαρακτηριστικά του ήχου ως είσοδο, αντί για μία σειρά από αναγνωρισμένες νότες.

Μια άλλη διαφορά μεταξύ των αλγορίθμων είναι το αν λαμβάνουν υπ’ όψη πληροφορίες συγχρονισμού (timing) στην αντιστοίχιση των γεγονότων: κάποιιοι το κάνουν, ενώ άλλοι αγνοούν εντελώς τον συγχρονισμό και αντιστοιχούν τα γεγονότα με βάση το ύψος.

Είναι αξιοσημείωτο το γεγονός ότι η χρήση της λειτουργίας αντιστοίχιση σε συστήματα μουσικής διδασκαλίας δεν έχει σημειωθεί παρά ελάχιστες φορές στην υπάρχουσα βιβλιογραφία. Η πιο σημαντική αναφορά βρίσκεται στο [DSJ+90] όπου περιγράφεται ένα σύστημα διδασκαλίας για σπουδαστές στο πιάνο, βασισμένο πλήρως στον υπολογιστή. Ωστόσο, δεν δίνεται συγκεκριμένη αναφορά στην μέθοδο για χρήση της για αντιστοίχιση, αν και μπορούμε αν υποθέσουμε ότι είναι το ίδιο ή παρόμοιο με αυτό που περιγράφεται από τον Dannenberg στις προηγούμενες δημοσιεύσεις του.

Από αυτά τα συστήματα μόνο κάποια σχετικά με τους στόχους του VEMUS παρουσιάζονται.

Οι [DM88] προτείνουν τη χρήση ενός επιπλέον “matcher”, ώστε να μελετούν περισσότερες από μία εναλλακτικές αντιστοιχίσεις παράλληλα, όταν η εκτέλεση δεν μπορεί να ερμηνευτεί μόνο κατά ένα τρόπο. Ο δεύτερος matcher συχνά προσπαθεί να κάνει άλμα στην παρτιτούρα αναφοράς προβλέποντας ένα σημείο στην παρτιτούρα όπου ο εκτελεστής θα μπορούσε να βρίσκεται αν δεν είχε κάνει κανένα λάθος. Κατά συνέπεια, κάποιος από τους matchers θα συγχρονιστεί με την εκτέλεση ξανά και θα αναγνωριστεί ως ο πρωτεύων matcher, ενώ ο άλλος θα απενεργοποιηθεί μέχρι ο πρωτεύων matcher να παρατηρήσει κάποια άλλη αμφισημία. Σ’ αυτό το πλαίσιο, ένας matcher είναι ένα «αντικείμενο», έχοντας μία εσωτερική κατάσταση (που αποτελείται από την θέση στην παρτιτούρα αναφοράς όπου οι νότες συγκρίνονται προς αντιστοιχία).

Οι [PL92] προτείνουν ένα απλό σύστημα για αντιστοίχιση που στηρίζεται στην ιδέα χρήσης μιας λίστας αλμάτων (“skip list”) για να αντιστοιχίσουν περασμένα γεγονότα. Μια “skip list” είναι μια λίστα που κρατά ένα πλήθος νοτών που είναι πριν την τρέχουσα θέση και που δεν έχουν ακόμα

αντιστοιχιστεί. Όταν υπάρχει μια νότα που δεν αντιστοιχίζεται με την αναμενόμενη, ο αλγόριθμος ψάχνει μία αντιστοιχία μέσα στην skip list. Όταν εκεί δεν βρίσκεται μια αντιστοιχία, υποθέτει πως ο εκτελεστής έχει χάσει μια νότα και ο αλγόριθμος ξεκινάει να ψάχνει στις επόμενες νότες.

Και οι δυο τελευταίοι matchers χρησιμοποιούν μια απλή μορφή ευρετικών αλγορίθμων, ψάχνοντας για αντιστοιχία μόνο εντός ενός πλαισίου γύρω από την αναμενόμενη νότα. Υποθέτουν δικαιολογημένα ότι ο εκτελεστής δεν συμπεριφέρεται ακανόνιστα, παίζοντας σε τυχαία σημεία της παρτιτούρας. Υποθέτουν ότι κάποια σειρά τηρείται, η οποία διαταράσσεται από αυθόρμητα λάθη.

Τέλος οι [HDH00] περιγράφουν ένα γενικό πλαίσιο για συστήματα αντιστοίχισης, δίνοντας μια ενοποιημένη περιγραφή διαφόρων υπάρχοντων συστημάτων. Στο πλαίσιό τους, η εκτέλεση κάθε αλγορίθμου αντιστοίχισης αναπαρίσταται σαν ένας κατευθυνόμενος γράφος, όπου οι κορυφές αναπαριστούν αντιστοιχίσεις και οι ακμές αναπαριστούν μεταβάσεις.

3.3 Η Βαθμίδα Αντιστοίχισης του VEMUS

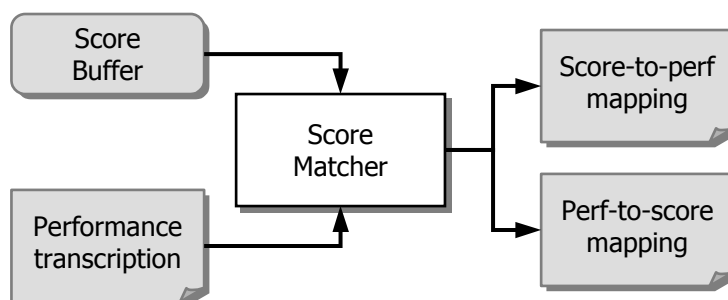
3.3.1 Εισαγωγή

Οι απαιτήσεις του IMUTUS/VEMUS, καθιστούν απαραίτητη την ανάπτυξη δύο διαφορετικών βαθμίδων: μιας σύγχρονης βαθμίδας πραγματικού χρόνου (score follower) και μιας ασύγχρονης μη-πραγματικού χρόνου (score matcher).

Τα παρακάτω σχήματα δείχνουν τις εισόδους και εξόδους των βαθμίδων.



Εικόνα 43 : Διάγραμμα εισόδων-εξόδων της σύγχρονης βαθμίδας αντιστοίχισης.



Εικόνα 44 : Διάγραμμα εισόδων-εξόδων του της ασύγχρονης βαθμίδας αντιστοίχισης.

Η σύγχρονη βαθμίδα αντιστοίχισης είναι υπεύθυνη για τον εντοπισμό της τρέχουσας θέσης της εκτέλεσης του μαθητή στην παρτιτούρα του μουσικού κομματιού. Η τρέχουσα θέση υπολογίζεται με δυναμική αντιστοίχιση της αναγνωρισμένης εκτέλεσης του μαθητή με την παρτιτούρα αναφοράς, δίνοντας μία ένδειξη για το πού «βρίσκεται» ο μαθητής. Το αποτέλεσμα σύγχρονη βαθμίδα αντιστοίχισης χρησιμοποιείται για να εκτελέσει σύγχρονες ενέργειες όπως:

- Παροχή οπτικής ένδειξης της τρέχουσας θέσης κατά τη διάρκεια της εκτέλεσης π.χ. κουνώντας έναν κέρσορα πάνω στην παρτιτούρα
- Αυτόματη ολίσθηση ή αλλαγή σελίδας όταν αυτό είναι απαραίτητο

Η ασύγχρονη βαθμίδα αντιστοίχισης είναι η ασύγχρονη έκδοση της σύγχρονης βαθμίδας αντιστοίχισης. Επιτελεί παρόμοιο έργο με αυτή αλλά, ενεργοποιείται αφού ο μαθητής έχει τελειώσει την εκτέλεσή του. Η ασύγχρονη βαθμίδα αντιστοίχισης έχει πρόσβαση στη καταγεγραμμένη εκτέλεση ολόκληρου του κομματιού και μπορεί μ' αυτόν τον τρόπο να πάρει πιο ενήμερες αποφάσεις για τον τρόπο που πρέπει να αντιστοιχιστεί με την παρτιτούρα αναφοράς.

Τα ζητήματα των βαθμίδων αντιστοίχισης παρουσιάζουν διαφορές. Η διαφορά έγκειται στο γεγονός ότι στο σύγχρονο πρόβλημα, η παρτιτούρα εισόδου αποτελείται από ένα μόνο τμήμα της εκτέλεσης. Αυτό συμβαίνει γιατί σε ένα δεδομένο σημείο κατά τη διάρκεια της εκτέλεσης ο αλγόριθμος μπορεί να έχει πρόσβαση μόνο στο τμήμα της παρτιτούρας που έχει εκτελεστεί μέχρι εκείνη τη στιγμή και δεν μπορεί να λάβει υπόψη του μελλοντικά γεγονότα. Λέμε ότι ο αλγόριθμος ακολουθεί την εκτέλεση, γιατί η νότα που αναμένεται να εκτελεστεί είναι η επόμενη από την τελευταία αντιστοιχισμένη. Ωστόσο, το ένα είναι ουσιαστικά μια ιδιαίτερη περίπτωση του άλλου. Κατά συνέπεια, θα δώσουμε ένα κοινό προσδιορισμό και στα δύο και το πρόβλημα θα αναφέρεται συνολικά ως αντιστοίχιση.

Η λειτουργικότητα της σύγχρονης βαθμίδας αντιστοίχισης στο IMUTUS/VEMUS είναι προφανής. Αρχικά, το σύστημα παρέχει οπτική βοήθεια στο μαθητή, δείχνοντάς του συνεχώς την επόμενη νότα ή το επόμενο μέτρο και γυρνώντας σελίδα στην παρτιτούρα όταν αυτό είναι απαραίτητο. Ο πλέον σημαντικός ρόλος όμως του είναι η παραγωγή δύο στοιχισμένων παρτιτούρων στη βαθμίδα αξιολόγησης της εκτέλεσης (Performance Evaluation Module).

Οι είσοδοι του προβλήματος ασύγχρονης αντιστοίχισης είναι:

- Μια παρτιτούρα αναφοράς, η οποία αποτελείται από μια πεπερασμένη ακολουθία μουσικών νοτών, καθεμιά εκ των οποίων έχει συγκεκριμένο ύψος και διάρκεια.
- Μια παρτιτούρα εισόδου, που ορίζεται όπως και η παρτιτούρα αναφοράς, αλλά το μήκος της μπορεί να είναι διαφορετικό από αυτό της αναφοράς.

Υποθέτουμε ότι ένας εκτελεστής (π.χ. ο μαθητής που χρησιμοποιεί το IMUTUS/VEMUS) διαβάζει την παρτιτούρα αναφοράς και προσπαθεί να την αναπαράγει. Η πραγματική εκτέλεση (όπως αναγνωρίζεται από την βαθμίδα αναγνώρισης) αποτελεί την παρτιτούρα εισόδου.

Η έξοδος του αλγορίθμου ασύγχρονης αντιστοίχισης είναι η αντιστοίχιση μεταξύ της παρτιτούρας αναφοράς και της παρτιτούρας εισόδου, όπου περιοχές της παρτιτούρας αναφοράς αντιστοιχίζονται με περιοχές της παρτιτούρας εισόδου, δηλαδή ακολουθίες νοτών.

Το νόημα αυτής της αντιστοίχισης είναι ότι αν δυο νότες είναι αντιστοιχισμένες, τότε η αναγνωρισμένη νότα από την παρτιτούρα αναφοράς είναι η νότα που ο εκτελεστής διάβασε και έπαιξε από την παρτιτούρα αναφοράς.

Ιδανικά, στην περίπτωση ενός ιδανικού εκτελεστή, οι δύο παρτιτούρες θα είναι ίσου μήκους και η αντιστοιχία θα είναι ένα προς ένα από την αρχή έως το τέλος των δύο ακολουθιών.

3.3.2 Ρόλος και Τεχνική Περιγραφή Βαθμίδας

Ο ρόλος των βαθμίδων αντιστοίχισης είναι να αντιστοιχίσουν γεγονότα νοτών από την (μονοφωνική) εκτέλεση με τα γεγονότα νοτών από την αναφορά. Στο τέλος, λαμβάνει τη σειρά των νοτών όπως έχουν αναγνωριστεί από την βαθμίδα αναγνώρισης και τη σειρά των νοτών από την αναφορά και υπολογίζει την βέλτιστη αντιστοίχιση μεταξύ τους.

Ωστόσο, στην περίπτωση του VEMUS που οι εκτελεστές είναι αρχάριοι είναι αναμενόμενο ένα πλήθος από λάθη. Κάποια από τα κύρια ζητήματα που οι βαθμίδες πρέπει να αντιμετωπίσουν είναι:

- Αποτυχία της βαθμίδας αναγνώρισης να αναγνωρίσει σωστά την νότα. Μια τέτοια αποτυχία αντιστοιχεί σε λανθασμένο ύψος
- αλλά θα μπορούσε να έχει την αιτία του σε πιο πολύπλοκα προβλήματα όπως:
 - (α) σπάσιμο μιας νότας σε πολλαπλές συνεχόμενες νότες (π.χ. όταν το ύψος ή η ένταση της νότας είναι πολύ ασταθής κατά τη διάρκεια της νότας)
 - (β) ένωση δυο γειτονικών νοτών σε μια (π.χ. συνεχείς νότες με το ίδιο ύψος)
- Λάθη κατά την εκτέλεση (λανθασμένες νότες)

Επίσης, στα πλαίσια ενός συστήματος εκμάθησης μουσικής επιπρόσθετα ζητήματα πρέπει να προσεχθούν:

- Φτωχή εκτέλεση νότας από τον σπουδαστή. Αυτό εισάγει διαστρεβλώσεις στον παραγόμενο ήχο και
- συνεπώς, αναγνωρίζεται μια διαφορετική νότα από αυτή που παίχτηκε. Για παράδειγμα, υπερβολικό φύσημα στη φλογέρα μπορεί να επηρεάσει σημαντικά την περιβάλλουσα που παραγόμενου ήχου, εισάγοντας μια εικονική μετατόπιση κατά μία οκτάβα.
- Η αντίδραση ενός μαθητή σε περίπτωση λάθους ποικίλει: θα μπορούσε να ξαναδοκιμάσει την ίδια νότα ή να αρχίσει να παίζει από την αρχή της γραμμής ή και ολόκληρο το κομμάτι. Αν και οι βαθμίδες αντιστοίχισης δεν είναι δυνατόν να ανταποκρίνονται επιτυχώς σε κάθε τυχαία συμπεριφορά, πρέπει να είναι αντιμετωπίζουν ικανοποιητικά κάποιες από τις πιο συνηθισμένες.
- Άλλο ένα γεγονός που συχνά περιπλέκει την κατάσταση για τις βαθμίδες στην περίπτωση άλματος, είναι τα επαναλαμβανόμενα μοτίβα που συχνά υπάρχουν σε ασκήσεις που χρησιμοποιούνται σαν υλικό μουσικής εκπαίδευσης.

Ο επόμενος πίνακας παρουσιάζει ένα περιεκτικό κατάλογο των βασικών λαθών που μπορεί να κάνει ένας εκτελεστής μαζί με κάποιους συνδυασμούς λαθών που παρατηρούνται συχνά στις εκτελέσεις των μαθητών. Η δεύτερη στήλη δείχνει την αντιστοιχία μεταξύ της παρτιτούρας αναφοράς και της παρτιτούρας εισόδου (στα αριστερά βρίσκεται η παρτιτούρα εισόδου). Η Τρίτη στήλη δείχνει την αντιστοιχία από άλλη άποψη, υπογραμμίζοντας την πιθανή επικάλυψη εξαιτίας του άλματος προς τα πίσω στην εκτέλεση και την επανάληψη κάποιων νοτών.

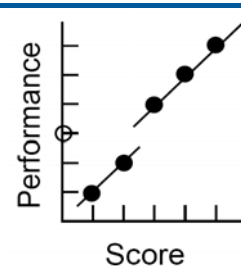
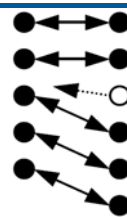
Σε μια ιδανική εκτέλεση η πρώτη εικόνα θα αποτελούνται από δυο παράλληλες 1-προς-1 αντιστοιχισμένες στήλες γεγονότων και η δεύτερη εικόνα θα ήταν μια απλή διαγώνια γραμμή.

Πολλά από τα λάθη που παρουσιάζονται στον επόμενο Πίνακα παρατηρούνται ανεξάρτητα, όπως όταν ένας μαθητής παίζει πολύ γρήγορα ή «χάνει» μία νότα. Ένας συνηθισμένος συνδυασμός τους παρατηρείται όταν ένας μαθητής παίζει μια λάθος νότα ή όταν χάνει μια νότα και ξεκινάει από την αρχή του μέτρου ή ακόμα και του προηγούμενου στην προσπάθειά του να διορθώσει το λάθος του.

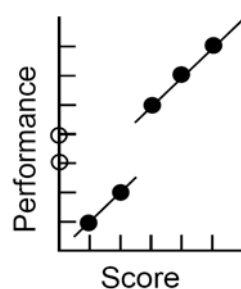
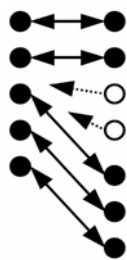
Table 1: Τύποι λαθών συχνά παρατηρούμενοι σε εκτελέσεις μαθητών

Τύπος λάθους

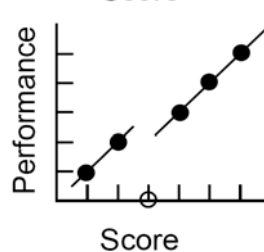
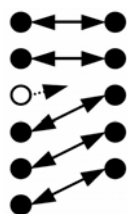
Εισαγωγή. Μια λανθασμένη νότα παίζεται μεταξύ δύο σωστών νοτών.



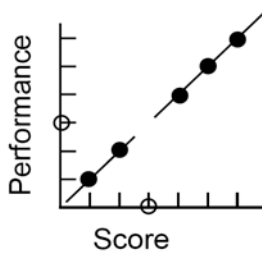
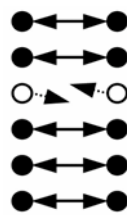
Πολλαπλές εισαγωγές. Δύο ή περισσότερες λανθασμένες νότες παίζονται μεταξύ δύο σωστών νοτών.



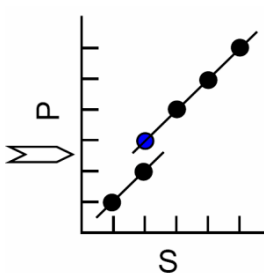
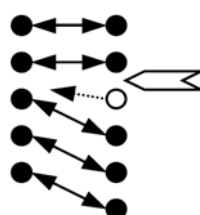
Διαγραφή. Μια νότα δεν παίζεται και η εκτέλεση συνεχίζει από την επόμενη.



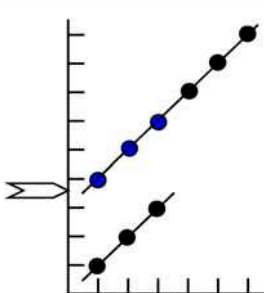
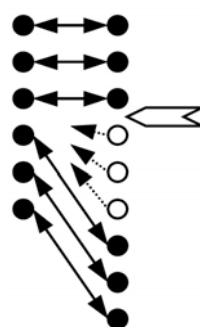
Λανθασμένη νότα. Μια λανθασμένη νότα παίζεται στη θέση της σωστής.



Επανάληψη της τελευταίας νότας. Η τελευταία παιγμένη νότα επαναλαμβάνεται, ίσως εξαιτίας δισταγμού του μαθητή.
(Επαναλαμβανόμενη νότα με μπλε χρώμα)



Επανάληψη ενός συνόλου από νότες.



Είναι βέβαιο το γεγονός ότι με την παρουσία λαθών εκτέλεσης στην ακολουθία εισόδου, κάποια αμφισημία εισάγεται στο επιθυμητό ταίριασμα. Δεν είναι πάντα δυνατή η εύρεση της νότας που ο μαθητής προσπάθησε να παίξει.

Ας δούμε το παρακάτω παράδειγμα όπου η παρτιτούρα αναφοράς είναι [...a b c d] και ο εκτελεστής έχει παίξει [...a c] θεωρώντας ότι ως αυτό το σημείο η αντιστοίχιση έχει πραγματοποιηθεί. Κάποιες από τις δυνατές υποθέσεις είναι οι παρακάτω:

- Ο μαθητής πέρασε τη νότα 'b' και συνέχισε στη νότα 'c'. Στην περίπτωση αυτή η επόμενη νότα είναι η 'd'.
- Ο μαθητής έπαιξε την 'c' αντί για την 'b', π.χ. εξαιτίας λάθους δακτυλισμού. Στην περίπτωση αυτή η επόμενη νότα είναι η 'c'.
- Ο μαθητής λανθασμένα εισήγαγε τη νότα 'c'. Στην περίπτωση αυτή η επόμενη νότα είναι η 'b'.

Αξίζει να σημειώσουμε ότι με περισσότερα δεδομένα η αμφισημία μπορεί να επιλυθεί.

- Αν η ακολουθία εισόδου συνεχίζει ως [...a b c] μπορούμε να υποθέσουμε την παράληψη μιας νότας
- Αν η ακολουθία εισόδου συνεχίζει ως [...a c c d] μπορούμε να υποθέσουμε την περίπτωση λάθους νότας
- Αν η ακολουθία εισόδου συνεχίζει ως [...a c b c] μπορούμε να υποθέσουμε την περίπτωση εισαγωγής μιας νότας.

Συνεπώς, είναι προφανές ότι χωρίς μελλοντικά δεδομένα, το πρόβλημα της σύγχρονης αντιστοίχισης είναι πιο δυσεπίλυτο από αυτό της ασύγχρονης. Αυθόρμητες αμφισημίες μπορούν επίσης να υπάρχουν και το τρόπο επίλυσης τους θα περιγραφεί στη συνέχεια.

Είναι προφανές ότι οι βαθμίδες αντιστοίχισης δουλεύουν στενά με τις βαθμίδες αναγνώρισης. Στην πραγματικότητα πρέπει να αντιμετωπίσουν οποιαδήποτε λάθη εισαχθήκανε από τις βαθμίδες αυτές. Η σύνδεση μεταξύ των βαθμίδων αντιστοίχισης και αναγνώρισης μπορεί να είναι ακόμα πιο στενή:

Γνωρίζοντας την τρέχουσα θέση στην παρτιτούρα, η σύγχρονη βαθμίδα αντιστοίχισης μπορεί να δώσει στοιχεία στον σύγχρονη βαθμίδα αναγνώρισης για την επόμενη αναμενόμενη νότα, βοηθώντας τον να βελτιώσει τον ρυθμό αναγνώρισης.

Έχοντας μια συνολική εικόνα του τρόπου με τον οποίο η εκτέλεση έχει αντιστοιχιστεί με την παρτιτούρα, η ασύγχρονη βαθμίδα αντιστοίχισης μπορεί να αναγνωρίσει νότες που λείπουν και κατατετηγμένες νότες. Η πληροφορία αυτή μπορεί να περάσει και σε άλλο ένα επίπεδο αναγνώρισης όπου ζητείται από την ασύγχρονη βαθμίδα αναγνώρισης να εξετάσει λεπτομερέστερα κάποια τμήματα της εκτέλεσης ώστε να βρει νότες που δεν έχουν εντοπιστεί ή να επανεξετάσει τις αποφάσεις του για το αν δυο ή περισσότερες αναγνωρισμένες νότες θα πρέπει να ενωθούν σε μία νότα.

Προφανώς, η εκτέλεση ενός μαθητή δεν μπορεί να θεωρηθεί πάντα ότι είναι συνεχής. Άλματα και επαναλήψεις είναι αναμενόμενες, π.χ. όταν ο μαθητής κάνει ένα λάθος. Η ασύγχρονη βαθμίδα αντιστοίχισης μπορεί να αναγνωρίσει τέτοια άλματα σε μία εκτέλεση. Ένα τέτοιο άλμα πρακτικά σημαίνει ότι υπάρχουν δύο ή και περισσότερες προσπάθειες να εκτελέσει ένα συγκεκριμένο τμήμα της παρτιτούρας. Επίσης, μια επαναπροσπάθεια σημαίνει ότι ο μαθητής γνωρίζει τα λάθη του και ξαναδοκιμάζει και δεν υπάρχει νόημα στην αξιολόγηση εκείνου του τμήματος της εκτέλεσης αφού ο ίδιος ο μαθητής το απέρριψε στην προσπάθειά του να βελτιωθεί. Έτσι, όταν το σύστημα θα πρέπει να αξιολογήσει αυτό το τμήμα δεν είναι εποικοδομητικό να αξιολογήσει όλες τις προσπάθειες και δεν είναι προφανές πώς οι διάφορες αξιολογήσεις μπορούν να συνδυαστούν ώστε η ανάδραση να είναι

εποικοδομητική και να βοηθά το μαθητή. Συνεπώς, η πιο προφανής αντίδραση του συστήματος στην περίπτωση αλμάτων και επανεκτελέσεων κάποιων τμημάτων θα ήταν να λάβει υπ' όψη την τελευταία προσπάθεια του μαθητή αγνοώντας όλες τις προηγούμενες. Άρα, πριν την αξιολόγηση της εκτέλεσης, γίνεται "κανονικοποίηση" αφαιρώντας τις επαναλήψεις, τα άλματα και τις επαναπροσπάθειες με βάση την πληροφορία που παρέχεται από την ασύγχρονη βαθμίδα αντιστοίχισης. (Πρέπει να σημειωθεί ότι η κανονικοποίηση δεν απαιτείται μόνο για την ηχογραφημένη εκτέλεση του μαθητή αλλά και την βιντεοσκόπηση της εκτέλεσης του μαθητή.)

3.3.3 Αλγοριθμική Περιγραφή

Οι δύο βαθμίδες μοιράζονται τον ίδιο αλγόριθμο με μικρές μετατροπές ώστε η σύγχρονη βαθμίδα αντιστοίχισης να λειτουργεί σε πραγματικό χρόνο χωρίς μελλοντικές πληροφορίες. Ο αλγόριθμος αυτός πρέπει να υπακούει στις ακόλουθες απαιτήσεις:

- Ευρωστία: ο αλγόριθμος θα πρέπει να είναι όσο το δυνατόν ανθεκτικός στα πιθανά λάθη των μαθητών, όσο και στα λάθη αναγνώρισης. Αποτυχία συμμόρφωσης σε αυτή την απαίτηση θα επηρέαζε την αξιολόγηση της απόδοσης και θα προκαλούσε σύγχυση στον μαθητή.
- Ταχύτητα: η σύγχρονη έκδοση του αλγορίθμου θα πρέπει να είναι αρκετά γρήγορη ώστε να ακολουθεί την εκτέλεση του μαθητή σε πραγματικό χρόνο. Επίσης, η ασύγχρονη έκδοση δεν θα πρέπει να εισάγει σημαντική καθυστέρηση στην συνολική διαδικασία.
- Είσοδος και έξοδος: ολόκληρη η παρτιτούρα αναφοράς είναι διαθέσιμη κάθε στιγμή στη μνήμη του συστήματος. Η παρτιτούρα εισόδου δημιουργείται σε πραγματικό χρόνο από την βαθμίδα αναγνώρισης και επίσης αποθηκεύεται σε αντίστοιχη θέση στη μνήμη. Στην ασύγχρονη έκδοση όλη η παρτιτούρα εισόδου είναι διαθέσιμη. Η έξοδος της σύγχρονης έκδοσης είναι η τρέχουσα θέση στην παρτιτούρα. Στην ασύγχρονη έκδοση, η αντιστοίχιση μεταξύ της παρτιτούρας εισόδου και της παρτιτούρας αναφοράς καταγράφεται σε έναν απομονωτή (Performance Matching Buffer).

Ανάμεσα στις εναλλακτικές τεχνικές που παρουσιάστηκαν στη βιβλιογραφία, η προσέγγιση time-wrapping επιλέχθηκε ως βάση για την υλοποίηση των βαθμίδων αντιστοίχισης.

Τα Hidden Markov Models (HMM) αποκλείστηκαν επειδή είναι πολύπλοκα και απαιτούν εκπαίδευση. Στις μεθόδους που χρησιμοποιούν HMMs, ένα ξεχωριστό μοντέλο πρέπει να εκπαιδευτεί για κάθε μουσικό κομμάτι. Οι [OD01] περιγράφουν μια διαδικασία εκπαίδευσης για τα HMMs χρησιμοποιώντας ένα σύνολο από εκτελέσεις, είτε ηχογραφημένες κατά τη διάρκεια των δοκιμών είτε δημιουργημένες μέσω υπολογιστή με λάθη εκτέλεσης σκοπίμως εισαγόμενα. Ωστόσο, αυτό είναι σε αντίθεση με τις απαιτήσεις των βαθμίδων αναγνώρισης και αξιολόγησης του IMUTUS, που πρέπει να δουλεύουν με κάθε διαθέσιμο κομμάτι, χωρίς επιπλέον παρέμβαση, αφού οι δάσκαλοι θα έχουν τη δυνατότητα να εισάγουν νέο υλικό.

Ο προτεινόμενος αλγόριθμος αγνοεί πληροφορίες συγχρονισμού και οι νότες αντιστοιχίζονται μόνο βάση του ύψους τους. Αφού η ομάδα-στόχος του VEMUS είναι αρχάριοι μαθητές με περιορισμένη εμπειρία στην κατανόηση του χρονικού συγχρονισμού και υψηλό ποσοστό λαθών συγχρονισμού, η συνεισφορά πληροφοριών συγχρονισμού στην αντιστοίχιση είναι σημαντικά μειωμένη. Το γεγονός ότι η αντιστοίχιση δεν μπορεί να βασίζεται σε πληροφορίες συγχρονισμού, δεν υποδηλώνει ότι τα αντίστοιχα λάθη του μαθητή θα αγνοηθούν. Η βαθμίδα αξιολόγησης της απόδοσης θα δώσει την κατάλληλη ανάδραση στην περίπτωση λανθασμένου συγχρονισμού.

Κατάλληλοι ευρετικοί κανόνες θα χρησιμοποιηθούν για να βοηθήσουν τον αλγόριθμο να αποφασίσει σε περιπτώσεις αμφισημίας. Αυτοί οι ευρετικοί κανόνες περιλαμβάνουν παραμέτρους όπως το πλαίσιο άλματος και τον μέγιστο αριθμό των συνεχόμενων λαθών που μπορεί να ανεχθεί και ανταποκρίνονται σε εύλογες υποθέσεις για την συμπεριφορά του εκτελεστή καθώς ακολουθεί μια μουσική παρτιτούρα. Οι παράμετροι θα βελτιωθούν βασιζόμενοι σε παρατήρηση των περιπτώσεων δοκιμής που έχουν συλλεχθεί και σε μουσική γνώση του συγκεκριμένου κάθε φορά οργάνου που ηχογραφείται (όπως δύσκολοι δακτυλισμοί, κλασικά λάθη στο δακτυλισμό και στο φύσημα κ.ά.).

Μια βασική έννοια του αλγορίθμου είναι αυτή ενός 'MatchNode'. Ένα MatchNode αντιπροσωπεύει μια επιβεβαιωμένη αντιστοιχία μεταξύ ενός τμήματος της παρτιτούρας εισόδου και ενός τμήματος της παρτιτούρας αναφοράς.

Η έννοια του MatchNode περιέχει τις θέσεις έναρξης και τέλους (σχετικές με την αρχή της παρτιτούρας αναφοράς) και, ενδεχομένως, μια αναφορά σε έναν κόμβο γονέων.

Χρησιμοποιώντας αυτές τις αναφορές, κάποιος μπορεί να κατασκευάσει μια γραμμική ακολουθία MatchNodes για να αντιπροσωπεύσει μια αντιστοιχία μεταξύ της παρτιτούρας εισόδου και της παρτιτούρας αναφοράς. Στην ιδανική περίπτωση, που είναι μια ακριβής 1 -1 αντιστοιχία, η λίστα περιέχει έναν μοναδικό κόμβο.

Ένα MatchNode θα γραφτεί ως {end_pos.. start_pos}. Για τα ακόλουθα παραδείγματα περιπτώσεων αντιστοίχισης παρτιτούρας, δίνεται η απεικόνισή τους ως λίστες MatchNode.

Όπως αναφέρθηκε προηγουμένως, θα υπάρξουν περιπτώσεις όπου περισσότερες από μία αντιστοιχίες είναι δυνατές. Σε αυτές τις περιπτώσεις, η λίστα των MatchNodes γίνεται ένα δέντρο, υποθέτοντας ότι ένα MatchNode μπορεί να έχει περισσότερα από ένα παιδιά. Διαβαίνοντας το δέντρο από τα φύλλα της προς τη ρίζα, μπορούμε να πάρουμε πιθανές αντιστοιχίες ολόκληρης της παρτιτούρας εισόδου.

Ο σκοπός του αλγορίθμου που περιγράφεται είναι να δημιουργήσει το δέντρο των MatchNodes και να αποφασίσει την πιο πιθανή αντιστοιχία. Η πιο πιθανή αντιστοιχία καθορίζεται ως η μεγαλύτερη αντιστοιχία, δηλ. η αντιστοιχία που περιλαμβάνει τις περισσότερες νότες από την παρτιτούρα εισόδου.

Εάν υπάρχουν πολλά λάθη που εισάγουν ασάφεια, το μέγεθος του δέντρου μπορεί να αυξηθεί εκθετικά. Προκειμένου να ελεγχθεί η αύξηση, διάφορα φύλλα του δέντρου κλαδεύονται, δηλ. οι δυνατότητές τους αντιστοιχίας τους δεν εξετάζονται περαιτέρω. Αυτή η περικοπή είναι βασισμένη σε ευρετικούς κανόνες, όπως: «εάν ένας κόμβος συνεχίζει να αντιστοιχίζει τους λίγους τελευταίους κόμβους άψογα, ενώ άλλοι κόμβοι συνεχώς αποτυγχάνουν και δημιουργούν παιδιά, κράτησε μόνο τον πρώτο κόμβο και ξεκίνησε με αυτή την καθαρή κατάσταση»

Ο σύγχρονος αλγόριθμος δουλεύει λαμβάνοντας τις νότες της παρτιτούρας εισόδου μία μία, στην προσπάθειά του να βρεθεί μια πιθανή αντιστοιχία για την τρέχουσα νότα στο τρέχον πλαίσιο και να ενημερώσει τη δομή του δέντρου αναλόγως.

Ο ασύγχρονος αλγόριθμος λειτουργεί με τον ίδιο τρόπο, αλλά δεν είναι απαραίτητο να εκτελεστεί σε πραγματικό χρόνο ούτε να αποφασίσει στις προσωρινές ασάφειες μέχρι το τέλος της παρτιτούρας εισόδου, οπότε πρέπει να γεμίσει τον απομονωτή αντιστοίχισης της απόδοσης και να τον παραδώσει στην ενότητα αξιολόγησης απόδοσης.

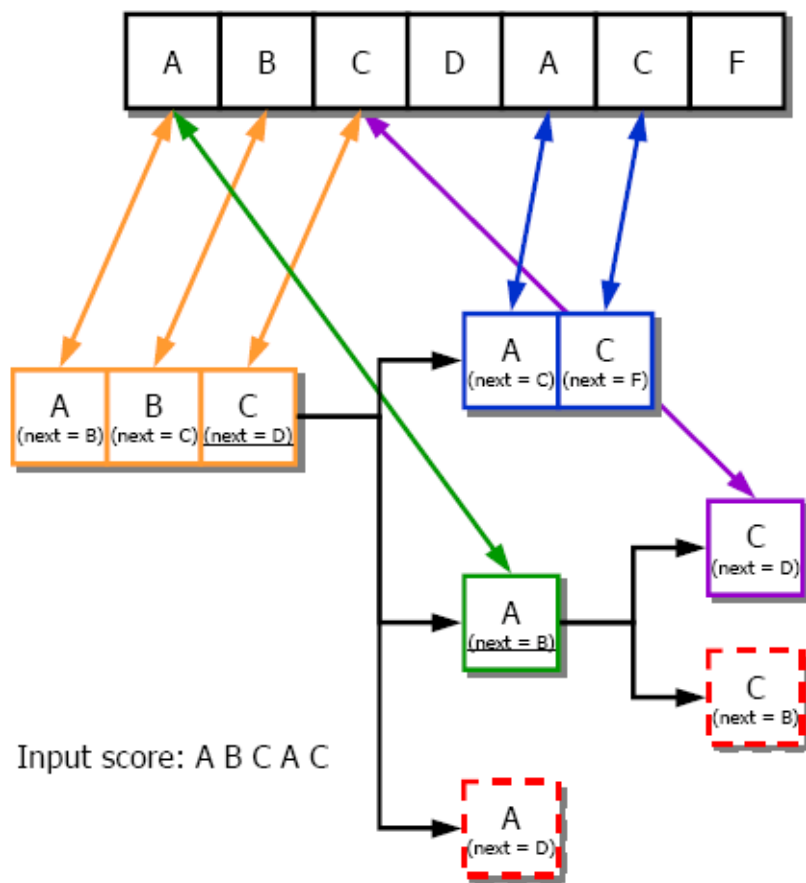
ΨΕΥΔΟΚΩΔΙΚΑΣ

```

Match {
  for each tree leaf node {
    if node is matching then update node
    else {
      delete node from leaf node set
      // search for possible occurrences around current position
      create child node for next note occurrence (in window)
      create child node for previous note occurrence (in window)
      // assume inserted/wrong note
      create child node matching to the wrong note
    }
  }
  select the position to report / most possible match (heuristic)
  possibly prune tree (heuristic)
}
  
```

Τα ακόλουθα σχήματα καταδεικνύουν τη λειτουργία του αλγορίθμου με μια ενδεικτική παρτιτούρα εισόδου και αναφοράς.

Το πρώτο παράδειγμα παρουσιάζει τη συμπεριφορά της σύγχρονης βαθμίδας αντιστοίχισης στην περίπτωση παράλειψης μιας νότας.



Εικόνα 45 :Συμπεριφορά της σύγχρονης βαθμίδας αντιστοίχισης στην περίπτωση παράλειψης μιας νότας.

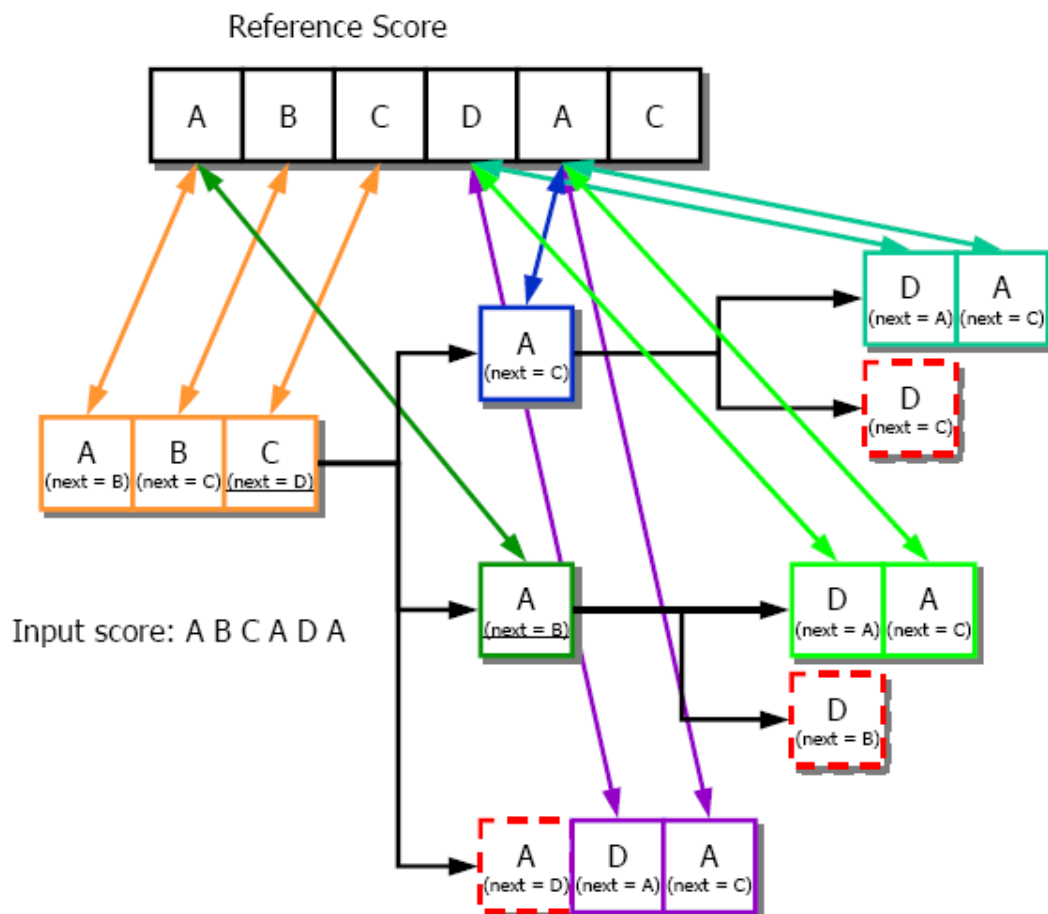
Ξεκινώντας από το πρώτο (πορτοκαλί χρώμα) MatchNode, οι πρώτες τρεις εισερχόμενες νότες, (A, B, C) αντιστοιχίζονται ακριβώς στην παρτιτούρα αναφοράς.

Εντούτοις, η τέταρτη νότα, που είναι η A, δεν ταιριάζει με την επόμενη αναμενόμενη νότα (D). Κατά συνέπεια, ο αλγόριθμος κοιτάζει προς τα εμπρός και προς τα πίσω για το A. Δύο A βρίσκονται, και τα αντίστοιχα MatchNodes (μπλε και πράσινο) δημιουργούνται. Επίσης, αυτή η νότα υποτίθεται ότι ήταν μια εισαγωγή, και ένα τρίτο MatchNode δημιουργείται, απεικονιζόμενο με μια διαστιγμένη γραμμή για να δηλώσει ότι αυτή δεν είναι μια πραγματική αντιστοιχία. Επίσης, δεδομένου ότι αυτή είναι μια εισαγωγή, αυτή η ταιριασμένη νότα αναμένει πάλι το D, ως γονέα της.

Τώρα το μπλε MatchNode αναμένει το C ως επόμενη νότα, ενώ το πράσινο αναμένει το B. Ο πρώτος θα προχωρήσει κανονικά, ταιριάζοντας την εισερχόμενη νότα C.

Ο δεύτερος κόμβος αποτυγχάνει και πρέπει να κοιτάξει πάλι για μια πιθανή αντιστοιχία. Ένας κόμβος παιδιού δημιουργείται (μοβ) βασισμένος στην πιθανή αντιστοιχία με το C στην Τρίτη θέση. Ακόμα ένας κόμβος δημιουργείται στην περίπτωση εισαγωγής.

Σε μια παραλλαγή του ανωτέρω παραδείγματος, παρουσιάζουμε τη συμπεριφορά του αλγορίθμου στην περίπτωση εισαγωγής μιας νότας:



Εικόνα 46: Συμπεριφορά της σύγχρονης βαθμίδας αντιστοίχισης στην περίπτωση εισαγωγής μιας νότας

Αυτό το παράδειγμα καταδεικνύει επίσης μια ενδιαφέρουσα ιδιότητα αυτού του αλγορίθμου: ανεξάρτητα από την ερμηνεία του λάθους (που ήταν μια εισαγόμενη νότα), ο αλγόριθμος μπορεί να συγχρονιστεί με τον εκτελεστή. Πράγματι, στο τέλος του παραδείγματός μας, όλα τα πιθανά μονοπάτια του δέντρου καταλήγουν με ένα MatchNode έχοντας το C ως επόμενη νότα. Είναι δυνατό

να κρατήσει μόνο έναν από αυτούς τους παρόμοιους κόμβους, απλοποιώντας τα επόμενα βήματα της διαδικασίας του ταιριάσματος και περικόπτοντας το δέντρο ταυτόχρονα.

Πρέπει επίσης να σημειωθεί ότι, όπως φαίνεται στα παραδείγματα, τα «προσωρινά» MatchNodes (που απεικονίζονται με τις διαστιγμένες γραμμές) δεν δίνουν παιδιά. Αντ' αυτού, παραμένουν στην ίδια θέση, περιμένοντας την ίδια νότα, η οποία μπορεί να εμφανιστεί μετά από έναν αριθμό εισαγομένων νοτών.

Οι βαθμίδες σύγχρονης και ασύγχρονης αντιστοίχισης του IMUTUS χρησιμοποιήθηκαν ως βάση για την ανάπτυξη των αντίστοιχων βαθμίδων του VEMUS.

Ωστόσο, μια νέα μηχανή αναγνώρισης θα αναπτυχθεί στο VEMUS. Μια διαφορετική έκδοση αυτής της μηχανής θα βελτιστοποιηθεί για κάθε υποστηριζόμενο όργανο. Δεδομένης της στενής σύνδεσης μεταξύ των διαδικασιών αναγνώρισης και αντιστοίχισης, τμήματα των βαθμίδων αντιστοίχισης θα πρέπει να ανά-κατασκευαστούν. Αυτό είναι απαραίτητο ώστε να ενσωματώσει τα χαρακτηριστικά των νέων βαθμίδων αναγνώρισης αλλά και αυτά των νέων υποστηριζόμενων οργάνων. Αυτό αναμένεται να επηρεάσει κυρίως τις υποθέσεις και τα αντίστοιχα βάρη παρά τον πυρήνα του αλγορίθμου αντιστοίχισης.

Ένα σύνολο από δεδομένα δοκιμών έχει συλλεχθεί. Αποτελείται από ένα πλήθος εκτελέσεων μαθητών, ηχογραφημένων υπό πραγματικές συνθήκες. Αυτό θα χρησιμοποιηθεί κατά την ανάπτυξη ως μέτρο σύγκρισης και διόρθωσης της απόδοσης των βαθμίδων αντιστοίχισης. Επιπρόσθετα δεδομένα αντλούμενα κατά την διάρκεια της αξιολόγησης με μαθητές, θα χρησιμοποιηθούν για την επαλήθευση και ρύθμιση των βαθμίδων.

3.3.4 Η Μηχανή Αντιστοίχισης

Ο σχεδιασμός της μηχανής αντιστοίχισης βασίζεται σε **καταστάσεις** και **υποθέσεις**. Ο αλγόριθμος αντιστοίχισης βασίζεται στη εύρεση ενός μονοπατιού ελάχιστου κόστους σε ένα γράφο όπου οι κόμβοι αναπαριστούν αντιστοιχισμένες καταστάσεις. Ένα σύνολο υποθέσεων γίνεται που αφορούν κάθε γεγονός νέας νότας που αναγνωρίζεται. Κάθε τέτοια υπόθεση φέρει ένα κόστος και οδηγεί σε μία νέα κατάσταση αντιστοίχισης. π.χ. ένα νέο κόμβο.

Αν In η ακολουθία εισόδου (η εκτέλεση) και Ref η ακολουθία αναφοράς (παρτιτούρα). Μια κατάσταση συμβολίζεται ως (i,j) εννοώντας πως σ' αυτή την κατάσταση, το i -στό αντικείμενο εισόδου πρέπει να αντιστοιχιστεί με το j -στο αντικείμενο αναφοράς.

Η αρχική κατάσταση είναι αυτή όπου το πρώτο αντικείμενο της ακολουθίας εισόδου πρέπει να αντιστοιχιστεί με το πρώτο αντικείμενο της ακολουθίας αναφοράς π.χ. $(0,0)$.

Κάθε κατάσταση περιέχει μια λίστα από υποθέσεις αντιστοίχισης, κάθε μία από τις οποίες οδηγεί σε μία νέα (στην ίδια) κατάσταση. Για παράδειγμα, για την κατάσταση (i,j) μια ή περισσότερες από τις παρακάτω υποθέσεις μπορούν να βρίσκουν εφαρμογή:

αν $In[i] = Ref[j]$, τότε μια υπόθεση «Αναμενόμενης νότας» προστίθεται στην κατάσταση (i,j) , και οδηγούμαστε στην κατάσταση $(i+1,j+1)$ αφού πρέπει να αντιστοιχηθούν τα επόμενα αντικείμενα από κάθε ακολουθία.

αν $In[i] \triangleleft Ref[j]$ αλλά $In[i]$ και $Ref[j]$ είναι παρόμοια¹, τότε μια υπόθεση “Λάθους Αναγνώρισης” προστίθεται στην κατάσταση (i,j) , οδηγώντας στην κατάσταση $(i+1,j+1)$.

Αν $In[i] \triangleleft Ref[j]$ αλλά $In[i]$ και $In[i-1]$ είναι παρόμοια, τότε μια υπόθεση “Σπασμένης νότας” προστίθεται στην κατάσταση (i,j) , και οδηγούμαστε στην κατάσταση $(i+1,j)$.

όμοια, ένα σύνολο από υποθέσεις γίνονται για ένα άλμα πίσω στην παρτιτούρα, μια εισαγόμενη νότα ή μια περασμένη νότα.

Είναι σημαντικό να σημειώσουμε ότι κάθε υπόθεση έχει και ένα αντίστοιχο κόστος. Μια αναμενόμενη νότα δεν έχει κανένα κόστος, ένα λάθος αναγνώρισης έχει ένα κόστος π.χ.0.1 ενώ ένα άλμα προς τα πίσω έχει ένα κόστος π.χ. 1.0 και ούτω καθεξής.

Με αυτόν τον τρόπο παράγεται ένας γράφος που ξεκινάει από τη μοναδική αρχική κατάσταση $(0,0)$ και συνεχίζει μέχρι το τελευταίο αντικείμενο εισόδου σε ένα σύνολο πιθανών τελικών καταστάσεων. Κάθε μονοπάτι από την αρχική κατάσταση μέχρι τις τελικές καταστάσεις έχει ένα συνολικό κόστος. Έτσι, μετά την παραγωγή του συνολικού γράφου, το καλύτερο μονοπάτι αναγνωρίζεται (minimum cost path). Αυτό το βέλτιστο μονοπάτι περιέχει καταστάσεις και υποθέσεις που οδηγούν από τη μία στην άλλη και αντιπροσωπεύει μια αντιστοίχιση από την ακολουθία εισόδου στην ακολουθία αναφοράς (που είναι ένα-προς-ένα).

Πρέπει επίσης να σημειωθεί ότι η μηχανή αντιστοίχισης λαμβάνει επίσης υπόψη κάποια χαρακτηριστικά των βαθμίδων αναγνώρισης όπως, για παράδειγμα, τις διαχωρισμένες νότες. Όπως έχουμε ήδη αναφέρει, είναι πολύ συχνό να «σπάει» κατά την διαδικασία αναγνώρισης μια νότα σε δύο ή περισσότερες νότες. Αυτό μπορεί να οφείλεται, για παράδειγμα, στη φτωχής εκτέλεσης μιας νότας από τον μαθητή και μπορεί να οδηγήσει σε διαδοχικά γεγονότα νοτών του ίδιου (ή διαφορετικού) ύψους. Αυτά τα γεγονότα νοτών ενώνονται κατά την διαδικασία αντιστοίχισης σε ένα μοναδικό **σύνολο** και αντιστοιχίζονται σε ένα μοναδικό γεγονός.

Βασισμένοι στην έξοδο της μηχανής αντιστοίχισης, η αρχική ακολουθία νοτών «καθαρίζεται», συγχωνεύοντας κάθε ένα από τα σύνολα νοτών που έχουν αναγνωριστεί σε μία μοναδική νότα (με τιμή του ύψους ισοδύναμη με το ύψος της νότας αναφοράς που έχει αντιστοιχιστεί με το σύνολο αυτό).

Η βαθμίδα αξιολόγησης της εκτέλεσης (Performance Evaluation Module) δέχεται ως είσοδο: (i) τα δύο MTX αρχεία (εκτέλεσης-αναφοράς και αναφοράς-εκτέλεσης, και (ii) τα δεδομένα ανάλυσης σε μορφή πινάκων του MATLAB. Οι επικεφαλίδες των αντιστοιχισμένων MTX αρχείων (time-signature, key-signature, tempo, sequence/track-name, midi) αντιγράφονται από το MTX αρχείο αναφοράς.

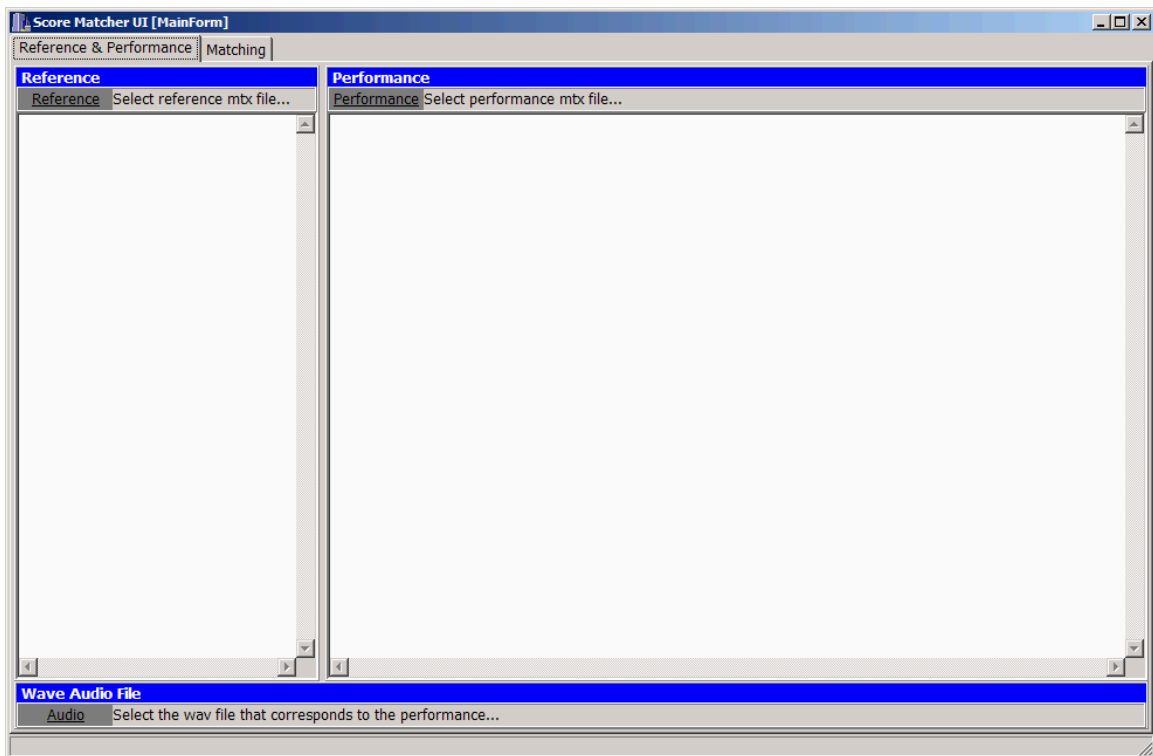
3.3.5 Περιβάλλον του Εργαλείου Αντιστοίχισης

Για τις ανάγκες μελέτης και ρύθμισης της απόδοσης των βαθμίδων αντιστοίχισης του έργου VEMUS, υλοποιήθηκε ένα αυτόνομο εργαλείο που προσφέρει εποπτεία της λειτουργίας των βαθμίδων αντιστοίχισης.

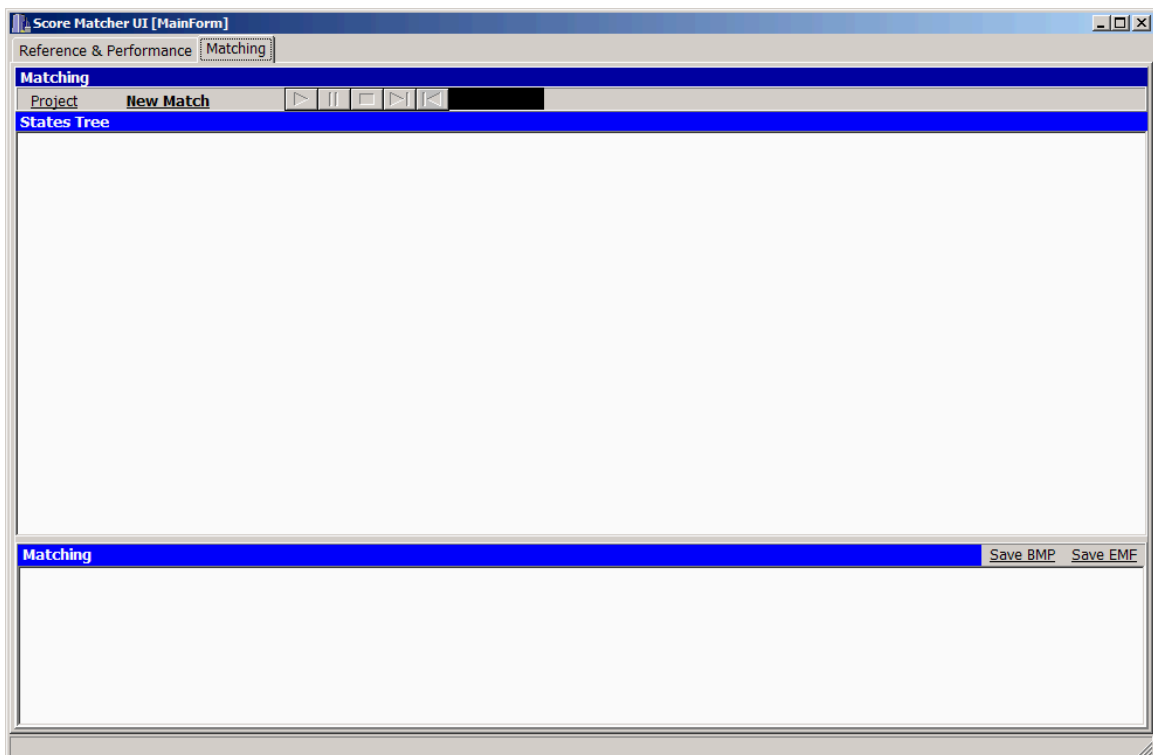
Το κύριο παράθυρο του ΕΡΓΑΛΕΙΟΥ έχει δύο κύρια πεδία: “Reference & Performance” («Αναφορά και Εκτέλεση») και “Matching” («Αντιστοίχιση»).

Στο πεδίο “Reference & Performance”, παρέχεται πληροφορία για τα δύο αρχεία MTX (αναφορά και εκτέλεση) και το ηχητικό αρχείο της εκτέλεσης.

¹ Παρόμοιες νότες είναι (i) ίδιες νότες (π.χ. C4/C4), (ii) νότες που διαφέρουν κατά ένα ημιτόνιο (π.χ. C#4/C4), και (iii) νότες που διαφέρουν κατά μία οκτάβα (π.χ. C4/C5).

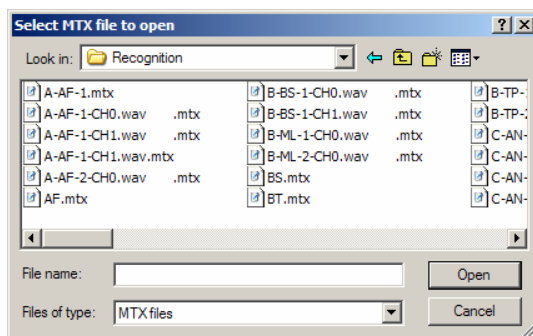


Εικόνα 47: Πεδία “Reference & Performance”



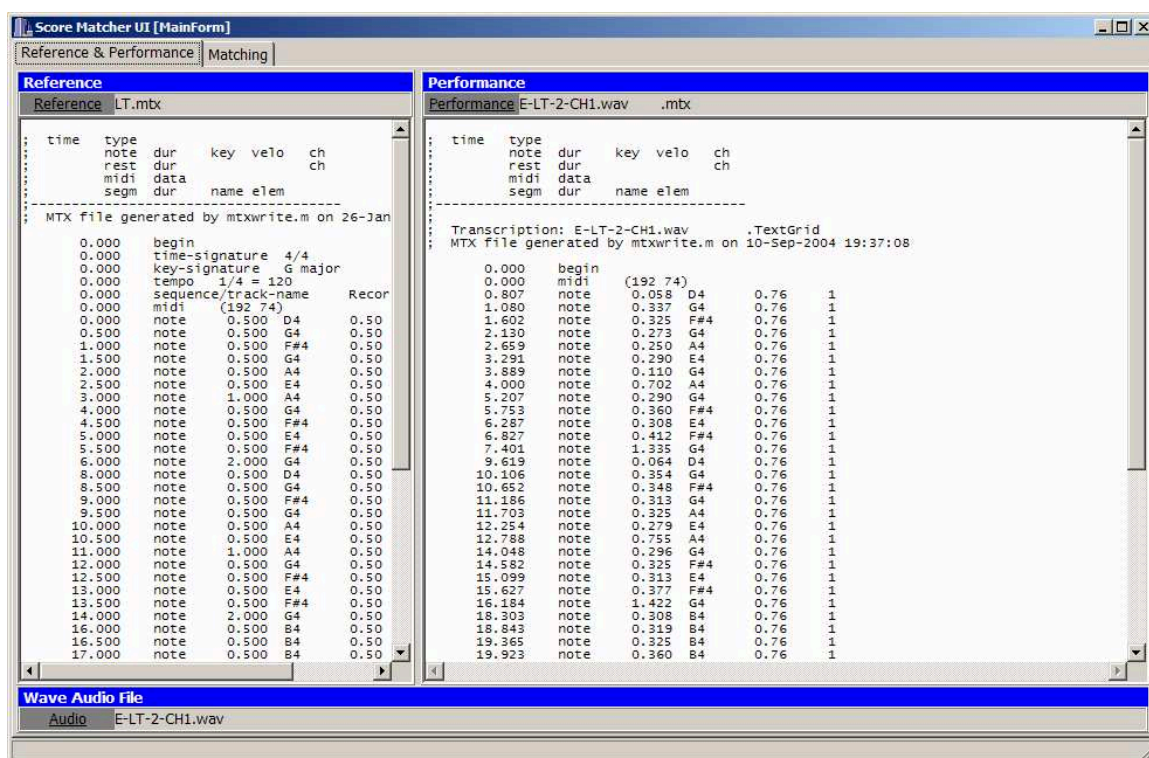
Εικόνα 48: Πεδίο “Mathing”

Πατώντας στα τρία γκρι κουμπιά με ονόματα “Reference”, “Performance” και “Audio”, εμφανίζεται ένας διάλογος επιλογής του αντίστοιχου αρχείου (.mtx ή .wav αντίστοιχα).



Εικόνα 49: Διάλογος επιλογής αρχείου

Παρακάτω φαίνεται το παράθυρο μετά την επιλογή αυτών των αρχείων.



Εικόνα 50: Παράθυρο μετά την επιλογή αυτών των αρχείων

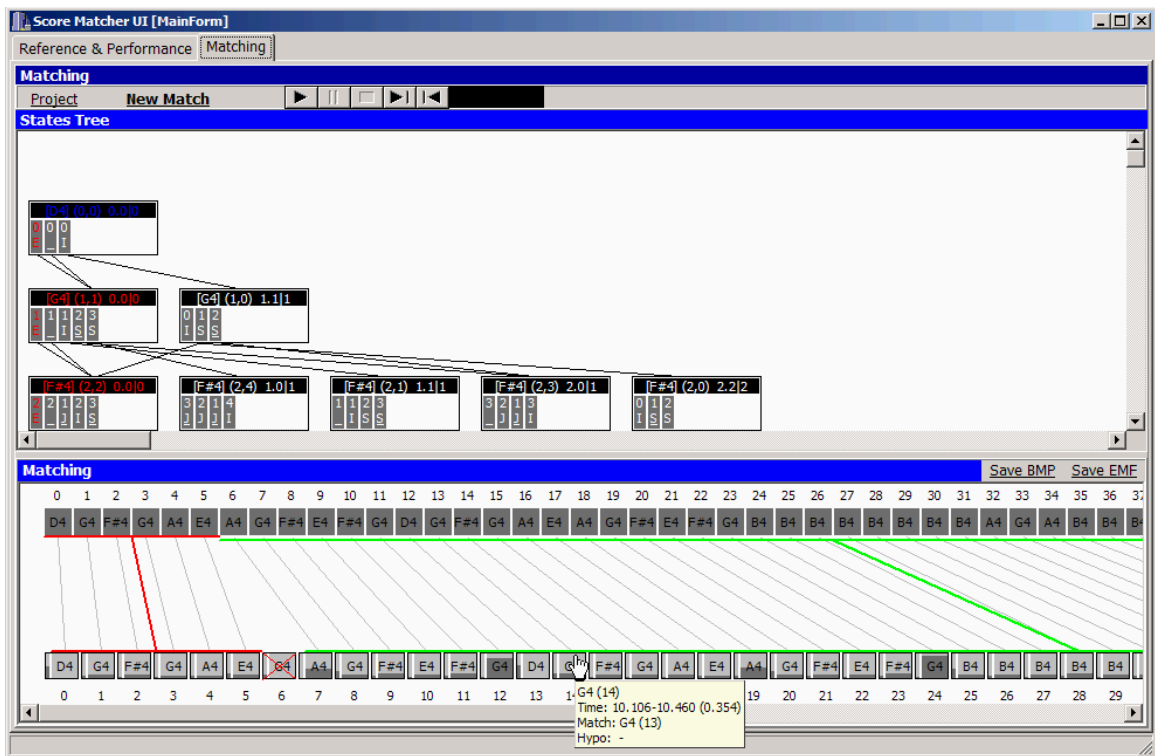
Για διευκόλυνση, μπορούν να χρησιμοποιηθούν “score matcher project files” («αρχεία έργων αντιστοίχισης παρτιτούρας»). Αυτά είναι αρχεία κειμένου με επέκταση “.smr” που περιέχουν τρεις γραμμές με το όνομα του MTX αρχείου αναφοράς, το όνομα του MTX αρχείου εκτέλεσης και το όνομα του αντίστοιχου WAV αρχείου της εκτέλεσης, π.χ.:

```
LT.mtx
RF-LT-mm104-1-CH0.mtx
RF-LT-mm104-1-CH0.wav
```

Έτσι, αυτά τα αρχεία μπορούμε εύκολα να ανοίξουμε αυτά τα αρχεία πατώντας το κουμπί “Project” και επιλέγοντας ένα .smr αρχείο.

Έχοντας τα κατάλληλα MTX και WAV αρχεία (χωριστά είτε σε ένα project), και πατώντας το κουμπί “New Match” ο matcher αρχίζει να δημιουργεί ένα δέντρο αποφάσεων με τις καταστάσεις και τις

υποθέσεις, αναγνωρίζει την καλύτερη αντιστοίχιση και δείχνει τα αποτελέσματα όπως φαίνεται στην παρακάτω εικόνα.

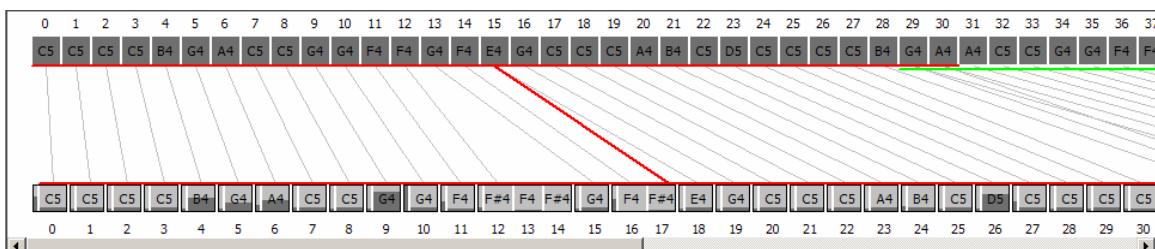


Εικόνα 51: Υποθέσεις και γράφος καλύτερης αντιστοίχισης

Η παραπάνω εικόνα δείχνει το γράφο και την καλύτερη αντιστοίχιση που βρέθηκε.

Στο παράθυρο όπου φαίνεται η καλύτερη αντιστοίχιση, η ακολουθία αναφοράς φαίνεται πάνω και η ακολουθία εισόδου (εκτέλεση) φαίνεται ακριβώς από κάτω. Κάθε αντικείμενο εισόδου και αναφοράς (γεγονός) απεικονίζεται με ένα γκρι τετράγωνο, με το ύψος της κάθε νότας να αναγράφεται μέσα σε αυτό. Η σκούρα γκρι περιοχή των αντικειμένων εισόδου παρέχει μια ένδειξη της (σχετικής) διάρκειας του αντίστοιχου γεγονότος νότας. Τα αντικείμενα εισόδου μπορούν να έπονται κάθετων γραμμών που αντιπροσωπεύουν διαστήματα σιγής. Και εδώ, οι σκούρες γκριζές περιοχές παρέχουν μια σχετική ένδειξη της διάρκειας της σιγής.

Με την κίνηση του ποντικιού πάνω σε ένα αντικείμενο εισόδου, εμφανίζεται ένα παράθυρο που παρέχει πληροφορίες για την αντίστοιχη νότα, τον αύξοντα αριθμό της στην ακολουθία εισόδου, τις χρονικές στιγμές αρχής και τέλους της, τη διάρκεια, το αντικείμενο και το δείκτη του αντικειμένου της ακολουθίας αναφοράς με το οποίο αντιστοιχίζεται καθώς και τις υποθέσεις που έγιναν για την πραγματοποίηση της συγκεκριμένης αντιστοίχισης.

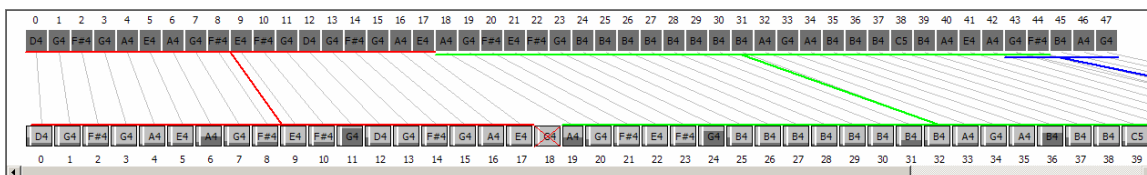


Εικόνα 52: Αντιστοίχιση ακολουθιών

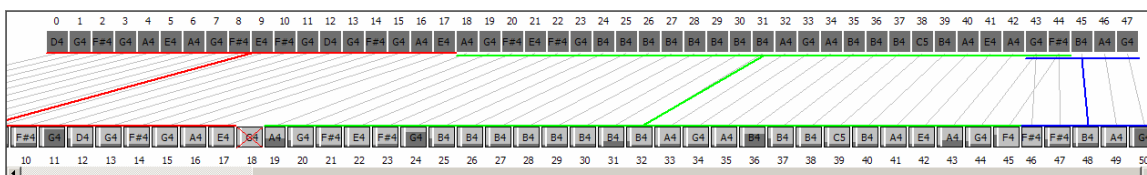
Στην παραπάνω εικόνα φαίνεται η περίπτωση μιας σπασμένης νότας. Τα αντικείμενα εισόδου με δείκτες 12, 13 και 14 αναγνωρίστηκαν από τον matcher ως τμήματα της ίδιας νότας (περίπτωση σπασμένης νότας). Ένα μαύρο περίγραμμα που περικλείει και τις τρεις αυτές νότες είναι ένδειξη του γκρουπ.

Με αριστερό κλικ σε ένα αντικείμενο εισόδου (ή σιγή) αναπαράγει το αντίστοιχο τμήμα του ηχητικού αρχείου, με δεξί κλικ αναπαράγει από το σημείο αυτό μέχρι το τέλος του αρχείου, ενώ με το πάτημα το μεσαίου κουμπιού του ποντικιού παίζει όλα τα αντικείμενα που ανήκουν σε ένα γκρουπ.

Χρησιμοποιώντας τη ροδέλα του ποντικιού, μπορούμε να μετακινήσουμε τη μπάρα δεξιά και αριστερά. Πατώντας Ctrl και χρησιμοποιώντας τη ροδέλα μετακινείται η ακολουθία αναφοράς δεξιά/αριστερά το οποίο είναι εξυπηρετικό όταν επεξεργαζόμαστε μεγάλες παρτιτούρες, όπως φαίνεται στην παρακάτω εικόνα.



Εικόνα 53



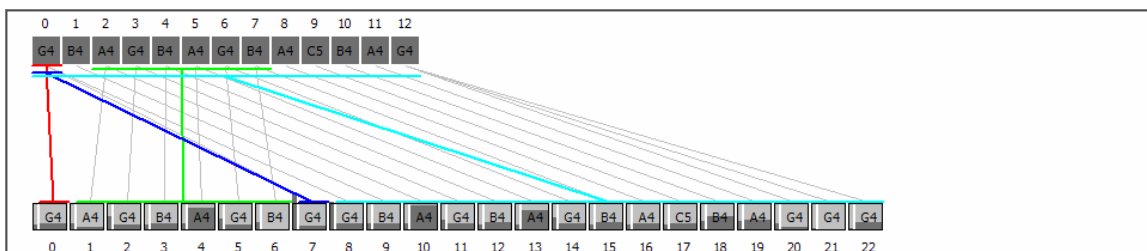
Εικόνα 54

Η αντιστοίχιση κάθε αντικειμένου απεικονίζεται με γκριζες γραμμές. Κάθε γκρουπ αντιστοιχίζεται με μία νότα στην ακολουθία αναφοράς και αντιμετωπίζεται πλέον ως ένα και μοναδικό αντικείμενο.

Ένα αντικείμενο εισόδου με κόκκινες διαγώνιες γραμμές (όπως το αντικείμενο 18 στην παραπάνω εικόνα) αντιπροσωπεύει εισαγόμενα αντικείμενα δηλ. Αντικείμενα στην ακολουθία εισόδου που δεν αντιστοιχούν σε κάποιο αντικείμενο στην ακολουθία αναφοράς.

Οι χονδρές έγχρωμες γραμμές αντιπροσωπεύουν μπλοκ, δηλαδή συνεχόμενα τμήματα στην ακολουθία εισόδου (εκτέλεση) που έχουν αντιστοιχιστεί με συνεχόμενα τμήματα στην ακολουθία αναφοράς. Μια εισαγόμενη νότα, ένα άλμα στην εκτέλεση ή μια παραλειπόμενη νότα, σπάει ένα μπλοκ.

Η παρακάτω Εικόνα δείχνει τα αποτελέσματα της αντιστοίχισης σε μια περίπτωση δοκιμής.”.



Εικόνα 55: Αποτελέσματα αντιστοίχισης σε περίπτωση δοκιμής”.

Εδώ ο μαθητής παίζει την πρώτη νότα και στη συνέχεια παραλείπει τη δεύτερη, παίζοντας την τρίτη. Μετά σταματάει την εκτέλεσή του και συνομιλεί με το δάσκαλό του (αναγνωρισμένη σιγή μεταξύ των αντικειμένων εισόδου 1 και 2) και προσπαθεί ξανά από την αρχή παίζοντας 5 ακόμα νότες. Ωστόσο, αυτές οι νότες είναι οι ίδιες με τις 3-7 της ακολουθίας αναφοράς και έτσι ο matcher κάνει αυτή τη σχέση. Άλλη μια διακοπή σπάει το μπλοκ της εκτέλεσης και ο μαθητής συνεχίζει τη συνομιλία του με το δάσκαλό του (σιγή μεταξύ των αντικειμένων εισόδου 6 και 7). Έπειτα, ακολουθεί άλλη μια προσπάθεια να ξεκινήσει από την αρχή (αντικείμενο εισόδου 7) και επιπλέον συζήτηση (σιγή μεταξύ των αντικειμένων εισόδου 7 και 8). Τελικά, ο μαθητής ξεκινά μια καθαρή εκτέλεση από την αρχή ως το τέλος σε ένα και μοναδικό (κυανό).

3.3.6 Αποθήκευση των MTX αρχείων των αποτελεσμάτων

Πατώντας το κουμπί “Save MTX”, τα αποτελέσματα του matcher σώζονται (ως MTX αρχεία). Τα δυο αρχεία θα σωθούν στον τρέχοντα φάκελο.

Για παράδειγμα αν το MTX της εκτέλεσης ονομάζεται ‘flickan.mtx’ τα αρχεία των αποτελεσμάτων θα ονομαστούν:

- r2p-flickan.mtx (αναφορά προς εκτέλεση)
 - p2r-flickan.mtx (εκτέλεση προς αναφορά)
-

3.3.7 Διεπαφή γραμμής εντολών

Μπορούμε επιπλέον να εκτελέσουμε το εργαλείο αντιστοίχισης σε μη διαδραστικό περιβάλλον με τον ακόλουθο τρόπο:

```
ScoreMatcherUI.exe <reference MTX file> <performance MTX file> <ref-2-perf MTX file> <perf-2-ref MTX file>
```

Οι δύο πρώτες παράμετροι είναι τα αρχεία εισόδου και οι άλλες δύο τα αρχεία εξόδου.

4

ΑΞΙΟΛΟΓΗΣΗ ΕΚΤΕΛΕΣΗΣ

Στο κεφάλαιο αυτό θα ασχοληθούμε με τη βαθμίδα αξιολόγησης της μουσικής εκτέλεσης του μαθητή, αναγνωρίζοντας και αναλύοντας ένα σύνολο από τυπικά λάθη που παρουσιάζονται σε εκτελέσεις αρχαρίων σπουδαστών. Τα λάθη αυτά περιγράφονται με όρους μεταβλητών της εκτέλεσης (pitch, tempo, loudness) σε σχέση με την εκτέλεση αναφοράς. Το σύνολο των τυπικών λαθών βασίζεται σε πληροφορίες έμπειρων δασκάλων στα συγκεκριμένα όργανα.

4.1 Βαθμίδα Αξιολόγησης Εκτέλεσης του VEMUS

Η Βαθμίδα Αξιολόγησης της Μουσικής Εκτέλεσης του VEMUS (PERFORMANCE EVALUATION MODULE) είναι υπεύθυνη για την αξιολόγηση της εκτέλεσης του μαθητή καθώς και για την παροχή κατάλληλης ανάδρασης δίνοντας μια συνολική βαθμολογία της εκτέλεσής του, συμπληρώνοντας τις στατιστικές των λαθών του συγκεκριμένου μαθητή και ανανεώνοντας το προφίλ του.

Η βαθμίδα αυτή αξιολογεί την απόδοση βάσει μιας συμβολικής αντιπροσώπευσης του καταγραμμένου ακουστικού σήματος. Η απόδοση του σπουδαστή συγκρίνεται με μια εκτέλεση αναφοράς που συντάσσεται από την αρχική παρτιτούρα.

Η Βαθμίδα Αξιολόγησης της Εκτέλεσης αναγνωρίζει ένα βασικό σύνολο χαρακτηριστικών λαθών στην εκτέλεση. Μια επιλογή των αναγνωρισμένων λαθών θα παρουσιαστεί στο μαθητή με σειρά προτεραιότητας. Επίσης, η Βαθμίδα Αξιολόγησης της Εκτέλεσης είναι σε θέση να ενθαρρύνει το σπουδαστή με θετική ανατροφοδότηση, παραδείγματος χάριν όταν δεν κάνει κανένα λάθος. Οι σχολιασμοί πάνω στην παρτιτούρα που εισάγονται από τον δάσκαλο μπορούν να καθοδηγήσουν την Βαθμίδα Αξιολόγησης της Εκτέλεσης στον προσδιορισμό και τον εντοπισμό πιθανών λαθών.

Ο βασικός στόχος της βαθμίδας είναι η αναγνώριση χαρακτηριστικών λαθών σε μια εκτέλεση. Τέτοια λάθη μπορούν να αφορούν στην απόδοση της μελωδίας (αν παίζει τις σωστές νότες), στη ρυθμική εκτέλεση, στην ποιότητα τόνου (επιθέσεις, σταθερότητα τόνου), στην προσωδία, στον έλεγχο της ροής αέρα, στο δακτυλισμό κ.ο.κ.. Μερικοί από τους τύπους λάθους είναι κοινοί για όλα τα όργανα και αφορούν *μουσικές δεξιότητες*, ενώ άλλοι είναι σαφώς εξαρτώμενοι από το συγκεκριμένο όργανο (*δεξιότητες οργάνων*).

Τα τυπικά λάθη μπορούν να χαρακτηριστούν με όρους των γνωρισμάτων της εκτέλεσης, που αναφέρονται ως *συμπτώματα*. Τα συμπτώματα μπορούν (α) να είναι παρόντα στην ακουστική περιοχή (π.χ. συμπτώματα σχετικά με την ποιότητα τόνου), ή (β) να εξάγονται από τη σύγκριση της απόδοσης με την παρτιτούρα αναφοράς (π.χ. μια λανθασμένη νότα ή ρυθμός). Η Βαθμίδα Αξιολόγησης της Εκτέλεσης Βάσει των συμπτωμάτων ανιχνεύει την ύπαρξη λαθών εκτέλεσης.

Οι σχολιασμοί που προστίθενται στην παρτιτούρα από τον δάσκαλο μπορούν να καθοδηγήσουν την Βαθμίδα Αξιολόγησης της Εκτέλεσης στην εύρεση των λαθών. Με την παροχή *σχολιασμών* ο δάσκαλος μπορεί να χαρακτηρίσει εκ των προτέρων τις δύσκολες μεταβάσεις όπου ένα λάθος ενός ορισμένου τύπου είναι πιθανόν να εμφανιστεί. Τέτοιοι σχολιασμοί μπορούν να αφορούν αλλοιώσεις, ρυθμούς, δύσκολες μεταβάσεις δακτυλισμών και ούτω καθεξής. Ένα άλλο πλεονέκτημα της χρήσης

των σχολιασμών είναι ότι η Βαθμίδα Αξιολόγησης της Εκτέλεσης θα είναι σε θέση να παρέχει θετική ανατροφοδότηση όταν παίζει ο σπουδαστής μια δύσκολη μετάβαση σωστά.

Η αξιολόγηση προσαρμόζεται στο επίπεδο ικανότητας και την προσωπικότητα του σπουδαστή, παρομοιάζοντας το *VEMUS με ένα Σχολικό σπίτι*. Κάθε πάτωμα αντιπροσωπεύει ένα συγκεκριμένο επίπεδο ικανότητας που είναι αντίστοιχο σε μια σχολή μουσικής με ένα έτος διδασκαλίας. Σε κάθε ένα πάτωμα υπάρχουν τρεις τύποι δασκάλων (καλός, κανονικός, ακριβής) με διαφορετική αποδοχή στα λάθη (*αυστηρότητα δασκάλων*). Το επίπεδο αποδοχής λάθους εφαρμόζεται με ένα σύνολο τιμών κατώτατων ορίων, συγκεκριμένο για κάθε δάσκαλο.

Η Βαθμίδα Αξιολόγησης της Εκτέλεσης θα δώσει την ανατροφοδότηση στο σπουδαστή ως σχολιασμούς στην παρτιτούρα. Τα αποτελέσματα της Βαθμίδας Αξιολόγησης της Εκτέλεσης επιδεικνύονται στο μαθητή πάνω στην παρτιτούρα, όπου φαίνεται ο τύπος λάθους και το σημείο του λάθους. Μια γενική βαθμολόγηση υπολογίζεται από το ποσοστό λάθους στην αξιολογημένη απόδοση. Το προφίλ των σπουδαστών, που χρησιμοποιείται για τον καθορισμό προτεραιοτήτων των ανιχνευμένων λαθών στις επόμενες αποδόσεις, ενημερώνεται σύμφωνα με την αξιολόγηση της τρέχουσας απόδοσης.

Η αξιολόγηση απόδοσης περιλαμβάνει την ακόλουθη ακολουθία:

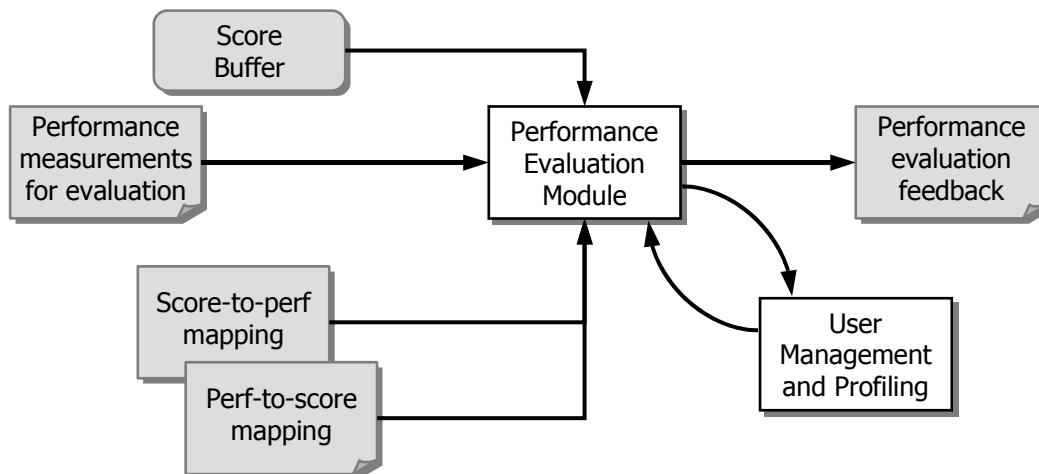
- Ανίχνευση των προκαθορισμένων βασικών λαθών σε μια απόδοση με:
 1. εύρεση και ανάλυση των συμπτωμάτων σε ακουστικό και συμβολικό επίπεδο
 2. έρευνα για τα αναμενόμενα λάθη στην παρτιτούρα (π.χ. τυχαία, μεγάλα άλματα)
 3. έρευνα για τα αναμενόμενα λάθη όπως υποδεικνύεται από τους σχολιασμούς στην παρτιτούρα (π.χ. δύσκολος ρυθμός)
 4. έρευνα για τα λάθη συγκεκριμένα για το εκάστοτε όργανο (π.χ. τριξίματα κλαρινέτων)

Η ανίχνευση χρησιμοποιεί ένα προεπιλεγμένο σύνολο κατώτατων τιμών που τίθενται από την παράμετρο *αυστηρότητας δασκάλων*.

- Συλλογή όλων των ανιχνευμένων λαθών στην απόδοση σε έναν κατάλογο. Η απουσία λαθών θα θεωρηθεί και θα χρησιμοποιηθεί ως ένδειξη της καλής/ σωστής εκτέλεσης και θα προστεθεί στον κατάλογο.
- Δεικτοδότηση του καταλόγου λαθών σύμφωνα με:
 1. μια προεπιλεγμένη προτεραιότητα για κάθε πάτωμα σπουδαστών και όργανο
 2. προτεραιότητα καθορισμένη από τις ενδείξεις στους σχολιασμούς της παρτιτούρας (υψηλή προτεραιότητα σε λάθη στα τμήματα που χαρακτηρίζονται ως «δύσκολα» από το δάσκαλο)
 3. δεδομένα από το προφίλ των σπουδαστών.
- Ταξινόμηση και φιλτράρισμα του καταλόγου με βάση τη δεικτοδότηση, και παραγωγή ενός μικρού αριθμού στοιχείων (χαρακτηριστικά τα κορυφαία τρία λάθη). Ένα από τα στοιχεία μπορεί να αντικατασταθεί από ένα θετικό σχόλιο σε διαστήματα που καθορίζονται από την *αυστηρότητα των δασκάλων*.
- Εξαγωγή των πληροφοριών λάθους για παραγωγή
 1. ενός κλειδιού για την πρόσβαση σε πληροφοριών που αποθηκεύονται στη βάση δεδομένων ανατροφοδότησης
 2. ενός κωδικού για την απεικόνιση του λάθους
 3. μιας γενικής βαθμολόγησης της απόδοσης, βασισμένης στο ποσοστό λαθών

- Επισύναψη του αρχικού καταλόγου λαθών στο ημερολόγιο των σπουδαστών συμπεριλαμβανομένης της βαθμολόγησης.
- Ενημέρωση του προφίλ των σπουδαστών με βάση τον τρέχοντα κατάλογο λαθών.

Η παρακάτω εικόνα δείχνει τις εισόδους και εξόδους των βαθμίδων.



Εικόνα 56: Διάγραμμα εισόδων-εξόδων της Βαθμίδας Αξιολόγησης της Εκτέλεσης.

4.2 Εξάρτηση Αξιολόγησης Απόδοσης από το Επίπεδο του Μαθητή

Ανάλογα με το επίπεδο των μαθητών και το μουσικό όργανο είναι διαφορετικά τα κριτήρια αξιολόγησης της εκτέλεσης των μαθητών.

Αρχικά σύμφωνα με ερωτηματολόγια που συμπληρώθηκαν από καθηγητές μουσικής σε διάφορα όργανα, ταξινομήθηκαν τα λάθη με βάση τη σημασία τους στο συγκεκριμένο επίπεδο που βρίσκεται ο εκάστοτε μαθητής. Παρακάτω παρουσιάζουμε ένα Πίνακα όπου φαίνονται ταξινομημένα τα κριτήρια αξιολόγησης ενός μαθητή για τα δύο πρώτα χρόνια μουσικών σπουδών στο συγκεκριμένο όργανο.

Rank order	Which of the following aspects are relevant in judging the student's performance in the first two years of study ?	average
1	Correct (or clean) attack of notes	4.4
2	Correct rhythm	4.2
2	Correct fingering	4.2
3	Quality of sound	4.1
4	Posture	4.0
4	Play in tune	4.0
5	Correct respiration	3.8
6	Keep time	3.7
7	Continuity of execution	3.6
8	Correct phrasing (legato/staccato)	3.4
9	Dynamics	3.2

Πίνακας 3 : Απαντήσεις 34 δασκάλων μουσικής στην ερώτηση: «Ποιες από τις παρακάτω πτυχές χρησιμοποιούνται για την αξιολόγηση ενός μαθητή στα δύο πρώτα χρόνια διδασκαλίας;» (κλίμακα 1= ΟΧΙ ΣΗΜΑΝΤΙΚΟ, 5= ΠΟΛΥ ΣΗΜΑΝΤΙΚΟ)

Κρίνοντας από τις παραπάνω απαντήσεις των ειδικών βλέπουμε ποιο είναι το σύνολο των λαθών μεγαλύτερης σημασίας (τιμή ≥ 4). Οι πιο σημαντικές πτυχές μιας εκτέλεσης είναι τα σωστά attacks, ο ρυθμός, ο δακτυλισμός, η ποιότητα του ήχου, ο σωστός τόνος και η σωστή στάση του σώματος.

Σε δεύτερο επίπεδο, ζητήθηκε από καθηγητές φλογέρας να περιγράψουν τα κύρια λάθη που παρουσιάζονται σε εκτελέσεις αρχαρίων κατά τα δύο πρώτα χρόνια διδασκαλίας στο συγκεκριμένο όργανο. Στον παρακάτω Πίνακα παρουσιάζονται τα πιο συχνά λάθη. Παρατηρούμε ότι τα κύρια λάθη σχετίζονται με κακό έλεγχο στο φύσημα στο στόμιο του οργάνου, με κακό δακτυλισμό, με φτωχά attacks και με λανθασμένο ρυθμό.

Rank order	Number of answers	Type of error
1	8	Poor control of blowing (air flow, generally blowing too hard)
1	8	Fingering technique (covering of tone holes, motoric difficulties)
1	8	Poor attacks (not using tongue correctly, 'tu' du')
1	8	Wrong rhythm
2	4	Posture & holding of the instrument
2	4	Reading the music (not the correct pitch, forgetting accidentals)
3	2	Playing wrong notes (not mapping the note to the correct fingering)
3	2	Shifting left and right hand
4	1	Intonation
4	1	Not keeping stable tempo
4	1	Phrasing (not keeping groups of notes together)

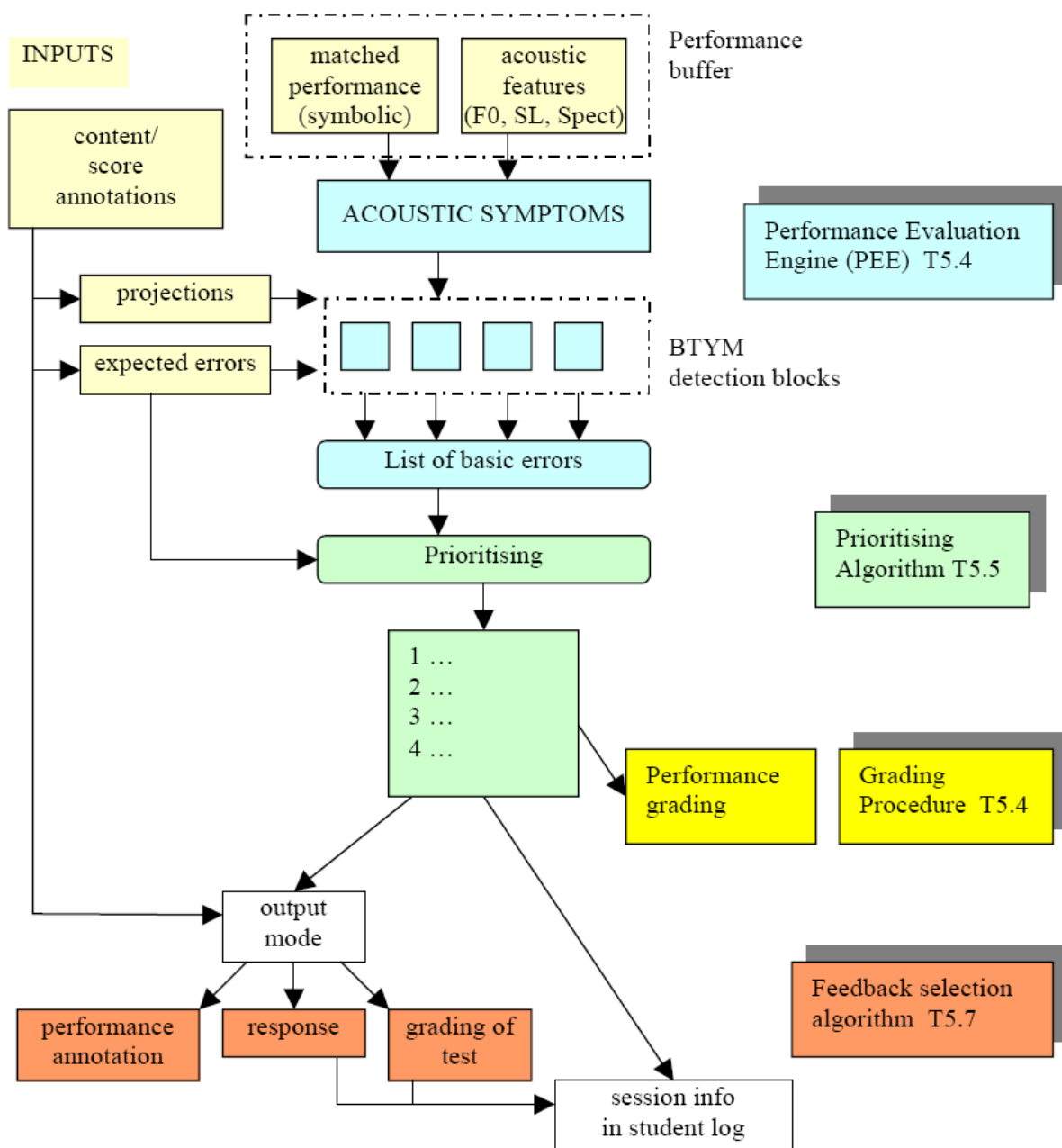
Πίνακας 4 Τυπικά βασικά λάθη σε μαθητές κατά τη διάρκεια των δύο πρώτων χρόνων διδασκαλίας

Τα λάθη που είναι υπογραμμισμένα με κίτρινο χρώμα ανιχνεύονται καλύτερα και διορθώνονται από τους καθηγητές σε παραδοσιακά μαθήματα σε αίθουσες διδασκαλίας. Ο δάσκαλος δείχνει τον σωστό τρόπο και ο μαθητής επαναλαμβάνει. Αν είναι απαραίτητο ο καθηγητής δίνει επιπλέον οδηγίες και συμβουλές. Τα λάθη που είναι υπογραμμισμένα με πράσινο χρώμα είναι αυτά που μπορούν να ανιχνευτούν στο ηχητικό σήμα, παρόλο που δεν είναι πάντα εφικτή η σωστή ερμηνεία της αιτίας του λάθους (υπογραμμισμένα με πράσινο και κίτρινο χρώμα στον Πίνακα). Στη συνέχεια του κεφαλαίου ακολουθεί μια μελέτη εξαγωγής παραμέτρων που μπορούν να χαρακτηρίζουν αυτό το σύνολο λαθών. Συγκεκριμένα εξάγουμε χαρακτηριστικά από προβληματικές νότες, στην προσπάθειά μας να εξάγουμε γενικούς κανόνες για την ποιότητα της νότας.

No	Type of mistake/error		Aspects considered by PEM	output level	Acoustical/musical correlates
1	Air flow	1	Pitch/ sound level	perf	F0 & sound level stability
2	Fingering	2	Hesitation	note	prolonged IOI & gap between notes
3	Rhythmic performance	3a	Reading score	note	timing (beats per note value)
		3b	Execution	note	relations between note values
4	Attack	4	Quality	perf	good ('du'/tu'), sharp ('tu'), blown (no tongue)
5	Melodic performance	5a	Reading score	note	insertions/ deletions
		5b	Reading/ fingering	note	wrong notes (pitches)
6	Tempo	6a	Speed	perf	overall tempo (BPM)
		6b	Stability	perf	tempo fluctuations
7	Intonation	7	High - low	note	deviation from tuning ref. & equal temperament
8	Phrasing	8	Breathing	note	gaps between notes at breathing marks
9	Articulation	9	Staccato - legato	perf	relative duration & gap

Πίνακας 5 Βασικό σύνολο τυπικών λαθών της Βαθμίδας Αξιολόγησης

4.3 Αρχιτεκτονική Βαθμίδας Αξιολόγησης της Εκτέλεσης



Εικόνα 57 : Διάγραμμα του σχεδιασμού της Βαθμίδας Αξιολόγησης της Εκτέλεσης.

4.4 Ποιότητα του Τόνου

Στα πνευστά όργανα, όπως και στα περισσότερα άλλα μουσικά όργανα, η ποιότητα του τόνου είναι ένα κύριο χαρακτηριστικό γνώρισμα του ήχου. Η ποιότητα του τόνου εξαρτάται έντονα από τον εκτελεστή και πώς ελέγχει τη ροή αέρα από τα χείλια ως το δακτυλισμό. Είναι πολύ σημαντικό για το νέο σπουδαστή να διδάσκεται τις σωστές τεχνικές στα πρώτα έτη. Ο κύριος τρόπος διδασκαλίας είναι συνεδρίες σύντομης διάρκειας με δάσκαλο συνήθως μία φορά την εβδομάδα. Θα ήταν επιθυμητό, ωστόσο, να είναι δυνατό να δίνεται βοήθεια στους σπουδαστές να εξασκούνται επίσης μόνοι στο σπίτι.

Το VEMUS προορίζεται να δίνει αυτή τη δυνατότητα με ηχογράφηση της εκτέλεσης του μαθητή και ανάλυση και αξιολόγηση του μουσικού σήματος.

Είναι σημαντικό για την αξιολόγηση αυτή να μπορούμε να διακρίνουμε την ποιότητα του τόνου και να διαχωρίσουμε κατηγορίες «κακών» ήχων.

Η ποιότητα του τόνου ή αλλιώς χροιά, μπορεί να περιγραφεί από την σχέση μεταξύ των φασματικών συστατικών του, τη σύνθεση των αρμονικών. Τόσο το πλήθος, όσο και η διαφορετικότητα των αρμονικών που προστίθενται στην θεμελιώδη συχνότητα, επηρεάζουν την ποιότητα του τόνου, τη φωτεινότητα και την έντασή του. Πολλές φορές όμως τα φασματικά συστατικά ενός ήχου δεν είναι αρμονικά συσχετισμένα μεταξύ τους (partials), το οποίο επηρεάζει την ποιότητα του τόνου. Τα partials αυτά μπορούν να παραχθούν με διάφορους τρόπους και όταν αυτό συμβαίνει παράγεται ένας προβληματικός ήχος.

Οι συχνότητες αυτές μπορούν να παραχθούν κατά τη διάρκεια του attack ή κατά τη διάρκεια του decay μιας νότας. Οι συχνότητες των transients των προβληματικών νοτών πλαταίνουν και σχηματίζουν τις κορυφές του spectrum καθώς ενώνονται με άλλα κοντινά partials, επηρεάζοντας με αυτόν τον τρόπο την τονική ποιότητα του ήχου.

Ποιοτικές ενδείξεις της τονικής ποιότητας είναι:

- **Spectral centroid:** συσχετίζεται με την αντιληπτή «φωτεινότητα» του ήχου. Δείχνει την συχνότητα όπου είναι συγκεντρωμένη η ενέργεια του ήχου.
- **Inharmonicity:** Παίζει σημαντικό ρόλο στην αναγνώριση του ήχου. Αποτελεί μέτρο του κατά πόσο τα partials είναι ακριβείς αρμονικές.
- **Irregularity:** Η διακύμανση της ενέργειας μεταξύ των partials.
- **Odd and even:** Ο λόγος των περιττών και άρτιων αρμονικών.
- **Roughness:** Ένα μέτρο της αντιληπτής ενόχλησης που προκαλείται από την παρουσία των partials.

4.5 Κατηγορίες προβληματικών τόνων

Τέσσερις είναι οι βασικές κατηγορίες προβληματικών τόνων που βρέθηκαν κατά τη διάρκεια της ανάλυσης των εκτελέσεων κλαρινέτων και που παρατηρήθηκαν επανειλημμένα από τους σπουδαστές. Είναι σημαντικό να αναφέρουμε ότι μερικοί από τους όρους που χρησιμοποιούνται για να περιγράψουν τους τόνους μπορούν να μην είναι ευρέως χρησιμοποιούμενοι. Ορισμένοι όροι εισάγονται ειδικά, όπως οι «διπλοί τόνοι» βασισμένοι στον τρόπο που ηχούν.

1. «Τριξίματα»

Τα «τριξίματα» (squeaks) έχουν συνήθως μια ποιότητα μεταλλικού, σκληρού και να διαπεραστικού ήχου και το ύψος τους είναι πολύ υψηλότερο από τον ακόλουθο πραγματικός τόνο. Τα τριξίματα προκαλούνται από κακό έλεγχο της επαφής των χειλιών με το επιστόμιο του μουσικού οργάνου ή της ροής αέρα, ή διαρροή στο όργανο (κάλυψη των τρυπών).

2. Ασταθείς Τόνοι

Οι ασταθείς τόνοι (unstable tones) δεν έχουν σταθερό ύψος με την πάροδο του χρόνου. Είναι συνήθως προκαλούμενοι από ασταθή επαφή των χειλιών με το επιστόμιο του μουσικού οργάνου ή την ασταθή πίεση αέρα. Τόνοι που πάσχουν από αυτό το είδος προβλήματος είναι συχνά μακροχρόνιοι τόνοι που εμφανίζονται στο τέλος μιας φράσης.

3. Κενοί Τόνοι



Οι άδειοι ή κενοί τόνοι (hollow/empty tones) χαρακτηρίζονται ως μην έχοντες σχήμα, σώμα, φωτεινότητα ή βάθος. Συχνά προκαλούνται από σπασμένη ή στεγνή γλωττίδα, ανεπαρκή πίεση ή επαφή στα χείλια ή πολύ ήπια attacks. Οι άδειοι τόνοι μπορούν επίσης να γίνουν αντιληπτοί ως αδύναμοι τόσο στην ηχηρότητα όσο και στο περιεχόμενο.

4. Διπλοί Τόνοι



Οι διπλοί τόνοι (double tones) γίνονται αντιληπτοί ως έχοντες δύο ύψη, το πρώτο πιο αδύναμο από το δεύτερο επιδιωκόμενο ύψος. Οι διπλοί τόνοι παράγονται όταν ο εκτελεστής δεν κατορθώνει να παίξει το σωστό τόνο από την αρχή, π.χ. λόγω προβλημάτων με το συντονισμό των δακτύλων.

4.6 Φασματική Ανάλυση των Προβληματικών Νοτών σε Εκτελέσεις Αρχαρίων

Στο σημείο αυτό εξετάζεται φασματικά ένα αντιπροσωπευτικό σύνολο από προβληματικές σε ποιότητα νότες. Βασισμένοι στην προηγούμενη ανάλυση των προβληματικών νοτών, είχαμε στη διάθεσή μας υλικό προς ανάλυση από επαγγελματία καθηγητή κλαρινέτου, από τον οποίον ζητήθηκε να αναπαράγει προβληματικές νότες των παραπάνω κατηγοριών καθώς και καλής ποιότητας αντίστοιχες νότες προς σύγκριση του φάσματός τους. Οι τόνοι που παρήχθησαν, εκτελέστηκαν σε *mezzo forte (mf)* το οποίο είναι σύντομο σε διάρκεια και πλούσιο σε αρμονικές.

Συγκεκριμένα 4 νότες επιλέχθηκαν ως οι αντιπροσωπευτικότερες νότες που συχνά οι αρχάριοι μαθητές κλαρινέτου αντιμετωπίζουν πρόβλημα στην εκτέλεσή τους. Οι νότες αυτές είναι οι F^{Low} , G^{Mid} , B^{Mid} , A^{Up} .

Στον Πίνακα παρουσιάζονται οι αντιστοιχίες των νοτών στο B_b κλαρινέτο και η θεμελιώδης συχνότητά τους με κούρδισμα $A_4=440$ Hz.

Written notes	Sounding notes	Fundamental frequency
F^{Low}	Eb_3	155.6 Hz
G^{Mid}	F_4	349.2 Hz
B^{Mid}	A_4	440.0 Hz
A^{Up}	G_5	784.0 Hz

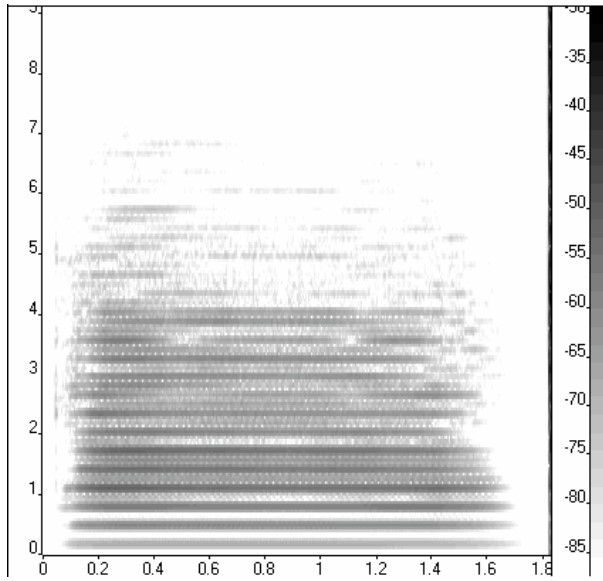
Πίνακας 6 Νότες και η θεμελιώδης συχνότητά τους

Για το F^{Low} (Εικόνα 58) ο *mf* τόνος έχει αρμονικές μέχρι τα 6.9 kHz περίπου (περισσότερες από 40 αρμονικές), συμπαγείς μέχρι τα 4.2 kHz.

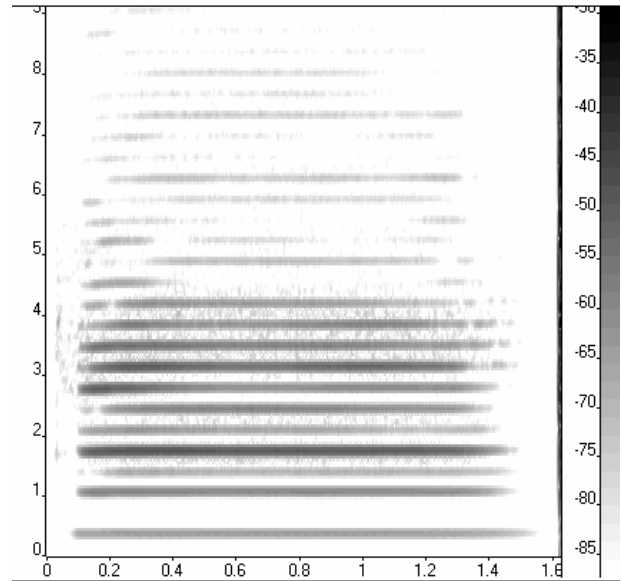
Για το G^{Mid} (Εικόνα 59) ο *mf* τόνος έχει αρμονικές μέχρι τα 8.2 kHz (περίπου 20 αρμονικές), συμπαγείς μέχρι τα 4.4 kHz.

Για το A^{Up} (Εικόνα 60) ο *mf* τόνος έχει αρμονικές μέχρι τα 6.5 kHz (8 αρμονικές). Επιπλέον ο *mf* τόνος έχει θόρυβο μεταξύ των 4 πρώτων αρμονικών.

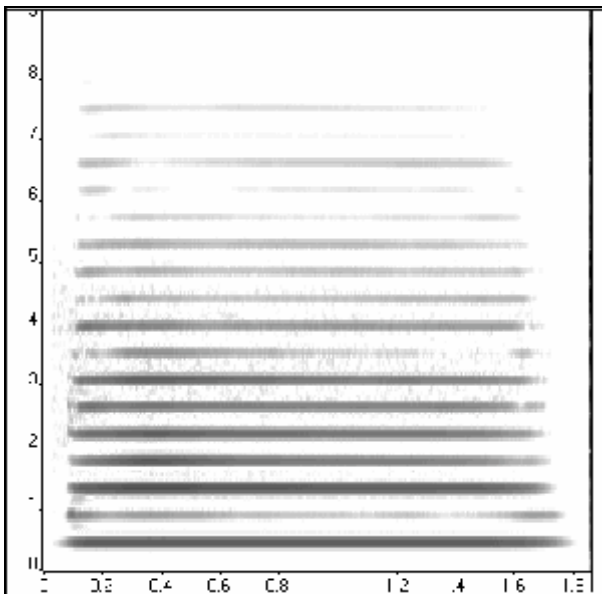
Για το **BMid** (Εικόνα 61) ο *mf* τόνος έχει συμπαγείς αρμονικές μέχρι τα 3.3 kHz και ψεύτο-αρμονικές μέχρι τα 7.6 kHz.



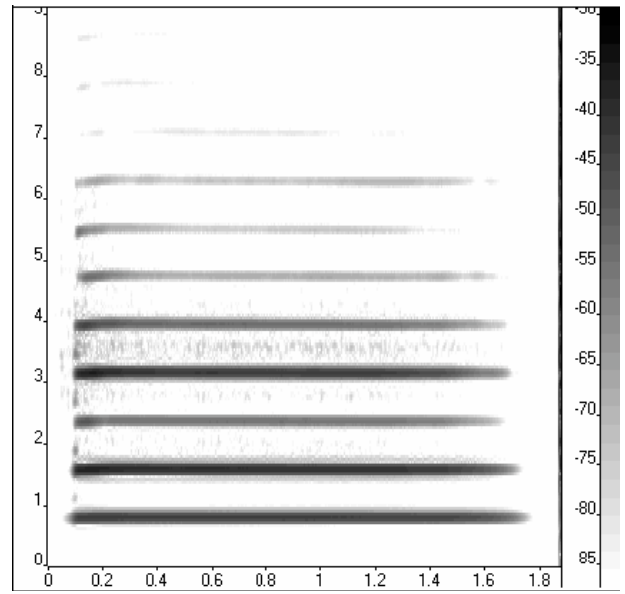
Εικόνα 58: FLow mf φασματογράφημα



Εικόνα 59 : GMid mf φασματογράφημα



Εικόνα 60 : AUr mf φασματογράφημα

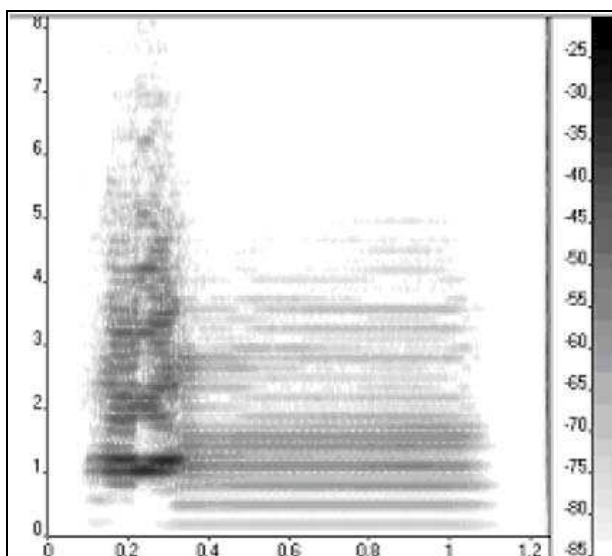


Εικόνα 61 :BMid mf φασματογράφημα

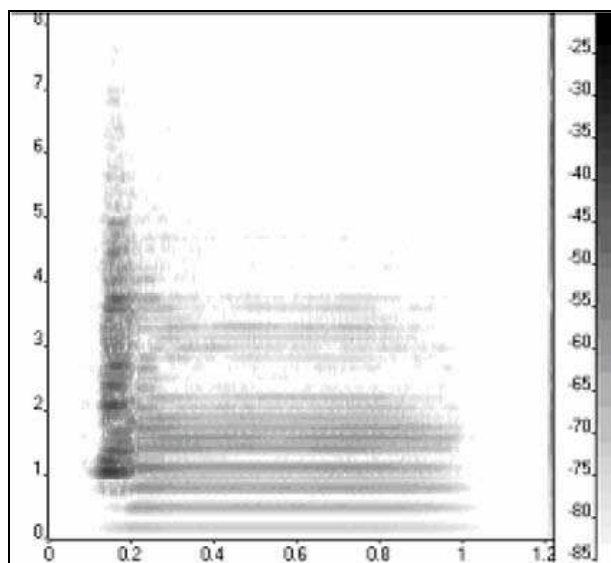
4.6.1 Τριζήματα

Στην ενότητα αυτή θα εξετάσουμε την πρώτη κατηγορία προβληματικών τόνων, τα τριζήματα. Τα τριζήματα μπορούν να οφείλονται είτε σε κακό συντονισμό των δακτύλων, είτε σε κακή κατάσταση του στομίου του οργάνου. Θα εξεταστούν δύο χαρακτηριστικές νότες που συχνά εμφανίζουν πρόβλημα σε εκτελέσεις αρχαρίων και θα σχολιαστούν διαφορές μεταξύ των καλής ποιότητας

αντίστοιχων νοτών και των προβληματικών νοτών, στην προσπάθεια να εξάγουμε κάποια συνολικά συμπεράσματα για αυτή την κατηγορία προβληματικής νότας.



Εικόνα 62 :Τριξίμο οφειλόμενο σε κακό συντονισμό των δακτύλων

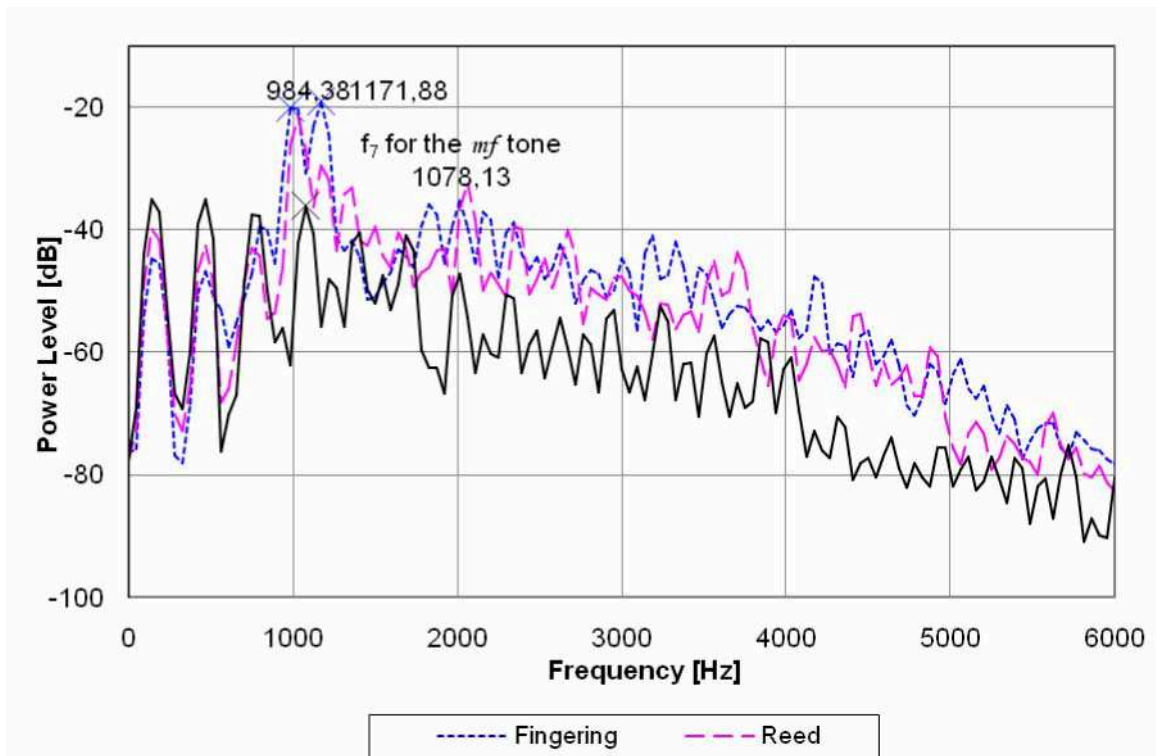


Εικόνα 63: Τριξίμο οφειλόμενο στο στόμιο

Στη νότα F^{Low} το squeak που οφείλεται τόσο σε κακό δακτυλισμό όσο και στην κατάσταση του στομίου του οργάνου παρατηρείται στην αρχή της κάθε νότας και κατά τη διάρκειά τους και χαρακτηριστική είναι η απουσία της θεμελιώδους συχνότητας καθώς και της πρώτης αρμονικής, όπως φαίνεται στα σχήματα 62 και 63. Επίσης, και στους δύο τύπους παρατηρείται από το φασματογράφημά τους ενέργεια μεταξύ των partials, που αντιστοιχεί στις αρμονικές του ακόλουθου, καλής ποιότητας ήχου.

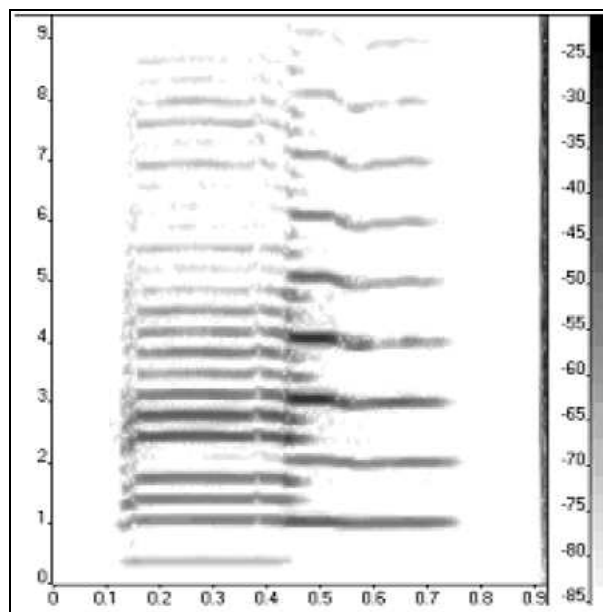
Τα τριξίματα, έχουν εντελώς διαφορετικό ύψος από το ακόλουθο καλής ποιότητας τμήμα, όπως φαίνεται από το γραμμικό φάσμα του σχήματος 58.

Στην Εικόνα 64 βλέπουμε το μέσο φάσμα του κανονικού τόνου F^{Low} μαζί με τα φάσματα των προβληματικών ήχων. Παρατηρούμε ότι οι προβληματικοί ήχοι έχουν κοινά χαρακτηριστικά με τον κανονικό τόνο. Οι αρμονικές f_0 , f_2 , f_4 των δύο squeak έχουν μόνο χαμηλότερη ένταση σχετικά με τον κανονικό τόνο, ενώ οι επόμενες αρμονικές είναι γενικά δυνατώτερες. Η διαφορά στη μορφή των προβληματικών από τον κανονικό τόνο παρατηρείται από 1000-1200 Hz, με κορυφές σε συχνότητα περίπου ίση με αυτή της έκτης αρμονικής του αντίστοιχου κανονικού τόνου. Χαρακτηριστικές είναι οι δύο κορυφές στην περίπτωση του squeak από δακτυλισμό, οι οποίες δημιουργούν την εντύπωση του διπλού τόνου.



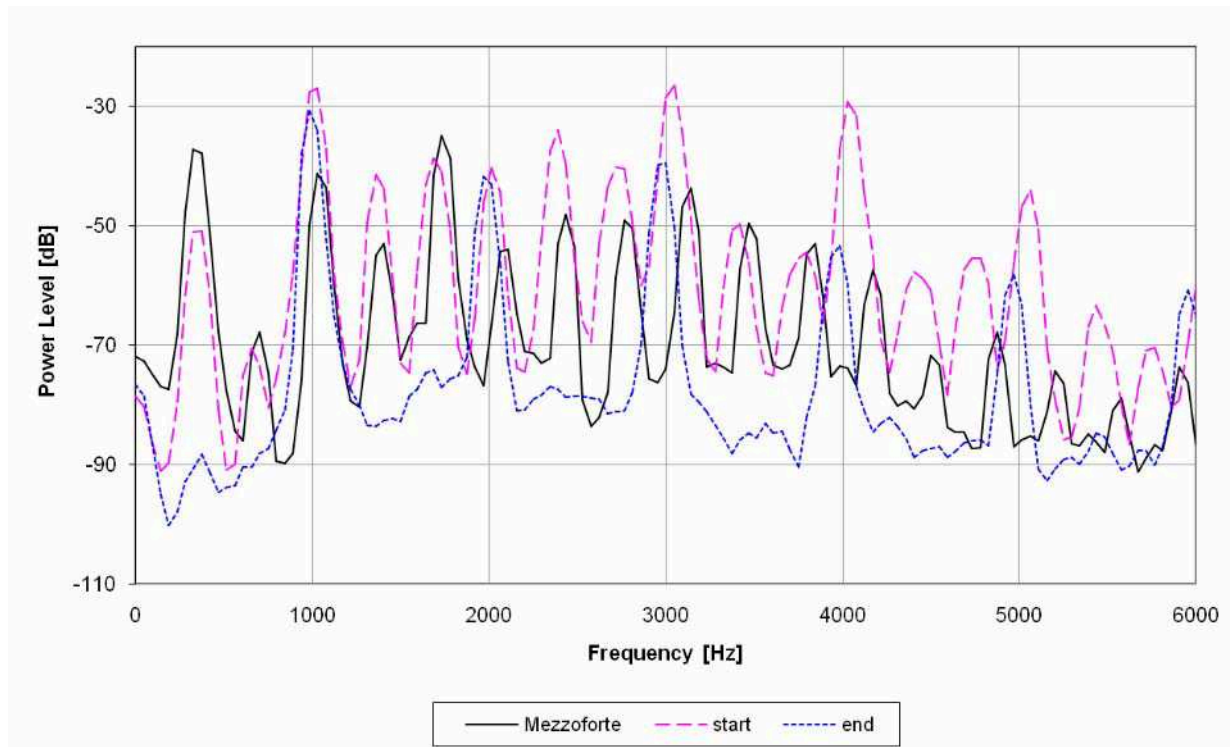
Εικόνα 64: Γραμμικό φάσμα του τόνου FLow mf καλής ποιότητας ήχου και του προβληματικού squeak τμήματος εξαιτίας κακού δακτυλισμού και κακού στομίου αντίστοιχα.

Στη νότα G^{Mid} το τρίξιμο (Εικόνα 65) δημιουργεί μια ευδιάκριτη αλλαγή στο φασματογράφημα του τόνου. Κατά τη διάρκειά του παρατηρείται μετατόπιση της θεμελιώδους συχνότητας κατά μία οκτάβα και $1/5$ πάνω από την f_0 του τόνου (987 Hz). Στην αρχή του τόνου και για περίπου 50ms οι προηγούμενες αρμονικές υπάρχουν αν και σταδιακά πέφτουν.



Εικόνα 65 G^{Mid} τρίξιμο

Στην Εικόνα 66 βλέπουμε το φάσμα του προβληματικού από τριζίμο G^{Mid} τόνου στην αρχή και στο τέλος του συγκριτικά με το μέσο γραμμικό φάσμα του κανονικού τόνου. Παρατηρούμε ότι στην αρχή όλες οι αρμονικές του τόνου υπάρχουν, αν και η συχνότητά τους έχει μετατοπιστεί προς τα κάτω. Επίσης η έντασή τους είναι ελαφρώς μεγαλύτερη συγκριτικά με εκείνη του κανονικού τόνου. Στο τέλος του προβληματικού τόνου παρατηρούμε μετατόπιση των μερικών του κατά μία οκτάβα και 1/5 υψηλότερα.



Εικόνα 66 : Φάσμα του προβληματικού από τριζίμο G^{Mid} τόνου στην αρχή και στο τέλος του συγκριτικά με το μέσο γραμμικό φάσμα του κανονικού τόνου

Συμπέρασμα

Από την παραπάνω μελέτη του φάσματος των προβληματικών από τριζίμο τόνων συνοψίζουμε κάποια συμπεράσματα τα οποία μπορούν να μας βοηθήσουν στην εξαγωγή συμπερασμάτων για το αν μια εκτέλεση εμφανίζει αυτό το είδος προβληματικής ποιότητας τόνου.

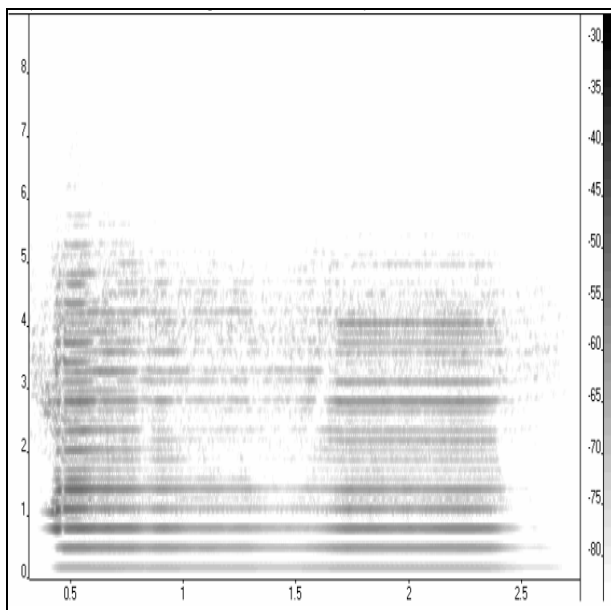
Τα τριζίματα μπορούν να εμφανιστούν στην έναρξη του τόνου, όπως στον τόνο FLow καθώς και στο τέλος όπως στον τόνο GMid.

Για τα τριζίματα στην έναρξη του τόνου μπορεί να συναχθεί το συμπέρασμα ότι τα φάσματα παρουσιάζουν πολύ υψηλότερα μέγιστα στις μερικές τους αλλά και λιγότερο έντονες εμβυθίσεις ενδιάμεσα. Επίσης, οι μερικές εμφανίζονται συχνά μετατοπισμένες. Επιπλέον, υπάρχουν αιχμές μεταξύ των πραγματικών αρμονικών του τόνου.

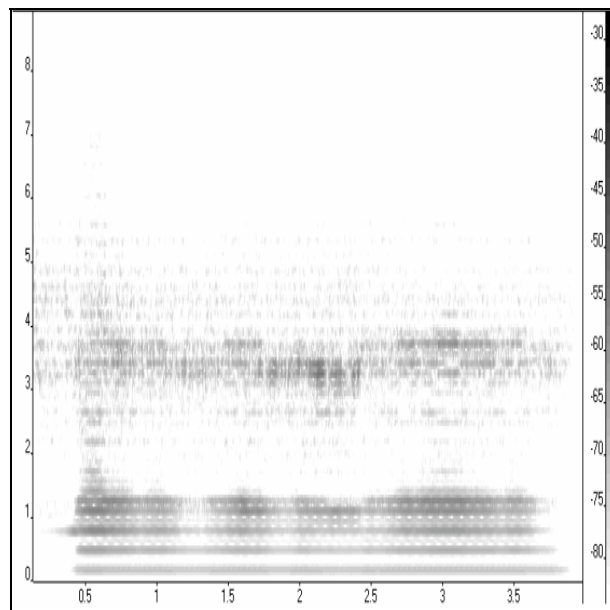
Για τα τριζίματα στο τέλος του τόνου, οι αρμονικές έχουν πολύ υψηλότερο επίπεδο απ' ότι κατά τη διάρκεια του καλού μέρους του τόνου, και εμφανίζονται μερικές μεταξύ των πραγματικών αρμονικών. Επιπλέον, αυτός ο τύπος στριγκλίσματος τείνει να θεωρηθεί ως ένας άλλος τόνος, δεδομένου ότι η θεμελιώδης συχνότητα του πραγματικού τόνου εξαφανίζεται.

4.6.2 Κενές Νότες

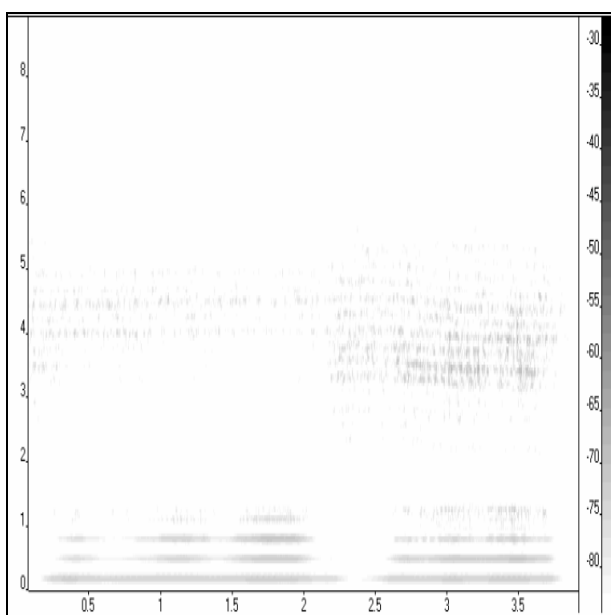
Στην ενότητα αυτή θα εξετάσουμε τη δεύτερη κατηγορία προβληματικών τόνων, τις κενές νότες. Αυτή η κατηγορία προβληματικής νότας έχει 3 διαφορετικά αίτια: κακή κατάσταση του στομίου του οργάνου, χαμηλή πίεση των χειλιών στο στόμιο και χαλαρή επαφή του στόματος με το στόμιο του οργάνου. Θα εξεταστούν δύο χαρακτηριστικές νότες που συχνά εμφανίζουν πρόβλημα σε εκτελέσεις αρχαρίων και θα σχολιαστούν διαφορές μεταξύ των καλής ποιότητας αντίστοιχων νοτών και των προβληματικών νοτών, στην προσπάθεια να εξάγουμε κάποια συνολικά συμπεράσματα για αυτή την κατηγορία προβληματικής νότας.



Εικόνα 67: Άδεια F^{Low} από κακή κατάσταση στομίου



Εικόνα 68 : Άδεια F^{Low} από ελλιπή πίεση των χειλιών



Εικόνα 69: Άδεια F^{Low} από κακή κατάσταση του στομίου

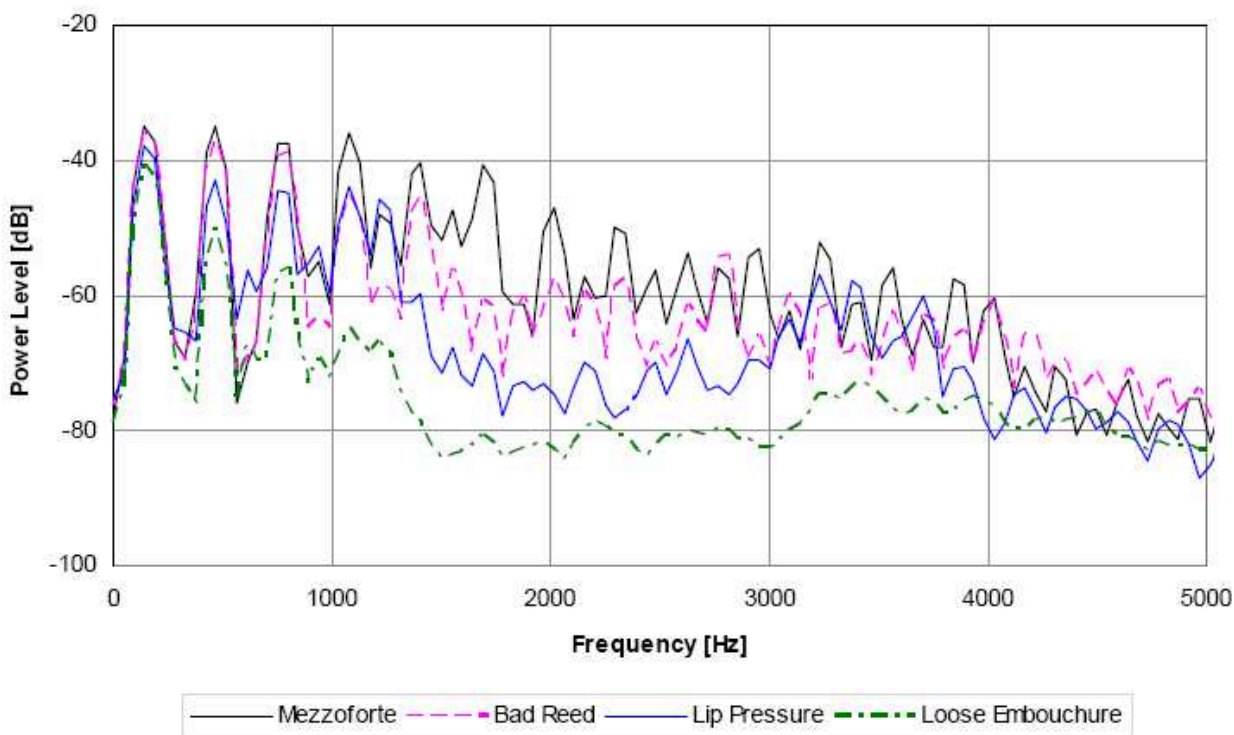
Στις Εικόνες 67, 68 και 69 παρουσιάζονται τα φασματογραφήματα του προβληματικού F^{Low} τόνου από τα 3 διαφορετικά αίτια που προαναφέραμε.

Στο φασματογράφημα της άδειας F^{Low} νότας που οφείλεται σε κακή κατάσταση του στομίου (Εικόνα 67) παρατηρούμε ομοιότητες με την κανονική νότα μέχρι την f_4 , αλλά οι υψηλότερες αρμονικές δεν είναι το ίδιο εμφανείς. Στις υψηλές συχνότητες οι αρμονικές εξαφανίζονται στο το φάσμα γεμίζει το θόρυβο.

Στο φασματογράφημα της άδειας F^{Low} νότας που οφείλεται σε λιγότερη πίεση των χειλιών (Εικόνα 68) μόνο η θεμελιώδης συχνότητα είναι εμφανής. Οι επόμενες αρμονικές είναι ασαφείς μέχρι τα 1,4 kHz, ενώ στις μεγαλύτερες συχνότητες μέχρι τα 5,5 kHz υπάρχει μόνο θόρυβος.

Στο φασματογράφημα της άδειας F^{Low} νότας που οφείλεται σε χαλαρή επαφή με το στόμιο του οργάνου (Εικόνα 69) εμφανίζονται οι αρμονικές έως και την f_2 οι οποίες όμως είναι ιδιαίτερα ασθενείς και διακόπτονται. Στις μεγαλύτερες συχνότητες δεν υπάρχουν καθόλου αρμονικές παρά μόνο θόρυβος μεταξύ 3,2 και 5,4 kHz. Στην χρονική διάρκεια 2,2-2,6 s όπου παρατηρείται η διακοπή των αρμονικών, ο ήχος περιέχει μόνο φύσημα και χάνεται ο τόνος.

Στην Εικόνα 70 παρουσιάζεται το μέσο γραμμικό φάσμα του F^{Low} τόνου στις 3 δυνατές περιπτώσεις άδειας νότας καθώς και του κανονικού F^{Low} .

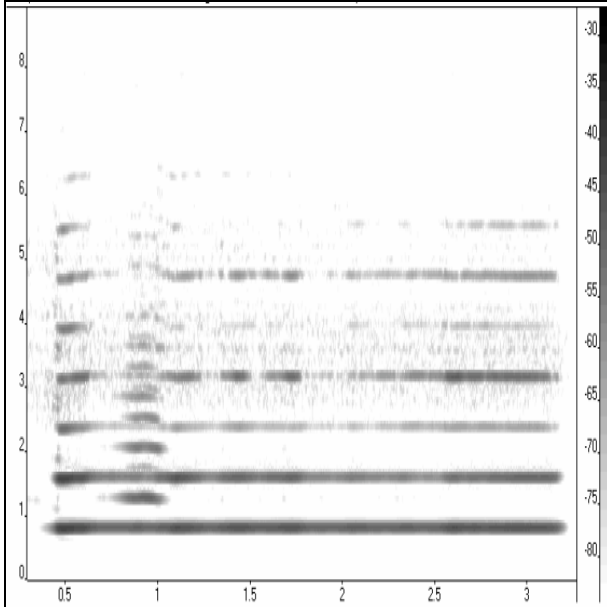


Εικόνα 70: Μέσο γραμμικό φάσμα του F^{Low} τόνου στις 3 δυνατές περιπτώσεις άδειας νότας και του κανονικού F^{Low}

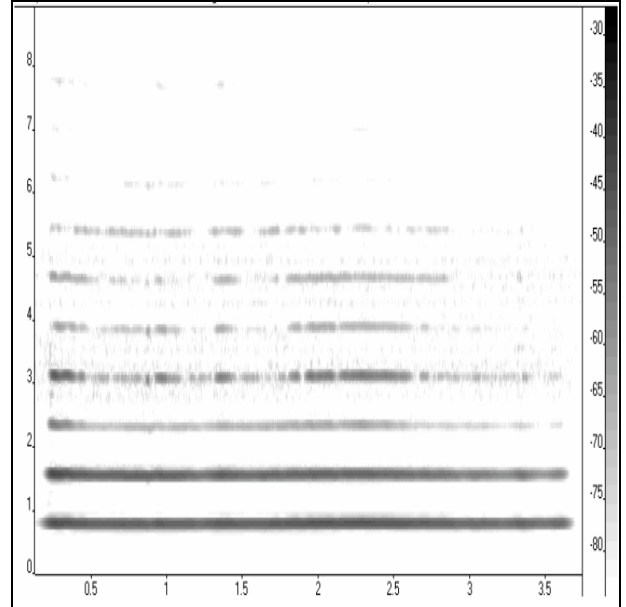
Παρατηρούμε από το παραπάνω διάγραμμα ότι οι αρμονικές πάνω από τα 1,1 kHz εμφανίζουν μεγάλη πτώση στην έντασή τους μέχρι περίπου τα 3kHz όπου η ένταση αυξάνει. Στην περίπτωση προβληματικής νότας F^{Low} από κακή κατάσταση στομίου οι f_0 f_2 f_4 έχουν την ίδια ένταση με τον κανονικό τόνο ενώ οι f_6 και f_8 έχουν μια πτώση 10Db. Στην περίπτωση που οφείλεται σε χαμηλότερη πίεση των χειλιών, παρατηρούμε ότι η θεμελιώδης συχνότητα έχει περίπου την ίδια ένταση με τον κανονικό τόνο, αλλά οι f_2 f_4 f_6 πέφτουν κατά περίπου 10 dB. Στην περίπτωση άδειας F^{Low} που οφείλεται σε κακή επαφή με το στόμιο του οργάνου εμφανίζονται μόνο οι 3 πρώτες αρμονικές αν και

η έντασή τους είναι χαμηλότερη κατά 15, 20 και 30 dB στις f_0 , f_2 , f_4 αντίστοιχα. Οι υψηλότερες αρμονικές, μετά το 1kHz δεν υπάρχουν.

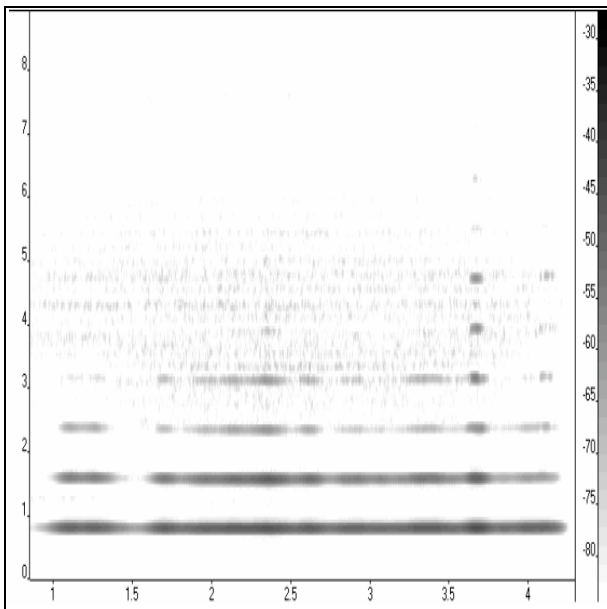
Στις παρακάτω εικόνες παρουσιάζονται τα αντίστοιχα φασματογραφήματα της νότας A^{Up} .



Εικόνα 71: Άδεια A^{Up} από κακή κατάσταση στομίου



Εικόνα 72: Άδεια A^{Up} από ελλιπή πίεση των χειλιών

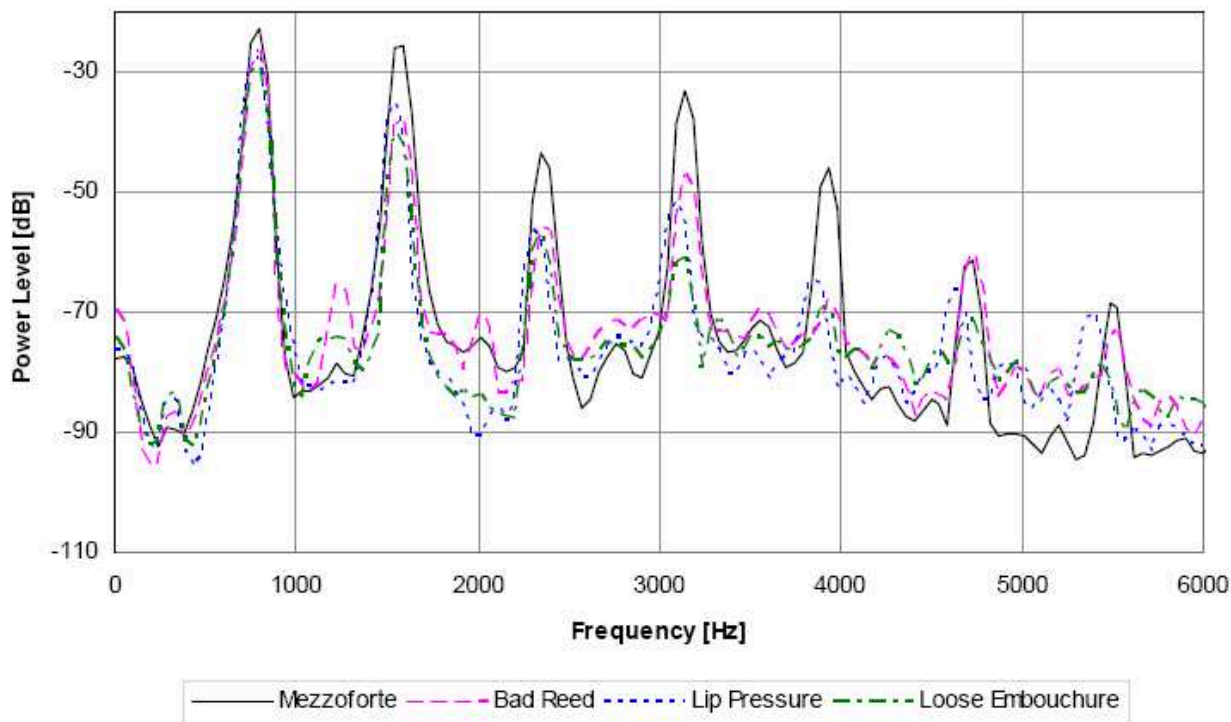


Εικόνα 73 : Άδεια A^{Up} από κακή κατάσταση του στομίου

Στην Εικόνα 71 παρουσιάζεται το φασματογράφημα της A^{Up} άδειας νότας που οφείλεται σε κακή επαφή με το στόμιο του οργάνου. Οι αρμονικές που φαίνονται εδώ, αν και δεν είναι καθαρές είναι οι f_0 - f_3 ενώ σε υψηλότερες συχνότητες υπάρχει μόνο θόρυβος.

Στην Εικόνα 72 που φαίνεται το φασματογράφημα της A^{Up} άδειας νότας που οφείλεται στην κακή κατάσταση του στομίου πρέπει να αγνοηθεί ότι συμβαίνει πριν το 1sec γιατί οφείλεται σε τρίζιμο. Ακολουθεί η άδεια νότα όπου είναι εμφανείς οι 4 πρώτες αρμονικές, η 5^η εξαφανίζεται, ενώ φαίνεται η 6^η αν και δεν είναι ευκρινής.

Στην Εικόνα 73 φαίνεται το φασματογράφημα της A^{Up} άδειας νότας που οφείλεται σε ελλιπή πίεση των χειλιών τόσο η θεμελιώδης συχνότητα όσο και οι αρμονικές είναι μετατοπισμένες προς μικρότερες τιμές όπως φαίνεται καθαρότερα στην Εικόνα 74.



Εικόνα 74: Μέσο γραμμικό φάσμα του A^{Up} τόνου στις 3 δυνατές περιπτώσεις άδειας νότας και του κανονικού A^{Up}

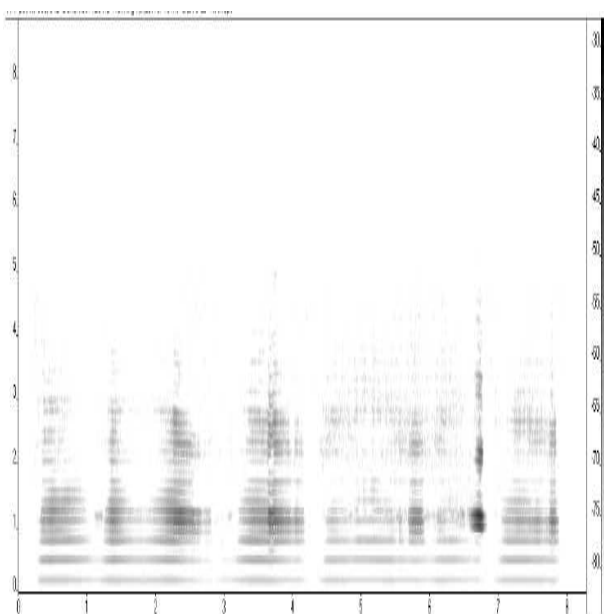
Στην Εικόνα 74 βλέπουμε το μέσο γραμμικό φάσμα του A^{Up} τόνου στις 3 δυνατές περιπτώσεις άδειας νότας καθώς και του κανονικού A^{Up} . Παρατηρούμε ότι η θεμελιώδης υπάρχει και στις 3 περιπτώσεις, αν και με ελαφρώς μικρότερη ένταση. Οι επόμενες αρμονικές σταδιακά πέφτουν σε ένταση και μετατοπίζονται προς μικρότερες συχνότητες, το οποίο γίνεται εμφανές κυρίως στις f_5 και f_6 .

Συμπεράσματα

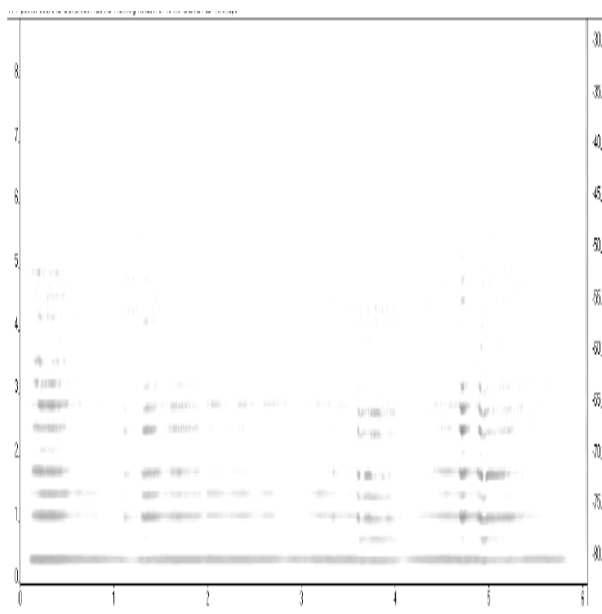
Όπως παρατηρούμε από τα φασματογραφήματα στις άδειες νότες η θεμελιώδης συχνότητα υπάρχει πάντοτε καθώς και οι πρώτες αρμονικές (συνήθως μέχρι την f_4), οι οποίες όμως εμφανίζονται με λιγότερη ένταση (10-15 dB) και βάθος σχετικά με αυτές του αντίστοιχου καλής ποιότητας τόνου και μετατοπισμένες προς μικρότερες συχνότητες όπως φαίνεται από τις Εικόνες 70 και 74 των γραμμικών τους φασμάτων. Οι υψηλές αρμονικές είναι συχνά περιορισμένες (ή εμφανίζονται κατά διαστήματα και με πολύ μικρότερη ένταση), ενώ στη θέση τους εμφανίζεται υψηλό ποσοστό θορύβου. Τέλος, εμφανίζονται διακυμάνσεις στην ένταση με το χρόνο όπως φαίνεται από τα φασματογραφήματα.

4.6.3 Ασταθείς τόνοι

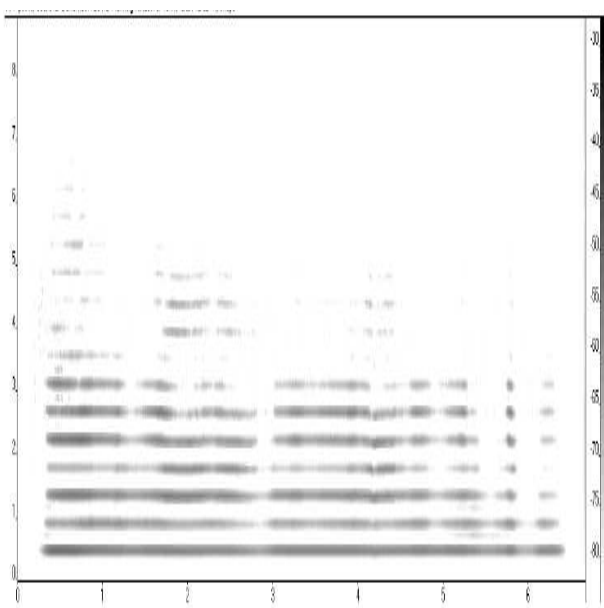
Στην ενότητα αυτή θα εξετάσουμε την τρίτη κατηγορία προβληματικών τόνων, τα τριξίματα. Θα εξεταστούν οι τέσσερις νότες που συχνά εμφανίζουν πρόβλημα σε εκτελέσεις αρχαρίων και θα σχολιαστούν διαφορές μεταξύ των καλής ποιότητας αντίστοιχων νοτών και των προβληματικών νοτών, στην προσπάθεια να εξάγουμε κάποια συνολικά συμπεράσματα για αυτή την κατηγορία προβληματικής νότας. Στις Εικόνες 75,76,77,78 βλέπουμε τα φασματογραφήματα των ασταθών Flow, GMid, BMid, AUr τόνων αντίστοιχα. Από τα φασματογραφήματα και των τεσσάρων τόνων μπορούμε να βγάλουμε κοινά συμπεράσματα. Εμφανίζουν διακοπές στις αρμονικές τους, κυρίως τις υψηλότερες. Επίσης έχουν φτωχό αρμονικό περιεχόμενο στις υψηλές συχνότητες. αν και οι 2 πρώτες αρμονικές συνήθως υπάρχουν σε όλη τη διάρκεια του τόνου, αν και εμφανίζονται με μικρότερη ένταση.



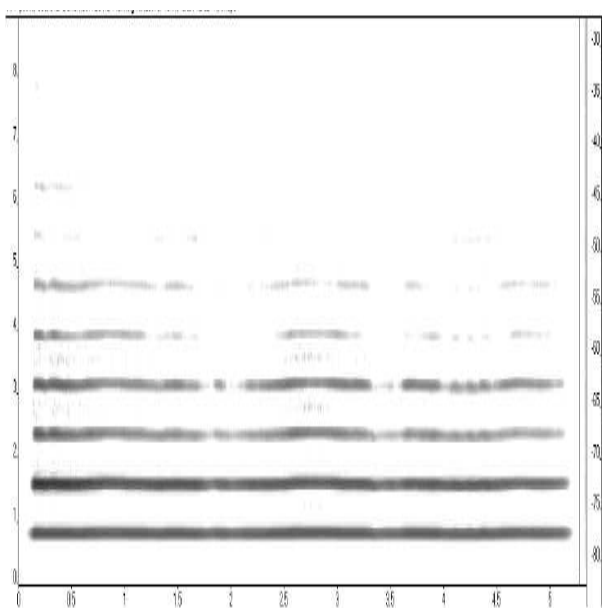
Εικόνα 75: Ασταθές Flow



Εικόνα 76 :Ασταθές GMid

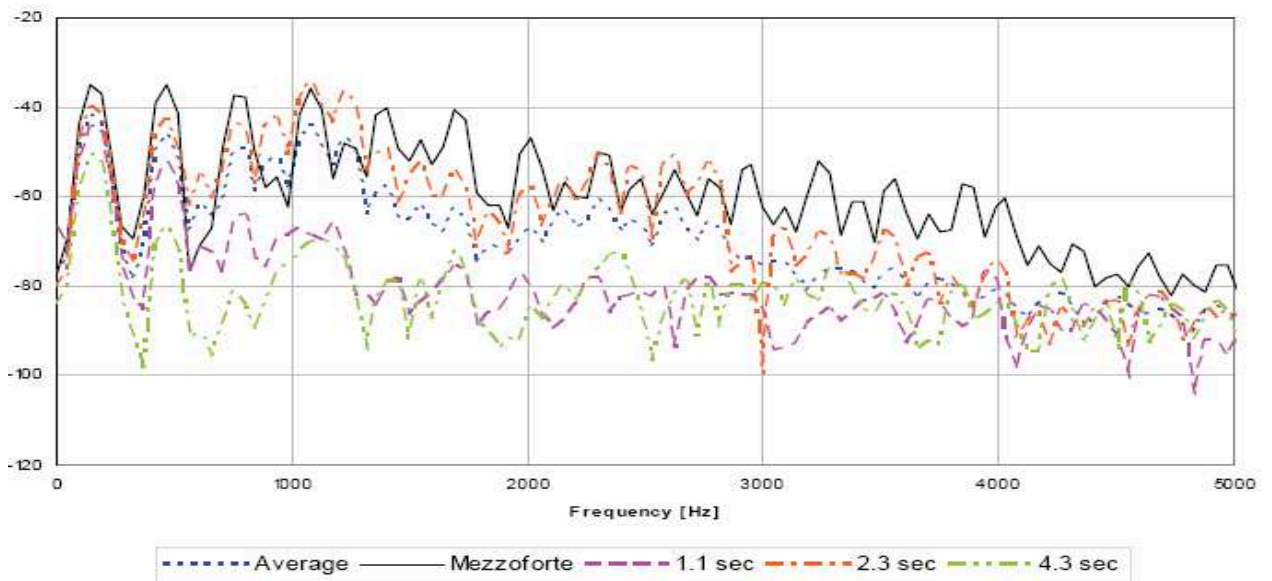


Εικόνα 77 :Ασταθές BMid

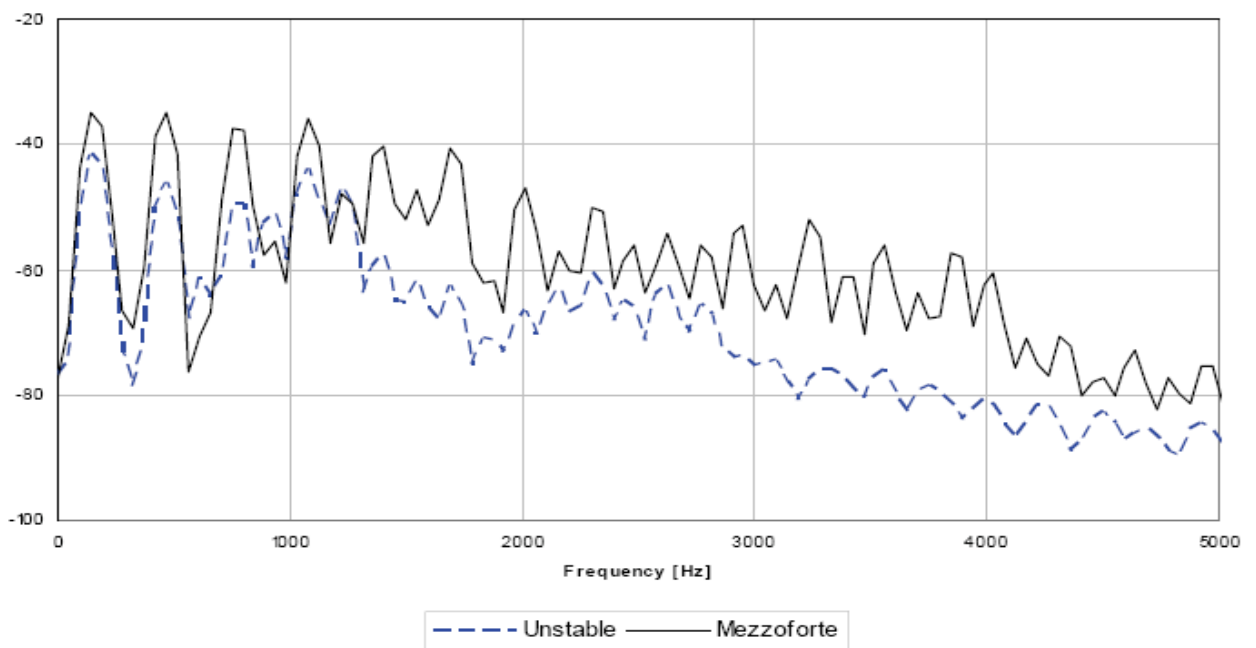


Εικόνα 78 :Ασταθές AUr

Παρακάτω παρουσιάζουμε τα γραμμικά φάσματα των ασταθών τόνων, αρχικά συγκρίνοντας τη μέση τιμή τους με αυτή του κανονικού αντίστοιχου τόνου και στη συνέχεια επιλέγουμε κάποιες συγκεκριμένες χρονικές στιγμές στη διάρκεια του τόνου και συγκρίνουμε το φάσμα τους με αυτό της μέσης τιμής του κανονικού τόνου.



Εικόνα 79 :Μέσο φάσμα του ασταθούς και του κανονικού F^{Low} καθώς και το φάσμα του ίδιου τόνου σε τρεις διαφορετικές χρονικές στιγμές

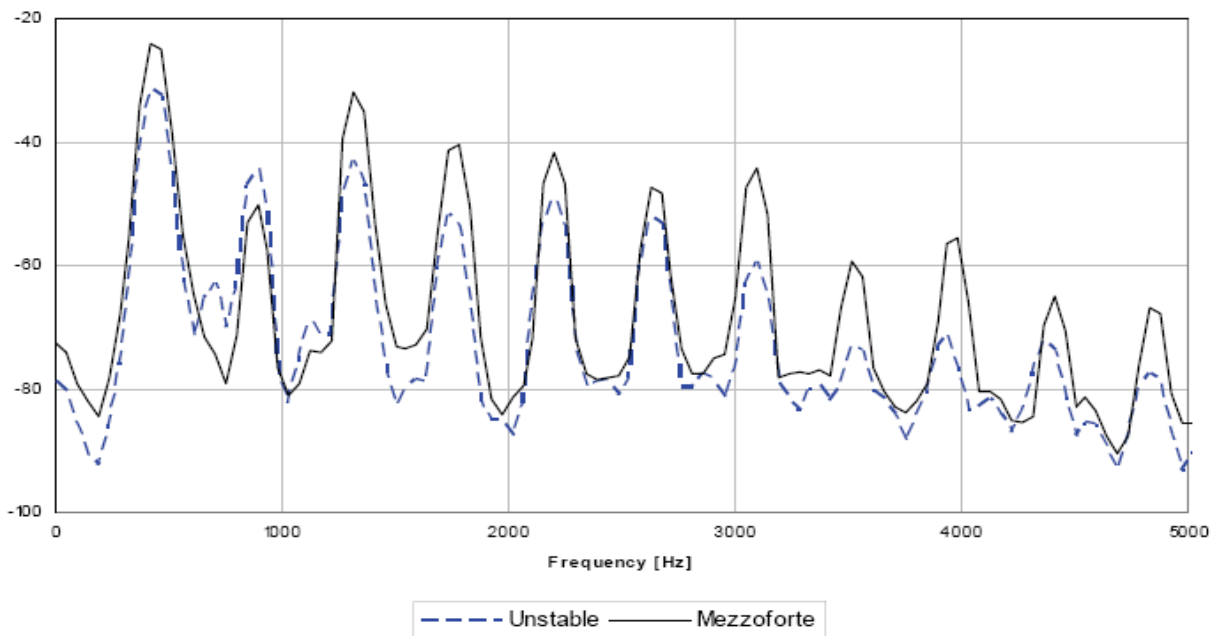


Εικόνα 80 :Μέσο φάσμα του ασταθούς F^{Low}

Παρατηρώντας τα μέσα φασματικά διαγράμματα παρατηρούμε ότι οι περισσότερες αρμονικές του κανονικού ήχου υπάρχουν αν και παρατηρείται πτώση στην έντασή τους. Όσο προχωράμε σε μεγαλύτερες συχνότητες οι αρμονικές δεν φαίνονται ευκρινώς.

Περισσότερες πληροφορίες μπορούμε να βγάλουμε συγκρίνοντας τα φάσματα σε διαφορετικές χρονικές στιγμές. Στην Εικόνα 79 παρουσιάζεται το μέσο φάσμα του ασταθούς και του κανονικού

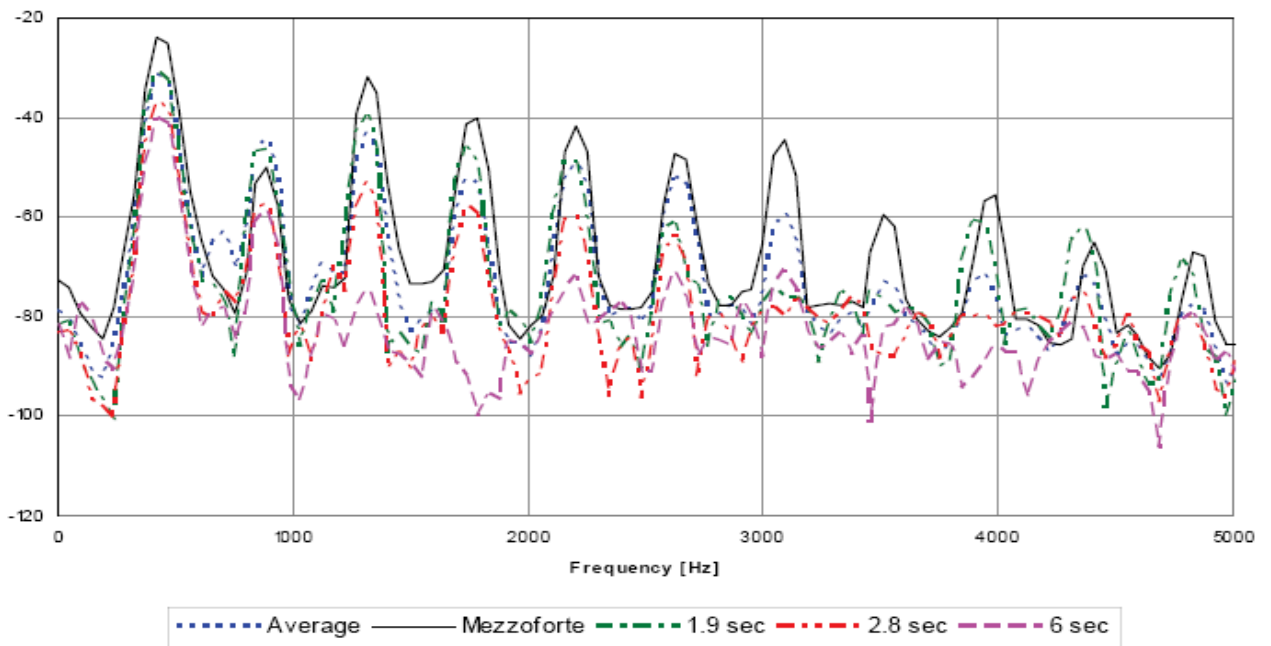
F^{Low} καθώς και το φάσμα του ίδιου τόνου σε τρεις διαφορετικές χρονικές στιγμές. Παρατηρούμε ότι στις τρεις διαφορετικές χρονικές στιγμές που παρουσιάζονται, υπάρχουν σημαντικές αυξομειώσεις στο φασματικό περιεχόμενο (π.χ. στο επίπεδο των μερικών). Η θεμελιώδης συχνότητα καθώς και η πρώτη αρμονική βλέπουμε ότι είναι παρούσες και στις τρεις χρονικές στιγμές αν και παρουσιάζουν αυξομειώσεις στο επίπεδο της έντασης. Οι άλλες αρμονικές χάνονται με το πέρασμα του χρόνου.



Εικόνα 81: Μέσο φάσμα του κανονικού και του ασταθούς B^{Mid}

Στην Εικόνα 81 βλέπουμε ότι το μέσο φάσμα του κανονικού και του ασταθούς B^{Mid} είναι παρόμοιο. Η μεγάλη διαφορά φαίνεται μετά την f_6 όπου οι αρμονικές του ασταθούς τόνου εμφανίζουν μεγάλη πτώση στο επίπεδο της έντασης.

Στην Εικόνα 82 βλέπουμε τις ίδιες αυξομειώσεις στο φασματικό περιεχόμενο του ασταθούς B^{Mid} όπως και στην Εικόνα 80. Παρατηρούμε ότι οι αρμονικές εξαφανίζονται με το πέρασμα του χρόνου μετά την f_2 και ότι το επίπεδο της έντασης των αρμονικών σταδιακά πέφτει.



Εικόνα 82: Φάσματα του αρχικού τόνου σε τρεις διαφορετικές χρονικές στιγμές BMid

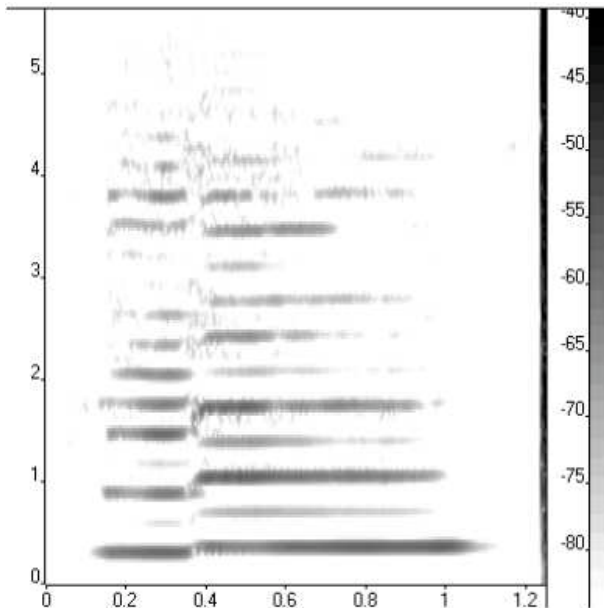
Συμπέρασμα

Η ανάλυση των παραπάνω ασταθών τόνων έδειξε σημαντικές αυξομειώσεις στο φάσμα τους μεταξύ των διαφορετικών χρονικών στιγμών. Επιπλέον, μία ή και περισσότερες αρμονικές χάνονται με αποτέλεσμα τον κακό ήχο των τόνων. Δεν είναι όμως δυνατόν να βρεθούν ποιες αρμονικές χάνονται ή ποιες πέφτουν σε ένταση.

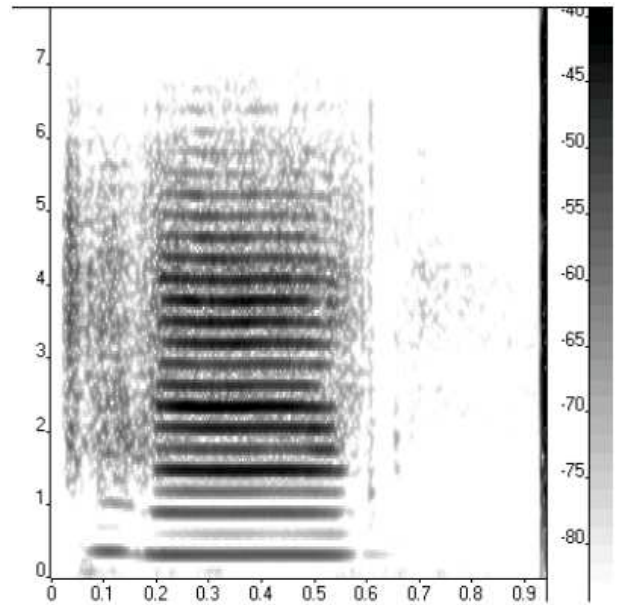
Για τον εντοπισμό των ασταθών τόνων, θα πρέπει να αναλυθεί η ένταση των αρμονικών και οι συχνότητές τους. Αν η έντασή τους παρουσιάζει αυξομειώσεις ή κάποιες αρμονικές χάνονται δεν υπάρχει σταθερότητα στον τόνο.

4.6.4 Διπλοί Τόνοι

Στην ενότητα αυτή θα εξετάσουμε την τέταρτη κατηγορία προβληματικών τόνων, τους διπλούς τόνους. Θα εξεταστούν δύο χαρακτηριστικές νότες που συχνά εμφανίζουν πρόβλημα σε εκτελέσεις αρχαρίων και θα σχολιαστούν διαφορές μεταξύ των καλής ποιότητας αντίστοιχων νοτών και των προβληματικών νοτών, στην προσπάθεια να εξάγουμε κάποια συνολικά συμπεράσματα για αυτή την κατηγορία προβληματικής νότας.

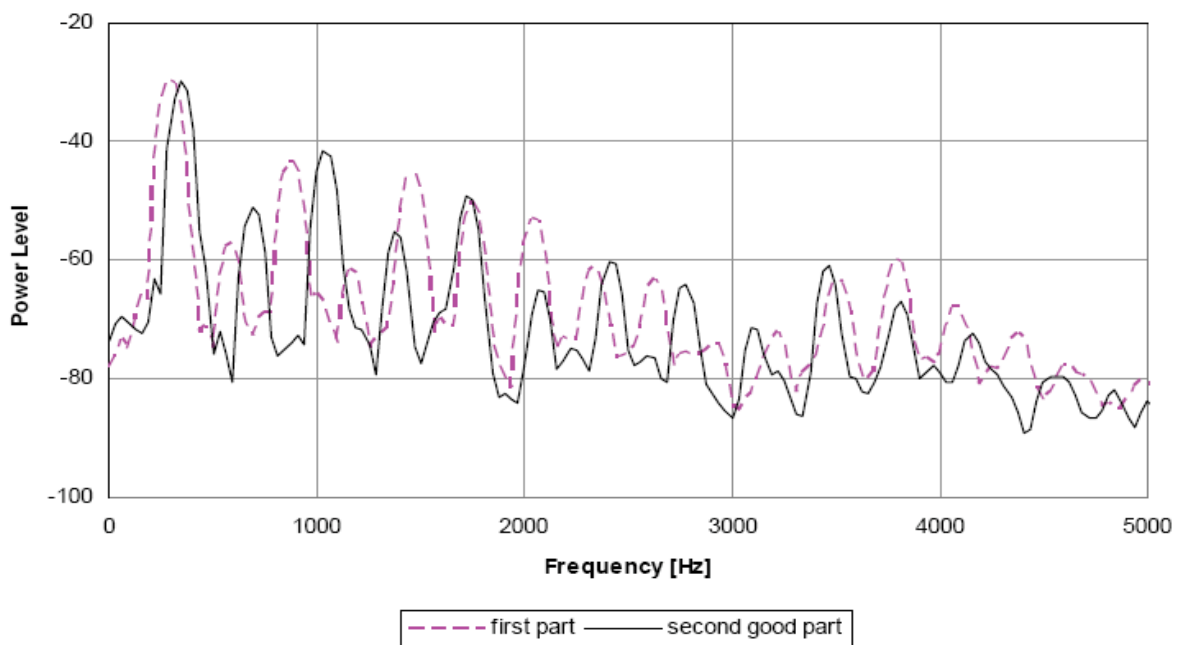


Εικόνα 83 :Διπλός G^{Mid} τόνος

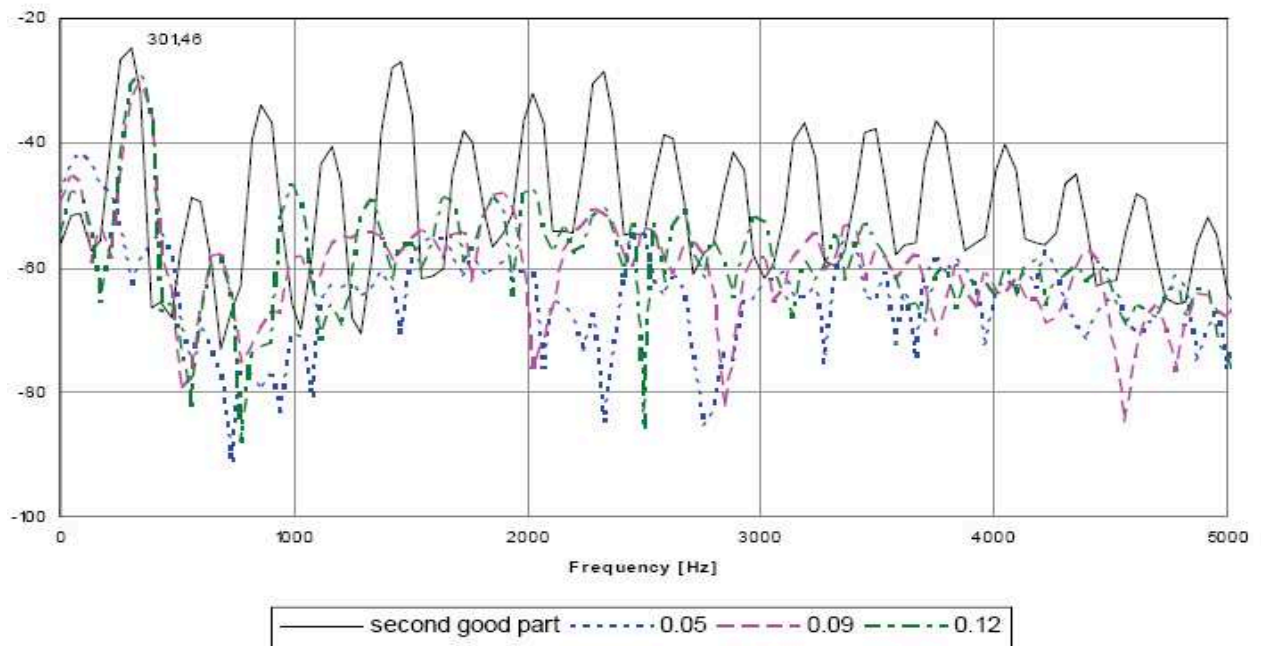


Εικόνα 84 :Διπλός E^{Mid} τόνος

Στις Εικόνες 83 και 86 παρουσιάζονται τα φασματογραφήματα των διπλών G^{Mid} και E^{Mid} τόνων. Χαρακτηριστικό των διπλών τόνων είναι ένας τόνος μικρότερης ή μεγαλύτερης θεμελιώδους συχνότητας και μικρής διάρκειας πριν από τον επιδιωκόμενο τόνο. Στην Εικόνα 83 ο αρχικός τόνος είναι χαμηλότερης θεμελιώδους συχνότητας από τον G^{Mid} , ενώ στην Εικόνα 84 ο αρχικός τόνος δεν είναι καθαρός και στο φασματογράφημα είναι εμφανής μόνο η θεμελιώδης συχνότητα η οποία είναι υψηλότερη από αυτή του προτιθέμενου τόνου. Στις υψηλότερες συχνότητες το φασματογράφημα είναι θολό και ο τόνος μοιάζει περισσότερο με θόρυβο.



Εικόνα 85 :Μέσο γραμμικό φάσμα του προβληματικού διπλού G^{Mid} τόνου



Εικόνα 86 :Γραμμικά φάσματα του αρχικού τόνου σε τρεις διαφορετικές χρονικές στιγμές και μέσο γραμμικό φάσμα του δεύτερου τμήματος, E^{Mid}

Στην Εικόνα 85 βλέπουμε το μέσο γραμμικό φάσμα του προβληματικού διπλού G^{Mid} τόνου. Παρατηρούμε ότι ο αρχικός τόνος είναι μετατοπισμένος προς μικρότερες συχνότητες με τη θεμελιώδη συχνότητα να είναι 281Hz, που είναι η f_0 του $D^{\#Mid}$ τόνου.

Στην Εικόνα 86 παρουσιάζονται τα γραμμικά φάσματα του αρχικού τόνου σε τρεις διαφορετικές χρονικές στιγμές μαζί με το μέσο γραμμικό φάσμα του δεύτερου τμήματος, E^{Mid} . Ο αρχικός τόνος δεν έχει ιδιαίτερες ομοιότητες με το δεύτερο μέρος του τόνου. Η θεμελιώδης συχνότητα στις χρονικές στιγμές 0,09 sec και 0,12 sec συμπίπτει στα 344 Hz, θεμελιώδη συχνότητα κοντά στον τόνο G^{Mid} . Από το φάσμα στα 0,05 sec βλέπουμε ότι μοιάζει περισσότερο με θόρυβο παρά με τόνο.

Συμπέρασμα

Οι διπλοί τόνοι γίνονται αντιληπτοί ως δύο συνεχόμενοι τόνοι με διαφορετικά ύψη, έναν τυχαίο χαμηλής έντασης και τον επιδιωκόμενο μεγαλύτερης έντασης τόνο. Από ανάλυση προέκυψε ότι οι διπλοί τόνοι οφείλονται σε κακό συντονισμό των δακτύλων. Οι διπλοί τόνοι εμφανίζουν δύο διαφορετικά τμήματα στο φασματογράφημά τους, ένα τόνο μικρότερης διάρκειας και διαφορετικής θεμελιώδους συχνότητας από τον επιδιωκόμενο τόνο και τον επιδιωκόμενο τόνο. Ο τόνος που προηγείται του επιδιωκόμενου μπορεί να έχει κάποιες από τις αρμονικές του επιδιωκόμενου τόνου ή να αποτελείται από τη θεμελιώδη συχνότητα και μόνο θόρυβο στις υψηλότερες συχνότητες.

4.7 Συμπεράσματα

Στο κεφάλαιο αυτό μελετήσαμε αρχικά τη λειτουργία και τη μέθοδο της βαθμίδας αξιολόγησης της εκτέλεσης. Στη συνέχεια μελετήσαμε κατηγορίες προβληματικών τόνων, στην προσπάθεια εντοπισμού των γενικών χαρακτηριστικών τους. Τα χαρακτηριστικά αυτά μπορούν να αξιοποιηθούν για τον αυτόματο εντοπισμό της ποιότητας του ήχου, χαρακτηριστικά που μπορεί να δέχεται ως είσοδο η βαθμίδα αξιολόγησης της απόδοσης για την οπτικοποίηση της τονικής ποιότητας.

Ορίζοντας ποσοτικές παραμέτρους για κάθε μία από τις παραπάνω κατηγορίες προβληματικών νοτών, όπως «hollowness», «squeakness», «doubleness» και «unstability» μπορούμε να ελέγξουμε σε έναν ήχο τις τιμές των παραμέτρων σε πραγματικό χρόνο και να τις χρησιμοποιήσουμε σε μια εφαρμογή οπτικοποίησης. Εκτός από τις παραμέτρους τονικής ποιότητας, και άλλες παράμετροι όπως το tempo και η ένταση της νότας μπορούν να ενσωματωθούν.

Οπτικοποιώντας την ποιότητα με ένα φιλικό προς το χρήστη τρόπο, είναι εύκολο για τον μαθητή να αντιληφθεί τα λάθη στην ποιότητα του ήχου. Αντιστοιχίζοντας χαρακτηριστικά του οπτικοποιημένου αντικειμένου (π.χ. χρώμα, βαθμός διαφάνειας, μέγεθος, ρυθμό αναπήδησης, κ.τ.λ.) με τις παραπάνω παραμέτρους παρέχουμε στο μαθητή ένα μέσο να αντιλαμβάνεται με εύκολο τρόπο κάποιες δύσκολες έννοιες.

5

Συμπεράσματα

Στο κεφάλαιο αυτό συνοψίζουμε τις μεθόδους και τα ευρήματα της συγκεκριμένης έρευνας, περιγράφοντας σύντομα τα διαφορετικά κεφάλαια που προηγήθηκαν.

Στο πρώτο κεφάλαιο δόθηκε μια περιγραφή της σύγχρονης κατάστασης της Μουσικής Τεχνολογίας. Αναλύθηκαν οι τομείς σύγχρονης έρευνας και οι εφαρμογές τους, κυρίως στην εκπαίδευση, όπου προορίζονται να χρησιμοποιηθούν κατά κύριο λόγο τα ευρήματα της έρευνας. Στη συνέχεια δόθηκε μια σύντομη θεωρητική περιγραφή βασικών εννοιών των ακουστικών σημάτων που χρησιμοποιούνται στην παρούσα εργασία. Περιγράφηκε ο σκοπός της εργασίας καθώς και η διαδικασία και μέθοδος που ακολουθήθηκε για την εξαγωγή των επιμέρους συμπερασμάτων.

Στο δεύτερο κεφάλαιο ασχοληθήκαμε με την αναγνώριση μονοφωνικής μουσικής. Δόθηκε μια σύντομη περιγραφή υπαρχουσών τεχνικών στον συγκεκριμένο τομέα και στην συνέχεια περιγράφηκαν αναλυτικά οι τεχνικές της εφαρμογής που χρησιμοποιήσαμε στα στάδια της εξαγωγής του ύψους, της ανίχνευσης του onset, τη διάκριση της μουσικής, την αναγνώριση της νότας και την ανίχνευση του τόνου. Ακολούθησε η αξιολόγηση της εφαρμογής, τα προβλήματα τα οποία παρουσίαζε και ερευνήθηκαν οι αιτίες και πιθανές βελτιώσεις.

Στο τρίτο κεφάλαιο μελετήσαμε το πρόβλημα του Score Following/ Matching. Μετά από μια γενική επισκόπηση και περιγραφή της υπάρχουσας έρευνας στο συγκεκριμένο πρόβλημα, αναλύθηκε η Μηχανή Score Matching που χρησιμοποιήσαμε.

Στο τέταρτο κεφάλαιο περιγράφηκε η Μονάδα Αξιολόγησης της εκτέλεσης του μαθητή, η αρχιτεκτονική της μονάδας και ερευνούνται τρόποι προσδιορισμού της τονικής ποιότητας. Μελετάμε ένα σύνολο από τυπικά λάθη και προβληματικούς ήχους σε εκτελέσεις αρχαρίων και ερευνούνται τρόποι αυτόματου προσδιορισμού των προβληματικών τόνων.

Τα παραπάνω εργαλεία μπορούν να χρησιμοποιηθούν τόσο για εκπαιδευτικές εφαρμογές μουσικής, όσο και ως βοηθήματα στη μουσική δημιουργία.

Στη διάρκεια της συγγραφής αντιληφθήκαμε τη σημασία της αξιολόγησης της απόδοσης των εφαρμογών που υλοποιούνται. Η αξιολόγηση είναι ένα απαραίτητο βήμα όχι μόνο για τη σύγκριση διαφορετικών υλοποιήσεων, αλλά και για τη βελτίωση μιας συγκεκριμένης υλοποίησης.

Παράρτημα – Το Ερευνητικό Έργο VEMUS

The poster for VEMUS (Virtual European Music School) features a central blue figure playing a flute, surrounded by musical notes and small photos of people. The text includes the website www.vemus.org, the title 'VEMUS Virtual European Music School', and a subtitle 'Exploration of new teaching practices in music'. It describes the project's aim to develop a multilingual music tuition framework for instruments like the flute, clarinet, saxophone, and trumpet. The poster also lists objectives, validation methods, and contact information for the Institute for Language and Speech Processing in Athens, Greece. Logos of partner institutions and project details are at the bottom.

www.vemus.org

VEMUS

Virtual European Music School

Exploration of new teaching practices in music

VEMUS aims at developing an open, highly interactive, and networked multilingual music tuition framework for popular wind instruments such as the flute, the clarinet, the saxophone and the trumpet.

Objectives

- Integration & pedagogical innovation through computer-aided environments and social features
- VEMUS will develop and validate a unified, innovative, pedagogically-tailored e-learning environment, already tested with traditional face-to-face lessons.

Environments

- self-practising environments
- adaptive performance evaluation
- constructive feedback
- maximizing the effectiveness of practice
- group classroom settings
- collaborative learning
- group sessions
- enriching learning experience
- a dynamic learning platform
- content reusability
- communication and progress monitoring tools
- virtual teaching, practicing and lesson coaching.

Validation

user-centred iterative approach

VEMUS will employ the validation activities to obtain more coherent and statistically meaningful results, and to identify the critical factors for wider acceptance, paying special focus on the new requirements and new potentials in the emerging Europe.

VEMUS adopts a strongly user-centred iterative approach, with user groups actively participating throughout the lifetime of the program. Their requirements, to build user-centred activities and tools.

Contact Information: Institute for Language and Speech Processing - www.ilsp.gr - Artemidos & EpidaurouGR-15123, Marousi, Athens, GREECE

Project number: FP6-0: Virtual European Music School - Ref. No.: 0127952 - Programme: INF-2004-0174
Website: IST-2004-2-4-13: Strengthening the Integration of ICT in e-Learning in the Educational Sector - Starting date: 1/10/2002 - Duration: 36 months - Website: www-ist.eu.org

Εικόνα 87: VEMUS

Το VEMUS ("Virtual European Music School") είναι ένα ερευνητικό και αναπτυξιακό έργο που συγχρηματοδοτείται από την Ευρωπαϊκή Επιτροπή στο πλαίσιο του Προγράμματος των Τεχνολογιών της Κοινωνίας της Πληροφορίας (IST - Information Society Technologies) του Έκτου Προγράμματος-Πλαισίου (FP6 - Sixth Framework Program).

Το VEMUS ξεκίνησε τον Οκτώβριο του 2005 με στόχο την ανάπτυξη και αξιολόγηση πρωτοποριακών τεχνολογικών στοιχείων σχεδιασμένων με βάση παιδαγωγικά κριτήρια. Πιο συγκεκριμένα, στοχεύει στην δημιουργία ενός πλαισίου μουσικής διδασκαλίας για δημοφιλή μουσικά όργανα που θα είναι ανοικτό, διαδικτυακό και πολύγλωσσο και θα προσφέρει υψηλή αλληλεπιδραστικότητα καθώς και ένα σύνολο από καινοτόμα εργαλεία ηλεκτρονικής μάθησης.

Μια από τις βασικές επιδιώξεις του έργου VEMUS είναι να συνεισφέρει στην ενίσχυση και την ολοκλήρωση του Ευρωπαϊκού ερευνητικού πεδίου στον τομέα των τεχνολογιών της Κοινωνίας της

Πληροφορίας, μέσω της ενεργούς συμμετοχής φορέων από τα Νέα Κράτη μέλη. Για τον σκοπό αυτό, στην Κοινοπραξία του Έργου συμμετέχουν 8 οργανισμοί από 6 χώρες, 3 από τις οποίες είναι Νέα ή Συνεργαζόμενα Κράτη.

Το VEMUS θα στηριχθεί στην γνώση και τα αποτελέσματα του επιτυχημένου έργου IMUTUS project που υλοποίησε ένα αποδοτικό περιβάλλον μουσικής εξάσκησης για την φλογέρα. Θα επεκτείνει την τεχνολογική και παιδαγωγική βάση του IMUTUS ώστε να καλύψει επιπρόσθετα δημοφιλή μουσικά όργανα όπως το φλάουτο, το κλαρινέτο, το σαξόφωνο και η φλογέρα, επιπρόσθετα περιβάλλοντα εκμάθησης και ενισχυμένα εκπαιδευτικά σενάρια. Το σύστημα απευθύνεται σε αρχάριους έως μέσου επιπέδου σπουδαστές. Τα προκαταρκτικά αποτελέσματα από την αξιολόγηση του IMUTUS κατέστησαν φανερή την ισχυρή δυναμική της προσέγγισης που υιοθετήθηκε. Το VEMUS θα συνεχίσει και θα κλιμακώσει τις δραστηριότητες αξιολόγησης, επιδιώκοντας αφενός την συλλογή πυκνότερων και στατιστικά πληρέστερων παρατηρήσεων και αφετέρου τον προσδιορισμό των κρίσιμων εκείνων παραγόντων που θα επιτρέψουν μια ευρύτερης κλίμακας αξιοποίηση του συστήματος.

Στο τρέχον πρόγραμμα VEMUS (εικονικό ευρωπαϊκό σχολείο μουσικής) αναπτύσσεται ένα περιβάλλον εικονικής άσκησης για πνευστά όργανα όπως φλογέρα, φλάουτο, κλαρινέτο, σαξόφωνο και τρομπέτα. Η ομάδα-στόχος χρηστών είναι σπουδαστές πνευστών οργάνων κατά τη διάρκεια των πρώτων 4 -5 ετών της μουσικής εκπαίδευσής τους. Το VEMUS θα μπορούσε να θεωρηθεί ως εικονικός βοηθητικός δάσκαλος που καθοδηγεί το σπουδαστή ενώ εξασκείται.

Το VEMUS καταγράφει το ακουστικό σήμα ενώ ο σπουδαστής παίζει μια μελωδία και αναλύει και αξιολογεί την απόδοση συγκρίνοντάς τη με την παρτιτούρα αναφοράς. Στη συνέχεια, δίνεται ανατροφοδότηση στο σπουδαστή για τα λάθη και προτείνονται τρόποι για να τα διορθώσει. Αυτή η ανατροφοδότηση περιορίζεται σε λίγα στοιχεία βασισμένοι σε παιδαγωγικές εκτιμήσεις.

Το περιβάλλον του VEMUS θα ενσωματώσει καινοτομικά, παιδαγωγικά στοιχεία ηλεκτρονικής μάθησης ώστε να επεκτείνει τη παραδοσιακή διδασκαλία της μουσικής σε διαφορετικά μαθησιακά περιβάλλοντα:

Αυτό-εξάσκηση: Ένα σύνολο εφαρμογών θα σχεδιαστεί, το οποίο θα ενισχύσει την αυτό-εξάσκηση, στο σπίτι ή στο εργαστήριο, ώστε να γίνονται οι συνεδρίες εξάσκησης περισσότερο επιμορφωτικές, δημιουργικές και ψυχαγωγικές. Η αυτόματη αξιολόγηση της εκτέλεσης και η υψηλού επιπέδου ανάδραση, προσαρμοσμένη σε κάθε σπουδαστή αναμένεται να μεγιστοποιούν την αποτελεσματικότητα της εξάσκησης στο σπίτι.

Αίθουσα διδασκαλίας Μουσικής: Το VEMUS θα ερευνήσει και θα αξιολογήσει καινοτομικά εργαλεία ώστε να υποστηρίξει τη διδασκαλία της μουσικής σε περιβάλλον ομάδας, με διδακτικά βοηθήματα που θα υποστηρίζουν τον δάσκαλο

την ομαδική μάθηση και τις ομαδικές δραστηριότητες. Αυτά τα εργαλεία θα αποτελέσουν τη βάση της εμφάνισης του VEMUS στην αίθουσα διδασκαλίας

ενισχύοντας τις διαδικασίες διδασκαλίας της μουσικής και την εμπειρία εκμάθησης του μαθητή.

Επεκτάσεις μάθησης εξ αποστάσεως : Το περιβάλλον του VEMUS θα παρέχει επεκτάσεις μάθησης εξ αποστάσεως, συντηρώντας και διαχειρίζοντας μια πηγή ανοιχτού περιεχομένου

προσφέροντας εργαλεία συγγραφής προς αύξησή του, παρέχοντας εργαλεία επικοινωνίας και επιτρέποντας διδασκαλία εξ αποστάσεως μαθητών και έλεγχο της προόδου τους με το πέρασμα του χρόνου. Αυτά τα γνωρίσματα θα βοηθήσουν στην περαιτέρω ενίσχυση της σχέσης δασκάλου-μαθητή

ενώ θα καθιστούν δυνατή τη συμμετοχή μαθητών που διαφορετικά θα ήταν αδύνατη εξαιτίας γεωγραφικών ή άλλων περιορισμών.

Απευθυνόμενο σ' αυτά τα τρία διαφορετικά αλλά συμπληρωματικά περιβάλλοντα εκμάθησης διαμέσου ενός περιβάλλοντος και διαλειτουργικών εργαλείων, το VEMUS θα συνδυάσει ομαλά την ηλεκτρονική μάθηση με τα παραδοσιακά μαθήματα. Αυτά θα παρέχουν την τεχνολογική και παιδαγωγική βάση για μια εικονική συνάντηση και χώρο εξάσκησης για σπουδαστές και δασκάλους σε όλη την Ευρώπη: ένα εικονικό ευρωπαϊκό μουσικό σχολείο. Η παρουσίαση καινοτομικών τεχνολογικών εργαλείων για διδασκαλία μουσικής θα επιτρέψει την έρευνα νέων μεθόδων διδασκαλίας για μουσική που θα εμπλουτίσουν τις παραδοσιακές διαδικασίες.

Το VEMUS θα στηριχθεί στα αποτελέσματα, τα εργαλεία και την πλατφόρμα που αναπτύχθηκαν και ελέγχθηκαν στο πλαίσιο του επιτυχημένου έργου **IMUTUS** που πρόσφατα ολοκληρώθηκε. Το IMUTUS (IST-2001-32270) αποτελεί ένα αποδοτικό περιβάλλον αυτό-εξάσκησης με αυτοματοποιημένα υψηλού επιπέδου χαρακτηριστικά αξιολόγησης της εκτέλεσης και μια υποδομή μάθησης εξ αποστάσεως για κοινόχρηστο αντικείμενο μάθησης και για επικοινωνία. Η αξιολόγηση και ο έλεγχος του IMUTUS έχουν αποδείξει τις μεγάλες δυνατότητες της καινοτομικής προσέγγισης για εξάσκηση με μουσικά όργανα.

Το VEMUS θα επεκτείνει αυτή την προσέγγιση σε νέα περιβάλλοντα εκμάθησης καλύπτοντας ανάγκες επιπλέον μαθησιακών τοποθετήσεων και ένα μεγαλύτερο κοινό. Το VEMUS θα αναπτυχθεί σε περιβάλλον μάθησης εξ αποστάσεως και θα εμπλουτίσει τα αντικείμενα μάθησης με επιπρόσθετα ουσιαστικά δεδομένα που θα οδηγήσουν στο περιεχόμενο χαρακτηρισμού βασισμένο σε βασικές ικανότητες ελέγχου μουσικής και οργάνων. Υψηλά καινοτομικά χαρακτηριστικά της αίθουσας διδασκαλίας όπως κοινή χρήση πηγών και αντικειμένων μάθησης θα ερευνηθούν ώστε να εξασφαλίσουν μια διαρκή επαφή μεταξύ της εξάσκησης στην αίθουσα διδασκαλίας και της εξάσκησης στο σπίτι με υλικό που συγκεντρώθηκε στην αίθουσα.

Η κοινοπραξία του VEMUS (VEMUS Consortium) αποτελείται από 6 χώρες, 3 από τις οποίες είναι νέα ή υπονήφια κράτη-μέλη. Το VEMUS συγκέντρωσε όλη την απαραίτητη πραγματογνωμοσύνη από όλους τους τομείς που ενέχονται και υιοθετεί καθαρά μια ευρωπαϊκή διάσταση με τη συμμετοχή εταίρων από 6 χώρες. Έχοντας τη μουσική ως κύριο όχημα και εμπλέκοντας ομάδες σε διαφορετικές χώρες και διάφορες γλώσσες, το VEMUS αναζητά να πλησιάσει ανθρώπους με διαφορετική ιστορία και κουλτούρα, δημιουργώντας πολιτιστικούς δεσμούς και διασφαλίζοντας μια μεγαλύτερη συμμετοχή στη γνώση και στα αποτελέσματα της έρευνας καθώς και ένα μεγαλύτερο αντίκτυπο.

Το VEMUS υιοθετεί μια προσέγγιση ισχυρά προσανατολισμένη στο χρήστη, με ομάδες χρηστών (user groups) που συμμετέχουν ενεργά σε όλη τη διάρκεια του έργου: από προϋποθέσεις, σε δοκιμές σε διάφορα πεδία, αξιολόγηση και άλλα. Η συμμετοχή ομάδων χρηστών στην κοινοπραξία επιβεβαιώνει μία ισχυρά προσανατολισμένη στο χρήστη διαδικασία εφαρμογής και δίνει την ευκαιρία στο έργο να αναπτύξει την προσέγγισή του σε ακόμη μεγαλύτερη κλίμακα και να καταπιαστεί με σημαντικούς παράγοντες που θα αποτελέσουν τη βάση για ευρύτερες μεταγενέστερες πρωτοβουλίες.

Με προγραμματισμένες σχετικά με την αγορά δραστηριότητες και τη συμμετοχή επιχειρήσεων, μια από τις οποίες είναι ένας σημαντικός οίκος μουσικής στην Ελλάδα με σημαντική και παρατεταμένη εμπειρία στον τομέα, το VEMUS θα είναι σε θέση επίσης να ενσωματώσει μια απευθυνόμενη στην αγορά διάσταση, ερευνώντας, προγραμματίζοντας και ακολουθώντας τις ευκαιρίες για την αξιοποίηση των αποτελεσμάτων του προγράμματος.

Αυτήν την περίοδο, δεν υπάρχει κανένα εμπορικό ή άλλο σύστημα που ενσωματώνει τα χαρακτηριστικά γνωρίσματα που προγραμματίζονται για το VEMUS. Τα υπάρχοντα συστήματα

καλύπτουν μεμονωμένα μέρη και τις περισσότερες φορές, χωρίς ικανοποιητικό αποτέλεσμα. Εάν το IMUTUS είναι μια καινοτομία στην αυτό-εξάσκηση, το VEMUS μπορεί να επιφέρει τεχνολογική επανάσταση στην εκπαίδευση της μουσικής παρουσιάζοντας μια πλήρη προσέγγιση που καλύπτει άτομα, ομάδες και τάξεις μουσικής, καθώς επίσης και οργανώσεις όπως σχολεία μουσικής, συνδυάζοντας την εξ αποστάσεως μάθηση και τις καινοτόμες τεχνολογικές βελτιώσεις με τα παραδοσιακά πρόσωπο με πρόσωπο μαθήματα σε μια πλήρες περιβάλλον εκμάθησης.

Μια από τις ομάδες χρηστών που συμμετέχουν ενεργά στο πρόγραμμα είναι οι μαθητές που με τις εκτελέσεις τους σε διάφορα μουσικά κομμάτια εμπλουτίζουν τη βάση δοκιμών του VEMUS, παρέχοντας μ' αυτόν τον τρόπο ένα ισχυρό εργαλείο για τη μελέτη των χαρακτηριστικών των ηχογραφημένων εκτελέσεων. Χαρακτηριστικά όπως ο περιβαλλοντικός θόρυβος σε πραγματικές αίθουσες διδασκαλίας μουσικής, οι επιδόσεις των σπουδαστών σε διαφορετικά επίπεδα εκμάθησης και διαφορετικής δυσκολίας μουσικά κομμάτια, τα διάφορα λάθη που κάνουν και ο τρόπος αντίδρασης τους σ' αυτές τις περιπτώσεις είναι κάποια από το σύνολο των γνωρισμάτων που έχουμε τη δυνατότητα να μελετήσουμε λεπτομερώς χάρη σ' ένα πλήθος πραγματικών εκτελέσεων που έχουμε στη διάθεσή μας. Οι εκτελέσεις αυτές πραγματοποιήθηκαν από διαφορετικούς μαθητές, διαφόρων επιπέδων, σε διαφορετικά μουσικά κομμάτια, με διαφορετικά μουσικά όργανα και με ένα σύνολο δασκάλων ώστε με το πλήθος αυτό των διαφορετικών συνθηκών να παρέχει ένα ικανοποιητικά εκτεταμένο σύνολο περιπτώσεων δοκιμής.

Παράρτημα – Πίνακας Μετάφρασης Επιλεγμένων Όρων

Score following/matching	Αντιστοίχιση των μουσικών συμβάντων (νοτών, παύσεων κ.λπ) της παρτιτούρας με τα μουσικά συμβάντα μιας εκτέλεσης όπως αυτά αναγνωρίζονται σε πραγματικό (following) ή μη πραγματικό (matching) χρόνο.
Centroid	Φασματικό κέντρο βάρους
Peak picking	Αλγόριθμος επιλογής μεγίστων
Partials	Μερικές
Harmonics	Αρμονικές
Pitch	Ύψος μουσικού ήχου. Εναλλακτικά, θεμελιώδης συχνότητα (f0).
Frame	Πλαίσιο
Onset (note ~)	Χρονικό σημείο εκκίνησης (νότας)
Offset (note ~)	Χρονικό σημείο τερματισμού (νότας)
Attack	Στην μουσική αναφέρεται και ως «ατάκα». Το εισαγωγικό τμήμα μιας νότας, η «έκρηξη» καθώς αυτή σχηματίζεται, το «μπάσιμό» της.
Transient	Μεταβατική συμπεριφορά νότας.
Hollow/empty (note)	Κενή/ άδεια (νότα).
Squeak	"Τρίξιμο".
Decay	Εξασθένιση νότας.

Ευρετήριο Εικόνων και Πινάκων

Εικόνα 1: Απεικόνιση διεπιστημονικότητας της Μουσικής Τεχνολογίας.....	9
Εικόνα 2 : Αραιώσεις και πυκνώσεις κατά τη διάδοση ενός κύματος	16
Εικόνα 3:Ο ήχος ως κύμα πίεσης	17
Εικόνα 4: Μήκος κύματος	17
Εικόνα 5 : Συχνότητες νοτών στο πιάνο	17
Εικόνα 6:Κυματομορφή στο πεδίο του χρόνου και στο πεδίο της συχνότητας.....	19
Εικόνα 7: Σύγκριση αρμονικών και οκτάβων. Οι αρμονικές είναι γραμμικά συσχετισμένες ενώ οι οκτάβες λογαριθμικά.....	20
Εικόνα 8: Μήκος κύματος και πλάτος κυματομορφής.	21
Εικόνα 9 : Καμπύλες ίσης ηχηρότητας.....	22
Εικόνα 10	23
Εικόνα 11 :Δεσμοί και Αντί-δεσμοί	25
Εικόνα 12: Αναπαράσταση της Μουσικής Καταγραφής.....	27
Εικόνα 13 :Στάδια επεξεργασίας που απαιτούνται για την μουσική αναγνώριση και οι μεταξύ τους αλληλεξαρτήσεις.....	28
Εικόνα 14:Λανθασμένα ανιχνευμένα μέρη σιωπής με τη χρησιμοποίηση των απλών τιμών φακέλων...	32
Εικόνα 15 :Απεικόνιση της παραμέτρου της έντασης και της αποδοτικότητάς του σε σήματα με θόρυβο	33
Εικόνα 16: Η ακολουθία αυτοσυσχέτισης ενός σήματος με φωνή και ενός χωρίς.	33
Εικόνα 17:Απεικόνιση της περιοδικότητας για την έκφραση της ανίχνευση.....	34
Εικόνα 18: Καθορισμός του κατώτατου ορίου έκφρασης από το ιστόγραμμα του επιπέδου περιοδικότητας.....	35
Εικόνα 19 : Κυματομορφή χωρίς υψηλότερες αρμονικές.....	36
Εικόνα 20: Κυματομορφή με υψηλότερες αρμονικές μικρότερης ισχύος.....	36
Εικόνα 21: Κυματομορφή με υψηλότερες αρμονικές μεγαλύτερης ισχύος.....	36
Εικόνα 22 :Κυματομορφή με υψηλότερες αρμονικές μικρότερης ισχύος και χωρίς θεμελιώδη συχνότητα	36
Εικόνα 23: FFT-based δομή υπολογισμού ύψους με εκτίμηση φάσης	38
Εικόνα 24:Αρμονικές και εκτίμηση της F_0	40
Εικόνα 25: Διάγραμμα ροής για το διαχωρισμό των μουσικών σημάτων σε γεγονότα νοτών.....	42
Εικόνα 26: Onset, attack, transient και decay μιας νότας.....	43
Εικόνα 27: Παραδείγματα ήχων όπου παρουσιάζονται τα onset	44
Εικόνα 28:Φάκελοι εύρους για τους τόνους που παράγονται από (από κάτω προς τα πάνω) βιολί, τρομπέτα και φλάουτο. Η αρχή του φλάουτου είναι σχεδόν στιγμιαία. (Πηγή [Mar98]).....	45
Εικόνα 29: Λανθασμένη νότα στο 1 ^ο στάδιο και σωστή στο 2 ^ο	46
Εικόνα 30: Γραφική παράσταση μηχανής απόφασης 3 επιπέδων	47
Εικόνα 31: Αυτόματη ανίχνευση τόνου της φλογέρας μέσω του PDF των τιμών ύψους	49
Εικόνα 32 : Περιβάλλον του προγράμματος Praat	52
Εικόνα 33: Αντιστοιχία νοτών και συχνοτήτων	53
Εικόνα 34 :Παράδειγμα διάσπασης νότας που οφείλεται σε μικρή διακύμανση του ύψους. Test-case «Martin_The Christmas song»_3.33 sec-3.699 sec.	55
Εικόνα 35 :Παράδειγμα διάσπασης νότας που οφείλεται σε μικρή διακύμανση του ύψους. Test-case «Martin_The Christmas song»_39.68 sec-39.129 sec.	56
Εικόνα 36 : Παράδειγμα διάσπασης νότας που οφείλεται σε διακύμανση της έντασης. Test-case «Martin_The Christmas song»_71.44 sec-71.53 sec.	57
Εικόνα 37: Παράδειγμα συγχώνευσης νοτών με διαφορά ενός ημιτονίου. Test-case «Sankta Lucia»_2.50 sec-3.23 sec.	58

Εικόνα 38 :Κομμάτι παρτιτούρας «Sankta Lucia»	58
Εικόνα 39 : Παράδειγμα συγχώνευσης νοτών με διαφορά ενός ημιτονίου. Test-case «Fredrik Jag vill ha en egen» 47.43 sec-48.03 sec.....	59
Εικόνα 40: Κομμάτι παρτιτούρας «Jag vill ha en egen»	59
Εικόνα 41: Παράδειγμα συγχώνευσης νοτών με διαφορά ενός τόνου Test-case «Johan Moritz julen er her» 47.43 sec-48.03 sec.....	60
Εικόνα 42: Κομμάτι παρτιτούρας julen er her.....	60
Εικόνα 43 : Διάγραμμα εισόδων-εξόδων της σύγχρονης βαθμίδας αντιστοίχισης.	65
Εικόνα 44 : Διάγραμμα εισόδων-εξόδων του της ασύγχρονης βαθμίδας αντιστοίχισης.	65
Εικόνα 45 :Συμπεριφορά της σύγχρονης βαθμίδας αντιστοίχισης στην περίπτωση παράλειψης μιας νότας.	72
Εικόνα 46: Συμπεριφορά της σύγχρονης βαθμίδας αντιστοίχισης στην περίπτωση εισαγωγής μιας νότας.....	73
Εικόνα 47: Πεδία “Reference & Performance”	76
Εικόνα 48: Πεδίο “Mathing”	76
Εικόνα 49: Διάλογος επιλογής αρχείου.....	77
Εικόνα 50: Παράθυρο μετά την επιλογή αυτών των αρχείων	77
Εικόνα 51: Υποθέσεις και γράφος καλύτερης αντιστοίχισης	78
Εικόνα 52: Αντιστοίχιση ακολουθιών.....	78
Εικόνα 53	79
Εικόνα 54	79
Εικόνα 55: Αποτελέσματα αντιστοίχισης σε περίπτωση δοκιμής.....	79
Εικόνα 56: Διάγραμμα εισόδων-εξόδων της Βαθμίδας Αξιολόγησης της Εκτέλεσης.	83
Εικόνα 57 : Διάγραμμα του σχεδιασμού της Βαθμίδας Αξιολόγησης της Εκτέλεσης.	85
Εικόνα 58: FLow mf φασματογράφημα.....	88
Εικόνα 59 : GMid mf φασματογράφημα	88
Εικόνα 60 : AUr mf φασματογράφημα	88
Εικόνα 61 : BMid mf φασματογράφημα.....	88
Εικόνα 62 :Τρίξιμο οφειλόμενο σε κακό συντονισμό των δακτύλων	89
Εικόνα 63: Τρίξιμο οφειλόμενο στο στόμιο	89
Εικόνα 64: Γραμμικό φάσμα του τόνου FLow mf καλής ποιότητας ήχου και του προβληματικού squeak τμήματος εξαιτίας κακού δακτυλισμού και κακού στομίου αντίστοιχα.	90
Εικόνα 65 G^{Mid} τρίξιμο	90
Εικόνα 66 : Φάσμα του προβληματικού από τρίξιμο G^{Mid} τόνου στην αρχή και στο τέλος του συγκριτικά με το μέσο γραμμικό φάσμα του κανονικού τόνου.....	91
Εικόνα 67: Άδεια F^{Low} από κακή κατάσταση στομίου	92
Εικόνα 68 : Άδεια F^{Low} από ελλιπή πίεση των χειλιών.....	92
Εικόνα 69: Άδεια F^{Low} από κακή κατάσταση του στομίου.....	92
Εικόνα 70: Μέσο γραμμικό φάσμα του F^{Low} τόνου στις 3 δυνατές περιπτώσεις άδειας νότας και του κανονικού F^{Low}	93
Εικόνα 71: Άδεια A^{Up} από κακή κατάσταση στομίου	94
Εικόνα 72: Άδεια A^{Up} από ελλιπή πίεση των χειλιών	94
Εικόνα 73 : Άδεια A^{Up} από κακή κατάσταση του στομίου.....	94
Εικόνα 74: Μέσο γραμμικό φάσμα του A^{Up} τόνου στις 3 δυνατές περιπτώσεις άδειας νότας και του κανονικού A^{Up}	95
Εικόνα 75: Ασταθές Flow.....	96
Εικόνα 76 :Ασταθές GMid.....	96
Εικόνα 77 :Ασταθές BMid	96
Εικόνα 78 :Ασταθές AUr.....	96

Εικόνα 79 : Μέσο φάσμα του ασταθούς και του κανονικού F^{Low} καθώς και το φάσμα του ίδιου τόνου σε τρεις διαφορετικές χρονικές στιγμές.....	97
Εικόνα 80 : Μέσο φάσμα του ασταθούς F^{Low}	97
Εικόνα 81: Μέσο φάσμα του κανονικού και του ασταθούς B^{Mid}	98
Εικόνα 82: Φάσματα του αρχικού τόνου σε τρεις διαφορετικές χρονικές στιγμές B^{Mid}	99
Εικόνα 83 : Διπλός G^{Mid} τόνος.....	100
Εικόνα 84 : Διπλός E^{Mid} τόνος	100
Εικόνα 85 : Μέσο γραμμικό φάσμα του προβληματικού διπλού G^{Mid} τόνου.....	100
Εικόνα 86 : Γραμμικά φάσματα του αρχικού τόνου σε τρεις διαφορετικές χρονικές στιγμές και μέσο γραμμικό φάσμα του δεύτερου τμήματος, E^{Mid}	101
Εικόνα 87: VEMUS.....	104

Πίνακας 1: Ερμηνεύοντας τους αριθμούς.....	22
Πίνακας 2 : Αρμονικές, δεσμοί και αντί-δεσμοί.....	25
Πίνακας 3 : Απαντήσεις 34 δασκάλων μουσικής στην ερώτηση: «Ποιες από τις παρακάτω πτυχές χρησιμοποιούνται για την αξιολόγηση ενός μαθητή στα δύο πρώτα χρόνια διδασκαλίας;» (κλίμακα 1= ΟΧΙ ΣΗΜΑΝΤΙΚΟ, 5= ΠΟΛΥ ΣΗΜΑΝΤΙΚΟ)	83
Πίνακας 4 Τυπικά βασικά λάθη σε μαθητές κατά τη διάρκεια των δύο πρώτων χρόνων διδασκαλίας... 84	84
Πίνακας 5 Βασικό σύνολο τυπικών λαθών της Βαθμίδας Αξιολόγησης.....	84
Πίνακας 6 Νότες και η θεμελιώδης συχνότητά τους.....	87

Βιβλιογραφία

- [Alp80] B. Alphonse. Music analysis by computer. *Computer Music Journal*, 4(2):26–35, 1980.
- [AP05] Jean-Julien Aucouturier and Francois Pachet. Ringomatic: A real-time interactive drummer using constraint-satisfaction and drum sound descriptors. In Proceedings of the International Symposium on Music Information Retrieval (ISMIR), pages 412–419, London, UK, September 2005.
- [APH04] Jean-Julien Aucouturier, Francois Pachet, and Peter Hanappe. From sound sampling to song sampling. In Proceedings of the International Symposium on Music Information Retrieval (ISMIR), pages 1–9, Barcelona, Spain, October 2004.
- [Bak89a] M. Baker. An artificial intelligence approach to musical grouping analysis. *Contemporary Music Review*, 3:43–68, 1989a.
- [Bak89b] M. Baker. A computational approach to modeling musical grouping structure. *Contemporary Music Review*, 3:311–325, 1989b.
- [Bak92] M. Baker. Design of an intelligent tutoring system for musical structure and interpretation. In K. E. M. Balaban and O. Laske, editors, *Understanding Music with AI: Perspectives on Music Cognition*, pages 467–489. The MIT Press, Cambridge, MA, 1992.
- [Bak96] Bakamidis Stelios G, (1996): “A New Frequency Domain Method for Amplitude and Frequency Demodulation”
- [Bal80] G. J. Balzano. The group-theoretic description of 12- fold and microtonal pitch systems. *Computer Music Journal*, 4(4):66–84, 1980
- [Bam74] J. Bamberger. Progress report: Logo music project. Technical report, A.I. Laboratory, Massachussets Institute of Technology, 1974.
- [Bam91] J. Bamberger. *The Mind Behind the Musical Ear*. Harvard University Press, Cambridge, Massachussets, 1991.
- [BBPa] Paul M. Brossier, Juan P. Bello, Mark D. Plumbley, fast labeling of notes in music signals Queen Mary College, University of London Centre for Digital Music
- [BBPb] Paul Brossier, Juan Pablo Bello and Mark D. Plumbley, Centre for Digital Music. Real time temporal segmentation of note objects in music signals, Queen Mary University of London
- [BDA+] Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler, A tutorial on onset detection in Music Signals, IEEE
- [Bel03] Juan-Pablo Bello. Towards the Automated Analysis of Simple Polyphonic Music. PhD thesis, Centre for Digital Music, Queen Mary University of London, London, UK, 2003.
- [Bent87] I. Bent. *The New Grove Handbooks in Music: ANALYSIS*. The Macmillan Press Ltd, London, 1987. With a glossary by W. Drebkin.
- [BHJ93] Bagshaw, P.C., S.M. Hiller and M.A. Jack (1993). Enhanced pitch tracking and the processing of f0 contours for computer aided intonation teaching, EuroSpeech’93, pp. 1003-1006. (<http://www.cstr.ed.ac.uk/~pcb/publications.html>).

- [Blo81] A. Blombach. An introductory course in computer-assisted music analysis: The computer and Bach chorales. *Journal of Computer-based Instruction*, 7(3), 1981.
- [Boe93] [Boersma, 1993] Boersma, P. (1993). Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound, IFA Proceedings 17, pp. 97-110, Institute of Phonetic Sciences, University of Amsterdam.
- [Bro06] P. Brossier, Automatic Annotation of Musical Audio for Interactive Applications, Centre for Digital Music, PhD thesis, Queen Mary, University of London, August 2006
- [BWP] Marcio Brandao, Geraint Wiggins and Helen Pain, Computers in music education, Division of Informatics, University of Edinburgh
- [Cai04] Carlos Caires. Irin: Micromontage in graphical sound editing and mixing tool. In Proceedings of the International Computer Music Conference (ICMC), pages 219–222, Miami, Florida, USA, November 2004.
- [Cas05] Michael A. Casey. Acoustic lexemes for organizing internet audio. *Contemporary Music Review*, 24(6):489–508, December 2005.
- [Cas94] Michael A. Casey. Understanding musical sounds with forward models and physical models. *Connection Science*, 6(2):355–371, 1994
- [CK01] Cheveigne Alain de H. Kawahara (2001): “Comparative Evaluation of F0 estimation algorithms”, Eurospeech 2001, Scandinavia
- [CK02] [Cheveigné and Kawahara, 2002] De Cheveigné, A. and H. Kawahara (2002). Yin, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.*, 111, pp. 1917-1930
- [Col04] Nick Collins. On onsets on-the-fly: Real-time event segmentation and categorization as a compositional effect. In Proceedings of Sound and Music Computing Conference (SMC 04), IRCAM, Paris, France, October 20–22 2004
- [Con04] Arshia Cont, Improvement of Observation Modeling for Score Following, Memoire de stage de DEA ATIAM annee 2003-2004, Universite Pierre et Marie Curie, PARIS VI, 2004
- [Coo93] J. Cook. Agent reflection in an intelligent learning environment architecture for musical composition. In M. Smith, A. Smaill, and G. Wiggins, editors, *Music Education: An Artificial Intelligence Approach*, Workshops in Computing, pages 3–23. Springer-Verlag, 1994. Proceedings of a workshop held as part of AIED 93, Edinburgh, Scotland, 25 August 1993.
- [Dan84] Dannenberg R.B, (1984): “An On-Line Algorithm for Real-Time Accompaniment”, International Music Conference, 1985
- [Dan85] Roger B. Dannenberg: An on-line algorithm for real-time accompaniment. In Proceedings of the 1984 International Computer Music Conference (ICMC), pages 193–198, IRCAM, Paris, France, June 1985.
- [Dan90] [Dannenberg et al., 1990] Dannenberg, Sanchez, Joseph, Capell, Joseph, Saul, “A Computer-Based Multi-Media Tutor for Beginning Piano Students,” *Interface – Journal of New Music Research*, 19(2-3), 1990, pp. 155-173.
- [Del00] de la Cuarda Patricio A. Master, Craig Sapp, (2000): “Efficient Pitch Detection Techniques for InteractiveMusic”
- [DG00] Durand N. E. Gomez (2000): “Periodicity Analysis using a Harmonic Matching Method and Bandwise Processing”

- [DGW04] Simon Dixon, Fabien Gouyon, and Gerhard Widmer. Towards characterisation of music via rhythmic patterns. In Proceedings of the International Symposium on Music Information Retrieval (ISMIR), pages 509–516, Barcelona, Spain, October 2004.
- [Dix97] Dixon, S. (1997). Beat induction and rhythm recognition, Proceedings of the Australian Joint Conference on Artificial Intelligence, pp. 311-320, (http://www.cs.flinders.edu.au/People/Simon_Dixon/pub/)
- [DM87] Roger B. Dannenberg and Bernard Mont-Reynaud: Following an improvisation in real time. In Proceedings of the International Computer Music Conference (ICMC), pages 241–248, University of Illinois at Champagne/Urbana, Illinois, USA, 1987.
- [DM88] Dannenberg R.B., H. Mikaino, (1988): “New Techniques for Enhanced Quality of Computer Accompaniment”, International Computer Music Conference, 1988
- [DML+03] De Mulder, T., J.P. Martens, M. Lesaffre, M. Leman, B. De Baets, H. De Meyer (2003). An Auditory Model Based Transcriber of Vocal Queries in: *Proceedings of the International Conference on Music Information Retrieval*, Baltimore, Maryland, USA and Library of Congress, Washington, DC, USA, October 26-30, 2003.
- [DP04] Matthew E. P. Davies and Mark D. Plumbley. Causal tempo tracking of audio. In Proceedings of the International Symposium on Music Information Retrieval (ISMIR), pages 164–169, Barcelona, Spain, October 2004.
- [DSJ+90] Dannenberg, Sanchez, Joseph, Capell, Joseph, Saul, “A Computer-Based Multi-Media Tutor for Beginning Piano Students,” *Interface – Journal of New Music Research*, 19(2-3), 1990, pp. 155-173.
- [Eve01] A. Everest, the Master Handbook of Acoustics, McGraw Hill, 4th Ed., 2001
- [FVC] Sam Ferguson, Andrew Vande Moere and Densil Cabrera, Seeing Sound: Real-time Sound Visualisation in Visual Feedback Loops used for Training Musicians, School of Architecture, Design Science and Planning, The University of Sydney
- [GD04] Fabien Gouyon and Simon Dixon. Dance music classification: a tempo based approach. In Proceedings of the International Symposium on Music Information Retrieval (ISMIR), pages 501–504, Barcelona, Spain, October 2004
- [GD97] Grubb & Dannenberg 97] Lorin Grubb and Roger B. Dannenberg. A Stochastic Method of Tracking a Vocal Performer. Proceedings of the ICMC, pages 301--308, 1997.
- [Ger03] David Gerhard, Pitch extraction and fundamental Frequency: history and current techniques, Technical Report TR-CS 2003-06, November, 2003
- [Got00] Goto Masataka, (2000): “A Robust Predominant F0 Estimation Method for Real-Time Detection of Melody and Bass Lines in CD Recordings”, ICASSP 2000
- [GPA03b] Emilia Gomez, Georges Peterschmitt, Xavier Amatriain, and Perfecto Herrera. Content-based melodic transformations of audio for a music processing application. In Proceedings of the International Conference on Digital Audio Effects (DAFx-03), pages 333–338, London, UK, 2003b
- [GR69] Gold B., Rabiner L. (1969). “Parallel processing techniques for estimating pitch periods of speech in the time domain”. *Journal of the Acoustical Society of America*, Volume 46, Number 2, pages 442-448.

- [Gro75] D. Gross. *A Set of Computer Programs to Aid in Music Analysis*. Ph.D. thesis, Indiana University, 1975.
- [HDH+00] [Heijink et al.] Heijink, H., Desain, P., Honing, H., & Windsor, L. “Make me a match: An evaluation of different approaches to score-performance matching.” *Computer Music Journal*, 24(1), (2000), 43–56 Score Following
- [Her88] Hermes, D. (1988). Measurement of pitch by subharmonic summation, *JASA* 83(1), pp. 257-264.
- [Hof75] F. Hoffstetter. GUIDO: an interactive computer-based system for improvement of instruction and research in ear-training. *Journal of Computer-Based Instruction*, 1(4):100–106, 1975.
- [Hof81] F. Hoffstetter. Computer-based aural training: The GUIDO system. *Journal of Computer-Based Instruction*, 7(3):84–92, 1981.
- [Hof88] F. T. Hoffstetter. *Computer Literacy for Musicians*. Prentice Hall, Englewood Cliffs, NJ, 1988.
- [Hol89] S. Holland. *Artificial Intelligence, Education and Music. The Use of Artificial Intelligence to Encourage and Facilitate Music Composition by Novices*. Ph.D. thesis, IET - Open University, UK, 1989.
- [Hol93] M. Smith and S. Holland. MOTIVE - the development of an AI tool for beginning melody composers. In M. Smith, A. Smaill, and G. Wiggins, editors, *Music Education: An Artificial Intelligence Approach*, Workshops in Computing, pages 41–55. Springer-Verlag, 1994. Proceedings of a workshop held as part of AIED 93, Edinburgh, Scotland, 25 August 1993.
- [Hur95] David Huron. *The Humdrum Toolkit: Reference Manual*. Center for Computer Assisted Research in the Humanities, Menlo Park, California, 1995. ISBN 0- 936943-10-6.
- [Jeh04] Tristan Jehan. Event-synchronous music analysis/synthesis. In Proceedings of the International Conference on Digital Audio Effects (DAFx-04), pages 561–567, Naples, Italy, 2004.
- [JR01] Florent Jaillet and Xavier Rodet. Improved modelling of attack transients in music analysis synthesis. In Proceedings of the International Computer Music Conference (ICMC), pages 30–33, Havana, Cuba, 2001.
- [Kir86] T. Kirshbaum. Using a touch-table as an effective, low-cost input device in a melodic dictation program. *Journal of Computer-Based Instruction*, 13(1):14–16, 1986.
- [Kla04] Anssi Klapuri. *Signal Processing Methods for the Automatic Transcription of Music*. PhD thesis, Tampere University of Technology, Tampere, Finland, 2004.
- [Kla99b] Anssi Klapuri. Sound onset detection by applying psychoacoustic knowledge. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), volume 6, pages 3089–3092, 1999b.
- [Kuh90] Kuhn, W.B. (1990). A real-time pitch recognition algorithm for music applications, *CMJ* 14(3), pp. 60-71.
- [KVJ93] Matti Karjalainen, Vesa Valimaki and Zoltan Janosy: Towards high-quality sound synthesis of the guitar and string instruments. In Proceedings of the International Computer Music Conference (ICMC), pages 56–63, Tokyo, Japan, August 1993

- [KVS01] Anssi Klapuri, Tuomas Virtanen, and Jarno Seppänen. Automatic transcription of musical recordings. In Proceedings of the Consistent and Reliable Acoustic Cues Workshop (CRAC-01), Aalborg, Denmark, 2001.
- [Lam82] M. Lamb. An interactive graphical modelling game for teaching musical concepts. *Journal of Computer-Based Instruction*, 1982. Autumn 1982.
- [LD99] Laroche J. and Dolson M. (1999). “Improved phase vocoder time scale modification of audio”. *IEEE Trans. On Speech and Audio Processing*, 7 (3):323-332
- [LeB97] A. LeBlanc, Y. C. Jin, M. Obert, and C. Siivola. Effect of audience on music performance anxiety. *Journal of Research in Music Education*, 45(3):480–496, 1997.
- [Lev85] D. A. Levitt. *A Representation of Musical Dialects*. Unpublished PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1985.
- [Mar98] Martin Keith D., (1998): “Toward Automatic Sound Source Recognition: Identifying Musical Instruments”, NATO Computational Hearing Advanced Study Institute, Il Ciocco, Italy, 1998
- [Mas96] Paul Masri. Computer modeling of Sound for Transformation and Synthesis of Musical Signal. PhD dissertation, University of Bristol, UK, 1996.
- [MH91]] Meddis, R. and M.J. Hewitt (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: pitch identification, *J. Acoust. Soc. Am.* 89(6), pp. 2866-2882.
- [Moo78] Moorer, J.A. (1977). On the transcription of musical sound by computer, *CMJ*, 1(4), pp. 32-38. Moorer, J.A. (1978). The use of the linear prediction of speech in computer music applications, *JAES*, 27(3), pp. 134-140.
- [MR97] Dirk Moelants and Christian Rampazzo. A computer system for the automatic detection of perceptual onsets in a musical signal. In A. Camurri, editor, *KANSEI - The Technology of Emotion*, pages 141–146, Genova: AIMI-DIST, 1997.
- [MW] Philip McLeod, Geoff Wyvill, Visualization of Musical Pitch, Department of Computer Science, University of Otago, New Zealand
- [Nar90] E. Narmour. *The Analysis and Cognition of Basic Melodic Structures*. University of Chicago Press, London, 1990.
- [NI86] Niihara, T. and S. Inokuchi (1986). Transcription of sung song, *ICASSP’86*, pp. 1277-1280.
- [Nol67] Noll, A.M. (1967). Cepstrum pitch determination, *J. Acoust. Soc. Am.* 41(2), pp. 293-309.
- [Nol69] Noll, A.M. (1969). “Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate”. Proceedings of the Symposium on Computer Processing in Communications, pp779-797.
- [OD01] N. Orio and F. Dechelle. Score Following Using Spectral Analysis and Hidden Markov Models. In Proceedings of the ICMC, Havana, Cuba, 2001.
- [OLS03] [Orio et al. 03] Nicola Orio, Serge Lemouton, Diemo Schwarz, “Score Following: State of the Art and New Developments” Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03), Montreal, Canada, 2003
- [OS83] T. O’Shea and J. Self. *Learning and Teaching with Computers*. Prentice-Hall, London, 1983

- [Pac02] Francois Pachet. Multimedia at work - playing with virtual musicians: the continuator in practice. *IEEE Multimedia*, 9(3):77–82, 2002.
- [PAZ98] Miller S. Puckette, Theodore Apel, and David D. Zicarelli. Real-time analysis tools for PD and MSP. In *Proceedings of the International Computer Music Conference (ICMC)*, Ann Arbor, University of Michigan, USA, 1998.
- [PG79a] Piszczalski, M. and B. Galler (1979). Predicting musical pitch from component frequency ratios, *J. Acoust. Soc. Am.* 66(3), pp. 710-720.
- [PG79b] Piszczalski, M. and B. Galler (1979). Computer Analysis and Transcription of Performed Music: a Project Approach, *Computers and the Humanities* 13, pp. 195-206.
- [PL92] [Puckette & Lippe 92] Miller Puckette and Cort Lippe. “Score Following in Practice”. *Proceedings of the ICMC*, pages 182--185, 1992
- [PPW05] Elias Pampalk, Tim Pohle, and Gerhard Widmer. Dynamic playlist generation based on skipping behavior. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, London, UK, September 2005.
- [Rab77] Rabiner, L.R. (1977). On the use of autocorrelation analysis for pitch detection, *IEEE Transactions on Acoustics, Speech and Signal Processing*, 25(1), pp. 24-33.
- [Rap01b] Christopher Raphael. Music Plus One: A system for expressive and flexible musical accompaniment. In *Proceedings of the International Computer Music Conference (ICMC)*, Havana, Cuba, 2001b.
- [Ras98] Raškinis, G. (1998). Preprocessing of folk song acoustic records for transcription into music scores, *Informatika* 9(3), pp. 343-364.
- [RCR+76] Prentice-Hall. Rabiner, L.R., M.J. Cheng, A.E. Rosenberg and C.A. McGonegal (1976). A comparative performance study of several pitch detection algorithms, *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(5), pp. 399-418.
- [Rob94] C. Robbie. *Implementing a Generative Grammar for Music*. MSc thesis, Department of Artificial Intelligence, University of Edinburgh, 80 South Bridge, Edinburgh, 1994.
- [RRS+98] Rossignol, S., X. Rodet, J. Soumagne, J.-L. Colette and P. Depalle. (1998). Feature extraction and temporal segmentation of acoustic signals, *Int. Computer Music Conf.* '98.
- [RSB05] Emmanuel Ravelli, Mark Sandler and Juan-Pablo Bello: Fast implementation for non-linear time scaling of stereo signals. In *Proceedings of the International Conference on Digital Audio Effects (DAFx-05)*, pages 182–185, Madrid, Spain, 2005.
- [SE03] Sheh, A. and D. P.W. Ellis (2003). Chord segmentation and recognition using em-trained hidden markov models. In *4th International Symposium on Music Information Retrieval ISMIR-03*, Baltimore
- [Ser05] X.Serra, TOWARDS A ROADMAP FOR THE RESEARCH IN MUSIC TECHNOLOGY, *Proceedings of the ICMC, 2005*
- [SH93] M. Smith and S. Holland. MOTIVE - the development of an AI tool for beginning melody composers. In M. Smith, A. Smaill, and G. Wiggins, editors, *Music Education: An Artificial Intelligence Approach*, Workshops in Computing, pages 41–55. Springer-Verlag, 1994. Proceedings of a workshop held as part of AIED 93, Edinburgh, Scotland, 25 August 1993.

- [SH94] M. Smith and S. Holland. MOTIVE - the development of an AI tool for beginning melody composers. In M. Smith, A. Smaill, and G. Wiggins, editors, *Music Education: An Artificial Intelligence Approach*, Workshops in Computing, pages 41–55. Springer-Verlag, 1994. Proceedings of a workshop held as part of AIED 93, Edinburgh, Scotland, 25 August 1993.
- [Shr98b] Eric D. Scheirer. Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America*, 103(1):588–601, 1998b.
- [SL90] Slaney, M. and R. F. Lyon (1990). A perceptual pitch detector. *ICASSP*, volume 1, pp. 357–360.
- [Smi95] M. Smith. *Towards an Intelligent Learning Environment for Melody Composition Through Formalisation of Narmour’s Implication-Realisation Model*. Ph.D. thesis, IET, Open University, 1995.
- [Smi96] Leslie S. Smith. Using an onset-based representation for sound segmentation. In Proceedings of the International Conference on Neural networks and their Applications (NEURAP), pages 274–281, Marseilles, France, March 20-22 1996.
- [Sor87] L. Sorisio. Design of an intelligent tutoring system in harmony. In *Proceedings of the 1987 International Computer Music Conference*, pages 356–363. Urbana, IL, 1987.
- [SS97] Schreier E., M. Slaney, (1997): “Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator”, ICASSP 1997
- [Ste99] Sterian, A.D (1999). Model-Based Segmentation of Time-Frequency Images for Musical Transcription. PhD thesis, University of Michigan.
- [Swa79] K. Swanwick. *A Basis for Music Education*. NFER Publishing, London, 1979.
- [SWK95] S. W. Smoliar, J. A. Waterworth, and P. R. Kellock. pianoforte: A system for piano education beyond notation literacy. In Proceedings of ACM Multimedia 95, pages 457–465. San Francisco, CA, 1995.
- [Tob88] J. C. Tobias. Knowledge representation in the harmony intelligent tutoring system. In *Proceedings of the First Workshop on Artificial Intelligence and Music*, pages 112–124. Minneapolis, Minnesota, 1988
- [Tza02] George Tzanetakis. Manipulation, Analysis and Retrieval for Audio Signals. PhD thesis, Faculty of Princeton University, Department of Computer Science, June 2002.
- [Ver04] Vincent Verfaillie. Effets audionumeriques adaptatifs: theories, mise en oeuvre et usage en creation musicale numerique, PhD thesis, Universite Aix-Marseille II, 2004.
- [Ver85] Barry L. Vercoe: The synthetic performer in the context of live performance. In Proceedings of the 1984 International Computer Music Conference (ICMC), pages 199–200, IRCAM, Paris, France, June 1985.
- [Wan03] Avery Wang. An industrial strength audio search algorithm. Invited speaker. In Proceedings of the International Symposium on Music Information Retrieval (ISMIR), Baltimore, Maryland, USA, October 2003.
- [Yin00] Ying G.S., L.H. Jamieson, C.D. Michell: “A Probabilistic Approach to AMDF Pitch Detection” [IRCAM Bibliography] Score Following Commented Bibliography, (<http://www.ircam.fr/equipes/temps-reel/suivi/bibliography.html>)

[Zla06] A. Zlatintsi, when the clarinet sounds bad, Master of Science Thesis in Music Acoustics, Stockholm 2006

[ZP01] Aymeric Zils and Francois Pachet. Musical mosaicing. In Proceedings of the International Conference on Digital Audio Effects (DAFx-01), pages 39–44, Limerick, Ireland, 2001.

<http://www.vemus.org>

www.users.globalnet.co.uk/~7Ebunce/sound.htm

<http://teachers.edenpr.org/~benanderson/Physics%20of%20Sound.htm>

road.uww.edu/road/herriotj/TME/chapt23part1.pdf