

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΣΧΟΛΗ ΜΟΥΣΙΚΗΣ & ΟΠΤΙΚΟΑΚΟΥΣΤΙΚΩΝ
ΤΕΧΝΩΝ

ΤΜΗΜΑ ΤΕΧΝΩΝ ΗΧΟΥ & ΕΙΚΟΝΑΣ
ΕΡΓΑΣΤΗΡΙΟ ΕΠΕΞΕΡΓΑΣΙΑΣ
ΟΠΤΙΚΟΑΚΟΥΣΤΙΚΩΝ ΣΗΜΑΤΩΝ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ ΜΕ ΤΙΤΛΟ:

*Εξόρυξη Συναισθηματικής Πληροφορίας από Ηχητικά
Γεγονότα*

Κωνσταντίνος Δρόσος

Επιβλέπων: Δρ. Ανδρέας Φλώρος

Κέρκυρα, 2015

Πιστοποίηση

Πιστοποιείται ότι η παρούσα διδακτορική διατριβή με θέμα:

«Εξόρυξη συναισθηματικής πληροφορίας από ηχητικά γεγονότα»

του Κωνσταντίνου Δρόσου, Πτυχιούχου Τεχνολογίας Ήχου και Μουσικών Οργάνων, MSc Sound & Vibration Studies του Institute of Sound & Vibration Research (I.S.V.R.) - Πανεπιστήμιο του Southampton, παρουσιάστηκε δημοσίως στο Τμήμα Τεχνών Ήχου και Εικόνας του Ιονίου Πανεπιστημίου στις 10 Ιουλίου 2015 και εξετάστηκε και εγκρίθηκε από την ακόλουθη Εξεταστική Επιτροπή:

1. **Ανδρέα Φλώρο**, Αναπληρωτή Καθηγητή της Σχολής Μουσικής και Οπτικοακουστικών Τεχνών του Ιονίου Πανεπιστημίου, Επιβλέπων Καθηγητής
2. **Ιωάννη Ζάννο**¹, Καθηγητή της Σχολής Μουσικής και Οπτικοακουστικών Τεχνών του Ιονίου Πανεπιστημίου
3. **Ιωάννη Μουρτζόπουλο**, Καθηγητή της Πολυτεχνικής Σχολής του Πανεπιστημίου Πατρών
4. **Νικόλαο - Γρηγόριο Κανελλόπουλο**, Καθηγητή της Σχολής Μουσικής και Οπτικοακουστικών Τεχνών του Ιονίου Πανεπιστημίου
5. **Ανδρέα Μνιέστη**, Αναπληρωτή Καθηγητή της Σχολής Μουσικής και Οπτικοακουστικών Τεχνών του Ιονίου Πανεπιστημίου
6. **Γεώργιο Τσιχριντζή**, Καθηγητή της Σχολής Τεχνολογιών Πληροφορικής και Επικοινωνιών του Πανεπιστημίου Πειραιώς
7. **Γεώργιο Καλλίρη**, Αναπληρωτή Καθηγητή της Σχολής Οικονομικών και Πολιτικών Επιστημών του Αριστοτέλειου Πανεπιστημίου Θεσσαλονίκης

Κέρκυρα, 10 Ιουλίου 2015

Ο Επιβλέπων Καθηγητής

Ο Πρόεδρος του Τμήματος

Ανδρέας Φλώρος

Ανδρέας Φλώρος

¹Απών λόγω ασθένειας

ΕΥΧΑΡΙΣΤΙΕΣ

Μέσω αυτής της ενότητας της διατριβής μου, θα ήθελα να ευχαριστήσω τους ανθρώπους οι οποίοι μου στάθηκαν και μου πρόσφεραν την αμέριστη και ανιδιοτελή βοήθειά τους. Ο πρώτος από αυτούς είναι ο Δρ. Ανδρέας Φλώρος. Ο Καθηγητής που με επέβλεψε όλα αυτά τα χρόνια και, πλέον, τον λογίζω ως τον επιστημονικό μου πατέρα. Χωρίς την επιστημονική και εν γένει καθοδήγηση και βοήθειά του δεν νομίζω ότι θα μπορούσα να τα είχα καταφέρει. Ο δεύτερος άνθρωπος είναι η Χαρά Σακελλάρη, όπου ήταν δίπλα μου, με υπόμενε και μου πρόσφερε ένα στέρεο τόπο για να στηριχτώ στις στιγμές που το είχα ανάγκη πιο πολύ. Έπειτα, είναι τα άλλα δύο μέλη της τριμελούς μου επιτροπής, ο Δρ. Ιωάννης Ζάννος και ο Δρ. Ιωάννης Μουρτζόπουλος. Αμφότεροι μου πρόσφεραν ευκαιρίες βελτίωσης της έρευνάς μου κατά την διάρκειά της. Τέλος, θα ήθελα να ευχαριστήσω τον Δρ. Ανδρέα Γιαννακουλόπουλο για την πολύτιμη συνεργασία του και στήριξή του, τον Δρ. Νικόλαο Κανελλόπουλο για την βοήθειά του και όλους όσους με άντεξαν και με βοήθησαν όλα αυτά τα χρόνια.

ΠΕΡΙΛΗΨΗ

Το ακουστικό κανάλι ένα από τα δύο συνηθέστερα κανάλια επικοινωνίας. Μέσω της ακοής, ο άνθρωπος αντιλαμβάνεται μία πληθώρα πληροφοριών σχετικά με το περιβάλλον του, μπορεί να λάβει γνώση από πληροφορίες που του μεταδίδει κάποιος άλλος άνθρωπος ή και να ψυχαγωγηθεί. Η γνώση ότι η ομιλία και η μουσική μπορούν να εγείρουν συναισθήματα στον ακροατή είναι κοινός τόπος στους ειδικούς και μη ειδικούς. Όμως, οι δύο αυτοί τύποι ηχητικού ερεθίσματος, ήτοι ομιλία και μουσική, είναι μόνο ένα μέρος του συνόλου των ήχων που μπορεί να ακούσει ένας άνθρωπος. Υπάρχει και μία άλλη κατηγορία, οι γενικοί ήχοι ή ηχητικά γεγονότα, οι οποίοι καταλαμβάνουν το μεγαλύτερο μέρος του πλήθους ηχητικών ερεθισμάτων που ακούει ένας άνθρωπος. Έχει ήδηδειχθεί, από δημοσιευμένες εργασίες, ότι τα ηχητικά γεγονότα μπορούν να επηρεάσουν το συναίσθημα του ακροατή τους, χωρίς όμως να έχει βρεθεί κάποια σχέση με τα τεχνικά χαρακτηριστικά του ηχητικού ερεθίσματος ή επιπλέον πληροφορίας που μεταφέρεται μέσω των ηχητικών γεγονότων. Λόγω της γενικής φύσης αυτού του τύπου ηχητικού ερεθίσματος, εάν βρεθεί μία σχέση μεταξύ των τεχνικών ηχητικών χαρακτηριστικών και του εγειρόμενου, στον ακροατή, συναίσθημα τότε είναι αρκετά πιθανό αυτό το τεχνικό χαρακτηριστικό να μπορεί να επιδράσει με τον συγκεκριμένο τρόπο και στις άλλες δύο, πιο ειδικές, μορφές ήχου, ήτοι την ομιλία και την μουσική. Το γεγονός αυτό θα επιτρέψει μία εποπτική έρευνα σχετικά με τα τεχνικά χαρακτηριστικά των εν γένει ηχητικών ερεθισμάτων με την συναισθηματική τους επίρεια στον ακροατή. Έτσι, η ανά χείρας διατριβή εστιάζει στην αναγνώριση συναισθήματος από ηχητικά γεγονότα. Έρευνά και παρουσιάζει αποτελέσματα σχετικά με την επίρεια του ρυθμού, της ενέργειας, του

σημασιολογικού περιεχομένου του ηχητικού γεγονότος αλλά και της χωρικής τοποθέτησης της πηγής στο συναίσθημα του ακροατή. Τα αποτελέσματα δείχνουν ότι ο ρυθμός αποτελεί ένα τεχνικό χαρακτηριστικό το οποίο μπορεί να επηρεάσει το συναίσθημα του ακροατή για όλους τους τύπους ηχητικών ερεθισμάτων αλλά και ότι η χωρική τοποθέτηση της πηγής μπορεί να τροποποιήσει το εγειρόμενο συναίσθημα από ένα ηχητικό ερέθισμα.

Λέξεις Κλειδιά

Αναγνώριση συναισθήματος. Ηχητικά γεγονότα. Συναισθηματική Ακουστική Οικολογία. Αμφιωτική ακρόαση. Ρυθμός. Ενέργεια. Χωρική τοποθέτηση πηγής.

ABSTRACT

Audio channel is one the two most common communication channels. Through hearing, humans perceive different information regarding their environs, can receive knowledge that another human communicates to them or can entertain their selfs. The concept that speech and music can convey emotions to the listener is common to experts and non-experts. But, these two types of audio stimuli, i.e. speech and music, are just a portion of the total amount of sounds that a man can hear. There is another type of sound stimulus, termed as general sounds or sound events, that occupy a bigger proportion of the total sound stimuli that we receive. Published works have shown that sound events can convey emotions to their listener, but they do not offer a relationship between the conveyed emotion and the technical characteristics of the sound events or any extra information that the latter can communicate. Due to the general nature of the sound events, if a relationship between the technical characteristics and the conveyed emotion is proven then is very possible that this characteristics can have a similar or same effect on emotions when the audio stimulus is speech or music. This fact will allow a higher level research focused on the effect of the sound's technical characteristics on emotions. Thus, this thesis focus on the emotion recognition from sound events. It examines and presents results with respect to the effect of rhythm, energy and semantic content of a sound event and source's spatial position on the listener's emotional state. Results portray that rhythm is a technical characteristic that affects listener's emotional state with all types of audio stimuli, whether energy is not. In addition, the source's spatial position can alter

the conveyed emotion by a sound event.

Keywords

Emotion recognition. Sound events. Affective Acoustic Ecology. Binaural hearing. Rhythm. Energy. Spatial positioning of source

ΠΕΡΙΕΧΟΜΕΝΑ

Ευχαριστίες	i
Περίληψη	iii
Abstract	v
Περιεχόμενα	ix
Κατάλογος Σχημάτων	xii
Κατάλογος Πινάκων	xvi
1 Εισαγωγή	1
1.1 Ήχος και συναίσθημα	3
1.2 Ερευνητικό ερώτημα και πιθανά οφέλη	4
1.3 Δομή τόμου	7
2 Θεωρητικό υπόβαθρο	9
2.1 Ερευνητικό υπόβαθρο	10
2.1.1 Μηχανική Μάθηση	12
2.1.2 Αλγόριθμοι κατηγοριοποίησης	14
2.1.3 Τι είναι το Συναίσθημα;	22
2.1.4 Διακριτά μοντέλα συναισθημάτων	23

2.1.5	Συνεχή μοντέλα συναισθημάτων	26
2.1.6	Αποτύπωση συναισθηματικής κατάστασης	28
2.1.7	Τεχνικά μεγέθη ήχου	30
2.2	Στάθμη της τεχνικής	44
2.2.1	Ανάκτηση μουσικής πληροφορίας, το ερευνητικό πεδίο MIR και εφαρμογές αυτού	45
2.2.2	Αναγνώριση συναισθήματος από μουσική	46
2.2.3	Αναγνώριση συναισθήματος από ομιλία	58
2.2.4	Το ηχητικό γεγονός	60
2.2.5	Διαθέσιμα σετ συναισθηματικά επισημασμένων ηχητικών γεγονότων	63
2.2.6	Αναγνώριση συναισθήματος από ηχητικά γεγονότα	65
3	Θεωρητική - Πειραματική προσέγγιση και μεθοδολογία	69
3.1	Ερευνητικά ερωτήματα	71
3.2	Θεωρητική προσέγγιση	72
3.2.1	Συναισθηματικά μοντέλα	73
3.2.2	Συσχέτιση υψηλής διέγερσης και χαμηλής ευχαρίστησης με τεχνικά χαρακτηριστικά	75
3.2.3	Ένταξη συναισθηματικής συνιστώσας στο μοντέλο ηχητικού περιβάλλοντος του ανθρώπου	77
3.2.4	Επήρεια χωρικής τοποθέτησης της πηγής στο συναίσθημα	79
3.2.5	Συσχέτιση τεχνικών χαρακτηριστικών και συναισθήματος	82
3.3	Πειραματική προσέγγιση	83
3.3.1	Αναγνώριση stress από φωνητικές εντολές	83
3.3.2	Αναγνώριση διέγερσης και ευχαρίστησης από ηχητικά γεγονότα	89
3.3.3	Διερεύνηση της επίρειας των ρυθμικών χαρακτηριστικών ηχητικών γεγονότων στην διέγερση	92
3.3.4	Διερεύνηση της επίρειας της χωρικής τοποθέτησης της πηγής στο συναίσθημα	98
3.3.5	Διερεύνηση της επίρειας των ρυθμικών χαρακτηριστικών ηχητικών γεγονότων στην ευχαρίστηση	106
4	Αποτελέσματα	111

4.1	Αναγνώριση stress από φωνητικές εντολές	112
4.2	Πρώτη προσπάθεια αναγνώρισης συναισθήματος στα πλαίσια της Συναισθηματικής Ακουστικής Οικολογίας	113
4.3	Συσχέτιση ρυθμού και Διέγερσης για ηχητικά γεγονότα	115
4.4	Επήρεια χωρικής τοποθέτησης της πηγής στο συναίσθημα του ακροατή	118
4.5	Συσχέτιση ρυθμού και Ευχαρίστησης	125
5	Συζήτηση και ανάλυση αποτελεσμάτων	129
5.1	Συσχ. υψηλής Διέγ/σης & χαμηλής Ευχ/σης με τεχ/κά χαρ/κά του ήχου	130
5.2	Αναγνώριση συναισθήματος από ηχητικά γεγονότα στα πλαίσια της Συναισθηματικής Ακουστικής Οικολογίας	133
5.3	Συσχέτιση ρυθμικών χαρακτηριστικών με διέγερση	136
5.4	Χωρική θέση ηχητικών πηγών και συναισθηματική κατάσταση του ακροατή	139
5.5	Συσχέτισης ρυθμικών χαρακτηριστικών με ευχαρίστηση	157
6	Συμπεράσματα και Μελλοντικές Προεκτάσεις	165
6.1	Επήρεια συναισθημάτων από τεχνικά χαρακτηριστικά ηχητικών γεγονότων	169
6.2	Επήρεια συναισθημάτων από χαρακτηριστικά της πηγής	172
6.3	Πρωτοτυπία της διατριβής	173
6.4	Μελλοντικές προεκτάσεις και εφαρμογές	174
	Βιβλιογραφία	177

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

2.1	Η γενική διαδικασία αναγνώρισης συναισθημάτων από ηχητικά δεδομένα . . .	11
2.2	Το παραγόμενο μοντέλο από έναν supervised αλγόριθμο (Random Tree) που χρησιμοποιεί δένδρα απόφασης	16
2.3	Ένα ANN για την κατηγοριοποίηση της εξόδου σε κλάσεις 0 και 1 βάσει εισόδου τριών δεδομένων: RMS, Pitch και Tempo.	19
2.4	Απεικόνιση των ομάδων στο μοντέλο “Λίστα της Hevner” (Hevner’s List) . .	25
2.5	Το αρχικό μοντέλο των συναισθηματικών καταστάσεων. Οι λεκτικές περιγραφές είναι στην αγγλική γλώσσα για πλήρη συμφωνία με την διεθνή βιβλιογραφία.	26
2.6	Ο χώρος της Ευχαρίστησης - Διέγερσης με αντιστοιχισμένα συναισθήματα .	27
2.7	Οι εικόνες SAM για τις τρεις συναισθηματικές καταστάσεις	29
2.8	Η ατάκα (attack) ενός ηχητικού σήματος. Ο οριζόντιος άξονας είναι ο χρόνος (Time) ενώ ο κάθετος η ενέργεια (Energy)	38
2.9	Οι τιμές της επισήμανσης για τους ήχους IADS και για τις διαστάσεις της Διέγερσης (Arousal) και Ευχαρίστησης (Valence)	64
3.1	Απεικόνιση της μοντελοποίησης της κατάστασης του stress σε σχέση με το μοντέλο Διέγερσης (Arousal) - Ευχαρίστησης (Valence) και των επισημάνσεων των ήχων IADS	76

3.2	Η χωρική τοποθέτηση της πηγής όπως επιλέχθηκε για την μελέτη της επή- ρειάς της στην συναισθηματική κατάσταση του ακροατή	82
3.3	Το σύστημα αναγνώρισης stress από φωνητικές εντολές	86
3.4	Τα όρια και οι μέσες τιμές για stress και χωρίς για την ενέργεια, θεμελιώδη συχνότητα και ρυθμό ομιλίας	88
3.5	Γραφική απεικόνιση της αντιστοίχισης τεταρτημορίων και κλάσεων συναι- σθηματικής κατάστασης που χρησιμοποιήθηκε	91
3.6	Το ιστόγραμμα των τιμών της Διέγερσης (Arousal) για τους ήχους IADS. Ως “Number of sound events” είναι το πλήθος των ηχητικών γεγονότων . . .	94
3.7	Το ιστόγραμμα των συχνοτήτων δειγματοληψίας (Sampling frequency) για τους ήχους IADS. Ως “Number of sounds” είναι το πλήθος των ηχητικών γεγονότων	95
3.8	Η διαδικασία των τεστ ακρόασης κατά την μελέτη της χωρικής τοποθέτησης της πηγής στο συναίσθημα του ακροατή	102
3.9	Οι διαθέσιμες επιλογές για την συναισθηματική επισήμανση των ηχητικών γεγονότων κατά την μελέτη της χωρικής τοποθέτησης της πηγής στο συναι- σθημα του ακροατή	103
3.10	Η κατανομή σε κλάσεις για την ευχαρίστηση και την διέγερση	109
4.1	Αποτύπωση της κανονικοποιημένης ενέργειας (Normalized energy) ενός η- χητικού γεγονότος σε σχέση με την επισήμανση Διέγερσης (Arousal) στους ήχους IADS	114
4.2	Μέσες τιμές Διέγερσης και Ευχαρίστησης για τα $S'_a(k)$, με $k \in [1, 5]$	121
5.1	Η αποτύπωση των συναισθηματικών επισημάνσεων, στο επίπεδο Διέγερσης (Arousal) – Ευχαρίστησης (Valence), για τα ηχητικά γεγονότα με χωρική τοποθέτηση και $\theta(1) = 0^\circ$	146
5.2	Το ποσοστό των ηχητικών γεγονότων που ανήκουν στις κλάσεις C_1 , C_2 και $C_1 \cup C_2$	151

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

2.1	Τα βασικά συναισθήματα	24
2.2	Συνοπτική παρουσίαση των μεγεθών που χρησιμοποιήθηκαν στην παρούσα διατριβή	31
2.3	Το είδος μουσικής που ανήκουν τα ηχητικά δεδομένα των εργασιών που αναφέρονται και έχουν ως αντικείμενο την αναγνώριση συναισθήματος από μουσική	50
2.4	Ο δείκτης των δημοσιευμένων εργασιών που αναφέρονται στην παρούσα διατριβή (Δ.Ε.), τα τεχνικά χαρακτηριστικά που χρησιμοποίησαν (Τ.Χ.) καθώς και την κατηγορία που ανήκουν τα τελευταία (Κατηγορία Τ.Χ.)	51
2.5	Ο δείκτης των δημοσιευμένων εργασιών που αναφέρονται στην παρούσα διατριβή (Δ.Ε.), ένδειξη για την χρήση προηγούμενης συναισθηματικής επισήμανσης και το συναισθηματικό μοντέλο που χρησιμοποιήθηκε	53
2.6	Οι αλγόριθμοι που χρησιμοποιήθηκαν, ο τύπος τους (Τ.Α.), η ευστοχία του παραγόμενου μοντέλου και ο αντίστοιχος δείκτης της εργασίας (Δ.Ε.). Μ.Ε. είναι αλγόριθμος χωρίς επίβλεψη (non-supervised) ενώ Ε. είναι με επίβλεψη (supervised). Για την στήλη της ευστοχίας, εάν από την αντίστοιχη εργασία προσφέρεται το ποσοστό του accuracy τότε απεικονίζεται αυτό. Εάν προσφέρονται τα ποσοστά ακρίβειας (precision) και ανάκλησης (recall) τότε απεικονίζονται αυτά ως: Ακρίβεια/Ανάκληση	53
2.7	Τα προφίλ της ομιλίας για διάφορα βασικά συναισθήματα	59

3.1	Τα τεχνικά χαρακτηριστικά που επιλέχθηκαν για εξαγωγή από τα ηχητικά δεδομένα φωνητικών εντολών	77
3.2	Τα στοιχεία του ηχητικού γεγονότος, όπως αυτό ορίζεται από την Συναισθηματική Ακουστική Οικολογία	78
3.3	Η μεταβολή της τιμής των τεχνικών χαρακτηριστικών που επιλέχθηκαν για την αναγνώριση stress σε σχέση με τα βασικά συναισθήματα. Με F_0 είναι η θεμελιώδης συχνότητα, ως S_R συμβολίζεται ο ρυθμός ομιλίας (speech rate) και E είναι η ενέργεια	84
3.4	Οι φωνητικές εντολές που χρησιμοποιήθηκαν στο σύστημα αναγνώρισης stress	85
3.5	Τα τεχνικά χαρακτηριστικά που χρησιμοποιήθηκαν στο πείραμα εξόρυξης συναισθηματικής πληροφορίας από τους ήχους IADS	90
3.6	Τα στατιστικά και μαθηματικά μεγέθη που χρησιμοποιήθηκαν στην εξόρυξη συναισθηματικής πληροφορίας από τους ήχους IADS	90
3.7	Τα τεχνικά χαρακτηριστικά και τα στατιστικά μεγέθη που χρησιμοποιήθηκαν για την συσχέτιση του ρυθμού και της διέγερσης	93
3.8	Η χρονική διάρκεια των πλαισίων που χρησιμοποιήθηκαν για την συσχέτιση των ρυθμικών χαρακτηριστικών με την διέγερση	93
3.9	Τα τεχνικά χαρακτηριστικά που χρησιμοποιήθηκαν στο πείραμα συσχέτισης ρυθμικών χαρακτηριστικών με την διέγερση και τα αντίστοιχα γράμματα για την γρήγορη δεικτοδότησή τους	95
3.10	Συσχέτιση τεχνικών χαρακτηριστικών κατά την αναγνώριση διέγερσης από ρυθμικά τεχνικά χαρακτηριστικά.	97
3.11	Η χώρα καταγωγής των συμμετεχόντων ανά ποσοστό του συνολικού αριθμού συμμετεχόντων	100
3.12	Η αντιστοιχία του δείκτη i' με το i καθώς και το σημασιολογικό περιεχόμενο των αντίστοιχων ηχητικών γεγονότων για το σετ ηχητικών γεγονότων BEADS103	
3.13	Η αντιστοιχία του i' με το i καθώς και το σημασιολογικό περιεχόμενο των τελικών ηχητικών γεγονότων	105
3.14	Τεχνικά χαρακτηριστικά και στατιστικά μεγέθη που χρησιμοποιήθηκαν κατά την συσχέτιση ρυθμικών χαρακτηριστικών και ευχαρίστησης	108
4.1	Η ευστοχία αναγνώρισης stress, δεδομένης μίας λέξης, με είσοδο λέξεις με και χωρίς stress	113

4.2	Αποτελέσματα ευστοχίας αναγνώρισης συναισθηματικής πληροφορίας	114
4.3	Λίστα κανονικοποιημένης ενέργειας, E , 7 ενδεικτικών ηχητικών γεγονότων με αύξουσα τιμή διέγερσης	115
4.4	Ιεράρχηση των τεχνικών χαρακτηριστικών βάσει της χρονικής διάρκειας του πλαισίου (w) και του αλγόριθμου ιεράρχησης που χρησιμοποιήθηκε. Η σειρά εμφάνισης είναι και η σειρά ιεράρχησης (πρώτα είναι τα σημαντικότερα)	116
4.5	Η ευστοχία των παραγόμενων μοντέλων για την συσχέτιση των ρυθμικών χαρακτηριστικών και της διέγερσης αναλόγως του χρονικού πλαισίου (διάρκειας w), του αλγόριθμου που χρησιμοποιήθηκε και της τεχνικής επαλήθευσης	117
4.6	Η μέση τιμή και η τυπική απόκλιση (σ) για το πλήθος των επισημάνσεων που έλαβε κάθε ήχος ανά γωνία και συνολικά	119
4.7	Η διαφορά των επισημάνσεων των συναισθηματικών επισημάνσεων από το πείραμα της επήρειας της χωρικής τοποθέτησης και των αρχικών ήχων IADS	119
4.8	Τα μέτρα και οι γωνίες των διανυσμάτων $ \vec{v}_a $ αναλόγως τις γωνιακές μεταβάσεις	123
4.9	Οι τιμές της ανάκλησης (R) και της ακρίβειας (P) για κάθε κλάση (C_1 και C_2), για όλους τους αλγόριθμους και για το σετ ήχων $R'_{i,x'}$. Με γκρι σκίαση είναι τα κελιά με τις μέγιστες τιμές ανά κλάση και ως “D. Stump” είναι ο “Decission Stump”.	126
4.10	Οι τιμές της ανάκλησης (R) και της ακρίβειας (P) για κάθε κλάση (C_1 και C_2), για όλους τους αλγόριθμους και για το σετ ήχων $R''_{i,x'}$. Με γκρι σκίαση είναι τα κελιά με τις μέγιστες τιμές ανά κλάση και ως “D. Stump” είναι ο “Decission Stump”.	127
4.11	Ευστοχία (A) και μέσο λάθος (E) για όλους τους αλγόριθμους και τα σετ $R'_{i,x'}$ και $R''_{i,x'}$. Με γκρι σκίαση είναι τα κελιά με τις μέγιστες τιμές ανά σετ ήχων.	128
5.1	Τα πρώτα δεκατρία χαρακτηριστικά κατά την αξιολόγησή τους στο πείραμα συσχέτισης τεχνικών ρυθμικών χαρακτηριστικών και διέγερσης	137
5.2	Σημασιολογικό περιεχόμενο, τιμές των $ D_a(i') $ και $ D_v(i') $ και του p από τα z-tests για τους ήχους με $ D_a(i') > 1$ και $ D_v(i') > 1$	143
5.3	Τα ηχητικά γεγονότα στις κλάσεις C_1 και C_2 για όλες τις γωνίες $\theta(k)$	148

5.4	Στατιστικά στοιχεία για την γωνία $ \phi(i', k) $ και το μέτρο $ \bar{v}^{\lambda}(i', k) $ για τα ηχητικά γεγονότα που είναι στις C_1 και C_2 . Οι τιμές για την γωνία $ \phi(i', k) $ είναι σε μοίρες ενώ για το μέτρο των διανυσμάτων σε θέσεις στο μοντέλο SAM151	
5.5	Μέσες τιμές των στοιχείων των διανυσμάτων ΔA για τις κλάσεις 1 και 2 . . .	154
5.6	Οι μέσες τιμές επισημάνσεων για την Διέγερση, $\bar{a}(i', k)$, και την Ευχαρίστηση, $\bar{v}(i', k)$, για τα ηχητικά γεγονότα με μουσικό σημασιολογικό περιεχόμενο	156
5.7	Ακρίβεια και ανάκληση για τον αλγόριθμο RepTree και για χρονικό παράθυρο διάρκειας 1.6 δευτερολέπτων	160
5.8	Ακρίβεια και ανάκληση των αλγορίθμων AdaBoost και SMO για χρονικό παράθυρο διάρκειας 1.8 δευτερολέπτων ($i = 6$)	162

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

Η επικοινωνία, μία βασική πτυχή της καθημερινότητας του ανθρώπου, μπορεί να θεωρηθεί ως άμεσα εξαρτώμενη από την λήψη και μετάδοση πληροφορίας μέσω του ακουστικού καναλιού. Αυτό το γεγονός αναδεικνύει την σπουδαιότητα της ακοής και την κατατάσσει ως τον πιο σημαντικό αισθητηριακό τρόπο (sensory modality) για την προαναφερόμενη λειτουργία [1]. Πλείστες πληροφορίες λαμβάνονται συνεχώς από τον άνθρωπο μέσω της ακοής και, σε αντίθεση με την όραση, αυτή η λήψη μπορεί να συμβεί ακόμα και εάν δεν υπάρχει ένα άμεσο και χωρίς εμπόδια μονοπάτι (direct path) μεταξύ της πηγής και του δέκτη [2]. Για παράδειγμα, κάποιος μπορεί να ακούσει ένα ποτήρι που σπάει στο διπλανό δωμάτιο ακόμα και εάν μεταξύ της πηγής, δηλαδή του ποτηριού που σπάει, και του δέκτη, δηλαδή του ανθρώπου, μεσολαβούν τοίχοι. Ενώ όμως μπορεί να ακούσει το ηχητικό γεγονός του σπασίματος δεν μπορεί να το δει. Αυτού του είδους η διάδραση χρησιμοποιείται καθημερινώς από τον άνθρωπο για την διαπροσωπική αλληλεπίδραση, π.χ. μέσω της ομιλίας, για ψυχαγωγικούς σκοπούς, π.χ. μουσική, και για την κάλυψη αναγκών επιβίωσης, όπως στην περίπτωση που κάποιος αντιλαμβάνεται μέσω της ακοής μία πιθανή απειλή και πράττει αναλόγως [3].

Ένα προφανές περιεχόμενο της επικοινωνίας διαμέσου του ακουστικού καναλιού είναι το σημασιολογικό/λεκτικό, το οποίο, συνήθως, μεταφέρεται με την ομιλία. Πέρα από αυτό,

υπάρχει ένα ευρύ φάσμα πληροφοριών οι οποίες μεταφέρονται μέσω του προαναφερθέντος καναλιού επικοινωνίας. Σε αυτές συμπεριλαμβάνονται μη εννοιολογικά χαρακτηριστικά όπως η θέση της ακουστικής πηγής συναρτήσει αυτής του δέκτη, η ταχύτητα της πηγής (εάν κινείται), το είδος και ο τρόπος παραγωγής ήχου (π.χ. κρούση, τριβή), το μέγεθος της πηγής κ.α. [4, 5]. Υπάρχουν όμως κάποια εννοιολογικά ή/και ανώτερου επιπέδου χαρακτηριστικά όπως το είδος του μουσικού οργάνου/μουσικής, αλλά και συναισθήματα τα οποία μπορούν να μεταδοθούν από το ακουστικό κανάλι [6, 7, 8, 9].

Αναφορικά με τα τελευταία, μπορεί να λεχθεί ότι αποτελεί ένα πεδίο των κοινωνικών ή/και ανθρωπιστικών επιστημών στο οποίο υπάρχει ακόμα μία ασυμφωνία των επιστημόνων σχετικά με ένα κοινό αποδεκτό ορισμό της έννοιας του συναισθήματος [10]. Παρόλα αυτά, με την χρήση φυσιολογικών (physiological) ενδείξεων του ανθρώπου και μοντέλων της συναισθηματικής κατάστασης και των συναισθημάτων είναι δυνατή τελικά η μέτρηση και αποτύπωση του συναισθήματος [11, 12]. Αυτό το γεγονός έχει επιτρέψει την έρευνα σχετικά με την χρήση τους ως έναν επιπλέον ανθρωποκεντρικό παράγοντα σε εφαρμογές, με αρκετές από αυτές να σχετίζονται ή να περιέχουν ηχητική διεπαφή. Ένα σημαντικό στάδιο στην έρευνα αυτή είναι η εξόρυξη της συναισθηματικής πληροφορίας από τα ηχητικά δεδομένα. Η ερευνητική κοινότητα έχει ήδη παρουσιάσει αξιοσημείωτο και ευρύ έργο σχετικά με την αναγνώριση συναισθημάτων από μουσική και ομιλία, όμως δεν συμβαίνει το ίδιο και με τους γενικούς ήχους όπου η έρευνα σχετικά με τα εγείρομενα συναισθήματα στον ακροατή παραμένει σε αρχικό στάδιο με εξαιρετικά λίγες επιστημονικές δημοσιεύσεις να εστιάζουν στο συγκεκριμένο θέμα [13].

Η εξόρυξη της συναισθηματικής πληροφορίας από ηχητικά δεδομένα, και δη γενικούς ήχους, είναι το αντικείμενο της συγκεκριμένης διδακτορικής διατριβής. Συγκεκριμένα, εξετάζεται η δυνατότητα ορθής αναγνώρισης συναισθημάτων από ηχητικά γεγονότα μέσω μηχανικής μάθησης. Στη συνέχεια, η δυνατότητα αυτή αξιολογείται βάσει της ευστοχίας του μοντέλου που παράγεται από τους αλγόριθμους κατηγοριοποίησης που χρησιμοποιήθηκαν. Το παρόν εισαγωγικό κεφάλαιο προσφέρει μία πρώτη παρουσίαση του υποβάθρου που οδήγησε στην ανάγκη για την έρευνα πάνω στο συγκεκριμένο αντικείμενο, μία πρώτη παρουσίαση του κυρίου ερευνητικού ερωτήματος καθώς και την οργάνωση του ανά χείρας τόμου.

1.1 Ήχος και συναίσθημα

Η μετάδοση των συναισθημάτων μέσω του ακουστικού καναλιού είναι ένας κοινός τρόπος για τους ειδικούς και μη-ειδικούς πάνω στο πεδίο του ήχου [14, 15, 16]. Το γεγονός αυτό μπορεί να παρατηρηθεί και στην ομιλία, όπου θεωρείται ότι ήταν το πρώτο ηχητικό ερέθισμα με το οποίο υλοποιήθηκε η ηχητική συναισθηματική αλληλεπίδραση. Αναφορικά με την μουσική και κατά τα πρώτα στάδια χρήσης αυτής, έρευνες δείχνουν ότι χρησιμοποιήθηκε για την επαύξηση της μετάδοσης συναισθημάτων από την ομιλία, συνοδεύοντας, μουσικά, το όποιο λεκτικό περιεχόμενο [14, 17]. Έπειτα, με την εξέλιξη της έρευνας πάνω στα πεδία που μελετούν τις φυσιολογικές (physiological) αντιδράσεις του ανθρώπου και της ψηφιακής επεξεργασίας σήματος, έγινε δυνατή η χρήση του συναισθήματος ως ένα παράγοντα για την επαύξηση της διάδρασης και εμπειρίας χρήσης σε πολλαπλά επίπεδα, προσφέροντας βελτιωμένη δυνατότητα αλληλεπίδρασης με μηχανισμούς και εφαρμογές, έχοντας ως παράμετρο τον συναισθηματικό χαρακτηρισμό του εκάστοτε περιεχομένου της ανάδρασης (feedback) για την αλληλεπίδραση/διάδραση ή του καθαυτού περιεχομένου, αντιστοίχως [2, 18, 19, 20]. Αυτή η παράμετρος μπορεί να χρησιμοποιηθεί για την γρήγορη και επικεντρωμένη στο περιεχόμενο αναζήτηση δεδομένων, για την υποβοήθηση και βελτίωση της εμπύθισης σε εικονικά περιβάλλοντα, για την αντιστοίχιση της συναισθηματικής πληροφορίας από το οπτικό στο ακουστικό κανάλι και εν γένει για την επαύξηση των δυνατοτήτων που προσφέρονται από οποιαδήποτε διεπαφή η οποία υποστηρίζει και ηχητική ανάδραση [3, 21, 22, 23].

Πάνω στον τομέα εξόρυξης πληροφορίας από ηχητικά δεδομένα έχουν αναπτυχθεί μερικά ερευνητικά πεδία όπως το πεδίο της ανάκτησης μουσικής πληροφορίας (Music Information Retrieval, MIR), που κυρίως ασχολείται με την εξόρυξη και ανάκτηση πληροφορίας από ηχητικά δεδομένα μουσικής, και το πεδίο της αναγνώρισης συναισθήματος από ήχο/ηχητικά δεδομένα (Audio Emotion Recognition, AER), το οποίο κυρίως εστιάζει στην αναγνώριση συναισθήματος από εν γένει ηχητικά δεδομένα [24, 25]. Επίσης, υπάρχει και ενεργό ενδιαφέρον από την ερευνητική κοινότητα που εστιάζει στην ψηφιακή επεξεργασία ομιλίας για την αναγνώριση συναισθήματος από ηχητικά δεδομένα ομιλίας [26, 27]. Από το πρώτο αναφερόμενο ερευνητικό πεδίο, ήτοι το MIR, έχουν αναφερθεί αρκετά υψηλά ποσοστά ευστοχίας στην αναγνώριση συναισθήματος από μουσική, τα οποία φτάνουν μέχρι και 85% [28, 29]. Μία από τις κύριες εφαρμογές που έχει επίσης προταθεί από το MIR είναι η χρήση του συναισθήματος ως χαρακτηριστικό κατηγοριοποίησης της μουσικής αντί

για το τυπικό “Είδος/Καλλιτέχνης/Άλμπουμ/Έτος” [3, 30]. Ακόμα, έχουν υλοποιηθεί μερικές ελεύθερα διαθέσιμες εφαρμογές, σε πρώιμο στάδιο όμως, οι οποίες προσφέρουν και την δυνατότητα χρήσης συναισθηματικής πληροφορίας σε δεδομένα μουσικής. Ένα τέτοιο παράδειγμα είναι η διαδικτυακή εφαρμογή allmusic², που προσφέρει μία κατηγοριοποίηση μουσικής αποκλειστικά και μόνο βάσει συναισθήματος. Επιπροσθέτως, έχουν αναφερθεί και ερευνητικές προσπάθειες για την αυτοματοποιημένη δημιουργία μουσικής η οποία στοχεύει στο να εγείρει συγκεκριμένα συναισθήματα στον ακροατή [31, 32, 33]. Αναφορικά με την αναγνώριση συναισθήματος από ομιλία, τα αποτελέσματα φτάνουν μέχρι και 92% [34, 35]. Επίσης, έχουν προταθεί εφαρμογές για τηλεφωνικά κέντρα για την ανίχνευση συναισθήματος ομιλητή σε τηλεφωνικές κλήσεις έτσι ώστε να λαμβάνονται μέτρα κατάλληλης προώθησης της κλήσης για την περαιτέρω διαχείρισή της [36]. Παρόλα αυτά, η έρευνα σχετικά με την αναγνώριση γενικών ήχων είναι ακόμα σε αρχικά στάδια [13], αφήνοντας έτσι ακόμα αναπάντητο το ερώτημα για την δυνατότητα εξόρυξης συναισθηματικής πληροφορίας από ηχητικά γεγονότα.

1.2 Ερευνητικό ερώτημα και πιθανά οφέλη

Τόσο η ομιλία όσο και η μουσική είναι μορφές ηχητικών δεδομένων που, σε μεγάλο ποσοστό, η διάρθρωση και η δόμησή τους υπακούει και τηρεί κανόνες και πρότυπα. Παρόλα αυτά, το μεγαλύτερο ποσοστό των ηχητικών ερεθισμάτων που δέχεται ένα μέσος άνθρωπος δεν είναι ομιλία ούτε μουσική. Μέσω των τελευταίων επικοινωνεί με και επηρεάζεται από το περιβάλλον του σε καθημερινή και συχνότερη βάση από ό,τι με την ομιλία και την μουσική [3, 30].

Έχοντας σαν δεδομένο και κοινή γνώση ότι μέσω της ομιλίας και της μουσικής μπορούν να μεταφερθούν και να διεγερθούν συναισθήματα, προκύπτει το ερώτημα εάν το ίδιο ισχύει και για τους γενικούς ήχους. Καθημερινά ο άνθρωπος γίνεται δέκτης πλείστων γενικών ήχων που προέρχονται από διάφορες πηγές και δραστηριότητες, όπως ανθρώπινες πράξεις (π.χ. βήξιμο, φωνές, κλάμα), φύση (π.χ. φύσημα ανέμου, φωνές ζώων) και αλληλεπιδράσεις μεταξύ αντικειμένων (π.χ. τριβή, κρούση). Αυτοί μεταφέρουν πληροφορία χαμηλού και υψηλού επιπέδου, π.χ. σχετικά με τα χαρακτηριστικά και θέση της πηγής και σχετικά με σημα-

²<http://www.allmusic.com>

σιολογικό περιεχόμενο αντιστοίχως, και είναι δυνατόν να επηρεάσουν τις αποφάσεις και τα συναισθήματα του ακροατή [3]. Επίσης, ο άνθρωπος ως μέρος της φύσης είναι αρκετά πιθανό να έχει αναπτύξει πρότυπα ακοής/επεξεργασίας ήχων με τα οποία να αντιλαμβάνεται πιθανές απειλές από γειτονικές δράσεις, αλλά και σημαντικές πληροφορίες χωρίς να απαιτείται άμεσο και ελεύθερο μονοπάτι μεταξύ της πηγής και του δέκτη (όπως το τρέξιμο μεγάλων ζώων, το κλάμα ενός μικρού ζώου, πτώση μεγάλων σωμάτων κτλ). Έτσι, δημιουργείται εύλογα το ερώτημα κατά πόσο τα ηχητικά συμβάντα (ήτοι, οι γενικοί ήχοι) που αντιλαμβάνεται ο άνθρωπος να μπορούν να εγείρουν συγκεκριμένα συναισθήματα.

Πράγματι, πριν την έναρξη της παρούσας διδακτορικής διατριβής υπήρχαν ήδη ενδείξεις για την συσχέτιση συγκεκριμένων ηχητικών χαρακτηριστικών με μερικά συναισθήματα. Σε προγενέστερη εργασία [37] παρουσιάζεται μία μελέτη σχετικά με την επίρεια της θέσης της πηγής στον φόβο. Αποδεικνύεται πως όταν η ηχητική πηγή είναι όπισθεν του ακουστικού δέκτη ο τελευταίος νιώθει μεγαλύτερη διέγερση και φόβο. Στο παραπάνω πλαίσιο, το ερώτημα που προσπαθεί να απαντήσει η ανά χείρας διατριβή είναι το κατά πόσο υπάρχει εγγενής πληροφορία στο ηχητικό σήμα, προερχόμενο από γενικούς ήχους, η οποία μπορεί να αξιοποιηθεί για την αναγνώριση του συναισθήματος που θα προκληθεί στον ακροατή. Για τον σκοπό αυτό επικεντρώνεται στα χαρακτηριστικά χαμηλού επιπέδου, προσπαθώντας να αποσυσχετίσει το σημασιολογικό περιεχόμενο ενός γενικού ήχου με τα πρότυπα που ενδέχεται αυτός να έχει και τα οποία να γίνονται αντιληπτά από τον άνθρωπο για την δημιουργία συναισθημάτων στον τελευταίο.

Για παράδειγμα, είναι κοινός τόπος ότι ο ρυθμός τόσο της ομιλίας όσο και της μουσικής επηρεάζει την διέγερση που νιώθει ο ακροατής. Μία γρήγορη ομιλία ή ένα γρήγορο μουσικό κομμάτι είναι αρκετά πιθανό να διεγείρουν τον ακροατή, επηρεάζοντας έτσι την συναισθηματική του κατάσταση. Συμβαίνει το ίδιο και με ήχους που δεν έχουν κάποιο σημασιολογικό περιεχόμενο ή με ηχητικά ερεθίσματα που δεν έχουν κάποιους συγκεκριμένους κανόνες δόμησης. Εάν ναι, τότε αυτό αποτελεί καθαρή ένδειξη ότι ο ρυθμός σε ένα ηχητικό σήμα αποτελεί ένα χαρακτηριστικό με καθολική ισχύ της συναισθηματικής του επίρειας, ανεξάρτητο από τον τύπο του καθαυτού σήματος (ήτοι ομιλία, μουσική ή ηχητικό γεγονός), το οποίο είναι ικανό να επιδράσει πάνω στην συναισθηματική κατάσταση του ανθρώπου με συγκεκριμένο και προβλέψιμο τρόπο. Ένα γεγονός που μπορεί να οδηγήσει στην εκμετάλλευση του χαρακτηριστικού αυτού για την επαύξηση τόσο του καλλιτεχνικού αποτελέσματος, π.χ.

στοχευμένα ρυθμικά χαρακτηριστικά στην ομιλία στο θέατρο, σε μουσικά ιδιώματα ή σε ηχητικά περιβάλλοντα που υπάρχουν σε εικονικούς και μη κόσμους με στόχο την στοχευμένη συναισθηματική διέγερση του ακροατή/θεατή, όσο και στην αξιολόγηση υφιστάμενων ηχητικών περιβαλλόντων και την επίρεια αυτών στον άνθρωπο, π.χ. ηχητικό περιβάλλον αστικής περιοχής. Όπως ήδη αναφέρθηκε, έχειδειχθεί πως όταν ένα ηχητικό ερέθισμα βρίσκεται όπισθεν του ακροατή του προκαλεί αυξημένη διέγερση και φόβο. Είναι μόνο η συγκεκριμένη θέση, δηλαδή ακριβώς πίσω από τον ακροατή, που έχει μία επίρεια στην συναισθηματική κατάσταση ή υπάρχει μεταβολή της τελευταίας αναλόγως την θέση της πηγής. Συμβαίνει το ίδιο με διάφορα ηχητικά γεγονότα ή μόνο με μία ομάδα εξ αυτών. Μία απάντηση στα τελευταία δύο ερωτήματα μπορεί να αποκαλύψει έναν ακόμα καθολικό παράγοντα συναισθηματικής επίρειας του ανθρώπου μέσω ηχητικών ερεθισμάτων. Όλες οι πηγές ήχου έχουν μία φυσική υπόσταση και, ως εκ τούτου, και μία χωρική τοποθέτηση σε σχέση με τον δέκτη.

Στον ανά χείρας τόμο δίνονται απαντήσεις στα παραπάνω. Παρουσιάζονται η ανάλυση και αξιολόγηση της συναισθηματικής επίρειας των χαρακτηριστικών χαμηλού επιπέδου των ηχητικών συμβάντων. Έτσι, αφενός γίνεται εφικτή η συναισθηματική περιγραφή και επισήμανση μη λεκτικών και μη μουσικών ήχων με την χρήση συναισθηματικής περιγραφής και αφετέρου αναδεικνύονται τεχνικά χαρακτηριστικά που ενυπάρχουν σε όλα τα ηχητικά ερεθίσματα και δύνανται να επηρεάσουν συναισθηματικά τον ακροατή. Ένα γεγονός που ενδέχεται να προσφέρει άρδην στην αξιοποίηση, επισήμανση και διαχείριση εν γένει της ηχητικής πληροφορίας και ειδικότερα μέσα από την στόχευση που πραγματοποιείται σε αυτή την διατριβή. Δηλαδή τα ρυθμικά χαρακτηριστικά και την χωρική τοποθέτησή των ηχητικών συμβάντων σε σχέση με τον δέκτη τους. Η ανάλυση και η αξιολόγηση της συναισθηματικής συσχέτισης και επίρειας των ηχητικών γεγονότων, καθώς και των τεχνικών χαρακτηριστικών αυτών, αναφορικά με τον άνθρωπο στοιχειοθετούν την πρωτοτυπία της ανά χείρας διατριβής και επιχειρούν να συμβάλλουν στην προαγωγή της έρευνας και της γνώσης σχετικά με την συνολική επίρεια και αλληλεπίδραση του ανθρώπου από και με τους γενικούς ήχους αλλά και τις ανάδειξης τεχνικών χαρακτηριστικών που έχουν καθολική ισχύ αναφορικά με την επίρειά τους στην συναισθηματική κατάσταση. Η προαναφερθείσα πρωτοτυπία θα αναλυθεί διεξοδικά στα επόμενα κεφάλαια της παρούσας διατριβής.

Ακόμα, τα παραγόμενα ερευνητικά αποτελέσματα εκτιμάται ότι μπορούν να βοηθήσουν στην αξιολόγηση του εν γένει ακουστικού περιβάλλοντος του ανθρώπου, προσφέροντας την δυ-

νατότητα για την συσχέτιση της συναισθηματικής κατάστασης και του ηχητικού περιβάλλοντος. Κάτι τέτοιο θα μπορούσε να επαυξήσει τόσο τις αρχιτεκτονικές δυνατότητες στον σχεδιασμό χώρων (π.χ. λαμβάνοντας υπόψη δομικά και γεωμετρικά χαρακτηριστικά του χώρου που θα μπορούσαν να επιδράσουν σε συγκεκριμένα τεχνικά χαρακτηριστικά του ηχητικού περιβάλλοντος), αλλά και τον καλλιτεχνικό σχεδιασμό ήχο-τοπίων, έτσι ώστε να εγείρονται συγκεκριμένα συναισθήματα στον ακροατή αναλόγως την τοποθεσία και εν γένει συγκυρία για την αναπαραγωγή του ήχο-τοπίου. Επίσης, τα αποτελέσματα θα μπορούσαν να συμβάλουν στην ανθρωποκεντρική κατηγοριοποίηση των γενικών ήχων προσφέροντας έναν επιπλέον ανθρωποκεντρικό παράγοντα κατηγοριοποίησης γενικών ήχων. Τέλος, τα αποτελέσματα της συγκεκριμένης διατριβής θα μπορούσαν να αξιοποιηθούν και σε μία περαιτέρω έρευνα σχετικά με την αυτοματοποιημένη σύνθεση ηχητικών γεγονότων και εν γένει ηχητικών περιβαλλόντων, προσφέροντας το υπόβαθρο για την ποσοτική συσχέτιση των τεχνικών χαρακτηριστικών των ηχητικών γεγονότων με τα εγείρομενα συναισθήματα στους ακροατές. Έτσι, μία μηχανή αυτόματης σύνθεσης θα μπορούσε να συνθέτει γενικούς ήχους με συγκεκριμένα τεχνικά χαρακτηριστικά με σκοπό την πρόκληση συγκεκριμένων συναισθημάτων.

1.3 Δομή τόμου

Η παρούσα διατριβή είναι οργανωμένη σε κεφάλαια. Πιο συγκεκριμένα, η οργάνωση του υπόλοιπου εγγράφου είναι η εξής: στο Κεφάλαιο 2 παρουσιάζεται μία δευτερογενής και βιβλιογραφική έρευνα σχετικά με το θεωρητικό υπόβαθρο που αφορά στην παρούσα διατριβή και με την υφιστάμενη στάθμη της τεχνικής στο πεδίο της αναγνώρισης συναισθημάτων από ηχητικά δεδομένα. Αναλυτικότερα, γίνεται αναφορά και παρουσίαση της διαδικασίας εξόρυξης συναισθηματικής πληροφορίας από ηχητικά δεδομένα, όπως αυτή μπορεί να αποτυπωθεί από τις έως τώρα δημοσιευμένες επιστημονικές εργασίες πάνω στο πεδίο της αναγνώρισης συναισθημάτων. Παράλληλα, παρουσιάζονται τα μοντέλα συναισθημάτων σύμφωνα με τους ορισμούς από τα συναφή επιστημονικά πεδία, καταδεικνύονται και αναλύονται οι υφιστάμενες τεχνικές επισήμανσης ηχητικών δεδομένων και παρουσιάζονται τα τεχνικά μεγέθη και οι αλγόριθμοι που χρησιμοποιούνται στην διαδικασία εξόρυξης συναισθηματικής πληροφορίας από ηχητικά γεγονότα καθώς και τα συνήθως χρησιμοποιούμενα τεχνικά μεγέθη και αλγόριθμοι

κατηγοριοποίησης.

Στο Κεφάλαιο 3 αναλύεται η πρωτότυπη προσέγγιση που υιοθετήθηκε για την εξόρυξη της συναισθηματικής πληροφορίας από γενικούς ήχους κατά την ολοκλήρωση της παρούσας διατριβής. Συγκεκριμένα, παρουσιάζονται τα επιμέρους ερευνητικά ερωτήματα και διαδικασίες που ακολουθήθηκαν προς την ολοκλήρωση αυτής. Περιλαμβάνονται η παρουσίαση των σετ ηχητικών δεδομένων που χρησιμοποιήθηκαν, οι τεχνικές επεξεργασίες που εφαρμόστηκαν στα ηχητικά δεδομένα, καθώς και η πορεία υλοποίησης έκαστου πειράματος στην πορεία της συνολικής ερευνητικής διαδικασίας. Στο Κεφάλαιο 4 παρουσιάζονται τα αποτελέσματα που παρήχθησαν από την έρευνα και συγκεκριμένα από τις διαδικασίες που περιγράφονται στο Κεφάλαιο 3. Αυτά περιλαμβάνουν τόσο τα ποσοτικά αποτελέσματα ορθής αναγνώρισης συναισθημάτων από ηχητικά δεδομένα όσο και ποιοτικά, όπως η επαύξηση υφιστάμενων πλαισίων εργασίας.

Τα αποτελέσματα συζητούνται και αναλύονται στο Κεφάλαιο 5 και στο Κεφάλαιο 6 εξάγονται συμπεράσματα αναφορικά με την διαδικασία εξόρυξης συναισθηματικής πληροφορίας από γενικούς ήχους. Η διατριβή καταλήγει με το Κεφάλαιο ;; όπου παρουσιάζετε μία σύνοψη των συμπερασμάτων που εξήχθησαν από τα αποτελέσματα, η συζήτηση τους και ο επίλογος της διατριβής. Επίσης, για μεγαλύτερη ευκολία ανάγνωσης, πριν από τον πίνακα περιεχομένων παρατίθενται λίστες με τους πίνακες και τα σχήματα που υπάρχουν στη διατριβή.

ΚΕΦΑΛΑΙΟ 2

ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

Τα ηχητικά γεγονότα (sound events) μπορούν να θεωρηθούν ως μία κατηγορία ήχου, της οποίας το περιεχόμενο δεν είναι ομιλία (non-verbal), δεν είναι μουσική (non-musical) και δεν διέπεται από κάποια οργάνωση. Πιο συγκεκριμένα, ως ηχητικό γεγονός μπορεί να οριστεί ο ήχος από τον οποίο ο δέκτης του, ήτοι ο άνθρωπος, αντιλαμβάνεται όχι τα καθαυτά ηχητικά χαρακτηριστικά (π.χ. ύψος συχνότητας, ρυθμός κτλ) αλλά την πηγή και το γεγονός που τον δημιούργησε [5]. Για παράδειγμα, κατά την ακρόαση μουσικής ο άνθρωπος προσλαμβάνει το τονικό ύψος (π.χ. υψηλές νότες), τον ρυθμό και άλλα καθαυτά τεχνικά χαρακτηριστικά. Όμως, σε συνθήκες μη ακρόασης μουσικής, ο άνθρωπος αντιλαμβάνεται τον ήχο ως προϊόν της πηγής του. Δηλαδή, ο κρότος σπασίματος ποτηριού γίνεται αντιληπτός ως ο ήχος ενός ποτηριού που σπάει [5]. Έτσι, βάσει της γενικότητάς τους, τα ηχητικά γεγονότα μπορούν να καταταγούν ως ένα υπερσύνολο των άλλων κατηγοριών ηχητικών δεδομένων, οι οποίες είναι η μουσική και η ομιλία [3]. Για τις τελευταίες έχουν δημοσιευθεί πλείστες έρευνες οι οποίες εστιάζουν στους μηχανισμούς και τους τρόπους μεταφοράς ή/και πρόκλησης συναισθήματος στον ακροατή [5]. Όμως, η έρευνα για την μεταφορά ή/και πρόκληση συναισθημάτων από ηχητικά γεγονότα είναι ένα αρκετά πρόσφατο πεδίο για το οποίο αναφέρεται ότι οι έρευνες είναι σπάνιες και όχι τόσο ισχυρά συνδεδεμένες, κυρίως λόγω της μη ύπαρξης ισχυρής

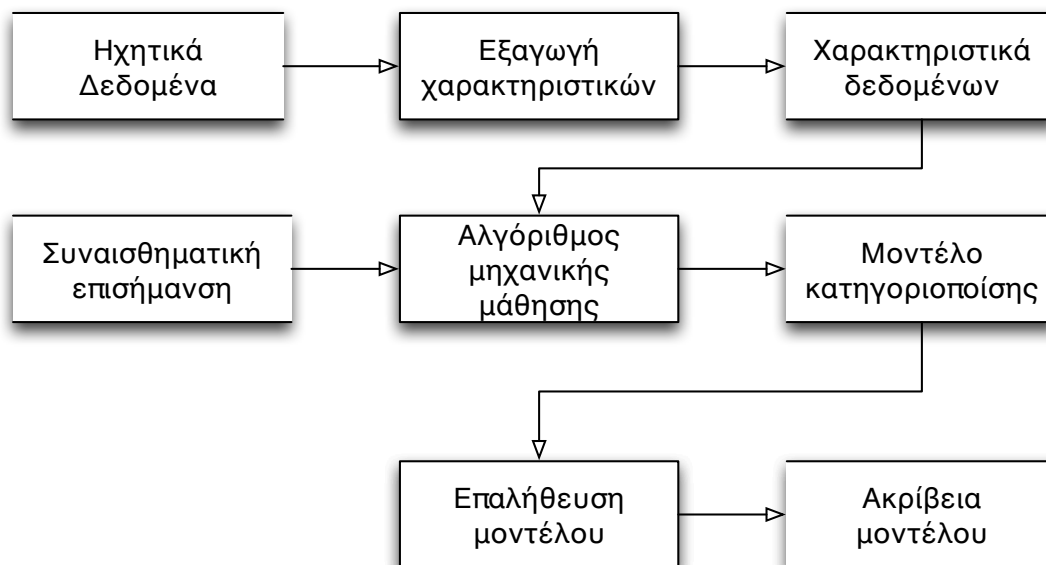
σύνδεσης και συνοχής των διαφορετικών αποτελεσμάτων μεταξύ των ερευνών [13].

Η παρούσα διατριβή στοχεύει ακριβώς στην κάλυψη του παραπάνω κενού. Έτσι, πέρα από την ενασχόλησή της με άλλες επιστήμες, αναγκαίες για την σφαιρική κατανόηση της μεταφοράς και πρόκλησης συναισθημάτων από ηχητικά ερεθίσματα (όπως η Ψυχολογία σχετικά με το συναίσθημα και τα συναισθηματικά μοντέλα, η Ιατρική σχετικά με τους μηχανισμούς βίωσης συναισθημάτων και η Μηχανική Μάθηση σχετικά με την δημιουργία μοντέλων κατηγοριοποίησης), αναγκαία εμβάθυνση και προσοχή έπρεπε να δοθεί στις υφιστάμενες έρευνες για την ανίχνευση συναισθήματος από ομιλία και μουσική. Λόγω της εγγενούς ομοιότητας αλλά και της μη χρήσης σημασιολογικού και λεκτικού περιεχομένου από την μουσική [14], υπήρξε περισσότερη εστίαση στην τελευταία.

Έτσι, σε αυτό το Κεφάλαιο θα παρουσιαστεί το ερευνητικό υπόβαθρο στο οποίο στηρίχθηκε η παρούσα διδακτορική διατριβή καθώς και η στάθμη της τεχνικής πάνω στον τομέα του σχετικού πεδίου έρευνας κατά την στιγμή της συγγραφής της διατριβής αυτής. Το πρώτο περιλαμβάνει την παρουσίαση της γενικής διαδικασίας αναγνώρισης συναισθήματος και των στοιχείων αυτής, την αναφορά στα συναισθήματα, τα μοντέλα αυτών και τους τρόπους μέτρησής τους από ηχητικά ερεθίσματα και τα συνήθως χρησιμοποιούμενα μεγέθη ήχου στην διαδικασία αναγνώρισης συναισθημάτων από ηχητικά δεδομένα. Το δεύτερο αφορά σε σύγχρονα ευρήματα και αποτελέσματα των πεδίων της αναγνώρισης συναισθημάτων από ομιλία, μουσική και ηχητικά γεγονότα.

2.1 Ερευνητικό υπόβαθρο

Η μεταφορά συναισθημάτων από μουσική είναι μία αντίληψη που ενυπάρχει τόσο σε ειδήμονες του χώρου της μουσικής ή/και των συναισθημάτων, όσο και σε ανθρώπους χωρίς εξειδίκευση σε σχετικά ερευνητικά πεδία [38]. Είναι επίσης και μία διαπίστωση βάσει εμπειρικών παρατηρήσεων, που είναι πολύ πιθανό να έχει κάνει ο μέσος άνθρωπος. Για την ακρίβεια, αναφέρεται ότι εν αρχή η μουσική χρησιμοποιήθηκε ως μέσο για την επαύξηση της μετάδοσης συναισθήματος από την ομιλία [14]. Έχουν δημοσιευθεί πλείστες έρευνες σχετικά με την σχέση μουσικής και συναισθημάτων, οι οποίες βασίζονται τόσο σε μουσικά, όσο και ψυχολογικά/συναισθηματικά δεδομένα [13].



Σχήμα 2.1: Η γενική διαδικασία αναγνώρισης συναισθημάτων από ηχητικά δεδομένα

Εν γένει, η αναγνώριση συναισθήματος από ηχητικά δεδομένα μπορεί να θεωρηθεί ως μία εφαρμογή Μηχανικής Μάθησης με στόχο την παραγωγή ενός μοντέλου αναγνώρισης συναισθήματος από τα δεδομένα ήχου. Σε αυτή την εφαρμογή χρησιμοποιείται ένα σετ συναισθηματικά επισημασμένων δεδομένων, μία ομάδα χαρακτηριστικών που σχετίζονται με τα δεδομένα αυτά και ένας αλγόριθμος κατηγοριοποίησης. Ο τελευταίος έχει σαν είσοδο τα συναισθηματικά επισημασμένα ηχητικά δεδομένα και τα τεχνικά χαρακτηριστικά και παράγει το μοντέλο κατηγοριοποίησης όπως φαίνεται στο **Σχήμα 2.1** [30]. Αν και σε αυτό το Σχήμα απεικονίζεται σε ένα αφαιρετικό επίπεδο η διαδικασία αναγνώρισης συναισθημάτων, μετέπειτα στο παρόν έγγραφο θα παρουσιαστεί μία πιο αναλυτική περιγραφή της εν λόγω διαδικασίας.

Όπως μπορεί να φανεί από το **Σχήμα 2.1**, σημαντικά στοιχεία στην υλοποίηση της προαναφερόμενης διεργασίας είναι:

1. Το συναισθηματικό μοντέλο που χρησιμοποιήθηκε για την αποτύπωση των συναισθημάτων, το οποίο επιδρά και σχετίζεται άμεσα με τις συναισθηματικές επισημάνσεις.
2. Η μέθοδος εκμαίευσης της συναισθηματικής κατάστασης για την συγκρότηση της συναισθηματικής επισήμανσης.
3. Τα χαρακτηριστικά που εξήχθησαν από τα ηχητικά δεδομένα και χρησιμοποιήθηκαν

ως είσοδος στον αλγόριθμο μηχανικής μάθησης.

4. Ο αλγόριθμος μηχανικής μάθησης που χρησιμοποιήθηκε για την κατηγοριοποίηση και την παραγωγή του μοντέλου.

Στις παρακάτω υπό-ενότητες αρχικά γίνεται μία παρουσίαση του πεδίου της Μηχανικής Μάθησης, εν συνέχεια παρουσιάζονται αναλυτικά τα προαναφερθέντα σημαντικά στοιχεία και τέλος στοιχειοθετούνται οι συγκεκριμένες επιλογές στην έρευνα που εκπονήθηκε στο πλαίσιο της παρούσας διατριβής. Πιο αναλυτικά, στην Ενότητα 2.1.1 παρουσιάζεται το πεδίο της Μηχανικής Μάθησης και βασικά στοιχεία αυτού. Έπειτα, στην Ενότητα 2.1.2 γίνεται αναφορά σε και αναλύονται οι αλγόριθμοι κατηγοριοποίησης που χρησιμοποιούνται σε εφαρμογές μηχανικής μάθησης εν γένει αλλά και σε αναγνώριση συναισθήματος συγκεκριμένα. Στην Ενότητα 2.1.3 είναι μία προσέγγιση του ορισμού του συναισθήματος ενώ στις Ενότητες 2.1.4 έως 2.1.6 παρουσιάζονται τα διακριτά και συνεχή μοντέλα συναισθημάτων ακολουθούμενα από τις διαδικασίες αποτύπωσης αυτών. Τέλος, στην Ενότητα 2.1.7 παρατίθενται τα τεχνικά μεγέθη ήχου που χρησιμοποιούνται στην εξόρυξη συναισθηματικής πληροφορίας από ηχητικά δεδομένα.

2.1.1 Μηχανική Μάθηση

Η Μηχανική Μάθηση είναι ένα επιστημονικό πεδίο το οποίο άρχισε να αναπτύσσεται από την δεκαετία του 1950, σχεδόν ταυτόχρονα με την εμφάνιση των ηλεκτρονικών υπολογιστών [39], και έχει τις βάσεις του στην επιστήμη των υπολογιστών [40]. Ο στόχος του είναι η χρήση και δημιουργία αλγόριθμων και διαδικασιών ικανών να παράγουν ή να βελτιστοποιήσουν υφιστάμενη γνώση βάσει δεδομένων που δέχονται ως είσοδο [39]. Πιο συγκεκριμένα και χρησιμοποιώντας το τυπικό και απλοποιημένο παράδειγμα της πρόβλεψης του καιρού, ας υποθεθεί ότι για μία συγκεκριμένη γεωγραφική περιοχή έχουν καταγραφεί τα εξής δεδομένα: α) εμφάνιση βροχής (ναι ή όχι), β) θερμοκρασία (σε °C), γ) μήνας του χρόνου (ονομαστικά), δ) ποσοστό υγρασίας (%) και ε) συννεφιά (καθόλου σύννεφα, λίγα, μέτρια και πολύ). Με την χρήση της Μηχανικής Μάθησης και έχοντας μία νέα πλειάδα δεδομένων που περιέχει καταχωρήσεις για τα προαναφερθέντα στοιχεία εκτός του α), κανείς θα μπορούσε να μάθει εάν (π.χ.) χρειάζεται να πάρει ομπρέλα (εμφάνιση βροχής) για την συγκεκριμένη μέρα όπου αναφέρεται η νέα πλειάδα. Για να επιτευχθεί το παραπάνω, γνώση για την χρήση ή όχι

ομπρέλας, απαιτείται να εκπαιδευτεί ένα σύστημα, να παραχθεί ένα μοντέλο και να γίνει η ερώτηση/παραγωγή γνώσης με νέα τα δεδομένα για την ημέρα ενδιαφέροντος.

Στο παραπάνω παράδειγμα οι αρχικές πλειάδες, όμως χωρίς να περιλαμβάνουν το στοιχείο α), καλούνται και παραδείγματα/περιπτώσεις (instances) και είναι αυτά τα οποία θα χρησιμοποιηθούν μαζί με το στοιχείο α) για να εκπαιδευθεί το σύστημα. Το στοιχείο α), δηλαδή η εμφάνιση ή όχι βροχής, χαρακτηρίζει ένα instance και καλείται και κλάση διότι αφενός είναι το στοιχείο με βάσει το οποίο θα εκπαιδευθεί το σύστημα και αφετέρου η μετέπειτα παραγωγή γνώσης θα εστιαστεί πάνω σε αυτό, δηλαδή στο εάν θα υπάρξει ή όχι εμφάνιση βροχής (ανάγκη για την χρήση ομπρέλας) δεδομένης μίας νέας πλειάδας/instance στοιχείων. Η εκπαίδευση του συστήματος γίνεται με την χρήση αλγορίθμων κατηγοριοποίησης οι οποίοι έχουν ως είσοδο τα instances και την κλάση για κάθε instance και ως έξοδο ένα μοντέλο με το οποίο μπορεί να γίνει πρόβλεψη για την κλάση ενός νέου instance. Η αξιολόγηση του συστήματος/παραγόμενης γνώσης και, συνεπώς, του μοντέλου που δημιουργήθηκε γίνεται με την χρήση instances τα οποία ενώ είναι γνωστή η κλάση τους δεν χρησιμοποιείται για την κατηγοριοποίησή τους παρά μόνο για την αξιολόγηση της κατηγοριοποίησης που έκανε το σύστημα. Το κατά πόσο το σύστημα μπορεί να προβλέψει σωστά την κλάση για μία νέα πλειάδα είναι η ευστοχία του και εξαρτάται τόσο από το πλήθος των παραδειγμάτων που χρησιμοποιήθηκαν κατά την εκπαίδευση όσο και από την σχέση που έχουν τα μεγέθη που αναπαριστώνται από τα instances με το μέγεθος/γνώση που αναπαριστά η κλάση. Δηλαδή, στο παραπάνω παράδειγμα και βάσει εμπειρικής γνώσης μπορεί να φανεί ότι τα μεγέθη που αναπαριστώνται από τα instances σχετίζονται με την κλάση. Όμως, θα μπορούσε κανείς να χρησιμοποιήσει ως μεγέθη στα instances, π.χ., και το τι χρώμα ρούχα φοράει, την τιμή μίας μετοχής κτλ. Σε μία τέτοια περίπτωση η ευστοχία του μοντέλου θα ήταν αμφίβολη.

Πλέον η Μηχανική Μάθηση χρησιμοποιείται σε πλείστες εφαρμογές με εύρος από την καθημερινή ζωή μέχρι και εξεζητημένα πεδία. Για παράδειγμα, η πολύ γνωστή λειτουργία των διαδικτυακών μηχανών αναζήτησης (π.χ. Google³) που μπορούν να καταλάβουν τι εννοούσαμε και τι θέλαμε να γράψουμε από μία σειρά λάθος λέξεων στο πεδίο αναζήτησης, είναι εφαρμογή Μηχανικής Μάθησης. Ακόμα, και εστιάζοντας σε συναφείς εφαρμογές με το πεδίο της παρούσας διατριβής, η λειτουργία που προσφέρουν κάποια λογισμικά ή υπηρεσίες αναπα-

³<http://google.com>

ραγωγής πολυμέσων/media players (π.χ iTunes⁴ και Spotify⁵) με την οποία το λογισμικό ή η υπηρεσία μπορεί να προτείνει στον χρήστη επόμενα τραγούδια που ο ίδιος ο χρήστης να θεωρεί ότι ταιριάζουν αισθητικά με αυτό ή αυτά που ήδη άκουγε, είναι εφαρμογή Μηχανικής Μάθησης. Τέλος, η ίδια η αναγνώριση συναισθήματος από ηχητικά δεδομένα είναι εφαρμογή μηχανικής μάθησης. Ως κλάση είναι το συναίσθημα και κάθε instance περιέχει τα χαρακτηριστικά στα οποία γίνεται εστίαση και μελέτη για την αναγνώριση συναισθήματος, π.χ. τεχνικά χαρακτηριστικά.

2.1.2 Αλγόριθμοι κατηγοριοποίησης

Όπως μπορεί να φανεί από την σύντομη παρουσίαση του πεδίου της Μηχανικής Μάθησης αλλά και το **Σχήμα 2.1**, ένα από τα βασικότερα στοιχεία στην αναγνώριση συναισθημάτων από ηχητικά γεγονότα, αλλά και οποιασδήποτε διαδικασίας Μηχανικής Μάθησης, είναι οι αλγόριθμοι που χρησιμοποιούνται για την κατηγοριοποίηση. Δηλαδή, οι αλγόριθμοι οι οποίοι δέχονται ως είσοδο τις κλάσεις και τα instances και προσφέρουν ως έξοδο ένα μοντέλο κατηγοριοποίησης ικανό να προσδιορίσει την κλάση για κάποιο άλλο και νέο instance.

Αν και στο απλό και τυπικό παράδειγμα της προηγούμενης ενότητας, 2.1.1, η κλάση κάθε instance ήταν γνωστή κατά την διάρκεια της εκπαίδευσης του συστήματος, υπάρχουν περιπτώσεις που αυτό δεν είναι δυνατό. Δηλαδή, ενώ υπάρχουν instances δεν είναι γνωστή η κλάση τους, δηλαδή δεν υπάρχει η γνώση εάν αυτά τα δεδομένα ανήκουν σε κάποια κατηγορία/κλάση, μπορούν να διαχωριστούν σε διαφορετικές κατηγορίες, πως μπορεί να γίνει ο διαχωρισμός ή/και πόσες είναι οι κατηγορίες που μπορούν να δημιουργηθούν. Στον αντίποδα είναι τα δεδομένα/instances για τα οποία είναι γνωστή η κλάση τους, όπως στο παραπάνω παράδειγμα. Σε κάθε περίπτωση ο στόχος σε μία διαδικασία Μηχανικής Μάθησης είναι κοινός. Δηλαδή, να μπορέσει να κατασκευαστεί ένα μοντέλο πρόβλεψης των κλάσεων για μελλοντικά και αγνώστου κλάσης instances. Ακολουθώντας τα προαναφερθέντα σενάρια γνώσης κατηγοριών/κλάσεων για τα διαθέσιμα δεδομένα, οι αλγόριθμοι κατηγοριοποίησης μπορούν να διαχωριστούν σε αντίστοιχες κατηγορίες [41]. Έτσι, υπάρχουν αλγόριθμοι που μπορούν να δημιουργήσουν μοντέλα κατηγοριοποίησης χωρίς να έχουν την κλάση για τα δεδομένα που δέχονται στην είσοδό τους. Επειδή δεν υπάρχει η πρότερη γνώση για την κλάση,

⁴<http://www.apple.com/itunes/>

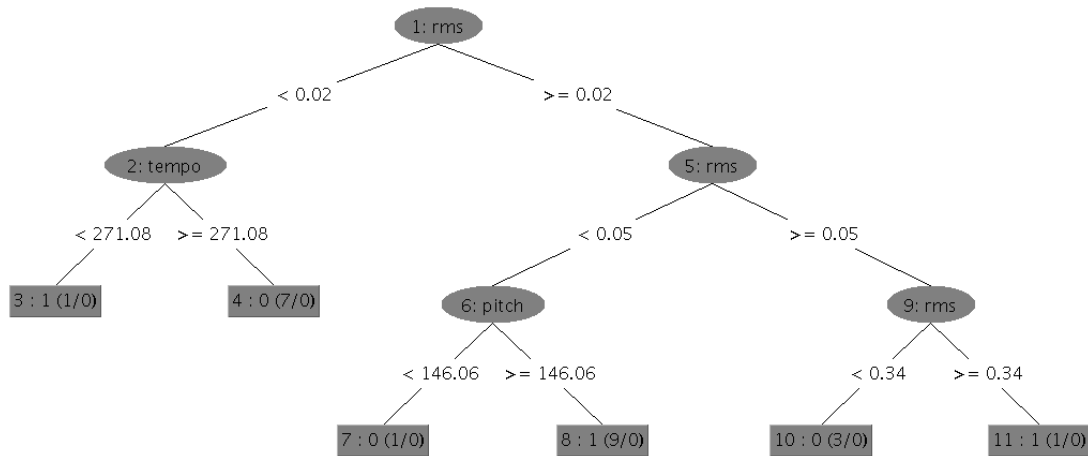
⁵<http://www.spotify.com/>

οι αλγόριθμοι αυτοί ονομάζονται μη-επιβλεπόμενοι (unsupervised). Σε αντίθεση, υπάρχουν εκείνοι οι αλγόριθμοι που μπορούν να παράγουν μοντέλο κατηγοριοποίησης χρησιμοποιώντας την κλάση για κάθε instance (εάν και εφόσον αυτή είναι γνωστή) και λέγονται επιβλεπόμενοι (supervised).

Έτσι, εάν X_1 είναι ένα instance με δεδομένα $\{x_{1,1}, x_{1,2}, x_{1,3}, \dots, x_{1,\nu}\}$ και y_1 η κλάση του X_1 , το πρώτο είδος αλγόριθμων, δηλαδή οι unsupervised, δέχονται σαν είσοδο μόνο το διάνυσμα X_1 και προσπαθούν να βρουν κάποια οργάνωση στα δεδομένα του X_1 και των υπόλοιπων διαθέσιμων instances. Η διαδικασία αυτή θεωρείται και είναι εύρεση και αναγνώριση προτύπων. Επίσης, οι unsupervised αλγόριθμοι χρησιμοποιούνται συνήθως σε προβλήματα δημιουργίας ομαδοποίησης (clustering) και μείωσης διαστάσεων (dimensionality reduction) [42]. Η ομαδοποίηση είναι η διαδικασία κατά την οποία ένας unsupervised αλγόριθμος ομαδοποιεί τα instances που δέχεται στην είσοδό του σε τέτοιες και τόσες ομάδες έτσι ώστε κάθε ομάδα να μπορεί να χαρακτηρίζεται από συγκεκριμένες τιμές ή συνδυασμούς τιμών των δεδομένων των instances. Η μείωση διαστάσεων είναι η διαδικασία κατά την οποία, δοθέντος ενός πλήθους instances με ν δεδομένα, ένας unsupervised αλγόριθμος μετασχηματίζει τα δεδομένα των instances έτσι ώστε αυτά να περιέχουν $\mu \leq \nu$ δεδομένα και χωρίς να γίνεται απώλεια πληροφορίας από την μείωση των δεδομένων. Δηλαδή, εάν X_1 το παραπάνω instance, μετά από την μείωση διαστάσεων αυτό θα περιέχει $\{x_{1,1}, x_{1,2}, \dots, x_{1,\mu}\}$ δεδομένα χωρίς όμως να έχει χαθεί πληροφορία που υπήρχε στο αρχικό σετ με πλήθος δεδομένων ν . Η τελευταία διαδικασία μπορεί να θεωρηθεί αρκετά χρήσιμη όταν στο σετ των τιμών (instances) υπάρχουν δεδομένα τα οποία έχουν μεγάλη συσχέτιση μεταξύ τους και, στην ουσία, περιγράφουν το ίδιο φυσικό μέγεθος ή διαδικασία. Τυπικά παραδείγματα αλγορίθμων αυτής της κατηγορίας αποτελούν οι: α') Ανάλυση πρωτευόντων στοιχείων (Principal Components Analysis), β') K κοντινότεροι γείτονες (K Nearest Neighbors, KNN), γ') αλγόριθμος Προσδοκίας-Μεγιστοποίησης (Expectation-Maximization algorithm), και δ') Παραγοντοποίηση μη-αρνητικής μήτρας (Non-negative Matrix Factorization) [42].

Από την άλλη, οι supervised αλγόριθμοι έχουν σαν είσοδο εκτός του X_1 και το y_1 . Δηλαδή, για κάθε instance έχουν και την κλάση στην οποία ανήκει αυτό και έτσι προσπαθούν, δεδομένων των στοιχείων των instances, να κατασκευάσουν ένα μοντέλο το οποίο για συγκεκριμένο instance να μπορεί να μαντέψει την κλάση στην οποία ανήκει [43]. Πιο συγκεκριμένα, οι supervised αλγόριθμοι προσπαθούν να βρουν μία συσχέτιση των τιμών των instances έτσι

ώστε συγκεκριμένες τιμές των δεδομένων να σηματοδοτούν την ύπαρξη συγκεκριμένης κλάσης. Ο τρόπος που επιτυγχάνεται το παραπάνω μπορεί να θεωρηθεί ότι είναι η δημιουργία μίας εξίσωσης η οποία για συγκεκριμένες τιμές των μεταβλητών της (δηλαδή, των τιμών για τα δεδομένα των instances) να έχει ως αποτέλεσμα την κλάση στην οποία ανήκει κάθε instance. Όμως, αυτό διαφέρει ελαφρώς ανάμεσα στους supervised αλγόριθμους. Έτσι, υπάρχουν εκείνοι που δημιουργούν ένα δένδρο απόφασης όπου κάθε κόμβος του είναι μία σύγκριση τιμών/απόφαση και κάθε φύλλο είναι μία κλάση. Π.χ., όπως φαίνεται στο **Σχήμα 2.2** όπου υπάρχουν τρία δεδομένα στα instances. Δηλαδή ένα instance εισόδου θα είναι της μορφής $\{x_1, x_2, x_3\}$ με x_1 να έχει ονομαστεί και να αντιστοιχεί στο rms, x_2 στο tempo και x_3 στο pitch.



Σχήμα 2.2: Το παραγόμενο μοντέλο από έναν supervised αλγόριθμο (Random Tree) που χρησιμοποιεί δένδρα απόφασης

Αναλόγως τις τιμές αυτών ακολουθείται μία συγκεκριμένη διαδρομή στο δένδρο απόφασης και βρίσκεται η κλάση που ανήκει το συγκεκριμένο instance. Η κλάση εικονίζεται στα φύλλα αμέσως μετά την άνω-κάτω τελεία. Δηλαδή, στο πρώτο από αριστερά προς τα δεξιά φύλλο υπάρχει η κλάση 1 και όπως μπορεί να φανεί από το εν λόγω Σχήμα, υπάρχουν δύο κλάσεις: η 1 και η 0. Έτσι, από την ρίζα του δένδρου ελέγχεται εάν η τιμή του δεδομένου που αντιστοιχεί στην rms τιμή είναι μικρότερη από 0.02 ή όχι. Εάν ναι, τότε το μοντέλο ελέγχει την τιμή του αριστερού κόμβου στο επόμενο επίπεδο, δηλαδή την τιμή του δεδομένου που αντιστοιχεί στο tempo. Εάν η τιμή αυτή είναι μικρότερη από 271.08 τότε το μοντέλο καταλήγει ότι το instance εισόδου ανήκει στην κλάση 1. Αν και το μοντέλο στο **Σχήμα 2.2** δεν έχει καλή ευστοχία (ίση με 54%), εντούτοις επιλέχθηκε ως παράδειγμα οπτικοποίησης της εξόδου ενός αλγόριθμου με δένδρο απόφασης ο οποίος έχει ως είσοδο instances με

λίγα δεδομένα. Ο αλγόριθμος που χρησιμοποιήθηκε είναι ο Random Tree [44] μέσα από το περιβάλλον εκτέλεσης πειραμάτων Μηχανικής Μάθησης WEKA [45]. Οι αλγόριθμοι αυτοί έχουν χρησιμοποιηθεί, μεταξύ άλλων, και σε περιπτώσεις αναγνώρισης εικόνας [46] και αναγνώρισης ομιλίας [47].

Πέρα από τους αλγόριθμους που χρησιμοποιούν δένδρα, υπάρχουν: α') τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks, ANN), β') οι αλγόριθμοι παλινδρόμησης (regression algorithms), γ') οι Support Vector Machines, και δ') οι αλγόριθμοι μετά-μάθησης. Οι πρώτοι, δηλαδή τα ANN, είναι μία οικογένεια αλγορίθμων με στατιστικό μοντέλο που προσομοιώνει την λειτουργικότητα των νευρώνων του εγκεφάλου και η λειτουργικότητά τους συνίσταται σε εξαγωγή γραμμικών συνδυασμών των δεδομένων ενός instance της εισόδου και κατηγοριοποίηση βάσει μη-γραμμικών συνδυασμών των δεδομένων αυτών [43, 48]. Πρόκειται για ένα δίκτυο νευρώνων, με πολλές εισόδους και εξόδους, όπου έχει τουλάχιστον 2 επίπεδα (ένα εισόδου και ένα εξόδου) και η μετάβαση μεταξύ των κόμβων/νευρώνων γίνεται μέσω σταθμισμένων (με βάρη) συνδέσεων. Η αλλαγή των τιμών των βαρών μπορεί να είναι προσαρμοζόμενη, καθιστώντας τα ANN ικανά να χρησιμοποιηθούν στην Μηχανική Μάθηση και να τροποποιήσουν τα βάρη αυτά έτσι ώστε να παράγουν την βέλτιστη λύση για το πρόβλημα που καλούνται να αντιμετωπίσουν. Με άλλα λόγια, να "μάθουν". Η προσαρμογή των βαρών γίνεται με αλγόριθμους μάθησης και έτσι τα ANN μπορούν να κατηγοριοποιηθούν αναλόγως την διαδικασία μάθησης που ακολουθούν [49]. Οι κατηγορίες αυτές είναι:

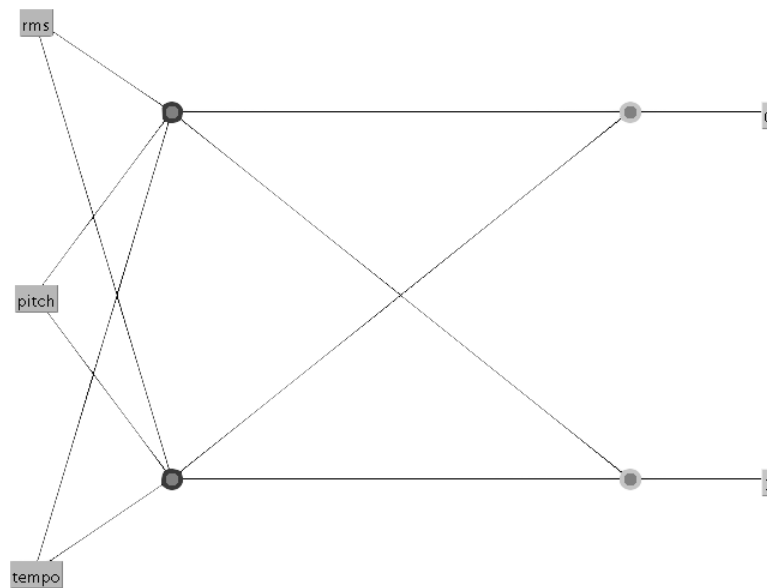
- Βάσει του κανόνα του Hebb [50], δηλαδή: *“Όταν ένας άξονας του κυττάρου A είναι αρκετά κοντά στο να διεγείρει ένα κύτταρο B και επαναλαμβανόμενα και συνεχώς συμμετέχει στην διέγερση αυτή, τότε λαμβάνει χώρα μία διαδικασία εξέλιξης ή αλλαγής σε ένα από τα δύο κύτταρα έτσι ώστε η αποτελεσματικότητα του A στο να διεγείρει το B αυξάνεται”*.
- Ενισχυόμενης μάθησης (Reinforced learning), όπου δεν υπάρχει εκ προοιμίου γνώση για τα δεδομένα αλλά μπορούν να παραχθούν κατόπιν αλληλεπίδρασης με το περιβάλλον. Η παρατήρηση του περιβάλλοντος μεταδίδεται στο ANN από την έξοδο προς την είσοδο.
- Ελαχίστου-Μεγίστου (Min-Max), όπου ως αλγόριθμος μάθησης χρησιμοποιείται ο Min-Max, γνωστός και ως Minimax [51]. Ακόμα, αναφέρονται και συνδυασμοί ANN

με ασαφή συστήματα και σύνολα όπου γίνεται χρήση του αλγόριθμου Min-Max και το σύστημα πλέον καλείται Fuzzy Min-Max Neural Network [52, 53].

- Στοχαστικά (Stochastic), όπου χρησιμοποιούνται στοχαστικές διαδικασίες για τον προσδιορισμό των βαρών και τον συναρτήσεων μεταφοράς από νευρώνα σε νευρώνα [54].
- Βασισμένα σε γενετικούς αλγόριθμους (Based on genetic algorithms), όπου οι γενετικοί αλγόριθμοι χρησιμοποιούνται είτε ως συναρτήσεις υπολογισμού βαρών είτε ως τρόπος δημιουργίας της τοπολογίας του νευρωνικού δικτύου [55].
- Βασισμένα σε τεχνητή ζωή (Based on Artificial Life), όπου η εκπαίδευση του δικτύου βασίζεται σε κανόνες που περιγράφουν την εκμάθηση και συμπεριφορά στα πλαίσια τεχνητής ζωής (artificial life) [49].
- Βάσει της αρχής διατήρησης μέγιστης πληροφορίας ή μέγιστης εντροπίας (Maximum Information Preservation, Maximum Entropy γνωστή και ως Infomax) [56], όπου ο στόχος των νευρώνων του δικτύου είναι να μεγιστοποιήσουν την αμοιβαία πληροφορία μεταξύ της εισόδου και της εξόδου [57].

Τα νευρωνικά δίκτυα έχουν χρησιμοποιηθεί σε πληθώρα εφαρμογών, όπως η αναγνώριση γραπτού κειμένου [58], αναγνώριση εικόνας [59] και έξυπνη επεξεργασία πολυμέσων [60]. Σε κάθε περίπτωση όμως, το ANN θεωρείται ένα μαύρο κουτί (black box). Δηλαδή, υπάρχει γνώση μόνο για την είσοδο και την έξοδο του συστήματος και έτσι δεν μπορεί να γίνει γνωστή η πραγματική διαδικασία επεξεργασίας των δεδομένων και εκπαίδευσης του συστήματος [52, 61]. Ένα παράδειγμα ενός ANN, όπως αυτό απεικονίζεται από το λογισμικό Μηχανικής Μάθησης WEKA [45], είναι στο **Σχήμα 2.3** όπου φαίνονται μόνο τα επίπεδα εισόδου και εξόδου με το σύστημα έχει 2 ενδιάμεσα επίπεδα και συνολικά 4 (2 τα εισόδου/εξόδου και 2 τα ενδιάμεσα). Το πλήθος των ενδιάμεσων επιπέδων στο συγκεκριμένο ANN υπολογίζεται από την εξίσωση: (πλήθος δεδομένων + πλήθος κλάσεων)/2.

Οι αλγόριθμοι παλινδρόμησης χρησιμοποιούν την στατιστική μέθοδο της παλινδρόμησης, όπου διερευνάται η σχέση μεταξύ μεταβλητών, πως μία ανεξάρτητη μεταβλητή επηρεάζει άλλες, εξαρτημένες, μεταβλητές [62]. Ως στόχο έχουν να συσχετίσουν και, συνεπώς, να προβλέπουν το αποτέλεσμα από τις διάφορες τιμές ενός σετ χαρακτηριστικών [63]. Δηλαδή,



Σχήμα 2.3: Ένα ANN για την κατηγοριοποίηση της εξόδου σε κλάσεις 0 και 1 βάσει εισόδου τριών δεδομένων: RMS, Pitch και Tempo.

και σύμφωνα με το προηγούμενο παράδειγμα σε αυτή την ενότητα, δεδομένου ενός instance ερευνάται ποια είναι η σχέση μεταξύ των διαφορετικών τιμών των δεδομένων που περιέχει το instance και της εξόδου, δηλαδή της κλάσης. Ως πρώτη προσέγγιση αναφέρεται η μέθοδος των ελαχίστων τετραγώνων (least square method) από δύο δημοσιευμένες εργασίες των Legendre και Gauss το 1805 και 1809, αντιστοίχως [64], η οποία πλέον χρησιμοποιείται και στην γραμμική παλινδρόμηση. Η μέθοδος της γραμμικής παλινδρόμησης, και στην απλή της μορφή, προσπαθεί να βρει μία εξίσωση/μοντέλο που να περιγράφει την σχέση μεταξύ μίας ή και περισσότερων ανεξάρτητων μεταβλητών και μίας εξαρτημένης μεταβλητής. Η διαδικασία αυτή υλοποιείται με την χρήση γραμμικών συναρτήσεων που καλούνται Linear Predictor Functions και οι οποίες υλοποιούν την διαδικασία συσχέτισης και πρόβλεψης της τιμής του αποτελέσματος από τα δεδομένα εισόδου στην διαδικασία της παλινδρόμησης. Επίσης, εάν υπάρχουν παραπάνω του ενός αποτελέσματα για τα οποία αναζητείται η σχέση με τα δεδομένα εισόδου, τότε η διαδικασία καλείται πολλαπλή γραμμική παλινδρόμηση (multiple linear regression) [62, 63]. Μία ακόμα και γνωστή μέθοδος παλινδρόμησης είναι η Logistic Regression όπου εξετάζει την συσχέτιση ενός δυαδικού αποτελέσματος με τα δεδομένα εισόδου. Δηλαδή, εάν υπάρχει ένα αποτέλεσμα το οποίο έχει δύο τιμές, δηλαδή κλάση 0 και κλάση 1 ή κλάση A και κλάση B, τότε η Logistic Regression εξετάζει την συσχέτιση των δεδομένων εισόδου με το αποτέλεσμα του συστήματος, δηλαδή την κλάση. Αυτό την καθιστά ως

προτιμητέα σε περιπτώσεις που το αποτέλεσμα έχει μοντελοποιηθεί ως τοποθέτηση μεταξύ δύο κλάσεων [65]. Οι αλγόριθμοι παλινδρόμησης έχουν επίσης χρησιμοποιηθεί για αρκετά προβλήματα κατηγοριοποίησης όπως αναγνώριση εκφράσεων προσώπου [66] και αναγνώριση ομιλίας [67].

Τις τεχνικές παλινδρόμησης μπορούν να χρησιμοποιήσουν ως αλγόριθμους εκμάθησης και οι Support Vector Machines (SVM). Πρόκειται για συστήματα τα οποία μετασχηματίζουν τον χώρο των δεδομένων σε έναν με περισσότερες διαστάσεις έτσι ώστε να βρουν ιδανικό τρόπο διαχωρισμού των δεδομένων για την πρόβλεψη του αποτελέσματος [68]. Δηλαδή, εάν τα instances έχουν ν πλήθος δεδομένων τότε τα SVM μετασχηματίζουν τα δεδομένα ώστε να είναι $x > \nu$. Ως ιδανικός τρόπος διαχωρισμού ορίζεται αυτός που παράγει το μικρότερο ποσοστό λάθους κατηγοριοποίησης των δεδομένων εκπαίδευσης [69]. Στην απλή τους μορφή, είναι ένα σύστημα το οποίο κατηγοριοποιεί τα δεδομένα εισόδου σε μία από τις δυο πιθανές εξόδου. Αυτό σημαίνει ότι η ιδανική τους χρήση είναι με προβλήματα Μηχανικής Μάθησης τα οποία έχουν μοντελοποιηθεί με τέτοιο τρόπο ώστε οι κλάσεις που θα χρησιμοποιηθούν για την κατηγοριοποίηση να είναι 2. Ακόμα, αναφέρεται πως σε σχέση με άλλες μεθόδους κατηγοριοποίησης και με μικρό πλήθος από instances εκπαίδευσης, οι SVM μπορούν να δημιουργήσουν μοντέλα με μεγαλύτερη ευστοχία από ότι άλλες μέθοδοι [70]. Τέλος, σε δημοσιευμένες έρευνες αναφέρεται πως οι SVM αντιμετωπίζουν καλύτερα το στάδιο εκπαίδευσης με το να μην υπέρ-εξειδικεύουν το παραγόμενο μοντέλο στα δεδομένα εκπαίδευσης ενώ καταφέρνουν και κρατάνε αρκετά υψηλά τις επιδόσεις ακρίβειας του μοντέλου [69]. Κάποια από τα πεδία που βρίσκουν εφαρμογή οι SVM είναι η αναγνώριση κίνησης [71], αναγνώριση προσώπου [72] και αναγνώριση ομιλίας [73].

Όλοι οι προηγούμενες αλγόριθμοι και μέθοδοι επιβλεπόμενης μάθησης μπορούν να χρησιμοποιηθούν ως κεντρικοί αλγόριθμοι σε συνδυασμό με αλγόριθμους/συστήματα μετά-μάθησης [74]. Τα τελευταία μπορούν να παραμετροποιούν αυτόματα τις συνθήκες εκτέλεσης για τους κεντρικούς αλγόριθμους έτσι ώστε να αυξήσουν την απόδοσή τους για το πρόβλημα που έχουν να αντιμετωπίσουν. Η παραμετροποίηση αυτή συμβαίνει με την επαναλαμβανόμενη εκτέλεση των κεντρικών αλγορίθμων και την παρακολούθηση, από τους αλγόριθμους μετά-μάθησης, του παραγόμενου μοντέλου και των επιδόσεών του. Οι δύο πιο συχνά χρησιμοποιούμενοι αλγόριθμοι μετά-μάθησης είναι ο Boosting και ο Bootstrap Aggregating (Bagging) [75]. Ο πρώτος είναι μία οικογένεια αλγορίθμων στην οποία χρησιμοποιείται έ-

νας κεντρικός αλγόριθμος κατ' επανάληψη με σκοπό το τελικό παραγόμενο μοντέλο να έχει αυξημένες επιδόσεις σε σχέση με το αποτέλεσμα κάθε αλγορίθμου που ανήκει στο σετ των κεντρικών [76]. Λειτουργεί χρησιμοποιώντας τον κεντρικό αλγόριθμο κατ' επανάληψη και βάσει των αποτελεσμάτων των επαναλήψεων προσπαθεί να ελαττώσει τα λάθη της κατηγοριοποίησης για τις υπόλοιπες εκτελέσεις του αλγορίθμου. Ο πιο γνωστός αλγόριθμος από αυτή την οικογένεια είναι ο Adaptive Boosting (AdaBoost) [76, 77] όπου οι συνολικές επιδόσεις του είναι ένα σταθμισμένο άθροισμα του κεντρικού αλγορίθμου. Στην διεθνή βιβλιογραφία αναφέρεται ότι οι αλγόριθμοι που χρησιμοποιούν δένδρα απόφασης αποτελούν την καλύτερη επιλογή για κεντρικοί αλγόριθμοι στον AdaBoost [43]. Ο αλγόριθμος Bagging διαφέρει από τον AdaBoost στο γεγονός ότι δεν χρησιμοποιεί επαναλήψεις και ότι μπορεί να λειτουργήσει με διαφορετικούς αλγορίθμους στο σετ των κεντρικών (αντί ενός, όπως συμβαίνει στον AdaBoost). Δημιουργεί υποσύνολα του δείγματος εκπαίδευσης τα οποία χρησιμοποιεί ως είσοδο στο σετ των κεντρικών αλγορίθμων. Το τελικώς παραγόμενο αποτέλεσμα είναι ένας συνδυασμός του αποτελέσματος από το σετ των κεντρικών αλγορίθμων. Οι αλγόριθμοι αυτοί (AdaBoost και Bagging) έχουν χρησιμοποιηθεί, μεταξύ άλλων, σε αναγνώριση φωνής και γένους ομιλητή [78], αναγνώριση προσώπου και εκφράσεων [79, 80], αναγνώριση γραπτού κειμένου [81, 82] και αναγνώριση εικόνας [83].

Τέλος, θα πρέπει να αναφερθεί ένας ακόμα αλγόριθμος ο οποίος αν και δεν ανήκει στους αλγορίθμους κατηγοριοποίησης χρησιμοποιείται συχνά σε διαδικασίες Μηχανικής Μάθησης και σε συνδυασμό με όλους του αλγορίθμους επιβλεπόμενης μάθησης που αναφέρθηκαν. Πρόκειται για τον αλγόριθμο Cross Fold Validation ο οποίος χρησιμοποιείται για την μέτρηση των επιδόσεων του μοντέλου που δημιουργήθηκε με τους αλγορίθμους επιβλεπόμενης μάθησης. Όπως μπορεί να παρατηρηθεί από τα παραπάνω, όλοι οι αλγόριθμοι supervised learning αλγόριθμοι χρειάζονται ένα στάδιο εκπαίδευσης. Σε αυτό χρησιμοποιείται ένα σετ δεδομένων. Έπειτα, ακολουθεί το στάδιο ελέγχου του μοντέλου που παράχθηκε. Η πιο ευθεία προσέγγιση αυτής της διαδικασίας είναι να χρησιμοποιηθούν διαφορετικά σετ δεδομένων για τα δύο στάδια, ήτοι εκπαίδευσης και ελέγχου. Όμως, πολλές φορές κάτι τέτοιο δεν είναι δυνατό, κυρίως λόγω δυσκολιών στην συγκέντρωση δεδομένων και, συνεπώς, για την διαδικασία της Μηχανικής Μάθησης υπάρχει μόνο ένα σετ δεδομένων. Με τον αλγόριθμο Cross Validation μπορεί να μετρηθεί η επίδοση του παραγόμενου μοντέλου χρησιμοποιώντας το μοναδικό σετ δεδομένων, το οποίο χρησιμοποιήθηκε και στην διαδικασία εκπαίδευσης, για το στάδιο του ελέγχου [84]. Το σετ δεδομένων χωρίζεται σε N υποσύνολα. Το στάδιο της

εκπαίδευσης και του ελέγχου εκτελείται N φορές χρησιμοποιώντας ως δεδομένα εκπαίδευσης $N - 1$ διαφορετικά, σε κάθε εκπαίδευση, παραδείγματα και τα υπόλοιπα ως δεδομένα ελέγχου. Επίσης, με N διαφορετικά υποσύνολα των παραδειγμάτων ο αλγόριθμος καλείται N Fold Cross Validation ενώ η πιο συνηθισμένη τιμή του N είναι 10.

Στην ανά χειράς διατριβή χρησιμοποιήθηκαν κυρίως αλγόριθμοι επιβλεπόμενης μάθησης, είτε ως βασικοί είτε ως κεντρικοί σε αλγόριθμους μετά-μάθησης. Ο λόγος της επιλογής αυτής είναι διττός. Από την μία το πρόβλημα Μηχανικής Μάθησης που αντιμετωπίστηκε βασίζεται πάνω σε απαντήσεις ανθρώπων. Πιο συγκεκριμένα, δεν υπήρχε η ερώτηση για την κατηγοριοποίηση των ήχων αναφορικά με πληροφορία, χαμηλού επιπέδου, που ενυπάρχει σε αυτούς αλλά βάσει της αντίληψης και συσχέτισης της εγγενούς, χαμηλού και υψηλού επιπέδου, πληροφορίας των ήχων από τον άνθρωπο. Αυτό το γεγονός όμως δεν σημαίνει όμως πως οι ήχοι δεν μεταφέρουν εγγενώς συναισθηματική πληροφορία. Η μεταφορά αυτή της συναισθηματικής πληροφορίας βασίζεται σε ανθρώπινες αντιληπτικές διαδικασίες. Δηλαδή, στην περίπτωση της μουσικής, για παράδειγμα, ο ίδιος ο άνθρωπος αντιλαμβάνεται συχνότητες συσχετίσεις και ακολουθίες ως λυπητερές ή χαρούμενες (ματζόρε ή μινόρε π.χ.) ενώ στην περίπτωση των ηχητικών γεγονότων, και πολύ πιθανώς μέσω εξελικτικών διαδικασιών, ο άνθρωπος έχει συσχετίσει διάφορα πρότυπα (patterns) της πληροφορίας του ήχου με καταστάσεις οι οποίες του εγείρουν συγκεκριμένα συναισθήματα. Έτσι, για να ερευνηθεί η αναγνώριση συναισθημάτων από γενικούς ήχους έπρεπε να υπάρχει η γνώση για το εγείρομενο συναίσθημα από κάθε ηχητικό γεγονός. Έτσι, αναδεικνύεται καθαρά πως η επιλογή που έπρεπε να γίνει ήταν οι αλγόριθμοι επιβλεπόμενης μάθησης. Από την άλλη, βάσει της διεθνούς επιστημονικής βιβλιογραφίας στο πεδίο της αναγνώρισης συναισθημάτων, φαίνεται ότι η παραπάνω άποψη ευσταθεί μιας και χρησιμοποιούνται σχεδόν αποκλειστικά αλγόριθμοι επιβλεπόμενης μάθησης για αυτή την διαδικασία. Ακόμα, λόγω περιορισμού σε διαθέσιμα σετ ηχητικών γεγονότων με συναισθηματική επισήμανση, ο αλγόριθμος Cross Validation χρησιμοποιήθηκε για την αξιολόγηση των παραγόμενων μοντέλων.

2.1.3 Τι είναι το Συναίσθημα;

Ο ορισμός του συναισθήματος είναι ακόμα ένα πεδίο σύγκρουσης απόψεων των επιστημόνων και ερευνητών από τα σχετικά με αυτό πεδία. Τι είναι το συναίσθημα και το πως αυτό

παράγεται δεν έχει οριστεί έως τώρα με έναν τρόπο που να είναι κοινώς αποδεκτός από ψυχιάτρους, ψυχολόγους και εν γένει θεωρητικούς επιστήμονες [10]. Εντούτοις, υπάρχουν κοινοί τόποι των επιστημόνων από τα προαναφερόμενα πεδία βάσει των οποίων είναι εφικτή η έρευνα σχετικά με την ανίχνευση συναισθήματος από εξωτερικά, προς το άνθρωπο, γεγονότα/δεδομένα.

Ένας τέτοιος κοινός τόπος είναι και η περιγραφή του συναισθήματος ως ένας “φυλογενετικός συνεχής μηχανισμός για ευέλικτη προσαρμογή” [11], θεωρούμενος ως ένα θεωρητικό κατασκεύασμα το οποίο αποτελείται από πέντε (5) στοιχεία: α) γνωσιακό (cognitive), β) περιφερική απαγωγή (peripheral efference), γ) κινητήριο (motivational), δ) μηχανική έκφραση (motor expression) και ε) υποκειμενική αίσθηση (subjective feeling) [85]. Επίσης, αναφέρεται και ως μία σύντομη αντίδραση του ανθρώπου σε εξωτερικά ερεθίσματα που μπορεί να εκδηλωθεί και να μετρηθεί από την “τριάδα της αντίδρασης” (reaction triad), ήτοι: α') την φυσιολογική διέγερση (physiological arousal), β') την μηχανική έκφραση (motor expression), και γ') την υποκειμενική αίσθηση [12].

Όμως, ο ορισμός του συναισθήματος είναι εκτός των ορίων και στων στόχων της παρούσας διατριβής. Στο υπόλοιπο κείμενο χρησιμοποιείται η παραπάνω περιγραφή που έχει δοθεί από διακεκριμένους επιστήμονες στον τομέα της ψυχολογίας και των συναισθηματικών επιστημών (affective sciences) και έχει δημοσιευθεί σε έγκριτα και έγκυρα διεθνή επιστημονικά περιοδικά. Έτσι, τη παρούσα ενότητα ακολουθεί η περιγραφή των μοντέλων συναισθημάτων, όπως αυτά δίνονται από δημοσιευμένες έρευνες σε διεθνή επιστημονικά περιοδικά, τις τεχνικές αποτύπωσης της συναισθηματικής κατάστασης, την περιγραφή και παρουσίαση των κυριότερων αλγορίθμων που χρησιμοποιούνται στην ανίχνευση συναισθημάτων από ηχητικά δεδομένα και τα συνήθως χρησιμοποιούμενα τεχνικά χαρακτηριστικά του ήχου στην προαναφερθείσα διαδικασία.

2.1.4 Διακριτά μοντέλα συναισθημάτων

Σε μία αφαιρετική προσέγγιση, τα μοντέλα συναισθημάτων μπορούν να ομαδοποιηθούν στις εξής δύο κατηγορίες [86]: α) Διακριτά, και β) Συνεχή - διαστατικά. Όπως προκύπτει από την ονομασία της κατηγορίας, τα διακριτά μοντέλα συναισθημάτων κατηγοριοποιούν τα συναισθήματα βάσει διακριτών λεκτικών περιγραφών τους, π.χ., “Λύπη” και “Χαρά”. Μπο-

Πίνακας 2.1: Τα βασικά συναισθήματα βάσει των [87, 88]

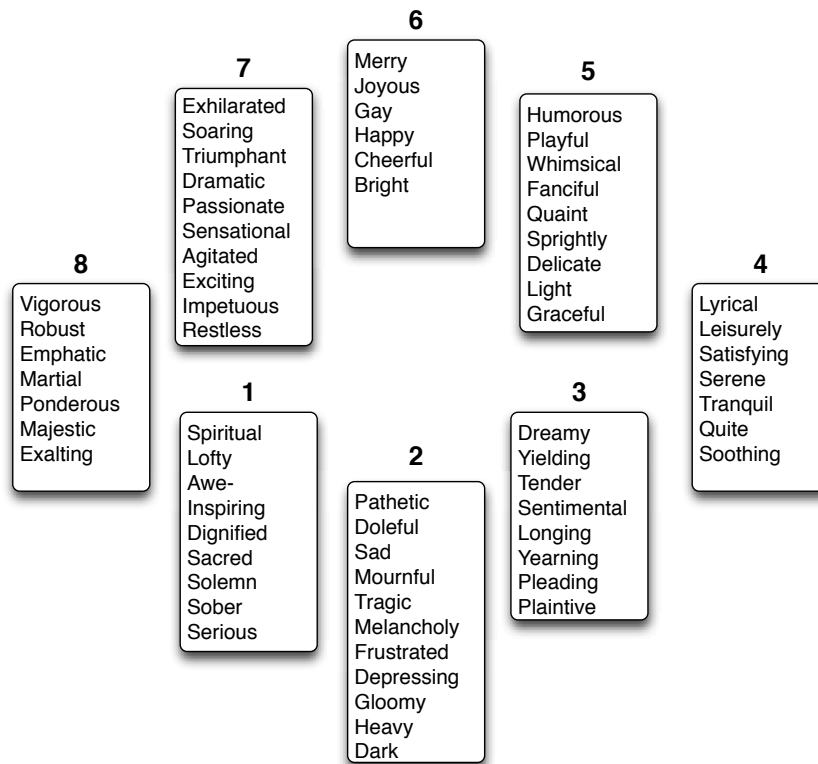
A/A	Συναίσθημα	A/A	Συναίσθημα
1	Χαρά	3	Λύπη
2	Φόβος	4	Θυμός

ρούν να θεωρηθούν και ως τα πρώτα που εμφανίστηκαν με απαρχή το μοντέλο Βασικών Συναισθημάτων. Το τελευταίο στηρίζεται στην θεωρία του Δαρβίνου για τα συναισθήματα κατά την οποία υπάρχει ένα σετ βασικών συναισθημάτων και όλα τα άλλα μπορούν να πηγάζουν από αυτά με κατάλληλους συνδυασμούς τους [87]. Τα προαναφερθέντα συναισθήματα εμφανίζονται στον **Πίνακα 2.1**.

Η επιλογή αυτών των συναισθημάτων μπορεί να δικαιολογηθεί λόγω μετέπειτα ευρημάτων της Νευρολογίας και Νευροφυσιολογίας [87]. Σε δημοσιευμένες έρευνες αναφέρεται σαφής σύνδεση της λειτουργίας τόσο των νευρικών συστημάτων του ανθρώπου, όσο και περιοχών του εγκεφάλου και της βίωσης ενός εκ των προαναφερθέντων συναισθημάτων [89, 90]. Έχουν αναφερθεί συσχετίσεις της διέγερσης στεφανιακών και παραστεφανιακών περιοχών του εγκεφάλου (π.χ. αμυγδαλή, ιππόκαμπος, φλοιός του προσαγωγίου), με την βίωση συναισθημάτων όπως φόβος, θυμός, χαρά και λύπη [91].

Αν και φαίνεται πως αυτό το μοντέλο είναι το κυρίως χρησιμοποιούμενο στις έρευνες που εστιάζουν στην σύνδεση λειτουργίας στοιχείων του ανθρώπου και βίωσης συναισθημάτων [87], η παραδοχή πως όλα τα άλλα συναισθήματα πηγάζουν από μία ομάδα βασικών οδήγησε στην κριτική και, αυξανόμενη, απόρριψή του [92]. Ως εναλλακτική προτάθηκε η από-συσχέτιση των συναισθημάτων από τις λεκτικές περιγραφές τους και η χρήση ενός χώρου για την αποτύπωση τους. Ο χώρος αυτός ορίζεται από διαστάσεις συναισθηματικών καταστάσεων από τις οποίες μπορούν να σχηματιστούν όλα τα συναισθήματα που μπορεί να βιώσει ο άνθρωπος [93].

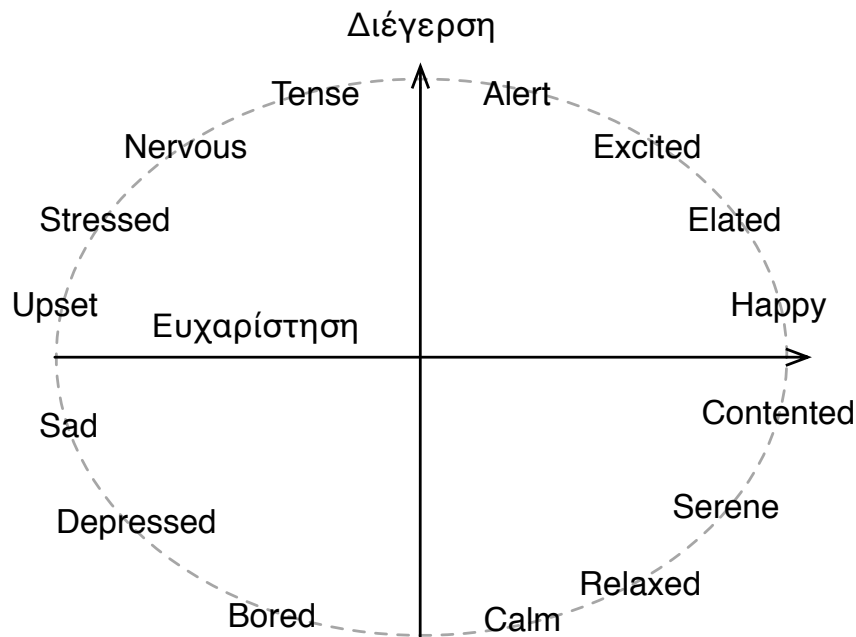
Λίγο πριν την εδραίωση της προαναφερθείσας, νέας, προσέγγισης, σε άλλη εργασία [94] παρουσιάζεται ένα ακόμα μοντέλο το οποίο χρησιμοποιεί μία συγκεκριμένη ομαδοποίηση των περιοχών του προαναφερθέντος χώρου και ονομάζεται “Λίστα (της) Hevner” (Hevner’s, Hevner List). Το πλήθος των ομάδων είναι 8 και για κάθε ομάδα χρησιμοποιείται ένα σύνολο συνωνύμων επιθέτων κατάλληλων για να περιγράψουν μία διακριτή συναισθηματική



Σχήμα 2.4: Απεικόνιση των ομάδων στο μοντέλο “Λίστα της Hevner” (Hevner’s List)

κατάσταση. Η τοποθέτηση των ομάδων επιθέτων ακολουθεί το συνεχές μοντέλο και έτσι αυτές βρίσκονται σε μία κυκλική διαρρύθμιση. Μία απεικόνιση των ομάδων αυτών, βάσει της σχετικής αρχικής πρότασης [95], φαίνεται στο **Σχήμα 2.4**. Έχουν προταθεί αλλαγές προσθήκες στο μοντέλο αυτό με προσθήκες επιπλέον ομάδων όπου ο συνολικός αριθμός ομάδων επιθέτων φθάνει στις 13 [96].

Τα διακριτά μοντέλα συναισθημάτων προσφέρονται για άμεση συσχέτιση των απαντήσεων του χρήστη με λεκτικές περιγραφές (εφόσον τέτοιες χρησιμοποιούνται και στο μοντέλο). Όμως, αυτή ακριβώς η ιδιότητά τους εισάγει ένα πρόβλημα κατά την έτερο-αξιολόγηση ερευνών. Ο λόγος είναι πως αφενός υπάρχουν αρκετές λεκτικές περιγραφές των οποίων η ερμηνεία δεν είναι πάντα σταθερή και, αφετέρου, δύο διαφορετικές λεκτικές περιγραφές μπορούν να εκληφθούν ως το ίδιο συναίσθημα ενώ στην πράξη αναφέρονται σε διαφορετικό. Οι λέξεις “Happy” και “Happiness”, “Joy” και “Joyful” κτλ. αποτελούν ένα τυπικό παράδειγμα [14]. Αυτό το πρόβλημα εμποδίζει την κοινή χρήση ή ανταλλαγή αποτελεσμάτων μεταξύ ερευνών που κάνουν χρήση διακριτών μοντέλων και δεν έχουν χρησιμοποιήσει τις ίδιες λεκτικές περιγραφές συναισθημάτων.

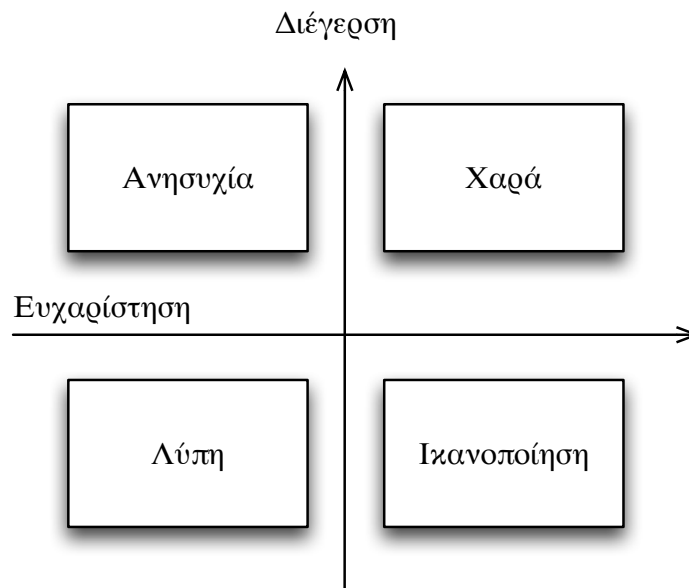


Σχήμα 2.5: Το αρχικό μοντέλο των συναισθηματικών καταστάσεων όπως περιγράφεται στο [97, 93]. Οι λεκτικές περιγραφές είναι στην αγγλική γλώσσα για πλήρη συμφωνία με την διεθνή βιβλιογραφία.

2.1.5 Συνεχή μοντέλα συναισθημάτων

Σε αντίθεση με τα διακριτά μοντέλα συναισθημάτων, τα συνεχή μοντελοποιούν το συναίσθημα ως συνισταμένη κάποιων συναισθηματικών καταστάσεων. Έτσι, το συναίσθημα αποδομείται σε συνιστώσες, ο συνδυασμός των οποίων μπορεί να αποδώσει και να αποτυπώσει όλα τα συναισθήματα [30]. Για την τελική αποτύπωση ενός συναισθήματος χρησιμοποιείται ένας χώρος διαστάσεων ίσων με τις συνιστώσες αυτές. Η αρχή αυτής της προσέγγισης φαίνεται ότι έγινε γύρω στο 1990 όπου ο Ortony, πέρα από την τεκμηρίωση της παραπάνω προσέγγισης, στοιχειοθέτησε και μία καλά τεκμηριωμένη κριτική για το μοντέλο των βασικών συναισθημάτων [92, 93].

Στο αρχικό μοντέλο, οι λεκτικές περιγραφές τοποθετούνταν πάνω στην περιφέρεια του κύκλου που σχηματίζεται από τον χώρο των συναισθηματικών καταστάσεων. Έτσι, η μετάβαση από μία λεκτική περιγραφή σε άλλη αναγνωριζόταν από την γωνία και την διεύθυνση του διανύσματος που έχει ως αρχή την τομή των αξόνων και τέλος το σημείο (V_V, V_A) , όπου V_A είναι η τιμή της Διέγερσης (Arousal) και V_V είναι η τιμή της Ευχαρίστησης (Valence). Ένα παράδειγμα απεικόνισης του μοντέλου φαίνεται στο **Σχήμα 2.5**.



Σχήμα 2.6: Ο χώρος της Ευχαρίστησης - Διέγερσης με αντιστοιχισμένα συναισθήματα

Ο αρχικός αριθμός των συνιστωσών ήταν 2 [93]. Αντιπροσωπεύουν τις συναισθηματικές καταστάσεις που αναφέρθηκαν προηγουμένως και οδηγούν στην δημιουργία συναισθημάτων, ήτοι: α) την Ευχαρίστηση (Valence), και β) την Διέγερση (Arousal). Αυτές είναι οι βασικές και ευρέως χρησιμοποιούμενες σε έρευνες σχετικά με την αναγνώριση συναισθημάτων από ηχητικά δεδομένα [3, 30]. Ένα παράδειγμα του χώρου/επιπέδου που σχηματίζεται από τις προαναφερθείσες συναισθηματικές καταστάσεις μαζί με την αντιστοιχία περιοχών και λεκτικών περιγραφών των συναισθημάτων δίνεται στο **Σχήμα 2.6**.

Στο Σχήμα αυτό, οι λεκτικές περιγραφές των συναισθημάτων μπορούν να αντιστοιχηθούν σε ομάδες τιμών για τις χρησιμοποιούμενες συνιστώσες. Αν και σε διάφορα άρθρα αναφέρεται αντιστοιχία τέτοιων λεκτικών περιγραφών με περιοχές στο πεδίο της Ευχαρίστησης και της Διέγερσης, η αντιστοίχιση αυτή είναι τελικά ποιοτική και όχι ποσοτική [98, 19, 93]. Δηλαδή, δεν υπάρχει αποδεδειγμένα (μέχρι τώρα) κάποια ποσοτική σχέση μεταξύ των λεκτικών περιγραφών των συναισθημάτων και των τιμών για τις συναισθηματικές καταστάσεις/συνιστώσες [99].

Η Bradley σε δημοσιευμένη εργασία της [6] αναφέρεται σε μία επαύξηση του χώρου των συναισθηματικών καταστάσεων, με την προσθήκη επιπλέον διαστάσεων. Αν και έχουν αναφερθεί κάποιες συναισθηματικές καταστάσεις, η επικρατέστερη φαίνεται να είναι η Κυριαρχία (Dominance). Έτσι, ο νέος χώρος είναι τρισδιάστατος [6, 100]. Λόγω όμως της αυξανόμενης πολυπλοκότητας [101], στις έρευνες που σχετίζονται με την αναγνώριση συναισθημάτων από

ηχητικά δεδομένα χρησιμοποιούνται δύο διαστάσεις. Εντούτοις, σε μερικά άρθρα αναφέρεται χρήση τριών διαστάσεων αντί για δύο [6, 100].

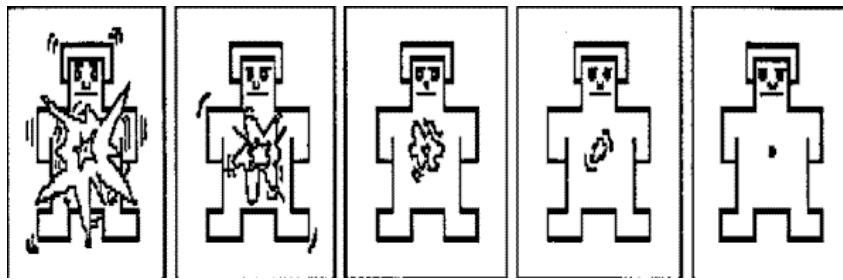
2.1.6 Αποτύπωση συναισθηματικής κατάστασης

Όπως φαίνεται και στο **Σχήμα 2.1**, κατά τη διαδικασία αναγνώρισης συναισθημάτων από ηχητικά δεδομένα, απαιτείται και επιπροσθέτως η συναισθηματική επισήμανση των δεδομένων. Αυτή λαμβάνεται με πειράματα στα οποία άνθρωποι καλούνται να επισημάνουν ένα σετ ήχων με το συναίσθημα που βιώνουν ή τους μεταδίδει κάθε ήχος ξεχωριστά [102]. Αναλόγως το συναισθηματικό μοντέλο που χρησιμοποιείται κατά τη μετέπειτα διαδικασία, υπάρχουν και διαφορετικές τεχνικές κωδικοποίησης της συναισθηματικής επισήμανσης [3]. Έτσι, όταν χρησιμοποιείται ένα διακριτό μοντέλο, τότε η συναισθηματική επισήμανση γίνεται με την κατάδειξη/επιλογή της λεκτικής περιγραφής από τους συμμετέχοντες στα προαναφερθέντα πειράματα. Δηλαδή, συνήθως κάθε συμμετέχων ακούει ένα σετ ήχων και για κάθε έναν καλείται να επιλέξει την λεκτική περιγραφή συναισθήματος που αρμόζει καλύτερα σε αυτό που άκουσε.

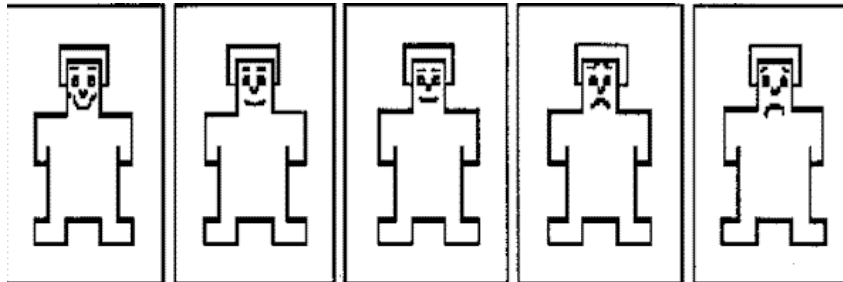
Στην περίπτωση όμως των διαστατικών μοντέλων, κάτι τέτοιο δεν είναι εφικτό. Σύμφωνα και με το μοντέλο, οι συμμετέχοντες θα πρέπει να καταδεικνύουν την συναισθηματική κατάσταση αντί για μία λεκτική περιγραφή. Έχουν δημοσιευθεί εργασίες όπου πραγματοποιείται μία απευθείας ανάθεση τιμών στο πεδίο του διαστατικού μοντέλου από τους συμμετέχοντες σε αντίστοιχο πείραμα. Δηλαδή, κάθε συμμετέχων καλούνταν να τοποθετήσει μία ένδειξη πάνω σε ένα δισδιάστατο πεδίο, διέγερσης και ευχαρίστησης, η οποία να ανταποκρίνεται στις αντίστοιχες συναισθηματικές του καταστάσεις [3].

Όμως, μία τέτοια μέθοδος είναι αρκετά πιθανό να παράγει συναισθηματικές αξιολογήσεις μη αξιοποιήσιμες μεταξύ ερευνών, ακυρώνοντας έτσι το μεγάλο πλεονέκτημα των διαστατικών συναισθηματικών μοντέλων λόγω της προφανούς έλλειψης κοινού μέτρου για τις τιμές στο προαναφερόμενο πεδίο. Δημοσιευμένες εργασίες αναφέρουν μία τυποποίηση ποσοτικής μέτρησης των τιμών για τρεις διαστάσεις του συναισθηματικού χώρου, ήτοι της Διέγερσης, της Ευχαρίστησης και της Κυριαρχίας [102].

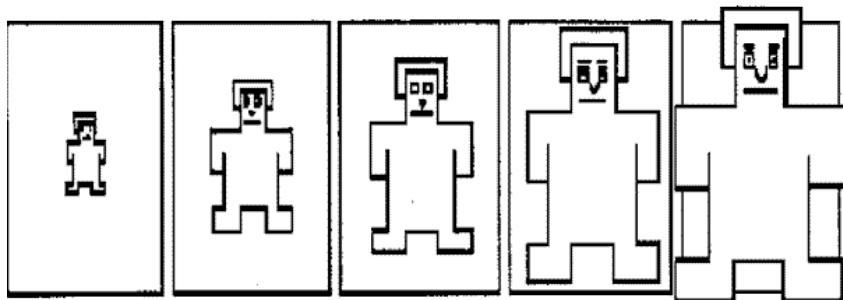
Η τυποποίηση αυτή γίνεται με την χρήση τριών ομάδων εικόνων που αντιπροσωπεύουν την



(α') Οι εικόνες SAM για την Διέγερση



(β') Οι εικόνες SAM για την Ευχαρίστηση



(γ') Οι εικόνες SAM για την Κυριαρχία

Σχήμα 2.7: Οι εικόνες SAM για τις τρεις συναισθηματικές καταστάσεις

μεταβολή των συναισθηματικών καταστάσεων και η τελική τιμή για κάθε μία από τις καταστάσεις είναι η σειρά της εικόνας που επιλέχθηκε, με την δυνατότητα της επιλογής και ενδιάμεσης κατάστασης. Το σετ εικόνων αυτό λέγεται “Ανδρείκελα αυτό-αξιολόγησης” (Self Assessment Manikin, SAM) και απεικονίζεται στα **Σχήματα 2.7α’** έως **2.7γ’** για την Διέγερση, Ευχαρίστηση και Κυριαρχία αντιστοίχως.

Για κάθε συναισθηματική κατάσταση υπάρχουν συνολικά 9 διαθέσιμες επιλογές. Πέντε που αντιστοιχούν απευθείας στις εικόνες και 4 για ενδιάμεσα από κάθε εικόνα. Αυτό σημαίνει πως για κάθε κατάσταση η μέγιστη τιμή είναι 9 και η ελάχιστη 1 ενώ, για παράδειγμα, η τιμή 2 αντιστοιχεί στην κατάσταση ανάμεσα στην πρώτη και την δεύτερη εικόνα για κάθε σειρά. Ο τρόπος αυτός συναισθηματικής αποτύπωσης έχει χρησιμοποιηθεί και στο πρώτο σετ συναισθηματικά επισημασμένων ηχητικών γεγονότων που έγινε διαθέσιμο στην ερευνητική

κοινότητα, τους International Affective Digital Sounds (IADS) ήχους όπου αποτελείται από 167 μονοφωνικά και συναισθηματικά επισημασμένα ηχητικά γεγονότα από διάφορες ηχητικές πηγές (π.χ. ανθρώπινες εκδηλώσεις και πράξεις, ηχοτοπία, ήχους από μηχανήματα και εκρήξεις, ήχους ζώων κ.α.) [103].

2.1.7 Τεχνικά μεγέθη ήχου

Η αναγνώριση συναισθήματος από ηχητικά δεδομένα υλοποιείται με με την χρήση δύο ειδών πληροφορίας: α') την συναισθηματική επισήμανση που έχουν τα ηχητικά δεδομένα, η οποία λαμβάνεται από πειράματα υποκειμενικής αξιολόγησης και σύμφωνα με τις διαδικασίες που αναφέρθηκαν προηγουμένως στην Ενότητα 2.1.6, και β') από χαρακτηριστικά του καθαυτού ηχητικού δεδομένου. Εστιάζοντας στο β'), τα χαρακτηριστικά αυτά θα πρέπει να μπορούν να είναι τέτοια ώστε αφενός να επιτρέπουν την γενίκευση ενός μοντέλου για την εφαρμογή του σε ήχους που δεν έχουν χρησιμοποιηθεί κατά το στάδιο εκπαίδευσης (όπως αυτό παρουσιάζεται στην Ενότητα 2.1.2 και αναλόγως τον αλγόριθμο κατηγοριοποίησης) και αφετέρου να μπορούν να υπολογιστούν ή να παρασχεθούν για το είδος των ηχητικών δεδομένων που εστιάζει η εκάστοτε διαδικασία αναγνώρισης συναισθήματος. Επιπροσθέτως, θα πρέπει αυτά τα χαρακτηριστικά να έχουν κάποια σύνδεση με το πρόβλημα της διαδικασίας Μηχανικής Μάθησης που αντιμετωπίζεται έτσι ώστε οι χρησιμοποιούμενοι αλγόριθμοι κατηγοριοποίησης να μπορούν να εξάγουν ένα μοντέλο με ικανοποιητικές επιδόσεις κατηγοριοποίησης.

Αν και έχουν αναφερθεί προσπάθειες αναγνώρισης συναισθήματος χρησιμοποιώντας μη τεχνικά χαρακτηριστικά, π.χ. χρήση στίχων από τραγούδια [104, 105], στην πληθώρα των εργασιών που εστιάζουν στην αναγνώριση συναισθήματος από ηχητικά δεδομένα χρησιμοποιούνται χαμηλού ή υψηλού επιπέδου τεχνικά χαρακτηριστικά του σήματος. Για παράδειγμα, σε δημοσιευμένες εργασίες που εστιάζουν στην αναγνώριση συναισθήματος από μουσική ή/και ομιλία έχει ήδη αναφερθεί ισχυρή σύνδεση της ενέργειας, της τραχύτητας και της διαφωνίας (dissonance) με την συναισθηματική κατάσταση της Διέγερσης, όπως παρουσιάζεται στην εκτενής ανασκόπηση εργασιών αναγνώρισης συναισθήματος των Julsin και Västfjäll [38]. Τα τεχνικά χαρακτηριστικά αυτά μπορούν να μετρηθούν από αντίστοιχα μεγέθη, π.χ. μέση ενέργεια (RMS energy), τραχύτητα (roughness) και μη-αρμονικότητα (inharmonicities), αντιστοίχως. Σε επόμενες ενότητες θα δοθεί παραπάνω βάρος στην συσχέ-

τιση χαρακτηριστικών του ήχου με το εγχειρόμενο συναίσθημα, όπου και θα παρουσιαστεί αναλυτικά η στάθμη της τεχνικής της αναγνώρισης συναισθήματος από ηχητικά δεδομένα.

Υπάρχει μία πληθώρα αντιστοιχιών των τεχνικών μεγεθών του ήχου με τα τεχνικά χαρακτηριστικά του. Στην παρούσα διατριβή τηρήθηκε αυτή που είναι συμβατή με το MIR Toolbox⁶ [106], ένα προγραμματιστικό εργαλείο εξαγωγής και υπολογισμού τέτοιων μεγεθών. Έτσι υπάρχουν τεχνικά μεγέθη συσχετιζόμενα με: α') την ενέργεια, β') την χροιά (timbre), γ') τον ρυθμό, και δ') το τονικό ύψος [24, 107]. Τα προαναφερθέντα χαρακτηριστικά αναφέρονται ως χαμηλού επιπέδου, λόγω του γεγονότος ότι αφορούν χαρακτηριστικά του ήχου τα οποία σχετίζονται άμεσα με το σήμα/κυματομορφή. Επίσης, υπάρχει και μία ομάδα τεχνικών χαρακτηριστικών τα οποία μπορούν να κατηγοριοποιηθούν ως ανωτέρου επιπέδου (High level), λόγω του ότι περιγράφουν συνολικές ή αφηρημένες ιδιότητες/χαρακτηριστικά του σήματος [24]. Τα μεγέθη που χρησιμοποιούνται κυρίως στην αναγνώριση συναισθήματος από ηχητικά δεδομένα συνοψίζονται στον **Πίνακα 2.2** και παρουσιάζονται στις επόμενες ενότητες.

Πίνακας 2.2: Συνοπτική παρουσίαση των μεγεθών που χρησιμοποιήθηκαν στην παρούσα διατριβή

A/A	Ονομασία	Κατηγορία
1	Μέση ενέργεια (rms)	Ενέργεια
2	Ποσοστό χαμηλής ενέργειας (low energy rate)	Ενέργεια
3	Διακύμανση (fluctuation)	Ρυθμός
4	Φάσμα κτύπων (beat spectrum)	Ρυθμός
5	Εμφανίσεις (onsets)	Ρυθμός
6	Πυκνότητα γεγονότων (event density)	Ρυθμός
7	Τέμπο (tempo)	Ρυθμός
8	Καθαρότητα παλμών (pulse clarity)	Ρυθμός
9	Παράγοντες συχνοτήτων Mel (MFCC)	Χροιά
10	Στατιστικές ροπές και μέσοι όροι του φάσματος	Χροιά
11	Τραχύτητα (roughness)	Χροιά

Πίνακας 2.2 – Συνέχεια σε επόμενη σελίδα

⁶<https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox>

Πίνακας 2.2 – Συνέχεια από προηγούμενη σελίδα

A/A	Ονομασία	Κατηγορία
12	Αλλαγή προσήμου (zero cross)	Χροιά
13	Φωτεινότητα (brightness)	Χροιά
14	Roll off	Χροιά
15	Κλίση ατάκας (attack slope)	Χροιά
16	Χρόνος ατάκας (attack time)	Χροιά
17	Άλμα ατάκας (attack leap)	Χροιά
18	Τονικό ύψος (pitch)	Τονικό ύψος - τονικότητα
19	Μη αρμονικότητα (inharmonic)	Τονικό ύψος - τονικότητα
20	Harmonic change detect function	Τονικό ύψος - τονικότητα

2.1.7.1 Μεγέθη συσχετιζόμενα με την ενέργεια του ηχητικού σήματος

Τα μεγέθη που σχετίζονται με την ενέργεια του ηχητικού σήματος αφορούν τον υπολογισμό αυτής σε απόλυτο νούμερο (δηλαδή, όχι ποσοστό) και υπολογίζονται απευθείας από την πρωτογενή (raw) ηχητικά δεδομένα. Υπάρχουν μεγέθη που αφορούν την ενέργεια που υπάρχει στο ηχητικό σήμα αλλά και την ενέργεια που αντιλαμβάνεται ο άνθρωπος, ανά συχνότητα, ανά ζώνη συχνοτήτων κτλ. Μία συγκεντρωτική παράθεση αυτών επιχειρείται στον

Πίνακα 2.2.

Το βασικότερο από την ομάδα μεγεθών που σχετίζονται με την ενέργεια του σήματος είναι η καθαυτή ενέργεια που περιέχεται στην ηχητική κυματομορφή. Αυτή μπορεί να υπολογισθεί τόσο από τα πρωτογενή δεδομένα του σήματος όσο και από τον κατάλληλο μετασχηματισμό αυτών στο πεδίο της συχνότητας. Ειδικότερα, η καθαυτή ενέργεια μπορεί να υπολογιστεί από τα δεδομένα στο πεδίο του χρόνου ως:

$$E_{rms} = \sqrt{\frac{1}{X} \sum_{x=1}^X s[x]^2} \quad (2.1)$$

όπου x είναι ο αύξων αριθμός του ηχητικού δείγματος (sample), s το raw ηχητικό σήμα, X είναι το πλήθος των δειγμάτων στο σήμα s και E_{rms} η καθαυτή ενέργεια.

Επίσης, σύμφωνα με την εργασία των Lie Lu, Dan Liu και Hong-Jiang Zhang [28], η ενέργεια μπορεί να υπολογιστεί για κάθε πλαίσιο σήματος από τα δεδομένα αυτού στο πεδίο της συχνότητας ως:

$$E(n) = \sum_{k=0}^K A(n, k) \quad (2.2)$$

Όπου $E(n)$ είναι η ενέργεια στο n -ιοστό πλαίσιο του σήματος, $A(n, k)$ είναι η απόλυτη τιμή για τον k -ιοστό συντελεστή της ανάλυσης Fourier του n -ιοστού πλαισίου του σήματος και K είναι το πλήθος των συντελεστών της ανάλυσης Fourier. Τα πλαίσια του σήματος λαμβάνονται από την τμηματοποίηση αυτού στο πεδίο του χρόνου. Μπορούν να είναι μη επικαλυπτόμενα ή επικαλυπτόμενα κατά ένα μέρος και για την εξαγωγή αυτών από το αρχικό σήματα χρησιμοποιείται μία κατάλληλη συνάρτηση. Η συνάρτηση αυτή καλείται συνάρτηση παραθυροποίησης και μπορεί να αναπαρασταθεί ως ένα σήμα με πλήθος δειγμάτων ίσο με το πλήθος δειγμάτων του πλαισίου. Οι τιμές κάθε δείγματος του σήματος της συνάρτησης παραθυροποίησης καθορίζονται από την συνάρτηση περιγραφής της. Για παράδειγμα, μία από τις πιο ευρέως χρησιμοποιούμενες τέτοιες συναρτήσεις, η hamming, είναι:

$$w[x] = a - b \cos\left(\frac{2\pi x}{X-1}\right) \quad (2.3)$$

όπου x είναι η τιμή/δείγμα του παραθύρου, X το σύνολο των τιμών/δειγμάτων και $a = 0.54$, $b = 1 - a$.

Έτσι, εάν s είναι το αρχικό σήμα, w_s το πλαίσιο και w η συνάρτηση για την τμηματοποίηση, τότε το πλαίσιο λαμβάνεται ως:

$$w_s = s \times w \quad (2.4)$$

Η τμηματοποίηση ενός σήματος s με X δείγματα σε N μη επικαλυπτόμενα πλαίσια μπορεί να γίνει ως ακολούθως:

$$s[n] = s[x + (i \times len_w)] \times w \quad (2.5)$$

όπου len_w είναι το πλήθος των δειγμάτων που θα έχει το πλαίσιο, $i \in [0N-1]$, n είναι ο αύξων αριθμός του πλαισίου με $n = i + 1$ και $N \times len_w \leq X$. Εάν πρόκειται για επικαλυπτόμενα παράθυρα, αντί από το αρχικό σήμα να λαμβάνονται διαδοχικά σύνολα δειγμάτων, κάθε ομάδα δειγμάτων x που θα επηρεαστεί από την διαδικασία τμηματοποίησης εμπεριέχει δείγματα από την προηγούμενη.

Συνεχίζοντας, η ενέργεια ενός σήματος είναι αρκετά πιθανό να μην μένει σταθερή κατά την εξέλιξη αυτού και έτσι κάποιες χρονικές στιγμές να υπάρχει/μεταδίδεται/μεταφέρεται περισσότερη ενέργεια. Ένας τρόπος για να εξεταστεί αυτό το ενδεχόμενο είναι η διακύμανση της ενέργειας σε διαδοχικές στιγμές και χρονικά πλαίσια του σήματος. Οι Τζανετάκης και Cook [108] αναφέρουν ότι η ίδια πληροφορία μπορεί να αναζητηθεί και από το ποσοστό των χρονικών πλαισίων του σήματος που έχουν ενέργεια μικρότερη από τον μέσο όρο του σήματος. Αυτό το μέγεθος καλείται ποσοστό χαμηλής ενέργειας (low energy rate ή low energy feature) και μπορεί να υπολογιστεί ως:

$$L_E(w_T) = \frac{N[RMS(w_A)_L]}{N[w_A]} \quad (2.6)$$

Όπου w_T είναι ένα χρονικό πλαίσιο διάρκειας 1 δευτ., w_A είναι ένα χρονικό πλαίσιο διάρκειας 23 ms με αρχή και τέλος μέσα στο w_T , $N[RMS(w_A)_L]$ είναι το πλήθος των w_A του w_T που έχουν μέση ενέργεια κάτω από τον μέσο όρο της μέσης ενέργειας του w_T και $N[w_A]$ είναι το πλήθος πλαισίων w_A που περιέχονται στο w_T .

Οι τιμές που μπορούν να ληφθούν από τις προηγούμενες εξισώσεις έχει βρεθεί ότι σχετίζονται με την διέγερση του ακροατή [3]. Έτσι, ένας ήχος για τον οποίο λαμβάνονται μεγάλες τιμές από τις εξισώσεις 2.1 και 2.2 και μικρές από την 2.6, είναι αρκετά πιθανό να συνεισφέρει στην αύξηση της διέγερσης του δέκτη. Θα πρέπει όμως να τονιστεί πως για την συγκριτική μελέτη διαφορετικών ηχητικών δεδομένων απαιτείται μία κανονικοποίηση της ενέργειας των σημάτων έτσι ώστε να υπάρχει μία κοινή βάση αναφοράς.

2.1.7.2 Μεγέθη που σχετίζονται με την χροιά του σήματος

Η χροιά ενός σήματος μπορεί να αναπαρασταθεί από πλήθος μεγεθών, όπως φαίνεται και στον Πίνακα 2.2. Ίσως ένα από τα πιο συνήθη μεγέθη είναι οι Mel Frequency Cepstral Coefficients (MFCC), όπου αφορούν την αντιστοίχιση του συχνοτικού περιεχομένου του σήματος στην κλίμακα MEL [109]. Μέγεθος που κυρίως χρησιμοποιείται στο πεδίο της ψηφιακής επεξεργασίας ομιλίας αλλά πιθανώς λόγω της χρήσης της κλίμακας *MEL*, η οποία αντιστοιχεί την υποκειμενική αντίληψη για το συχνοτικό ύψος στο αντικειμενικά μετρούμενο, έχει βρει ευρεία εφαρμογή και στην επεξεργασία ήχου εν γένει [110]. Η μετατροπή συχνότητας από *Hz* σε *MEL* γίνεται ως εξής:

$$f_{mel} = 2595 \log_{10} \left(1 + \frac{f_{Hz}}{700} \right) \quad (2.7)$$

όπου, f_{mel} είναι η συχνότητα σε κλίμακα *MEL* και f_{Hz} είναι η συχνότητα σε *Hz*. Η εύρεση των MFCC, που στην πράξη είναι ένα σύνολο παραμέτρων από-συσχετισμένων μέσω διακριτού μετασχηματισμού συνημιτόνου (DCT), γίνεται με την χρήση μίας σειράς υπολογισμών και μετασχηματισμών. Ξεκινάει από την επιβολή K τριγωνικών πλαισίων στο πεδίο της συχνότητας για τον υπολογισμό ενός σταθμισμένου αθροίσματος των παραγόντων ισχύος (power spectrum coefficients), $|V(n)|^2$, των πλαισίων του σήματος. Το πλήθος των παραγόντων για κάθε πλαίσιο είναι:

$$\eta_{kn}; k = 0, 1, \dots, K - 1; n = 0, 1, \dots, N/2 \quad (2.8)$$

όπου N είναι το πλήθος των συντελεστών από τον μετασχηματισμό στο πεδίο της συχνότητας. Έτσι, οι συντελεστές του φάσματος *MEL* μπορούν να υπολογιστούν από τον τύπο:

$$G(k) = \sum_{n=0}^{N/2} \eta_{kn} |V(n)|^2 \quad k = 0, 1, \dots, K - 1 \quad (2.9)$$

Για τον μετασχηματισμό των $G(k)$ σε συντελεστές cepstral και λόγω της συμμετρίας του φάσματος *MEL*, εφαρμόζεται ένας μετασχηματισμός ημιτόνου:

$$c(q) = \sum_{k=0}^{K-1} \log(G(k)) \cos\left(\frac{\pi q(2k+1)}{2K}\right); \quad q = -, 1, \dots, Q-1 \quad (2.10)$$

όπου Q είναι το πλήθος συντελεστών cepstral στην κλίμακα MEL για το οποίο γίνεται υπολογισμός. Τέλος, υπολογίζονται η πρώτη και η δεύτερη παράγωγος των συντελεστών για να ενδυναμωθεί η αποτύπωση των χρονικών και δυναμικών χαρακτηριστικών του σήματος βάσει των εξισώσεων 2.11 και 2.12 αντιστοίχως:

$$\Delta c_m(q) = c_{m+\tau(q)} - c_{m-\tau(q)}; \quad q = 0, 1, \dots, Q-1 \quad (2.11)$$

$$\Delta\Delta c_m(q) = \Delta c_{m+\tau(q)} - \Delta c_{m-\tau(q)}; \quad q = 0, 1, \dots, Q-1 \quad (2.12)$$

όπου m είναι ο δείκτης της χρονικής στιγμής/χρονικού πλαισίου και τ είναι το πλήθος των χρονικών δεικτών για τους οποίους θα υπολογιστεί η παράγωγος.

Αν και η πληροφορία που εμπεριέχεται μέσα στους MFCC υπάρχει ήδη στο σήμα και μπορεί να εξαχθεί από έναν μετασχηματισμό Fourier (FT), με την χρήση του παραπάνω μεγέθους θα μπορούσε να γραφτεί ότι αυξάνονται οι πιθανότητες για καθαρότερη εμφάνιση των συσχετισμών μεταξύ εξωγενών χαρακτηριστικών (π.χ. των κατηγοριών/κλάσεων, όπως στην περίπτωση κατηγοριοποίησης) και φασματικών χαρακτηριστικών του σήματος [109, 111].

Ένα άλλο μέγεθος που χρησιμοποιείται για την υπολογιστική περιγραφή της χροιάς είναι οι στατιστικές ροπές του σήματος στο πεδίο της συχνότητας (ασυμμετρία και κύρτωση) καθώς και οι μέσοι όροι (centroid). Χρησιμοποιούνται για την υπολογιστική περιγραφή του σχήματος του φάσματος και, συνεπώς, της χροιάς του σήματος. Ο υπολογισμός τους γίνεται με την χρήση του σήματος στο πεδίο του χρόνου και την εφαρμογή των τύπων 2.13, 2.14 και 2.15 αντιστοίχως [108].

$$Sk_n = \frac{\frac{1}{F_{bin}} \sum_{f_{bin}=1}^{F_{bin}} (M_n(f_{bin}) - \overline{M_n})^3}{\left(\sqrt{\frac{1}{F_{bin}} \sum_{f_{bin}=1}^{F_{bin}} (M_n(f_{bin}) - \overline{M_n})^2} \right)^3} \quad (2.13)$$

$$K_n = \frac{1}{F_{bin}} \frac{\sum_{f_{bin}=1}^{F_{bin}} (M_n(f_{bin}) - \overline{M_n})^4}{\left(\frac{1}{F_{bin}} \sum_{f_{bin}=1}^{F_{bin}} (M_n(f_{bin}) - \overline{M_n})^2 \right)} \quad (2.14)$$

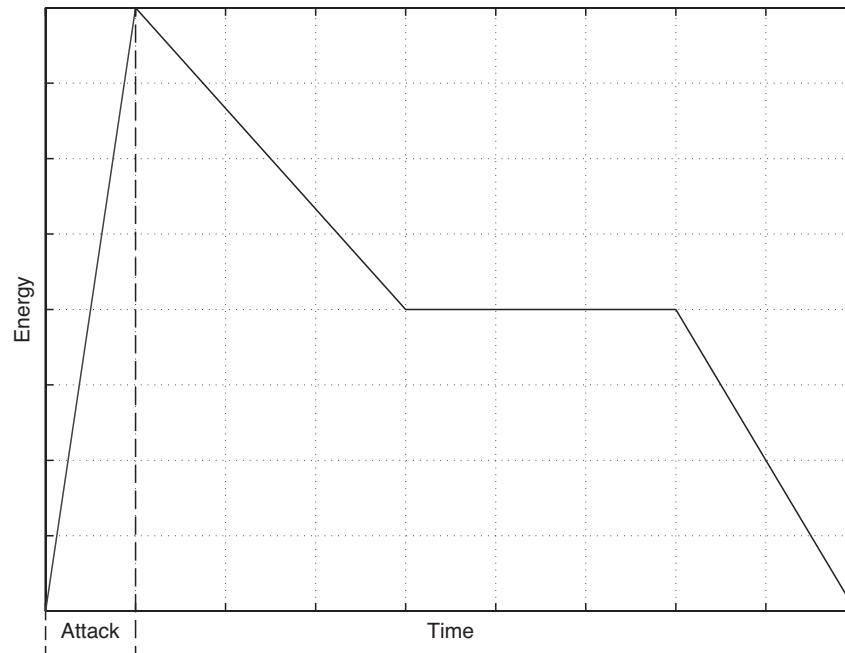
$$C_n = \frac{\sum_{f_{bin}=1}^{F_{bin}} M_n[f_{bin}] \times f_{bin}}{\sum_{f_{bin}=1}^{F_{bin}} M_n[f_{bin}]} \quad (2.15)$$

όπου F_{bin} είναι το πλήθος των συντελεστών μετασχηματισμού Fourier, M_n είναι το πλαίσιο n , $\overline{M_n}$ είναι η μέση τιμή των συντελεστών Fourier για το πλαίσιο M_n και Sk_n , K_n και G_n είναι η ασυμμετρία, η κύρτωση και το centroid για το πλαίσιο M_n .

Εάν οι τιμές της ενέργειας για κάθε ζώνη συχνοτήτων (frequency bin) θεωρηθούν ως σημεία μίας καμπύλης, τα στατιστικά αυτά μεγέθη δείχνουν το σχήμα αυτής. Συνεπώς, μπορούν να αποκαλύψουν την κατανομή ενέργειας ανά συχνοτικές περιοχές (π.χ., χαμηλές, μεσαίες και υψηλές συχνότητες). Έτσι, υψηλότερες τιμές του spectral centroid φανερώνουν την ύπαρξη περισσότερων υψηλών συχνοτήτων ενώ οι τιμές από την κύρτωση και την ασυμμετρία την κατανομή σε μεσαίες, χαμηλές και υψηλές.

Εκτός όμως από τα παραπάνω, συχνά χρησιμοποιούνται μεγέθη που σχετίζονται με την ατάκα (attack) του σήματος. Ως ατάκα ορίζεται το διάστημα από την αρχική (πιθανότητα και μηδενική) τιμή του σήματος μέχρι το πρώτο μέγιστο αυτού και μία αναπαράσταση αυτής φαίνεται στο **Σχήμα 2.8**. Για αυτό το σκοπό υπολογίζονται η μέση κλίση της αύξησης του πλάτους (slope), ο χρόνος που διαρκεί η ατάκα (attack time) καθώς και το άλμα (leap) αυτής, το οποίο ορίζεται ως η διαφορά μεταξύ της αρχικής και τελικής τιμής του σήματος κατά την ατάκα.

Ένα ακόμα μέγεθος που χρησιμοποιείται για τον υπολογισμό και ανίχνευση πληροφορίας σχετικά με την χροιά ενός σήματος είναι η σχέση συχνοτικών περιοχών/συχνότητας με την ενέργεια του σήματος. Υπάρχουν δύο μεγέθη που προσεγγίζουν αυτό το πρόβλημα, κάθε ένα από διαφορετική σκοπιά. Ένα είναι το roll off, το οποίο δείχνει την συχνότητα όπου μέχρι εκείνη υπάρχει συγκεκριμένο ποσοστό της ενέργειας του σήματος, και το άλλο είναι το brightness, το οποίο μετράει το ποσοστό ενέργειας πάνω από μία συγκεκριμένη



Σχήμα 2.8: Η ατάκα (attack) ενός ηχητικού σήματος. Ο οριζόντιος άξονας είναι ο χρόνος (Time) ενώ ο κάθετος η ενέργεια (Energy)

συχνότητα.

Έτσι, το roll off μπορεί να υπολογιστεί ως:

$$\sum_{f_{bin}=1}^{F_{bin}lim} M_n[f_{bin}] = X \times \sum_{f_{bin}=1}^{F_{bin}} M_n[f_{bin}] \quad (2.16)$$

όπου όλα τα σύμβολα όπως έχουν οριστεί προηγουμένως, $F_{bin}lim$ είναι το πλήθος των συντελεστών μετασχηματισμού Fourier και X το ποσοστό ενέργειας του σήματος που υπάρχει μέχρι την $F_{bin}lim$. Τα κύρια ποσοστά ενέργειας που χρησιμοποιούνται είναι 0.85 [108] και 0.95 [112].

Αν και το roll off είναι μία τεχνική περιγραφή της κατανομής της ενέργειας για ένα σήμα στο πεδίο της συχνότητας, το brightness ανταποκρίνεται και σε υποκειμενική αντίληψη [113, 114]. Μπορεί να μετρηθεί με τον απλό υπολογισμό του ποσοστού ενέργειας που υπάρχει στο σήμα πάνω από μία ορισμένη συχνότητα (π.χ., χρησιμοποιώντας την Εξίσωση 2.2). Αυτή η συχνότητα ορίζεται είτε ως $3000Hz$ [114] είτε ως $1000Hz$ [115].

Ένα ακόμα μέγεθος που σχετίζεται με την χροιά ενός ήχου είναι και η εκτίμηση της αισθη-

τηριακής διαφωνίας (sensory dissonance), που συνδέεται με την αίσθηση του διακροτήματος και συμβαίνει όταν ένα ζεύγος ημιτονοειδών σημάτων έχουν μικρή διαφορά στις συχνότητές τους. Περιγράφηκε αναλυτικά από τους Kameoka και Kuriyagawa [116] και, για την απλή περίπτωση δύο συχνοτήτων, μπορεί να υπολογιστεί ως:

$$D = \begin{cases} k_0(100^{\frac{2+\log((f_2-f_1)/f_1)}{2+\log(f_b/f_1)}}), & f_2 - f_1 \leq f_b, \text{ και} \\ k_0(90^{\frac{\log((f_2-f_1)/f_1)}{\log(f_b/f_1)}} + 10 + C_0), & f_2 - f_1 > f_b, \text{ και } f_2 \leq 2f_1 \end{cases} \quad (2.17)$$

όπου, C_0 και k_0 είναι σταθερές και ίσες με 65 και 1.0 αντιστοίχως, f_2 είναι η μεγαλύτερη εκ των δύο συχνοτήτων και f_1 η μικρότερη, ενώ η f_b είναι ίση με:

$$f_b = 2.27\left(\frac{L_1 - 57}{40}\right)f_1^{0.477} \quad (2.18)$$

όπου L_1 είναι στάθμη ακουστότητας της f_1 .

Τέλος, ένα ακόμα μέγεθος που μπορεί να περιγράψει την χροιά ενός σήματος είναι ο ρυθμός αλλαγής προσήμου των δειγμάτων του, γνωστό και ως zero crossing rate. Υπολογίζεται με την απλή καταμέτρηση του πλήθους των αλλαγών προσήμου των δειγμάτων ανά μονάδα χρόνου, όπως στην Εξίσωση 2.19

$$Z_{cr} = \frac{1}{N} \sum_{n=1}^N F\{S[n] \times S[n-1] < 0\} \quad (2.19)$$

όπου, N είναι ο συνολικός αριθμός δειγμάτων του σήματος, n είναι το εκάστοτε δείγμα και $F\{x\}$ είναι μία συνάρτηση που επιστρέφει 0 εάν το όρισμά της είναι ψευδές και 1 στην αντίθετη περίπτωση. Με το zeros crossing rate προσφέρεται μία εκτίμηση του πόσο θορυβώδες είναι ένα σήμα λόγω της μικρότερης περιόδου του θορύβου σε σχέση με των άλλων, ωφέλιμων, ήχων (π.χ. φωνή, μουσικό όργανο κτλ) [108]. Για παράδειγμα, ένας θόρυβος θα έχει μεγαλύτερο zero crossing rate από ότι η ομιλία.

Με τα παραπάνω μεγέθη μπορεί να γίνει μία εκτίμηση της αντιλαμβανόμενης ή/και αντικειμενικής χροιάς του σήματος. Έχουν χρησιμοποιηθεί και χρησιμοποιούνται στην εξόρυξη πληροφορίας από ηχητικά δεδομένα εν γένει, όμως, μέχρι τώρα, δεν έχει αναφερθεί κάποιος

απευθείας συσχετισμός τους με κάποια συναισθηματική διάσταση ή λεκτική περιγραφή συναισθήματος.

2.1.7.3 Μεγέθη που σχετίζονται με τα ρυθμικά χαρακτηριστικά του σήματος

Όπως συμβαίνει και με την χροιά, τα ρυθμικά χαρακτηριστικά ενός ηχητικού σήματος μπορούν εξίσου αποτελεσματικά να περιγραφούν από ένα σύνολο τεχνικών μεγεθών του ηχητικού σήματος. Βάσει κοινής εμπειρικής γνώσης θα μπορούσε να γραφτεί ότι μετά την ενέργεια ο ρυθμός είναι αρκετά σημαντικός για την μεταφορά συναισθήματος μέσω ηχητικών δεδομένων. Για παράδειγμα, σχεδόν όλοι έχουμε νιώσει να μεταβάλλεται η διέγερση ακούγοντας ένα γρήγορο, ρυθμικά, μουσικό κομμάτι και, συνήθως, αυτό έρχεται να το επιβεβαιώσει ένα κούνημα του χεριού ή του ποδιού. Το ίδιο συμβαίνει και με γενικούς ήχους αλλά και χρησιμοποιείται σε ανθρώπινες εκδηλώσεις, π.χ. γρήγορη και ρυθμική χρήση της κόρνας αμαξιών σε γάμους. Αυτή την συνήθη εμπειρική γνώση, δηλαδή ότι ο ρυθμός του ήχου δύναται να επηρεάσει την διέγερση του ακροατή, επιβεβαιώνουν δημοσιευμένες έρευνες. Συγκεκριμένα, έχει αναφερθεί απευθείας συσχέτιση του ρυθμού με την συναισθηματική κατάσταση της Διέγερσης αναφορικά με σήματα ομιλίας και μουσικής [30, 38].

Σε όλα τα προηγούμενα μεγέθη που παρουσιάστηκαν, ο υπολογισμός τους γινόταν μέσω της απευθείας χρήσης μίας εξίσωσης. Για τα μεγέθη όμως που σχετίζονται με τον ρυθμό αυτό διαφέρει. Τα τελευταία συνήθως υπολογίζονται μέσω μίας σειράς διαδικασιών και δεν μπορούν να περιγραφούν από μία εξίσωση. Από αυτά, ίσως το πιο κοινό και γνωστό είναι το τέμπο (tempo). Για τον υπολογισμό του θα πρέπει πρώτα να βρεθούν οι περιοδικότητες σε ένα ηχητικό σήμα και έπειτα, από αυτές, να υπολογιστεί αυτή με την μεγαλύτερη συχνότητα εμφάνισης. Έχουν προταθεί διάφορες υλοποιήσεις για αυτόν τον υπολογισμό. Σε κάθε περίπτωση, η τελική τιμή του υπολογισμού είναι το tempo του ηχητικού σήματος σε χτύπους ανά λεπτό (beats per minute, bpm).

Ένας ακόμα τρόπος για τον προσδιορισμό της ταχύτητας ενός μουσικού σήματος, tempo, είναι με τον υπολογισμό των εξάρσεων ενέργειας στο σήμα, που καλούνται onsets. Για τον υπολογισμό τους πρέπει να καταταμηθεί το σήμα σε χρονικά πλαίσια και για κάθε ένα θα πρέπει να εφαρμοστεί η Εξίσωση 2.1. Έτσι, αναλόγως την κατάτμηση, μπορεί να υπολογιστεί

η ενέργεια του σήματος σε διαδοχικές χρονικές στιγμές. Η καμπύλη που ενώνει αυτές τις τιμές καλείται *onsets curve* και μπορεί να δείξει σε ποιες χρονικές στιγμές υπάρχουν εξάρσεις ενέργειας, οι οποίες ανταποκρίνονται σε διαδοχικούς παλμούς του σήματος. Εάν εφαρμοστεί κάποιος αλγόριθμος επιλογής μεγίστων σε αυτή την καμπύλη, μπορεί να αποκαλύψει τις χρονικές στιγμές εμφάνισης νοτών ή κτύπων (*beats*). Ακόμα, ο υπολογισμός της μέσης συχνότητας των *onsets* καλείται πυκνότητα γεγονότων (*event density*). Συνίσταται στον υπολογισμό του μέσου πλήθους των *onsets* σε ένα δευτερόλεπτο και μπορεί να φανερώσει, σε κάποιο βαθμό, αφενός την ταχύτητα του ηχητικού σήματος και αφετέρου εάν υπάρχει πολύπλοκο ή όχι ρυθμικό σχήμα.

Ο ρυθμικός χαρακτήρας ενός ηχητικού σήματος μπορεί να αναπαρασταθεί και από την ρυθμική περιοδικότητα χρησιμοποιώντας την ομοιότητα (*similarity*) μεταξύ των χρονικών πλαισίων. Αυτό το μέγεθος συχνά καλείται και φάσμα κτύπων (*beat spectrum*). Για τον υπολογισμό της ομοιότητας χρησιμοποιείται η αναπαράσταση του συχνοτικού περιεχομένου του σήματος στην κλίμα *MEL*, χρησιμοποιώντας μόνο τους πρώτους 13 - 20 συντελεστές των *MFCC*, με εύρος χρονικού πλαισίου $23ms$ και με επικάλυψη πλαισίων ίση με 50%. Έπειτα, εάν \vec{c}_i και \vec{c}_j είναι δύο διανύσματα που περιέχουν τους συντελεστές *MFCC* από δύο χρονικά πλαίσια του σήματος, μπορεί να υπολογιστεί η ομοιότητα $sim(i, j)$, ως:

$$sim(i, j) = (\vec{c}_i - \vec{c}_j)^T (\vec{c}_i - \vec{c}_j) \quad (2.20)$$

Η αναπαράσταση της τιμής $sim(i, j)$ για κάθε συνδυασμό χρονικών πλαισίων ονομάζεται και πίνακας ομοιότητας (*similarity matrix*). Από το $sim(i, j)$ μπορεί να υπολογιστεί το άθροισμα για ένα χρονικό εύρος T και χρονικών διαφορών D (διαγωνίων στον πίνακα) σύμφωνα με την Εξίσωση 2.21. Το μέγεθος $BS_m(d)$ είναι το φάσμα κτύπων [117, 118].

$$BS_m(d) = - \sum_{i=mT}^{(m+1)T-1} d(i, i+d); \quad d = 0, 1, \dots, D-1 \quad (2.21)$$

Ακόμα, οι Pampalk, Rauber και Merkl [119] προτείνουν τη χρήση της εκτίμησης του φάσματος για ζώνες, έπειτα από μετασχηματισμό του τελευταίου βάσει του ακουστικού μοντέλου (*auditory model*), για τον προσδιορισμό της ρυθμικότητας ενός ηχητικού σήματος έτσι ώστε να ευρεθούν πληροφορίες σχετικά με το πόσο εμφανής και γρήγορος είναι ο ρυθμός ανά

ζώνη συχνοτήτων. Η ονομασία του παραγόμενου μεγέθους είναι διακύμανση (fluctuation). Σύμφωνα με δημοσιευμένες εργασίες [119, 120], ήχοι με διαφορετική τιμή για τη fluctuation παρουσιάζουν διαφορετικές ακουστικές ποιότητες αναφορικά με τα ρυθμικά τους χαρακτηριστικά. Για την περίπτωση ενός ημιτονοειδώς διαμορφωμένου θορύβου, το fluctuation ορίζεται ως:

$$F \sim \frac{\Delta L}{\frac{f_{mod}}{4Hz} + \frac{4Hz}{f_{mod}}} \quad (2.22)$$

όπου F είναι η fluctuation, ΔL είναι η διαφορά της μέγιστης και ελάχιστης στάθμης και f_{mod} είναι η συχνότητα διαμόρφωσης.

Αν και η fluctuation μπορεί να δώσει μία αίσθηση διακύμανσης, σε πολύ υψηλές τιμές του ταυτίζεται με την τραχύτητα ή, αλλιώς, roughness, η οποία αναφέρθηκε στην Ενότητα 2.1.7.2.

2.1.7.4 Μεγέθη που σχετίζονται με το τονικό ύψος του σήματος

Η τελευταία ομάδα χαρακτηριστικών που παρουσιάζονται στον Πίνακα 2.2 είναι αυτή του τονικού ύψους. Κατά το καλύτερο της γνώσης του συγγραφέα της ανά χείρας διατριβής, και κατά την περίοδο συγγραφής της τελευταίας, δεν έχει δημοσιευθεί κάποια εργασία που να μελετά αποκλειστικά την σχέση αυτής της ομάδας τεχνικών μεγεθών με το εγειρόμενο συναίσθημα στον ακροατή. Παρ' όλα αυτά, τα τεχνικά μεγέθη που μπορούν να κατηγοριοποιηθούν σε αυτή την ομάδα χρησιμοποιούνται σε έρευνες στον τομέα της εξόρυξης συναισθηματικής πληροφορίας από ηχητικά δεδομένα. Το βασικότερο και, ίσως, το πιο ευρέως και εμπειρικά γνωστό είναι το καθαυτό τονικό ύψος ενός ήχου. Καλείται pitch και ισούται με την θεμελιώδη συχνότητα ενός ήχου.

Με την εύρεση της θεμελιώδους συχνότητας γίνεται εφικτός ο υπολογισμός της μη αρμονικότητας (inharmonicities). Το μέγεθος αυτό ισούται με την ενέργεια που υπάρχει στις συχνότητες που δεν είναι ακέραια πολλαπλάσια της θεμελιώδους. Έτσι μπορεί να υπολογιστεί η ενέργεια που περιέχεται στα ηχητικά δεδομένα και συμβάλει αντίθετα στην δημιουργία αίσθησης αρμονίας.

Τέλος, με το μέγεθος Harmonic Change Detect Function (HCDF) είναι δυνατός ο υπολογισμός των αλλαγών του τονικού μέσου. Αυτό γίνεται με την χρήση ενός διανύσματος, ζ , που περιέχει τα τονικά κέντρα (centroid) για το χρονικό πλαίσιο, n , του σήματος και ένα πίνακα μετασχηματισμού Φ :

$$\zeta_n(d) = \frac{1}{\|c_n\|_1} \sum_{l=0}^{11} \Phi(d, l) c_n(l), \quad 0 \leq d \leq 5, \quad 0 \leq l \leq 11 \quad (2.23)$$

όπου το l είναι ο δείκτης για την κλάση του τονικού ύψους (ένας για κάθε διαφορετική νότα) και το d είναι ο δείκτης της διάστασης του ζ . Ο πίνακας μετασχηματισμού Φ ορίζεται ως:

$$\Phi = [\phi_0, \phi_1, \dots, \phi_{11}] \quad (2.24)$$

και:

$$\phi(l) = \begin{bmatrix} \Phi(0, l) \\ \Phi(1, l) \\ \Phi(2, l) \\ \Phi(3, l) \\ \Phi(4, l) \\ \Phi(5, l) \end{bmatrix} = \begin{bmatrix} \sin l \frac{7\pi}{6} \\ \cos l \frac{7\pi}{6} \\ \sin l \frac{3\pi}{2} \\ \cos l \frac{3\pi}{2} \\ 0.5 \sin l \frac{2\pi}{3} \\ 0.5 \cos l \frac{2\pi}{3} \end{bmatrix} \quad (2.25)$$

Έτσι, το HCDF (ξ_n) υπολογίζεται ως:

$$\xi_n = \sqrt{\sum_{d=0}^5 [\zeta_{n+1}(d) - \zeta_{n-1}(d)]^2} \quad (2.26)$$

Καταλήγοντας, με τα παραπάνω τεχνικά μεγέθη είναι δυνατός ο υπολογισμός των θεμελιωδών χαρακτηριστικών του σήματος, ήτοι: α) ενέργεια, β) ρυθμός, γ) χροιά, και δ) τονικό ύψος/τονικότητα [24]. Οι τιμές αυτών αφενός περιγράφουν ποσοτικά τα τεχνικά χαρακτηριστικά του ηχητικού σήματος και αφετέρου επιτρέπουν την χρήση των τελευταίων ως δεδομένα εισόδου στην διαδικασία που περιγράφεται από το **Σχήμα 2.1** και συγκεκριμένα από

το στοιχείο με όνομα “Χαρακτηριστικά δεδομένων”.

2.2 Στάθμη της τεχνικής

Όπως αναφέρθηκε, η έρευνα σχετικά με την αναγνώριση συναισθήματος αφενός είναι νέο πεδίο και αφετέρου σχετίζεται με την έρευνα στο πεδίο της αναγνώρισης συναισθημάτων από μουσική. Το τελευταίο είναι παλαιότερο από το πρώτο, γεγονός αναμενόμενο εάν κανείς αναλογιστεί αφενός την σημαντικότητα της μουσικής ως ένα μέσο και δίοδο πολιτισμικής και καλλιτεχνικής έκφρασης αλλά και την πρώτη χρήση της. Δηλαδή, την επαύξηση της συναισθηματικής επήρειας της ομιλίας [14, 17]. Ως εκ τούτου, υπάρχουν αρκετές δημοσιευμένες εργασίες που εστιάζουν στην εξόρυξη συναισθηματικής πληροφορίας από μουσική. Σε αντίθεση, στο πεδίο της αναγνώρισης συναισθήματος από ηχητικά γεγονότα υπάρχουν λίγες δημοσιευμένες εργασίες (στις οποίες συγκαταλέγονται και αυτές του συγγραφέα του ανά χειρας τόμου) και ακόμα δεν διαφαίνεται κάποια ισχυρή σύνδεση μεταξύ τους [13]. Παρόλα αυτά, η μουσική αποτελεί μία μορφή ήχου, με τις υπόλοιπες να είναι η ομιλία και οι γενικοί ήχοι. Οι τελευταίοι έχουν χρησιμοποιηθεί και ως δομικά συστατικά της καλλιτεχνικής έκφρασης από την εποχή του κινήματος της *Musique Concrète*, με κύριο εκφραστή τον Pierre Schaeffer, και την ηλεκτροακουστική μουσική, όπως αυτή είναι στις μέρες μας [121, 122]. Για αυτούς τους λόγους, στην εκπόνηση της έρευνας που περιγράφεται από την παρούσα διατριβή έγινε χρήση αρκετών ευρημάτων από το ερευνητικό πεδίο της εξόρυξης πληροφορίας, εν γένει, αλλά και συναισθηματικής, ειδικά, από μουσική. Για την σφαιρική παρουσίαση των διαδικασιών που σχετίζονται με την παρούσα διατριβή πρώτα παρουσιάζεται η ανάγκη-ση μουσικής πληροφορίας, το ερευνητικό πεδίο της ανάκτησης πληροφορίας από μουσική (*Music Information Retrieval, MIR*) και εφαρμογές αυτού, ακολουθούμενα από μία επισκόπηση των εργασιών του *MIR* που εστιάζουν στην εξόρυξη συναισθηματικής πληροφορίας. Στην επισκόπηση αυτών των εργασιών, δίνεται πρώτα μία συγκεντρωτική παρουσίαση των εργασιών αναφορικά με τα είδη μουσικής που χρησιμοποιήθηκαν και τεχνικά χαρακτηριστικά, την χρήση συναισθηματικών μοντέλων και, αντίστοιχης, συναισθηματικής επισήμανσης και, τέλος, των αλγορίθμων κατηγοριοποίησης που χρησιμοποιήθηκαν. Την παρουσίαση αυτή ακολουθεί ανάλυση σχετικά με την στάθμη της τεχνικής στην αναγνώριση συναισθήματος από μουσική. Έπειτα, γίνεται μία επισκόπηση των δημοσιευμένων εργασιών οι οποίες εξε-

τάζουν στην αναγνώριση συναισθήματος από ομιλία και η παρούσα ενότητα καταλήγει με την παρουσίαση της υφιστάμενης στάθμης της τεχνικής πάνω στο πεδίο αναγνώρισης συναισθημάτων από ηχητικά γεγονότα με την παρουσίαση του ορισμού του ηχητικού γεγονότος βάσει σύγχρονων δημοσιευμένων εργασιών, των ελεύθερα διαθέσιμων σετ συναισθηματικά επισημασμένων ηχητικών γεγονότων και την επισκόπηση των δημοσιευμένων εργασιών πάνω στην αναγνώριση συναισθήματος από ηχητικά γεγονότα.

2.2.1 Ανάκτηση μουσικής πληροφορίας, το ερευνητικό πεδίο MIR και εφαρμογές αυτού

Η ανάκτηση μουσικής πληροφορίας είναι μία εξειδίκευση της διαδικασίας ανάκτησης πληροφορίας (Information Retrieval) [123]. Για την τελευταία διαδικασία, υπάρχει ένας χρήστης και ένα σύστημα που υλοποιεί το έργο της ανάκτησης. Ο πρώτος, αλληλεπιδρά με το σύστημα για να προσπελάσει πληροφορία σχετικά με το αντικείμενο που τον ενδιαφέρει. Για αυτό το λόγο, το σύστημα θα πρέπει να λαμβάνει υπόψη τις προτιμήσεις του χρήστη αναφορικά με αυτό που επιθυμεί να βρει αλλά και αναφορικά με την διαδικασία αναζήτησης [124]. Στην ανάκτηση μουσικής πληροφορίας εστιάζει το ερευνητικό πεδίο MIR. Ένα διεπιστημονικό πεδίο με συνεισφορά από τεχνολογικές και μη επιστήμες, όπως Ψυχολογία, Γνωσιακές Επιστήμες (Cognitive Sciences), Μουσικολογία και θεωρητική Μουσική, Επιστήμη των Υπολογιστών κ.α. [123]. Άρχισε να ωριμάζει κατά το 1990 και ο κύριος στόχος του είναι να προσφέρει την δυνατότητα σε όλους να έχουν άμεση και απρόσκοπτη πρόσβαση σε μουσική πληροφορία [123, 125]. Η ανάγκη αυτού του στόχου μπορεί να αναδειχθεί περισσότερο εάν κανείς αναλογιστεί την πληθώρα ψηφιακών, και μη, μουσικών δεδομένων που υπάρχουν. Με την έξαρση της τεχνολογίας η οποία κατέστησε προσβάσιμα εξελιγμένα εργαλεία παραγωγής μουσικής σε οικιακά περιβάλλοντα αλλά και με την εξάπλωση της χρήσης της ψηφιακής μορφής ήχου και του διαδικτύου, παρατηρείται ένα μεγάλο πλήθος μουσικών δεδομένων διαθέσιμων από πολλές διαφορετικές πηγές και υπηρεσίες, π.χ. διαδίκτυο, ψηφιακά μέσα αποθήκευσης, υπηρεσίες παροχής μουσικής κατ' αίτηση (on demand) κ.α. Ήδη υπάρχουν διαθέσιμες εφαρμογές οι οποίες μπορούν να ανακτήσουν μουσική πληροφορία βασιζόμενες όχι σε τυπικά χαρακτηριστικά, για παράδειγμα όνομα τραγουδιού ή είδος μουσικής, αλλά σε αναπαράσταση της μελωδίας με μурμουρητό⁷ ή βάσει τις προτιμήσεις του χρήστη, π.χ. τι

⁷<http://www.midomi.com>

τραγούδια άκουσε πριν και ποιο από αυτά του άρεσε⁸. Εν γένει, οι εφαρμογές του MIR εστιάζουν στην δημιουργία της δυνατότητας για την προσπέλαση και ανάκτηση μουσικής πληροφορίας με τρόπους που να είναι διασθητικοί, εύκολοι για τον χρήστη και, στο μέτρο του δυνατού, έξυπνοι (intelligent). [123, 125].

Οι βασικές προσεγγίσεις του MIR για την ανάκτηση πληροφορίας είναι δύο: α) αναζήτηση βάσει μετά-δεδομένων, και β) αναζήτηση βάσει περιεχομένου [123]. Η πρώτη πρόκειται, κυρίως, για διαχείριση πληροφορίας σε μορφή κειμένου. Δηλαδή, αναζήτηση, διαχείριση και συσχέτιση πληροφοριών όπως όνομα και ιδιότητες καλλιτεχνών (π.χ. τραγουδιστής, συνθέτης), είδη μουσικής (π.χ. ονόματα και συσχετίσεις ειδών), αναζήτηση στίχων και εν γένει διαχείριση πληροφορίας που περιγράφει τα μουσικά δεδομένα ή υπάρχει σε μορφή κειμένου για τα μουσικά δεδομένα. Η δεύτερη αφορά την ανάκτηση και διαχείριση πληροφορίας που υπάρχει εγγενώς στα μουσικά αρχεία ήχου. Δηλαδή, αναζήτηση πληροφορίας βάσει των χαρακτηριστικών του ήχου που περιγράφηκαν στην προηγούμενη ενότητα, ήτοι ρυθμός, ενέργεια, χροιά, τονικό ύψος/τονικότητα, τα οποία είναι και αυτά που χρησιμοποιεί και ο άνθρωπος για να εξάγει πληροφορία από τα ηχητικά σήματα [123]. Ένα είδος πληροφορίας που μπορεί να εξαχθεί από τα μουσικά δεδομένα είναι και το συναίσθημα το οποίο έχει αρκετή συσχέτιση με τα τεχνικά χαρακτηριστικά του ήχου [126].

2.2.2 Αναγνώριση συναισθήματος από μουσική

Η καθιερωμένη αναζήτηση μουσικής πληροφορίας βάσει του σχήματος “Καλλιτέχνης-Έτος-Δίσκος” είναι σχεδόν παραχωμένη μιας και εστιάζει στον δημιουργό. Μία από τις εναλλακτικές προτάσεις για προσπέλαση, αναζήτηση και ανάκτηση μουσικής πληροφορίας είναι η χρήση δεδομένων συναισθηματικού χαρακτηρισμού της μουσικής, τα οποία εστιάζουν σε πληροφορία που ενυπάρχει στο σήμα και που περιγράφει το περιεχόμενο (content oriented) [24, 127]. Ως εκ τούτου, μία αναζήτηση για μουσική που εγείρει συγκεκριμένα συναισθήματα δεν απαιτεί την γνώση για το όνομα ή την ονομασία των τραγουδιών του εκάστοτε καλλιτέχνη. Έτσι, ο ακροατής μπορεί να προσπελάσει εύκολα πληροφορία που, εάν υποτεθεί ότι δεν γνώριζε κάποιον συγκεκριμένο καλλιτέχνη, δεν θα είχε αλλιώς την δυνατότητα βάσει του τυπικού σχήματος κατηγοριοποίησης που αναφέρθηκε στην αρχή αυτής της ενότητας. Την αναγνώ-

⁸<https://www.spotify.com>

ριση και εξαγωγή συναισθηματικής πληροφορίας από μουσική έχουν ως κύριο ενδιαφέρον πληθώρα ερευνητικών πεδίων. Τα πεδία αυτά μπορούν να διαχωριστούν σε δύο αφαιρετικές κατηγορίες, δηλαδή: α) Συσχετιζόμενα με τα τεχνικά χαρακτηριστικά του σήματος, και β) Συσχετιζόμενα με τις ανθρώπινες φυσιολογικές (physiological) αντιδράσεις [29].

Στην πρώτη ομάδα τα κυρίως πεδία είναι το MIR και το πεδίο αναγνώρισης συναισθήματος από μουσική (Music Emotion Recognition, MER) [3]. Η κύρια ενασχόλησή τους είναι η εύρεση του τρόπου με τον οποίο θα μπορούσε να εξαχθεί και να προβλεφθεί με ακρίβεια το εγχειρόμενο συναίσθημα στον δέκτη κατά την ακρόαση μουσικής, χρησιμοποιώντας τεχνικά μεγέθη για την περιγραφή των μουσικών κομματιών και αλγόριθμους κατηγοριοποίησης για την δημιουργία και επαλήθευση μοντέλων ταξινόμησης [24, 29]. Τα παραγόμενα αποτελέσματα δείχνουν αφενός την ακρίβεια των μοντέλων που παράχθηκαν και αφετέρου αναδεικνύουν τα τεχνικά χαρακτηριστικά του σήματος τα οποία μπορούν να χρησιμοποιηθούν για τον σκοπό αυτό. Στην δεύτερη ανήκουν πεδία όπως αυτό της Νευροεπιστήμης, την Ψυχολογίας και της Ψυχιατρικής. Η έρευνα των πεδίων αυτών πάνω στο συναίσθημα εστιάζει κυρίως στην τεκμηρίωση της φύσης του συναισθήματος, στην δημιουργία μοντέλων συναισθημάτων, στις φυσιολογικές αντιδράσεις του ανθρώπου κατά την βίωση συναισθημάτων και την συσχέτιση των αντιδράσεων αυτών με συγκεκριμένα συναισθήματα κατά την ακρόαση μουσικής [128, 129]. Στην συνέχεια, η ενότητα θα επικεντρωθεί στα πεδία MIR και MER και θα παρουσιαστούν τα ευρήματα εκείνα τα οποία αφενός αναδεικνύουν την βίωση διαφορετικών συναισθημάτων από διαφορετικά μουσικά ηχητικά ερεθίσματα και αφετέρου προσφέρουν αποτελέσματα που μπορούν να χρησιμοποιηθούν και χρησιμοποιήθηκαν στην έρευνα για την εξόρυξη ίδιας πληροφορίας από ηχητικά γεγονότα. Μέσω της επισκόπησης των δημοσιευμένων εργασιών θα ερευνηθούν διάφορα ερευνητικά ερωτήματα, όπως το: “γιατί βιώνονται συναισθήματα κατά την ακρόαση μουσικής”. Δηλαδή, ποιοι ψυχολογικοί και βιολογικοί μηχανισμοί εμπλέκονται στην διαδικασία έγερσης συναισθημάτων κατά την ακρόαση μουσικής και ποια τεχνικά χαρακτηριστικά μπορούν να επηρεάσουν την συναισθηματική κατάσταση του ακροατή. Ως αρχή επιλέχθηκε το άρθρο του Juslin [38] λόγω του ότι είναι μία ερευνητική εργασία επισκόπησης της τότε (2008) υφιστάμενης βιβλιογραφίας. Σε αυτό το άρθρο αναδεικνύονται οι βαθύτεροι και υποκειμενικοί μηχανισμοί που εμπλέκονται στην δημιουργία ή/και μεταφορά συναισθημάτων κατά την ακρόαση μουσικής και ομιλίας. Οι βασικότεροι εξ αυτών είναι:

- Αντανακλαστικά εγκεφαλικού στελέχους (brainstem reflexes), και αφορά την διαδικασία κατά την ένα ή περισσότερα τεχνικά χαρακτηριστικά του ηχητικού σήματος προσλαμβάνονται από τον εγκέφαλο ως ενδείξεις σημαντικών ή επειγόντων συμβάντων
- Αξιολογούμενη συνθήκη (evaluate conditioning), το οποίο συμβαίνει όταν κάποιο χαρακτηριστικό του ηχητικού σήματος έχει συνδεθεί με πρότερα βιώματα, αναμνήσεις και συναισθήματα και έτσι τα επαναφέρει
- Συναισθηματική μετάδοση (emotional contagion), όπου το λαμβανόμενο συναίσθημα του ηχητικού ερεθίσματος γίνεται αντικείμενο μίμησης από τον ακροατή και έτσι τον οδηγεί στο να μετέχει της ίδιας συναισθηματικής κατάστασης
- Οπτικοποίηση (visual imagery), όπου ο ακροατής βιώνει ένα συναίσθημα επειδή το ηχητικό ερέθισμα τον οδηγεί στο να οπτικοποιήσει συγκεκριμένες εικόνες οι οποίες του προκαλούν την βίωση συναισθημάτων
- Επεισοδιακή μνήμη (episodic memory), όπου το ηχητικό ερέθισμα εγείρει στον ακροατή συγκεκριμένες μνήμες, επαναφέρονται και βιώνοντας τα συναισθήματα που είχε κατά εκείνες τις στιγμές, και
- Μουσικό προσδόκιμο (musical expectancy), όπου ένα συγκεκριμένο χαρακτηριστικό του ηχητικού ερεθίσματος αλληλεπιδρά (π.χ. καταργεί, επιβεβαιώνει κτλ) με την προσδοκία που έχει ο ακροατής για την εξέλιξη του ερεθίσματος αυτού

Σύμφωνα με τα παραπάνω, γίνεται φανερό πως κατά την έγερση συναισθημάτων στον ακροατή από ηχητικά ερεθίσματα εμπλέκονται τρεις κύριοι παράγοντες, οι οποίοι μπορούν να επιδρούν αυτόνομα ή και σε συνδυασμό με άλλους, δηλαδή:

1. Το σήμα καθαυτό, συμπεριλαμβανομένων και των τεχνικών χαρακτηριστικών του, π.χ. στα αντανακλαστικά του εγκεφαλικού στελέχους
2. Το σημασιολογικό περιεχόμενο του σήματος, του οποίου η εμπλοκή μπορεί να φανεί καθαρά στους μηχανισμούς: αξιολογούμενη συνθήκη, συναισθηματική μετάδοση και οπτικοποίηση
3. Το πολιτιστικό, πολιτισμικό και κοινωνικό υπόβαθρο του ακροατή, όπου επιδρά στους

μηχανισμούς: οπτικοποίηση, επεισοδιακή μνήμη και μουσικό προσδόκιμο

Από τους παράγοντες αυτούς, ο κύρια ερευνηθείς είναι ο πρώτος μιας και οι άλλοι δύο δεν μπορούν να αξιοποιηθούν σε μία έρευνα εξόρυξης συναισθηματικής πληροφορίας η οποία θα μπορεί να παρέχει αποτελέσματα αξιοποιήσιμα σε παγκόσμια και ευρεία κλίμακα ανθρώπων. Ο κύριος λόγος για το τελευταίο είναι η μεγάλη διαφοροποίηση της υποκειμενικής κρίσης για το σημασιολογικό περιεχόμενο, π.χ. όπως στην περίπτωση του κρύου θυμού (cold anger) που εξαρτάται από τις κοινωνικές συμβάσεις του ακροατή κατά την ερμηνεία του σημασιολογικού περιεχομένου [130, 131], και η μεγάλη διαφορετικότητα των πολιτιστικών και πολιτισμικών προφίλ ανά την υφήλιο. Αναφορικά με τις εργασίες που αξιοποιούν μόνο τα τεχνικά χαρακτηριστικά, μπορεί να παρατηρηθεί ένα μοτίβο το οποίο περιγράφεται στο **Σχήμα 2.1** και χρησιμοποιείται ευρέως και σχεδόν αποκλειστικά. Έτσι, οι περισσότερες δημοσιευμένες εργασίες χρησιμοποιούν συναισθηματικά προ-επισημασμένα ηχητικά δεδομένα από τα οποία εξάγουν ένα πλήθος τεχνικών μεγεθών. Τα τελευταία τα χρησιμοποιούν ως είσοδο μαζί με την συναισθηματική επισήμανση στους αλγόριθμους μηχανικής μάθησης όπου και λαμβάνουν ένα μοντέλο κατηγοριοποίησης. Η προσαρμογή και υιοθέτησή αυτής της διαδικασίας γίνεται σχεδόν από όλες της δημοσιευμένες εργασίες πάνω στον τομέα της εξόρυξης συναισθηματικής πληροφορίας από ηχητικά δεδομένα και το τελικό αποτέλεσμα αξιολογείται από την ευστοχία του μοντέλου που παράχθηκε. Έτσι, υπάρχουν έρευνες που χρησιμοποιούν τα διαστατικά μοντέλα συναισθημάτων και έρευνες που χρησιμοποιούν τα διακριτά. Στις πρώτες μπορούν να βρεθούν και έρευνες που ενώ μεν χρησιμοποιούν διαστατικά μοντέλα, εντούτοις αντιστοιχούν τις περιοχές του μοντέλου σε διακριτά συναισθήματα και πραγματοποιούν την κατηγοριοποίηση βάσει των τελευταίων. Σε κάθε περίπτωση όμως, η ακολουθούμενη διαδικασία είναι όμοια με εκείνη που περιγράφεται στο **Σχήμα 2.1** και αναφέρθηκε προηγουμένως.

Στην αναγνώριση συναισθήματος από μουσική τα ηχητικά δεδομένα είναι ολόκληρα μουσικά κομμάτια ή αποσπάσματα από αυτά. Αυτά μπορούν να είναι ένα συγκεκριμένο είδος, π.χ. κλασική/λόγια μουσική, ποπ και jazz μουσική [28, 127, 132]. Ακόμα, μπορούν να προέρχονται από μία συγκεκριμένη χώρα ή γεωγραφική τοποθεσία, π.χ. Κίνα και Ιαπωνία [132], ή να μην ανήκουν σε κάποια από τις παραπάνω κατηγορίες [133]. Επίσης, στην εργασία του Shiliang [127] αναφέρεται και η χρήση μουσικών βίντεο από το διεθνές κανάλι μουσικής Music Television (MTV). Στον **Πίνακα 2.3** παρατίθεται μία λίστα με τις δημοσιευμένες

Πίνακας 2.3: Το είδος μουσικής που ανήκουν τα ηχητικά δεδομένα των εργασιών που αναφέρονται και έχουν ως αντικείμενο την αναγνώριση συναισθήματος από μουσική

Δείκτης έρευνας	Είδος μουσικής
[28, 96, 134, 135]	Κλασική
[127, 133, 136, 137, 138, 139, 98]	pop
[9, 95, 96, 132, 133, 135, 140, 141, 142]	Άλλο, π.χ. jazz, ambient, soundtracks κτλ., ή άγνωστο/δεν αναφέρεται

εργασίες και το είδος μουσικής που ανήκουν τα δεδομένα που αυτές χρησιμοποίησαν. Οι εργασίες αυτές ανήκουν στον τομέα της αναγνώρισης συναισθήματος. Για λόγους ομαδοποίησης, στον προαναφερθέν Πίνακα δεν αναφέρονται αναλυτικά τα υπό-είδη μουσικής. Για πιο αναλυτικές πληροφορίες σχετικά με τα μουσικά δεδομένα που χρησιμοποιήθηκαν από τις εργασίες που αναφέρονται ο αναγνώστης ενθαρρύνεται να ανατρέξει στην αντίστοιχη δημοσιευμένη εργασία.

Τα τεχνικά χαρακτηριστικά του μουσικού σήματος που εξάγονται ανήκουν σε όλες τις κατηγορίες που αναφέρθηκαν στην Ενότητα 2.1.7. Στον **Πίνακα 2.3**, στον Πίνακα 2.4 συνοψίζονται τα τεχνικά χαρακτηριστικά που εξήχθησαν από τις ίδιες δημοσιευμένες εργασίες. Όπως μπορεί να παρατηρηθεί, σχεδόν καμία εργασία δεν αξιοποιεί χαρακτηριστικά που ανήκουν αποκλειστικά σε μία ομάδα, π.χ. ενέργεια ή ρυθμός. Οι περισσότερες χρησιμοποιούν χαρακτηριστικά από όλες τις ομάδες που φαίνονται στον **Πίνακα 2.2**. Επιπλέον, μερικές έρευνες δεν αναφέρουν ακριβώς τι τεχνικά χαρακτηριστικά χρησιμοποίησαν αλλά αναφέρουν το λογισμικό που χρησιμοποιήθηκε για την εξαγωγή των τεχνικών χαρακτηριστικών και έτσι είναι δυνατός ο προσδιορισμός των ευρύτερων κατηγοριών (βάσει του **Πίνακα 2.2**) που ανήκουν τα χαρακτηριστικά. Κάποιες άλλες εργασίες χρησιμοποίησαν μουσικά δεδομένα σε συμβολική μορφή, ήτοι M.I.D.I., και έτσι είχαν την δυνατότητα να εξάγουν στοιχεία που αφορούν δομικά μέρη της μουσικής, όπως νότες, ταχύτητα (velocity), αριθμό νοτών, αριθμό οργάνων, vibrato κτλ. [9]. Για περαιτέρω πληροφορίες ο αναγνώστης ενθαρρύνεται να ανατρέξει στις αντίστοιχες εργασίες.

Πίνακας 2.4: Ο δείκτης των δημοσιευμένων εργασιών που αναφέρονται στην παρούσα διατριβή (Δ.Ε.), τα τεχνικά χαρακτηριστικά που χρησιμοποίησαν (Τ.Χ.) καθώς και την κατηγορία που ανήκουν τα τελευταία (Κατηγορία Τ.Χ.)

Δ.Ε.	Όνομα Τ.Χ.	Κατηγορία Τ.Χ.
[28, 95, 127, 133, 134, 136, 98, 141, 142]	Ενέργεια ανά χρονικό πλαίσιο / Μέση ενέργεια	Ενέργεια
[9, 139]	Γενικά / Δεν αναφέρεται συγκεκριμένο	
[28, 95, 134]	Brightness	
[28, 134, 137, 142]	Roll off	
[95, 28, 134, 136, 137, 140, 142]	Στατιστικά μεγέθη φάσματος	Χροιά
[28, 134, 137, 140]	MFCC	
[127, 140]	zero crossing rate	
[142]	Roughness	
[9, 96, 139]	Γενικά / Δεν αναφέρεται συγκεκριμένο	
[28, 127, 134, 135, 136, 138, 98, 140, 141, 142]	Tempo	
[28, 127, 133, 134, 142]	Rhythm regularity	Ρυθμός
[28, 127, 134]	Beat strength	
[98]	Beat spectrum	
[135]	Event density	
[135]	Fluctuation	
[140]	Onsets	
[96, 139]	Γενικά / Δεν αναφέρεται συγκεκριμένο	
[127, 135, 136, 95, 140, 141]	Pitch	
[95, 98]	Inharmonicity	Τονικότητα / Pitch

Πίνακας 2.4 – Συνέχεια από προηγούμενη σελίδα

Δ.Ε.	Όνομα Τ.Χ.	Κατηγορία Τ.Χ.
[140]	HCDF	
[9, 96, 139]	Γενικά / Δεν αναφέρεται συγκεκριμένο	
[135, 141]	Δομικά μουσικά	M.I.D.I.

Τα εξαγόμενα χαρακτηριστικά χρησιμοποιούνται σε συνδυασμό με την συναισθηματική επισήμανση για την δημιουργία του μοντέλου κατηγοριοποίησης. Η επισήμανση αυτή προσφέρει την αναγκαία πληροφορία για την κατηγοριοποίηση των ηχητικών δεδομένων σε κλάσεις/συναισθήματα και χρησιμοποιείται στην διαδικασία εκπαίδευσης των αλγορίθμων. Η επισήμανση αυτή συνήθως λαμβάνεται από υποκειμενικά πειράματα αξιολόγησης όπου οι συμμετέχοντες αποδίδουν στα δεδομένα τις επισημάνσεις που αφενός ταιριάζουν στο δικό τους εγχειρόμενο συναίσθημα και αφετέρου ορίζονται και μπορούν να χρησιμοποιηθούν από το υιοθετημένο μοντέλο συναισθημάτων της εκάστοτε εργασίας. Δηλαδή, σε εργασίες που έχει χρησιμοποιηθεί ένα μοντέλων διακριτών συναισθημάτων, η συναισθηματική επισήμανση περιέχει τις λεκτικές περιγραφές των συναισθημάτων αυτών. Αυτές οι περιγραφές μπορούν να χρησιμοποιηθούν ως διαφορετικές κλάσεις στην διαδικασία κατηγοριοποίησης. Από την άλλη, σε εργασίες που γίνεται χρήση ενός διαστατικού μοντέλου μπορεί να χρησιμοποιηθούν οι τιμές για κάθε διάσταση του μοντέλου, π.χ. τιμή για την Διέγερση και τιμή για την Ευχαρίστηση. Στην διεθνή επιστημονική βιβλιογραφία αναφέρονται τόσο προσεγγίσεις όπου χρησιμοποιούν υφιστάμενες επισημάνσεις [140, 142] όσο και εργασίες κατά τις οποίες πραγματοποιήθηκε επισήμανση ειδικά για τους σκοπούς τις εκάστοτε έρευνας [9, 28, 95]. Αναφορικά με τα προηγούμενα στάδια που παρουσιάζονται στο **Σχήμα 2.1** και χρησιμοποιήθηκαν για την παρουσίαση της βιβλιογραφίας, έτσι και για την επισήμανση δίνεται μία συνοπτική παρουσίαση της χρήσης επισημάνσεων στον **Πίνακα 2.5** όπως αυτή μπορεί να αποδελτιωθεί από την διεθνή βιβλιογραφία που χρησιμοποιήθηκε στην παρούσα διατριβή. Ακόμα, στον **Πίνακα 2.5** παρουσιάζεται και η κατηγορία συναισθηματικού μοντέλου που χρησιμοποιήθηκε.

Για την διαδικασία κατηγοριοποίησης οι αλγόριθμοι που χρησιμοποιούνται υπακούουν στην ταξινόμηση που αναφέρθηκε στην Ενότητα 2.1.2. Όπως μπορεί να φανεί και από τον **Πίνα-**

Πίνακας 2.5: Ο δείκτης των δημοσιευμένων εργασιών που αναφέρονται στην παρούσα διατριβή (Δ.Ε.), ένδειξη για την χρήση προηγούμενης συναισθηματικής επισήμανσης και το συναισθηματικό μοντέλο που χρησιμοποιήθηκε

Δ.Ε.	Προ-υπάρχουσα επισήμανση	Συναισθηματικό μοντέλο
[95, 96, 132, 138]	Όχι	Διάκριτα συναισθήματα
[140, 142]	Ναι	
[28, 133, 134, 136, 139, 141]	Όχι	Διαστατικό με διάκριτα συναισθήματα
[9, 127, 135, 137, 98]	Όχι	Αμιγώς διαστατικό

κα 2.6, οι περισσότεροι ανήκουν στην κατηγορία των αλγορίθμων που υλοποιούν επιβλεπόμενη μάθηση. Το αποτέλεσμα της ευστοχίας του παραγόμενου μοντέλου επίσης φαίνεται στον **Πίνακα 2.6**. Θα πρέπει να σημειωθεί πως σε όσες εργασίες δεν δίνεται η παραγόμενη ευστοχία (accuracy) αλλά δίνεται η ακρίβεια (precision) και η ανάκληση (recall), τότε τα τελευταία μεγέθη απεικονίζονται στον **Πίνακα 2.6**.

Πίνακας 2.6: Οι αλγόριθμοι που χρησιμοποιήθηκαν, ο τύπος τους (T.A.), η ευστοχία του παραγόμενου μοντέλου και ο αντίστοιχος δείκτης της εργασίας (Δ.Ε.). M.E. είναι αλγόριθμος χωρίς επίβλεψη (non-supervised) ενώ E. είναι με επίβλεψη (supervised). Για την στήλη της ευστοχίας, εάν από την αντίστοιχη εργασία προσφέρεται το ποσοστό του accuracy τότε απεικονίζεται αυτό. Εάν προσφέρονται τα ποσοστά ακρίβειας (precision) και ανάκλησης (recall) τότε απεικονίζονται αυτά ως: Ακρίβεια/Ανάκληση

Δ.Ε.	Αλγόριθμος	Μέγιστη ευστοχία	Τύπος αλγορίθμου
[28]		86.3%	
[133]	Gaussian Mixture Model	80.0%	M.E.
[134]		86.3%	
[96]		0.54/0.35	
[132]	Support Vector Machines	83.7%	E.
[141]		74.11%	

Πίνακας 2.6 – Συνέχεια από προηγούμενη σελίδα

Δ.Ε.	Αλγόριθμος	Μέγιστη ευστοχία	Τύπος αλγορίθμου
[135]	Hierarchical clustering	N/A	M.E.
[127]	Affinity propagation	0.79/0.79	M.E.
[136]	K Nearest Neighbors	78.3%	M.E.
[95]		82.02%	
[9]		81.55%	
[137]	Regression	79.3%	E.
[139]		0.98, r^2	
[140]		0.69, <i>corr</i>	
[138]	Artificial Neural Networks	0.67/0.65	E.
[98]	Fuzzy inferencing	12.06	E.
[142]	Random forests	0.67/0.59	E.

Συμπερασματικά, μπορεί να φανεί πως τα συνηθέστερα τεχνικά χαρακτηριστικά είναι αφενός μέσα σε αυτά που παρουσιάζονται στον Πίνακα 2.2 και αφετέρου ανήκουν σε αυτά που μπορούν να εξαχθούν από τα πρωτογενή ψηφιακά ηχητικά δεδομένα. Το δεύτερο είναι ένα αναμενόμενο γεγονός μιας και η πληθώρα των δεδομένων μουσικής είναι σε αυτή την μορφή και όχι σε συμβολική (M.I.D.I.). Εξ αυτών, φαίνεται ότι η ενέργεια έχει πρωταγωνιστικό ρόλο. Αρκετές από τις παρατιθέμενες εργασίες, χρησιμοποιούν την ενέργεια ως ένα κύριο χαρακτηριστικό για την ανίχνευση της διέγερσης [28, 98, 133, 134, 139, 141]. Γεγονός που επιβεβαιώνει και κοινή εμπειρική γνώση, δηλαδή ότι ένα μουσικό κομμάτι με αρκετή ενέργεια είναι πολύ πιθανό να διεγείρει τον ακροατή του. Έτσι, στις εργασίες των Lu και Lie [28, 134] η ενέργεια χρησιμοποιείται ως το μόνο τεχνικό χαρακτηριστικό για τον αρχικό διαχωρισμό των μουσικών κομματιών σε αυτά που μεταφέρουν χαμηλή και υψηλή Διέγερση. Το ίδιο συμβαίνει και με την εργασία του Muyuan όπου η ενέργεια χρησιμοποιείται για τον διαχωρισμό μουσικών κομματιών σε “ήρεμα” και μη [141], ενώ στην εργασία του Yeh η ενέργεια είναι το μόνο τεχνικό χαρακτηριστικό που επιλέχθηκε για την αναγνώριση της Διέγερσης [133]. Τέλος, στην εργασία του Sanghoon [98] η ενέργεια ήταν ένα από τα βασικά χαρακτηριστικά για τον διαχωρισμό σε κλάσεις υψηλής και χαμηλής Διέγερσης ενώ ο McDorman στην εργασία του [139] δείχνει την σχέση μεταξύ υποκειμενικής στάθμης

ακουστότητας και Διέγερσης.

Επίσης, σε πολλές εργασίες χρησιμοποιείται ο ρυθμός ως τεχνικό χαρακτηριστικό. Ένα γεγονός που επίσης είναι αναμενόμενο μιας και ο τελευταίος σχετίζεται ισχυρά με την διέγερση του ακροατή. Αυτή η συσχέτιση, δηλαδή του ρυθμού με την Διέγερση, είναι μία κοινή εμπειρική αλλά και επιστημονικά αποδεδειγμένη γνώση [30]. Επίσης, μερικές εργασίες χρησιμοποιούν τον ρυθμό και για την ανίχνευση της Ευχαρίστησης. Από τις παρατιθέμενες εργασίες φαίνεται πως ο ρυθμός δεν χρησιμοποιείται ως μόνο τεχνικό χαρακτηριστικό για την ανίχνευση της Διέγερσης αλλά ο σημαντικός του ρόλος σε αυτή δείχνεται από όλες όσες αναφέρονται στον **Πίνακα 2.4** και χρησιμοποιούν τεχνικά μεγέθη συσχετιζόμενα με τον ρυθμό. Ενώ, για την περίπτωση της Ευχαρίστησης, ο ρυθμός χρησιμοποιείται σε συνδυασμό με άλλα χαρακτηριστικά. Έτσι, οι Lu και Lie αναδεικνύουν την συσχέτιση του ρυθμού με περιοχές στο διαστατικό μοντέλο που χρησιμοποιούν αναφέροντας ότι ο ρυθμός είναι ισχυρά συνδεδεμένος με την κατηγοριοποίηση των μουσικών ηχητικών δεδομένων σε κλάσεις με χαμηλή και υψηλή Διέγερση ενώ τον χρησιμοποιούν και για την κατηγοριοποίηση των δεδομένων αναφορικά με την Ευχαρίστηση, χωρίς ωστόσο να αναφέρουν κάποια σημαντική σχέση του ρυθμού με την τελευταία. Πιο συγκεκριμένα, αναφέρουν ότι ένας σταθερός και δυνατός (strong) ρυθμός σχετίζεται ισχυρά με την κλάση υψηλής διέγερσης και ευχαρίστησης ενώ ένας αργός ρυθμός με την κλάση της χαμηλής Διέγερσης [28, 134]. Ο Zhang στην εργασία του [127] χρησιμοποιεί τον ρυθμό για την ανίχνευση της Διέγερσης και της Ευχαρίστησης χωρίς να αναφέρει κάποια σχέση μεταξύ του ρυθμού και των δύο αυτών συναισθηματικών καταστάσεων. Όμως, από τα 3 τεχνικά μεγέθη του ρυθμού που χρησιμοποιεί, τα δύο εξ αυτών συμμετέχουν στην κατηγοριοποίηση της Διέγερσης, και το άλλο στην κατηγοριοποίηση της Ευχαρίστησης, δείχνοντας έτσι την σύνδεση, και στην εργασία αυτή, του ρυθμού και της Διέγερσης. Οι Bartoszewski, Yang, Muyuan και Ling, σε διαφορετικές εργασίες, χρησιμοποιούν τον ρυθμό αλλά χωρίς να αναφέρουν κάποια συγκεκριμένη σύνδεση μεταξύ των τεχνικών μεγεθών και των συναισθημάτων [135, 136, 141, 142]. Από την άλλη, ο Feng χρησιμοποιεί τον ρυθμό ως ένα από τα δύο τεχνικά χαρακτηριστικά για την ανίχνευση συναισθήματος και για την δύο συναισθηματικές καταστάσεις, δηλαδή της Διέγερσης και της Ευχαρίστησης [138] ενώ ο Sanghoon χρησιμοποιεί τον ρυθμό, μαζί με την ενέργεια, ως πρωτεύοντα τεχνικά χαρακτηριστικά για την αναγνώριση Διέγερσης [98]. Ο Lauriel, στην εργασία του, τονίζει την σχέση ρυθμού και συναισθημάτων ειδικά αναφερόμενος στο συναίσθημα της Χαράς [140]. Τέλος, ο Yeh τονίζει την σημαντικότητα του ρυθμού

αναφέροντας ότι ο ρυθμός, μαζί με την ενέργεια, έχει πρωτεύοντα ρόλο στην αναγνώριση συναισθημάτων [133].

Για την χροιά και την τονικότητα, αν και αναφέρονται κάποιες συσχετίσεις με συναισθηματικές καταστάσεις, δεν υπάρχει κάποια ισχυρή σύνδεση όπως με την ενέργεια και τον ρυθμό για να χρησιμοποιηθεί σε μία πρώτη κατηγοριοποίηση. Δηλαδή, ενώ στην περίπτωση της ενέργειας υπάρχουν εργασίες που την χρησιμοποιούν ως μόνο χαρακτηριστικό για την κατηγοριοποίηση της Διέγερσης και έπειτα χρησιμοποιούν άλλα χαρακτηριστικά για τον προσδιορισμό της Ευχαρίστησης, για την χροιά και την τονικότητα δεν αναφέρεται κάτι τέτοιο. Όμως, αυτά τα δύο χαρακτηριστικά, η χροιά και η τονικότητα, χρησιμοποιούνται ευρέως για την τελική απόφαση και κατηγοριοποίηση του συναισθήματος. Για παράδειγμα, στην εργασία του McDorman [139] φαίνεται καθαρά ότι η τονικότητα έχει σημαντικό ρόλο στην αναγνώριση της Διέγερσης. Ήχοι με υψηλή τονικότητα, εάν έχουν και υψηλή ενέργεια, επηρεάζουν την Διέγερση περισσότερο από αυτούς με την ίδια ενέργεια και χαμηλή τονικότητα. Ακόμα, στις εργασίες των Lu και Lie, η χροιά έχει σημαντικό ρόλο στην αναγνώριση ευχαρίστησης. Ήχοι χαρακτηριζόμενοι ως "φωτεινοί", δηλαδή με υψηλό brightness (τεχνικό μέγεθος που ανήκει στην κατηγορία της χροιάς), και εάν έχουν κατάλληλο ρυθμό και ενέργεια, χαρακτηρίζονται ως να έχουν υψηλότερη Ευχαρίστηση από άλλους με τις ίδιες τιμές για τα μεγέθη της ενέργειας και του ρυθμού.

Συνοψίζοντας την στάθμη της τεχνικής αναφορικά με την χρήση τεχνικών χαρακτηριστικών στην αναγνώριση συναισθήματος από μουσική, μπορεί να φανεί ότι υπάρχουν δύο κυρίαρχα χαρακτηριστικά: η ενέργεια και ο ρυθμός. Αυτά τα δύο χρησιμοποιούνται από δημοσιευμένες εργασίες για την καταρχήν κατηγοριοποίηση των δεδομένων στις κλάσεις του προβλήματος Μηχανικής Μάθησης που αντιμετωπίζεται σε έκαστη εργασία. Από την άλλη, η χροιά και η τονικότητα, αν και δεν έχουν ένα πρωτεύοντα ρόλο, περιλαμβάνονται στην διαδικασία αναγνώρισης συναισθήματος και χρησιμοποιούνται σε συνδυασμό με τα άλλα χαρακτηριστικά. Αυτό το γεγονός φανερώνει, αφενός, την σημαντικότητα της ενέργειας και του ρυθμού στην συναισθηματική επήρεια που ασκούν τα μουσικά δεδομένα αλλά και αναδεικνύει την επίδραση των άλλων δύο, ήτοι χροιά και τονικότητα, στην τελική συναισθηματική κατάσταση του ακροατή.

Συνεχίζοντας και συνδυάζοντας την πληροφορία από τους **Πίνακες 2.4** και **2.5**, μπορεί να φανεί ότι η επιλογή των τεχνικών χαρακτηριστικών είναι ανεξάρτητη του συναισθηματι-

κού μοντέλου που επιλέγεται. Ακόμα, από τα χρησιμοποιούμενα συναισθηματικά μοντέλα φαίνεται ότι προτιμότερο είναι ο συνδυασμός ενός διαστατικού μοντέλου με διάκριτα συναισθήματα. Σε αυτό τον συνδυασμό μοντέλων για την επισήμανση των ηχητικών δεδομένων δεν χρησιμοποιείται μία τιμή στο επίπεδο ενός διαστατικού μοντέλου αλλά μία λεκτική περιγραφή η οποία αντιστοιχείται σε μία περιοχή τιμών στο προαναφερθέν επίπεδο. Με αυτό τον τρόπο από την μία δίνεται η δυνατότητα για την χρήση υφιστάμενων επισημάνσεων αλλά από την άλλη θα πρέπει αυτές οι επισημάνσεις (λεκτικές περιγραφές συναισθημάτων) να είναι οι ίδιες που χρησιμοποιούνται και στον συνδυασμό των μοντέλων. Δηλαδή, εάν μία εργασία Α έχει χρησιμοποιήσει ένα σετ διακριτών συναισθημάτων Β, τότε η εργασία Γ, η οποία θα χρησιμοποιήσει τον συνδυασμό διαστατικού μοντέλου με διάκριτα συναισθήματα, θα πρέπει και αυτή να υιοθετήσει το σύνολο συναισθημάτων Β και να το αντιστοιχήσει στο διαστατικό μοντέλο που χρησιμοποιεί. Έτσι, αν και θα μπορούσε να λεχθεί ότι αυτός ο συνδυασμός συνδυάζει τα προτερήματα και των δύο μοντέλων, εντούτοις, και επειδή ακριβώς χρησιμοποιεί λεκτικές περιγραφές για τα συναισθήματα, αφενός δεν αφήνει την επιλογή για εξερεύνηση των συναισθηματικών καταστάσεων και την παραγωγή ενός αναλυτικού και ποσοτικού μοντέλου συσχετισμού των τελευταίων με τιμές από τα τεχνικά χαρακτηριστικά και αφετέρου ενδέχεται να δημιουργεί προβλήματα στην αξιοποίηση αποτελεσμάτων άλλων ερευνών, όπως αναφέρθηκε και στις Ενότητες 2.1.4 και 2.1.5 [14]. Από την άλλη, αυτή η έλλειψη ενός αναλυτικού και ποσοτικού μοντέλου φαίνεται ότι είναι η κύρια αιτία για την επιλογή ενός διαστατικού μοντέλου με αντιστοίχιση περιοχών του σε λεκτικές περιγραφές συναισθημάτων. Έτσι, η προαναφερθείσα έλλειψη καλύπτεται με ποιοτική αντιστοίχιση των περιοχών του διαστατικού μοντέλου σε λεκτικές περιγραφές συναισθημάτων. Επίσης, αρκετά υψηλά στην προτίμηση είναι και τα αμιγώς διαστατικά μοντέλα. Αν και η κύρια προτίμησή τους είναι σε εργασίες που σχετίζονται με την έρευνα βιολογικών λειτουργιών (π.χ. λειτουργία εγκεφάλου) κατά την βίωση συναισθημάτων λόγω της απευθείας συσχέτισης συγκεκριμένων συναισθημάτων με τις λειτουργίες αυτές, όπως αναφέρθηκε στην Ενότητα 2.1.4, φαίνεται ότι υπάρχουν εργασίες στο πεδίο της αναγνώρισης συναισθημάτων από μουσική όπου εστιάζουν σε αυτή την απευθείας συσχέτιση συναισθημάτων και μουσικών δεδομένων, όπως φαίνεται στον **Πίνακα 2.4**.

Τέλος, οι χρησιμοποιούμενοι αλγόριθμοι κατηγοριοποίησης φαίνεται ότι δεν έχουν άμεση σχέση με το συναισθηματικό μοντέλο. Επιπλέον, η προαναφερθείσα επιλογή βασίζεται αποκλειστικά πάνω στην επίδοση που έχει ένας συγκεκριμένος αλγόριθμος ή συγκεκριμένοι

αλγόριθμοι στο σετ τεχνικών χαρακτηριστικών που χρησιμοποιείται [142]. Ως προς το είδος, δηλαδή αλγόριθμοι με επίβλεψη και χωρίς, από τον **Πίνακα 2.6** μπορεί να φανεί πως συνήθως χρησιμοποιούνται αλγόριθμοι με επίβλεψη. Το ποσοστό ευστοχίας των παραγόμενων μοντέλων δείχνει ότι είναι της τάξης του 87%. Θα πρέπει να σημειωθεί πως η εμφανιζόμενη ευστοχία είναι η συνολική. Αυτό σημαίνει πως σε πολλές περιπτώσεις η ευστοχία του μοντέλου για συγκεκριμένες κλάσεις του προβλήματος κατηγοριοποίησης είναι μεγαλύτερη, για παράδειγμα στην εργασία του Lu η ευστοχία φτάνει μέχρι και 90% για μία από τις κλάσεις που χρησιμοποιήθηκαν [28]. Από τις παρατιθέμενες εργασίες και εστιάζοντας σε αυτές που χρησιμοποιούν κάποιο διαστατικό μοντέλο, η συμπερίληψη της διάστασης της Ευχαρίστησης ή περιοχών που ανήκουν σε άρδην διαφορετικές τιμές Ευχαρίστησης έχει ως αποτέλεσμα την μείωση του συνολικού ποσοστού ευστοχίας του μοντέλου. Κύριος λόγος για αυτό φαίνεται να είναι η δυσκολία που υπάρχει στην αναγνώριση της διάστασης της Ευχαρίστησης σε σύγκριση με αυτή της Διέγερσης, σύμφωνα με τα αποτελέσματα των παρατιθέμενων εργασιών για την διάσταση της Ευχαρίστησης συγκεκριμένα.

2.2.3 Αναγνώριση συναισθήματος από ομιλία

Στην αναγνώριση συναισθημάτων από μουσική, γίνεται εστίαση στην πρόσβαση και ανάκτηση μουσικής πληροφορίας βάσει ανθρωποκεντρικών και εστιασμένων στο περιεχόμενο παραγόντων, ήτοι το συναίσθημα. Στην αναγνώριση συναισθήματος από ομιλία η εστίαση αλλάζει και μετακινείται στην επαύξηση της αλληλεπίδρασης ανθρώπου και μηχανής ή υπολογιστή (Human Machine Interaction, HMI ή Human Computer Interacton, HCI) [143, 144]. Με την υφιστάμενη τεχνολογία υπάρχει η δυνατότητα αναγνώρισης ομιλητή καθώς και αναγνώρισης ομιλίας και φωνητικών εντολών. Με την αναγνώριση συναισθήματος από ομιλία όμως, οι παραπάνω δυνατότητες επαυξάνονται με την αναγνώριση και του τρόπου που ειπώθηκαν οι λέξεις. Δηλαδή, με την αναγνώριση του συναισθήματος του ομιλητή το οποίο ισοδυναμεί με την δυνατότητα αναγνώρισης του εάν ειπώθηκαν οι λέξεις με άγχος, φόβο ή όποιο άλλο συναίσθημα. Η εν γένει διαδικασία δεν διαφέρει από την αυτή της αναγνώρισης συναισθήματος όπως παρουσιάστηκε στο **Σχήμα 2.1** και αναλύεται στις προηγούμενες ενότητες. Έτσι, αν και η αναγνώριση συναισθήματος από ομιλία στοχεύει σε διαφορετικό αποτέλεσμα, ήτοι επαύξηση της HMI/HCI, έχει αρκετά κοινά στοιχεία με την αναγνώριση συναισθήματος από μουσική και από ηχητικά γεγονότα. Σε αυτή την ενότητα θα παρουσιαστούν οι πρόσφατες

Πίνακας 2.7: Τα προφίλ της ομιλίας για διάφορα βασικά συναισθήματα

Συναίσθημα	Τεχνικά χαρακτηριστικά		
	Τονικό ύψος	Ενέργεια	Ρυθμός ομιλίας
Θυμός	Υψηλή μέση τιμή	Αύξηση	Αύξηση
Ευτυχία	Αύξηση στην μέση τιμή	Αύξηση	Αύξηση
Λύπη	Κανικό ή χαμηλότερο από το κανονικό	Μείωση	Αργός

εξελίξεις στο πεδίο της αναγνώρισης συναισθήματος από ομιλία.

Παρόλο που η εν γένει διαδικασία δεν διαφέρει, τα επιμέρους στοιχεία παρουσιάζουν διαφορές. Κύρια αιτία είναι ο τεχνικός στόχος της αναγνώρισης συναισθήματος από ομιλία, δηλαδή να μπορέσει να απομονώσει και να αναλύσει τα στοιχεία εκείνα τις ομιλίας από τα οποία γίνεται η μεταφορά συναισθήματος στον ακροατή. Αυτά διαφέρουν σε διάρκεια αλλά και σε μεθόδους προσέγγισης. Πιο συγκεκριμένα, η διάρκεια των τμημάτων του σήματος ομιλίας που χρησιμοποιούνται στην αναγνώριση ομιλίας είναι ίση με λέξεις ή με αρθρώσεις (utterances) και ως βασικά τεχνικά στοιχεία χρησιμοποιούνται η ενέργεια, η τονικότητα και η χροιά [145]. Επιπροσθέτως, από τις δημοσιευμένες εργασίες στο πεδίο αυτό έχουν ήδη αναπτυχθεί προφίλ συναισθηματικής ομιλίας, δηλαδή συσχέτιση τεχνικών χαρακτηριστικών με συγκεκριμένα συναισθήματα [146, 147]. Στον **Πίνακα 2.7** παρατίθενται τα προφίλ αυτά σύμφωνα με σύγχρονες εργασίες επισκόπησης του πεδίου της αναγνώρισης συναισθήματος από ομιλία [146].

Στον **Πίνακα 2.7**, το μέγεθος του ρυθμού ομιλίας είναι το ποσοστό του της διάρκειας του ηχητικού κύματος που έχει ομιλία προς την συνολική. Οι αλγόριθμοι που χρησιμοποιούνται στην αναγνώριση συναισθήματος από ομιλία περιλαμβάνουν αλγόριθμους επιβλεπόμενης μάθησης αλλά και μη επιβλεπόμενης. Πέρα από όσους αναφέρθηκαν στην Ενότητα 2.1.2, ένας από τους συχνά χρησιμοποιούμενος είναι και τα κρυμμένα μοντέλο Markov (Hidden Markov Models) [144]. Η ευστοχία των μοντέλων αγγίζει μέχρι και το 85% [145]. Ακόμα, στο πεδίο αναγνώρισης συναισθημάτων από ομιλία υπάρχουν πάνω από 20 βάσεις δεδομένων με συναισθηματικά επισημασμένα ομιλία [143, 144, 145], κάτι που έρχεται σε αντίθεση με τα διαθέσιμα συναισθηματικά επισημασμένα δεδομένα στη αναγνώριση συναισθήματος από ηχητικά γεγονότα όπου είναι 3. Ο κύριος λόγος είναι η διαδικασία δημιουργίας ομιλίας ικανής

να διεγείρει συγκεκριμένο συναίσθημα. Η δυσάρεστη συναισθηματική κατάσταση ενός ανθρώπου μπορεί να γίνει εύκολα αντιληπτή από τους συνομιλητές του όπως και κανείς μπορεί να προσποιηθεί μία αντίστοιχη συναισθηματική κατάσταση κατάλληλα τροποποιώντας τον τρόπο ομιλίας του. Έτσι, υπάρχουν βάσεις δεδομένων με συναισθηματικά επισημασμένα δεδομένα ομιλίας όπου αυτά έχουν δημιουργηθεί από ηθοποιούς. Οι τελευταίοι, καταγράφονται κατά την εκφορά λέξεων ή/και προτάσεων ενώ προσποιούνται την βίωση συγκεκριμένων συναισθημάτων. Επίσης, αναφέρονται βάσεις δεδομένων όπου η συναισθηματική κατάσταση του ομιλητή έχει προκληθεί από εξωτερικούς παράγοντες. Έτσι, η συνηθισμένη λειτουργία αναγνώρισης της συναισθηματικής κατάστασης από τον ομιλητή και η συνήθης σύνδεσή της με συγκεκριμένα συναισθήματα (π.χ. δεν λέγεται συχνά ότι κάποιος βιώνει χαμηλή διέγερση και χαμηλή ευχαρίστηση αλλά ότι είναι στεναχωρημένος) δείχνει να είναι και ένας από τους λόγους που οι περισσότερες εργασίες σε αυτό το πεδίο κάνουν χρήση μοντέλων με διακριτά συναισθήματα [143, 144].

Τέλος, τα ερευνητικά αποτελέσματα από το πεδίο αυτό βρίσκουν εφαρμογή σε περιπτώσεις που η φωνητική αλληλεπίδραση με τον χρήστη έχει πρωτεύοντα ρόλο. Έτσι, αναφέρονται εφαρμογές σε τηλεφωνικά κέντρα όπου η διαχείριση της τηλεφωνικής κλήσης εξαρτάται και από το συναίσθημα που βιώνει ο καλών [36]. Επίσης, αναφέρονται εφαρμογές σε έξυπνα συστήματα διδασκαλίας (Intelligent Tutoring Systems και Affective Tutoring Systems) όπου ένας ηλεκτρονικός δάσκαλος μιλάει με κατάλληλα συναισθηματικά επιφορτισμένη ομιλία έτσι ώστε να βελτιώσει την διαδικασία μάθησης [148, 149]. Επίσης, η αναγνώριση συναισθήματος από ομιλία χρησιμοποιείται για την ανίχνευση ψεύδους αλλά και την αναγνώριση της συναισθηματικής κατάστασης σε περιπτώσεις όπου μπορεί να αυξήσει την ασφάλεια, για παράδειγμα στην ανίχνευση της συναισθηματικής κατάστασης ενός οδηγού [144].

2.2.4 Το ηχητικό γεγονός

Πριν την επισκόπηση της στάθμης της τεχνικής στην αναγνώριση συναισθήματος από ηχητικά γεγονότα, θα παρουσιαστεί ένας ορισμός του ηχητικού γεγονότος ο οποίος δίνεται εντός των πλαισίων της Συναισθηματικής Ακουστικής Οικολογίας. Η τελευταία, όπου θα παρουσιαστεί αναλυτικότερα σε επόμενες ενότητες, είναι ένα νέο σκεπτικό Ακουστικής Οικολογίας το οποίο επαυξάνει το υφιστάμενο με το να περιλαμβάνει και την συναισθηματική

αλληλεπίδραση του ανθρώπου/δέκτη με τις πηγές του ήχου σε ένα ηχητικό περιβάλλον. Τόσο στο τυπικό σκεπτικό της Ακουστικής Οικολογίας όσο και στο νέο, τα ηχητικά γεγονότα είναι το βασικό δομικό συστατικό τους. Επίσης, τα ηχητικά γεγονότα είναι και το κύριο στοιχείο αυτής της διατριβής.

Η βάση για αυτό τον ορισμό του ηχητικού γεγονός είναι ο προσδιορισμός της “καθημερινής ακρόασης” από τον Gaver και την εργασία του με τίτλο “What in the world do we hear? an ecological approach to auditory event perception ” [5]. Στην εργασία αυτή προσδιορίζονται δύο τύποι ακρόασης. Αυτός της καθημερινής (everyday listening) και αυτός της μουσικής (musical listening) ακρόασης. Στον δεύτερο, ο ακροατής επικεντρώνεται σε τεχνικά χαρακτηριστικά του ήχου, δηλαδή στον ρυθμό, το τονικό ύψος, την χροιά και την ενέργεια. Για παράδειγμα, όταν κανείς ακούει ένα μουσικό κομμάτι παρατηρεί τις νότες, την μελωδία, τον ρυθμό κ.α. Όμως, υπάρχει και η καθημερινή ακρόαση όπου δεν γίνεται εστίαση στα παραπάνω στοιχεία αλλά στα χαρακτηριστικά που μπορεί να μεταδώσει ο ήχος, στο καθαυτό ηχητικό γεγονός. Δηλαδή, όταν κανείς περνάει από μία διάβαση πεζών σε κάποια διασταύρωση και ακούσει μία κόρνα αμαξίου δεν θα εστιάσει στην μελωδία, τον ρυθμό και τα άλλα τεχνικά χαρακτηριστικά αλλά στην σχετική με αυτόν θέση με το αυτοκίνητο, στον εκλαμβανόμενο μέγεθος του αυτοκινήτου και την ταχύτητα αυτού. Το ίδιο μπορεί να συμβεί ακόμα και όταν το ηχητικό γεγονός περιλαμβάνει μουσική ή/και ομιλία αλλά ο ακροατής εστιάζει στα χαρακτηριστικά του ηχητικού γεγονός, δηλαδή δεν ασκεί μουσική ακρόαση αλλά καθημερινή.

Τα χαρακτηριστικά που μπορούν να μεταδοθούν μέσω του ηχητικού γεγονότος περιλαμβάνουν ιδιότητες της πηγής. Ακούγοντας ένα ηχητικό γεγονός με την καθημερινή ακρόαση, γίνονται αντιληπτά η διαδικασία παραγωγής ήχου, η φυσική κατάσταση των σωμάτων που παράγουν τον ήχο, το μέγεθος της πηγής, η θέση της πηγής σχετικά με τον δέκτη καθώς και η κίνησή της σχετικά με τον ακροατή. Δηλαδή, κανείς μπορεί να αντιληφθεί εάν ο ήχος παράγεται με χρούση, τριβή ή ταλάντωση, π.χ. αντικείμενο που τρίβεται πάνω σε ένα άλλο, αντικείμενα που χρούονται μεταξύ τους ή αντικείμενο που ταλαντώνεται. Επίσης, μπορεί να καταλάβει την φύση των αντικειμένων αυτών, δηλαδή εάν είναι υγρά στερεά κτλ. Για παράδειγμα, είναι εύκολα αντιληπτό εάν μία σταγόνα υγρού πέφτει πάνω σε ένα στέρεο σώμα ή σε ένα άλλο υγρό. Ακόμα, το πού γίνεται αυτή η πτώση της σταγόνας επίσης μπορεί να το καταλάβει ο ακροατής (θέση της πηγής). Γίνεται πίσω του; Γίνεται στα πλάγια; Κοντά

ή μακριά του; Μήπως ακούγονται σταγόνες που πέφτουν και πλησιάζουν (κίνηση της πηγής); Εάν ναι, πλησιάζουν γρήγορα (ταχύτητα της πηγής); Όλα αυτά τα χαρακτηριστικά επηρεάζουν τις αποφάσεις του ακροατή. Στην περίπτωση όπου ένα αυτοκίνητο πλησιάζει έναν πεζό, η ταχύτητα, η θέση και το μέγεθος του αυτοκινήτου προφανώς παίζουν καθοριστικό ρόλο στις αποφάσεις του ακροατή για τις επόμενες κινήσεις του. Ακόμα, κατά την ακρόαση του αυτοκινήτου και την λήψη αποφάσεων, ο ακροατής βιώνει συναισθήματα που εγείρονται από το καθαυτό ηχητικό γεγονός και την σχέση του ακροατή με αυτό, χωρική ή/και χρονική [3].

Με βάση τα παραπάνω, το ηχητικό γεγονός εντός του πλαισίου της Συναισθηματικής Ακουστικής Οικολογίας, δηλαδή εντός ενός πλαισίου Ακουστικής Οικολογίας που λαμβάνει υπόψη και την συναισθηματική αλληλεπίδραση του ηχητικού περιβάλλοντος και του ακροατή, το ηχητικό γεγονός ορίζεται ως μία δομή (structure) που έχει [3]:

- μία ηχητική κυματομορφή (sound waveform)
- εκδήλωση της στιγμιαίας χωρικής τοποθέτησης της πηγής
- διάρκεια
- ένδειξη του τρόπου παραγωγής ήχου (τριβή, κρούση, ταλάντωση ή ό,τι άλλο)
- ένδειξη της φυσικής κατάστασης των σωμάτων που μετέχουν στην παραγωγή ήχου (π.χ. στερεά, υγρά)
- σημασιολογικό περιεχόμενο

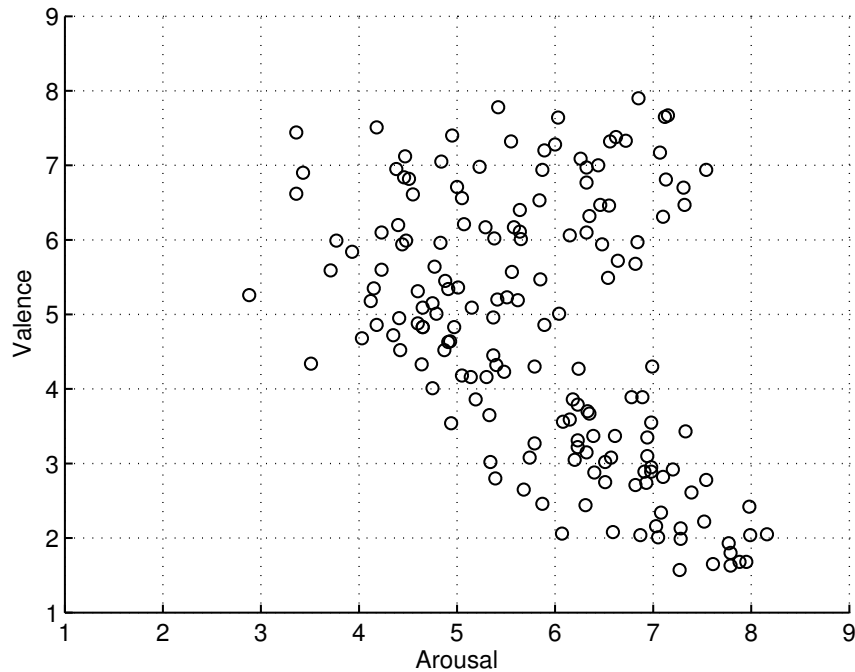
Από κάθε ένα χαρακτηριστικό δυνητικά μπορεί να συμβεί τόσο αναγνώριση του ηχητικού γεγονότος αλλά και της συναισθηματικής επίρρειας αυτού στον ακροατή. Για την αναγνώριση όμως του συναισθήματος που δύναται να εγείρουν τα ηχητικά γεγονότα, απαιτείται η συναισθηματική επισήμανση αυτών έτσι ώστε να χρησιμοποιηθεί στις διαδικασίες Μηχανικής Μάθησης. Σε αντίθεση με τα άλλα δύο παρουσιαζόμενα πεδία αναγνώρισης συναισθήματος από ηχητικά δεδομένα, δηλαδή της μουσικής και της ομιλίας, και κατά την στιγμή της συγγραφής του ανά χείρας τόμου υπάρχουν μόνο 3 διαθέσιμα σετ ηχητικών γεγονότων με συναισθηματική επισήμανση, εκ των οποίων το ένα έχει δημιουργηθεί εντός του πλαισίου της έρευνας της παρούσας διατριβής. Θα πρέπει να τονιστεί ότι ως “διαθέσιμα” ορίζονται τα σετ

δεδομένων που μπορούν να χρησιμοποιηθούν ελεύθερα τουλάχιστον από την επιστημονική κοινότητα. Έτσι, πριν την παρουσίαση της υφιστάμενης στάθμης της τεχνικής στο πεδίο της αναγνώρισης συναισθημάτων από ηχητικά γεγονότα, θα γίνει μία επισκόπηση των 2 διαθέσιμων σετ με συναισθηματικά επισημασμένα ηχητικά γεγονότα. Το τρίτο, θα παρουσιαστεί αναλυτικά σε επόμενες ενότητες εφόσον αποτελεί και αποτέλεσμα της διατριβής αυτής.

2.2.5 Διαθέσιμα σετ συναισθηματικά επισημασμένων ηχητικών γεγονότων

Το πρώτο σετ δημιουργήθηκε από τους Bradley και Lang το 1999 [103]. Ονομάζεται International Affective Digital Sounds (IADS) και αποτελείται από 167 ήχους με σημασιολογικό περιεχόμενο το οποίο καλύπτει πληθώρα ανθρώπινων εκδηλώσεων (π.χ. βήξιμο, λαχάνιασμα, ήχους συνουσίας, εμετό κτλ) μέχρι ήχους από μηχανές (ήχους γραναζιών που γυρίζουν, ήχο από γραφομηχανή κτλ), ζώα (π.χ. γάβγισμα, νιαούρισμα κτλ) και άλλους (π.χ. εκρήξεις, κόρνες κτλ). Η συναισθηματική επισήμανση κάθε ήχου στο σετ αυτό έγινε κατά μέσο όρο από 100 ανθρώπους και βασίζεται πάνω σε ένα αμιγώς διαστατικό μοντέλο. Προς αυτό τον σκοπό, στην επισήμανση των ήχων που περιέχονται στο σετ αυτό χρησιμοποιήθηκε το μοντέλο SAM [102]. Οι τιμές της επισήμανσης για κάθε διάσταση του μοντέλου και για κάθε ήχο του IADS είναι στο διάστημα [1, 9], σύμφωνα με την παρουσίαση της διαδικασίας της συναισθηματικής επισήμανσης που έγινε στην Ενότητα 2.1.6. Στο **Σχήμα 2.9** είναι οι τιμές όλων των ήχων για τις συναισθηματικές καταστάσεις της Διέγερσης και Ευχαρίστησης. Από το εν λόγω μπορεί να φανεί ένας περιορισμός στα εγείρομενα συναισθήματα από τα ηχητικά γεγονότα. Ο τελευταίος διατυπώνεται επίσης και από τους δημιουργούς του σετ και είναι η αδυναμία να νιώθει κάποιος δυσάρεστα από ένα ηχητικό γεγονός και ταυτόχρονα μη εγείρομενος [103]. Μία πιθανή εξήγηση για αυτό το φαινόμενο είναι η απώλεια σημασιολογικού περιεχομένου υψηλού επιπέδου, π.χ. λέξεις σε αντίθεση με κλάμα, όπου με τις λέξεις μεταδίδεται ένα σημασιολογικό περιεχόμενο μέσα από την σημασία των λέξεων ενώ από το κλάμα είναι το ίδιο ηχητικό γεγονός που μεταδίδει το σημασιολογικό περιεχόμενο.

Ένα νέο σετ συναισθηματικά επισημασμένων ηχητικών γεγονότων μαζί με μία εργασία πάνω στην εξόρυξη συναισθηματικής πληροφορίας από ηχητικά γεγονότα παρουσιάστηκε από τον Schuller το 2012 [150]. Σε αυτή την ενότητα θα παρουσιαστεί το σετ ενώ στην επόμενη η



Σχήμα 2.9: Οι τιμές της επισήμανσης για τους ήχους IADS και για τις διαστάσεις της Διέγερσης (Arousal) και Ευχαρίστησης (Valence)

δημοσιευμένη εργασία εν γένει. Το εν λόγω σύνολο γενικών ήχων δημιουργήθηκε από την online βάση ηχητικών δεδομένων FindSounds.com⁹. Επιλέχτηκαν 390 ηχητικά γεγονότα από σύνολο 10000 και από τις κατηγορίες “Ζώα”, “Μουσικά όργανα”, “Φύση”, “Θορυβοποιόι (Noisemakers)”, “Ανθρωποι”, “Αθλήματα”, “Εργαλεία και οχήματα”. Όλες οι κατηγοριοποιήσεις παρασχέθηκαν από την προαναφερθείσα βάση δεδομένων. Η επισήμανση έγινε με την χρήση διαστατικού συναισθηματικού μοντέλου, 2 διαστάσεων και συγκεκριμένα των: α) Διέγερση, και β) Ευχαρίστηση, και τεσσάρων συμμετεχόντων με μόνον ένα να είναι γένους θηλυκού. Για την γενίκευση της ισχύος των αποτελεσμάτων της επισήμανσης έγινε χρήση του εργαλείου Evaluator Weighted Estimator [151]. Το τελευταίο επιτρέπει την εξομάλυνση των απαντήσεων κάθε αξιολογητή λαμβάνοντας υπόψη τις διάφορες ενοχλήσεις που μπορεί να έχει. Αυτό γίνεται μέσω της χρήσης στατιστικών μεγεθών που φανερώνουν τυχόν συνδέσεις και εξαρτήσεις των συνολικών απαντήσεων των αξιολογητών με τα ηχητικά ερεθίσματα.

⁹<http://www.findsounds.com>

2.2.6 Αναγνώριση συναισθήματος από ηχητικά γεγονότα

Η αναγνώριση συναισθήματος από ηχητικά δεδομένα είναι επίσης μία εφαρμογή Μηχανικής Μάθησης. Επίσης, είναι ένα σχετικά νέο πεδίο, κάτι που αποδεικνύει και το γεγονός ότι προς το παρόν υπάρχουν λίγες δημοσιευμένες εργασίες που να εστιάζουν στο συγκεκριμένο θέμα, οι οποίες είναι ασθενώς συνδεδεμένες μεταξύ τους [13]. Δηλαδή, τα αποτελέσματα των εργασιών δεν παρουσιάζουν μία ερευνητική συνέχεια η οποία καλύπτει ερευνητικά κενά αλλά και έχει θέσει στιβαρές βάσεις για μετέπειτα έρευνα, γεγονός αναμενόμενο δεδομένου του πλήθους των δημοσιευμένων εργασιών σε αυτόν τον τομέα.

Παρόλα αυτά, και όπως δείχθηκε στις προηγούμενες ενότητες, τα ηχητικά γεγονότα μπορούν να εγείρουν συναισθήματα στον ακροατή. Έτσι, στην αναγνώριση συναισθήματος από ηχητικά γεγονότα ερευνάται ο τρόπος με τον οποίο γίνεται η παραπάνω έγερση συναισθημάτων. Ένας από τους παράγοντες που, μέχρι τώρα, έχει βρεθεί να επηρεάζει το συναίσθημα του ακροατή είναι η θέση της πηγής. Αν και δεν έχει μελετηθεί αναλυτικά, υπάρχουν δύο δημοσιευμένες εργασίες οι οποίες εστιάζουν στην σχέση της θέσης πηγής και του ακροατή, χωρίς όμως να παράγουν ένα αναλυτικό μοντέλο συσχέτισης της θέσης της πηγής και του συναισθήματος. Μία από αυτές της εργασίες ανήκει στον Ekman [37], όπου πραγματοποιήθηκε μία προσπάθεια αρχικής ανάλυσης της επίρειας της θέσης της πηγής στο συναίσθημα του ακροατή. Προς αυτό το σκοπό χρησιμοποιήθηκε ένα σετ 24 ηχείων για την αναπαραγωγή ήχων στο εμπρός και πίσω μέρος του ακροατή. Το συναισθηματικό μοντέλο που χρησιμοποιήθηκε ανήκει στα διακριτά και, συγκεκριμένα, θεωρήθηκε μόνο το συναίσθημα του Φόβου. Τα αποτελέσματα έδειξαν πως οι ήχοι που παράγονται στο πίσω μέρος του ακροατή γίνονται αντιληπτοί ως πιο τρομαχτικοί. Αντί για την χωρική τοποθέτηση της πηγής αλλά με επίσης εστίαση στο Φόβο, ο Hagman στην εργασία του [152] μελετά την σχέση απόστασης πηγής, μεγέθους δωματίου στο οποίο είναι η πηγή και εγειρόμενων συναισθημάτων στον ακροατή. Οι απαντήσεις από τους συμμετέχοντες σε πειράματα που διεξήχθησαν στα πλαίσια της προαναφερθείσας εργασίας έδειξαν ότι ο όγκος του δωματίου και η εγγύτητα της πηγής έχουν ανάλογη σχέση με την βίωση του συναισθήματος του Φόβου.

Επίσης, από τις λίγες υφιστάμενες δημοσιευμένες εργασίες σχετικά με την αναγνώριση συναισθήματος από ηχητικά γεγονότα, έχει γίνει προσπάθεια εξόρυξης συναισθηματικής πληροφορίας από τα τεχνικά χαρακτηριστικά του ηχητικού δεδομένου. Το αποτέλεσμα μίας

τέτοιας έρευνας μπορεί να αναδείξει κυρίως τεχνικά χαρακτηριστικά στην αναγνώριση συναισθήματος από ηχητικά δεδομένα εν γένει. Δηλαδή, εφόσον οι γενικοί ήχοι είναι μία μορφή ήχου χωρίς συγκεκριμένους κανόνες δόμησης, όπως συμβαίνει στην ομιλία και την μουσική, εάν βρεθεί ότι κάποιο ή κάποια τεχνικά χαρακτηριστικά μπορούν να επηρεάσουν το συναίσθημα του ακροατή τότε είναι αρκετά πιθανό τα ίδια αυτά τεχνικά χαρακτηριστικά να βρίσκουν εφαρμογή και στις περιπτώσεις της ομιλίας και της μουσικής.

Προς αυτή την κατεύθυνση κινήθηκαν οι εργασίες από τους Schiller, Hantke, Weninger, Wenjing, Zhang και Narayanan [13, 150]. Στην πρώτη τους εργασία το 2012 και με τίτλο “Automatic recognition of emotion evoked by general sound events” [150] εκτός από την παρουσίαση του σετ συναισθηματικά επισημασμένων ηχητικών γεγονότων που παρουσιάστηκε στην προηγούμενη ενότητα, διενεργήθηκε και ένα πείραμα εξόρυξης συναισθηματικής πληροφορίας από ηχητικά γεγονότα. Για την εξόρυξη της πληροφορίας χρησιμοποιήθηκε ο επιβλεπόμενος αλγόριθμος REPTree σε συνδυασμό με αλγόριθμους μετά-μάθησης και τα αποτελέσματα δόθηκαν υπό μορφή συντελεστή συσχέτισης. Για την Διέγερση η μέγιστη τιμή αυτού ήταν 0.61 και για την Ευχαρίστηση 0.49. Πιο συγκεκριμένα, χρησιμοποιήθηκαν τα τεχνικά χαρακτηριστικά ρυθμός, ενέργεια, χροιά και τονικό ύψος. Για αυτά, έγινε υπολογισμός αρκετών τεχνικών μεγεθών. Οι τιμές αυτών των μεγεθών καθώς και στατιστικά μεγέθη της μεταβολής των μεγεθών χρησιμοποιήθηκαν για την εύρεση συσχέτισης του συναισθήματος με τα τεχνικά μεγέθη και, συνεπώς, τα αντίστοιχα χαρακτηριστικά. Για παράδειγμα, η μέση τιμή της ενέργειας και η τυπική της απόκλιση κατά την εξέλιξη ενός ηχητικού γεγονότος.

Στην επόμενη εργασία, το έτος 2013 και με τίτλο “On the Acoustics of Emotion in Audio: What Speech, Music and Sound have in Common” [13], επιχειρήθηκε μία σύνδεση των τριών μορφών ήχου για την εξόρυξη συναισθηματικής πληροφορίας. Αναφορικά με τα ηχητικά γεγονότα, χρησιμοποιήθηκε η βάση δεδομένων που παρουσιάστηκε στην προηγούμενη εργασία τους [150], για την ομιλία η βάση δεδομένων Geneva Multimodal Expression Corpus for Experimental Research on Emotion Perception [153] ενώ για την μουσική ένα σετ δεδομένων για το οποίο πραγματοποιήθηκε επισήμανση από τους ίδιους συμμετέχοντες που έκαναν την επισήμανση για τα ηχητικά γεγονότα. Από τις τρεις αυτές ομάδες δεδομένων εξήχθησαν ηχητικά χαρακτηριστικά και έγινε η σύγκριση για την απόδοσή τους και στις τρεις κατηγορίες ταυτόχρονα. Τα τεχνικά χαρακτηριστικά που εξήχθησαν ανήκουν και στις

4 κατηγορίες που φαίνονται στον **Πίνακα 2.2**. Η διαδικασία αναγνώρισης πραγματοποιήθηκε με τον επιβλεπόμενο αλγόριθμο support vector regression και τα αποτελέσματα έδειξαν μέγιστη τιμή συντελεστή συσχέτισης για την διέγερση ίση με 0.65 ενώ για την ευχαρίστηση ίση με 0.44.

Συνοψίζοντας, από το πλήθος των εργασιών που παρατέθηκαν παραπάνω είναι φανερό ότι στο πεδίο της εξόρυξης συναισθηματικής πληροφορίας από ηχητικά γεγονότα υπάρχουν ελάχιστες δημοσιευμένες εργασίες. Όπως αναφέρθηκε και στην προηγούμενη ενότητα, υπάρχουν δύο ελεύθερες βιβλιοθήκες με συναισθηματικά επισημασμένα ηχητικά γεγονότα, εκ των οποίων η μία έχει χρησιμοποιήσει μόνο 4 συμμετέχοντες για την επισήμανση και κατόπιν έκανε χρήση μαθηματικών εργαλείων για την επαύξηση της γενικότητας των επισημάνσεων. Ακόμα, από τις προαναφερθείσες εργασίες, μόνο 2 [13, 150] εστιάζουν στην σχέση των τεχνικών και χαμηλού επιπέδου χαρακτηριστικών του ήχου (π.χ. ενέργεια, MFCC, κτλ) και συναισθήματος ενώ άλλες 2 [37, 152] στη σχέση της χωρικής τοποθέτησης και των χαρακτηριστικών του χώρου με το εγειρόμενο συναίσθημα στον ακροατή. Από αυτές μπορεί να φανεί μία αρχική συσχέτιση του συναισθήματος του ακροατή με την χωρική τοποθέτηση της πηγής αλλά και με τα τεχνικά χαρακτηριστικά του καθαυτού ήχου, χωρίς όμως να δίνεται σε καμία κάποιο αναλυτικό μοντέλο που θα μπορούσε να χρησιμοποιηθεί για την διαδικασία εξόρυξης συναισθηματικής πληροφορίας από ηχητικά γεγονότα. Ακόμα, μπορεί να φανεί η τάση για αποκλειστική χρήση διαστατικών μοντέλων εκτός εάν υπάρχει εστίαση σε ένα μόνο συγκεκριμένο συναίσθημα, κάτι που ίσως είναι οριακά σε αντίθεση με τον τομέα της εξόρυξης συναισθηματικής πληροφορίας από μουσική ή/και ομιλία.

Η παρούσα διατριβή εστιάζει στην εξόρυξη της συναισθηματικής πληροφορίας από ηχητικά γεγονότα. Για αυτό το σκοπό χρησιμοποιήθηκε η βάση δεδομένων IADS [103], κυρίως λόγω του μεγάλου αριθμού συμμετεχόντων κατά την διαδικασία επισήμανσης των ηχητικών γεγονότων. Επίσης, χρησιμοποιήθηκε διαστατικό μοντέλο δύο διαστάσεων, της διέγερσης και της ευχαρίστησης, και ερευνήθηκε η σχέση μεγάλου αριθμού τεχνικών χαρακτηριστικών και με τις δύο προαναφερθείσες διαστάσεις. Επίσης, κατά την διάρκεια εκπόνησης της έρευνας πραγματοποιήθηκε και δημοσιεύθηκε εργασία σχετικά με την επήρεια της θέσης της πηγής και του εγειρόμενου συναισθήματος. Ακόμα, δημιουργήθηκε και διανέμεται ελεύθερα από το οικείο εργαστήριο, κατά την συγγραφή του ανά χείρας τόμου, του Τμ. Τεχνών Ήχου και Εικόνας το οποίο περιέχει γενικούς ήχους με χωρική τοποθέτηση και συναισθηματική επισή-

μανση. Η θεωρητική και πειραματική προσέγγιση και υλοποίηση της έρευνας παρουσιάζεται και αναλύεται στα επόμενα κεφάλαια.

ΚΕΦΑΛΑΙΟ 3

ΘΕΩΡΗΤΙΚΗ - ΠΕΙΡΑΜΑΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΑ

Όπως αναλύθηκε στα προηγούμενα κεφάλαια, η αναγνώριση συναισθήματος από ηχητικά δεδομένα έχει αρκετά κοινά στοιχεία κατά την υλοποίησή της στις τρεις διαφορετικές κατηγορίες δεδομένων, ήτοι: α) ομιλία, β) μουσική, και γ) γενικοί ήχοι. Συγκεκριμένα, και στις τρεις κατηγορίες εφαρμόζεται η διαδικασία που περιγράφεται από το **Σχήμα 2.1**, δηλαδή η χρήση ενός σετ επισημασμένων ηχητικών δεδομένων από τα οποία γίνεται εξαγωγή τεχνικών μεγεθών όπου τα τελευταία, ήτοι τα τεχνικά μεγέθη, μαζί με τις επισημάνσεις των δεδομένων χρησιμοποιούνται ως είσοδο σε αλγόριθμους Μηχανικής Μάθησης. Η έξοδος από του αλγορίθμους είναι το μοντέλο κατηγοριοποίησης. Συγκεκριμένα για την μουσική και τους γενικούς ήχους, τα τεχνικά χαρακτηριστικά και οι αλγόριθμοι που μπορούν να χρησιμοποιηθούν ανήκουν στις ίδιες κατηγορίες, δηλαδή τα τεχνικά χαρακτηριστικά είναι ο ρυθμός, η χροιά, το τονικό ύψος και η ενέργεια ενώ οι αλγόριθμοι είναι επιβλεπόμενης μάθησης.

Όπως αναφέρθηκε προηγουμένως, ενώ στην εξόρυξη συναισθηματικής πληροφορίας από μουσική υπάρχει ένα υφιστάμενο ερευνητικό υπόβαθρο, για τα ηχητικά γεγονότα παρατηρείται μία έλλειψη αυτού. Ακόμα, από τις λίγες δημοσιευμένες εργασίες στο πεδίο της ανα-

γνώρισης συναισθήματος από ηχητικά γεγονότα μπορεί να παρατηρηθεί πως η ουσιαστική διαφορά με την αναγνώριση συναισθήματος από μουσική είναι η χρήση διαφορετικού τύπου ηχητικών δεδομένων. Αν και το τελευταίο σημαίνει μία διαφορετική προσέγγιση αναφορικά με τα χρησιμοποιούμενα τεχνικά μεγέθη, την διαχείριση του σημασιολογικού περιεχομένου και τα μοντέλα συναισθήματος που ενδέχεται να αρμόζουν περισσότερο, παράλληλα, και σε συνδυασμό με την εγγενή ομοιότητα της διαδικασίας αναγνώρισης συναισθήματος, υποστηρίζει την εφαρμογή μεθόδων από την εξόρυξη συναισθηματικής πληροφορίας από μουσική για ηχητικά γεγονότα. Έτσι, στο πλαίσιο της παρούσας διατριβής η έρευνα βασίστηκε στο υφιστάμενο ερευνητικό υπόβαθρο του πεδίου της αναγνώρισης συναισθήματος από μουσική. Πραγματοποιήθηκε αρχικά μία θεωρητική προσέγγιση, λαμβάνοντας υπόψη αφενός τα διαθέσιμα συναισθηματικά μοντέλα και αφετέρου την ιδιαιτερότητα των γενικών ήχων, η οποία στόχευσε στις εγγενείς διαφορές των ηχητικών γεγονότων σε σχέση με την μουσική (π.χ. όχι αρμονία, πιθανό σημασιολογικό περιεχόμενο, σύντομη διάρκεια κτλ). Έπειτα, ακολούθησε η πειραματική προσέγγιση όπου διεξήχθησαν πειράματα υποκειμενικής αξιολόγησης αλλά και εξόρυξης συναισθηματικής πληροφορίας.

Η εκπόνηση των πειραμάτων έγινε στο προγραμματιστικό περιβάλλον MATLAB χρησιμοποιώντας το ελεύθερο πλαίσιο εργασίας (framework) MIRToolbox [106]. Το τελευταίο προσέφερε την δυνατότητα για εξαγωγή πλείστων τεχνικών χαρακτηριστικών ήχου τα οποία χρησιμοποιούνται σε δημοσιευμένες έρευνες πάνω στον τομέα της αναγνώρισης συναισθήματος από ηχητικά δεδομένα, χρησιμοποιώντας κώδικα που έχει ελεγχθεί για την ορθότητά του και βασίζεται σε σύγχρονα αποτελέσματα ερευνών. Ακόμα, για την εκτέλεση των αλγόριθμων κατηγοριοποίησης χρησιμοποιήθηκε το λογισμικό WEKA, το οποίο θεωρείται ως το πλέον έγκυρο λογισμικό πάνω σε θέματα Μηχανικής Μάθησης και Τεχνητής Νοημοσύνης [45].

Στις παρακάτω ενότητες παρουσιάζονται τα ερευνητικά ερωτήματα που τέθηκαν ακολουθούμενα από την θεωρητική προσέγγιση που ακολουθήθηκε για την απάντηση των ερωτημάτων. Έπειτα παρατίθεται η μεθοδολογία που ακολουθήθηκε στην πειραματική διαδικασία (πειραματική προσέγγιση), οι αλγόριθμοι κατηγοριοποίησης και τα τεχνικά μεγέθη ήχου που χρησιμοποιήθηκαν καθώς και η χρήση της συναισθηματικής επισήμανσης. Τα αποτελέσματα από τις προαναφερθείσες διαδικασίες θα παρουσιαστούν στο Κεφάλαιο 4 ενώ η συζήτηση αυτών θα γίνει στο Κεφάλαιο 5.

3.1 Ερευνητικά ερωτήματα

Όπως έγινε φανερό στις προηγούμενες ενότητες, τα ηχητικά γεγονότα μπορούν να εγείρουν συναισθήματα στον ακροατή. Επίσης, το μικρό πλήθος δημοσιευμένων εργασιών πάνω στην εξόρυξη συναισθηματικής πληροφορίας από ηχητικά γεγονότα αφήνει αρκετά ερωτήματα απάντητα αναφορικά με την συναισθηματική αλληλεπίδραση των ηχητικών γεγονότων και του ακροατή. Για παράδειγμα, στις δημοσιευμένες εργασίες πάνω στην εξόρυξη συναισθηματικής πληροφορίας από ηχητικά δεδομένα εν γένει αναφέρονται διάφορα μοντέλα συναισθημάτων που χρησιμοποιούνται, τα οποία παρατέθηκαν και στην Ενότητα 2.1. Επίσης, ήδη αναφέρθηκε σε προηγούμενη ενότητα ότι στην αναγνώριση συναισθήματος από μουσική η ενέργεια έχει συσχετιστεί ισχυρά με την Διέγερση ενώ στην αναγνώριση συναισθήματος από ομιλία υπάρχουν ήδη προφίλ τεχνικών χαρακτηριστικών που έχουν συσχετιστεί με συγκεκριμένα συναισθήματα. Με τα ηχητικά γεγονότα τι ισχύει; Τέλος, πώς ακριβώς θα μπορούσε να επαυξηθεί το υφιστάμενο μοντέλο της Ακουστικής Οικολογίας έτσι ώστε να περιλαμβάνει και την συναισθηματική συνιστώσα της αλληλεπίδρασης του ανθρώπου με το ηχητικό του περιβάλλον;

Συνοψίζοντας τα παραπάνω, μπορεί κανείς να δει κάποια συγκεκριμένα ερευνητικά ερωτήματα να τίθενται. Αυτά, στα οποία εστίασε και η παρούσα διατριβή, είναι:

- Ποιες οι διαφορές των τύπων των συναισθηματικών μοντέλων και ποιο αρμόζει καλύτερα στην διαδικασία εξόρυξης συναισθηματικής πληροφορίας από ηχητικά δεδομένα;
- Ποια είναι η ειδοποιός διαφορά στα τεχνικά χαρακτηριστικά μεταξύ ακραίων συναισθηματικών καταστάσεων;
- Πώς εντάσσεται και σχετίζεται η συναισθηματική συνιστώσα των γενικών ήχων στο ευρύτερο πλαίσιο του ηχητικού περιβάλλοντος;
- Πώς τα τεχνικά χαρακτηριστικά ενός ηχητικού γεγονότος ή/και της πηγής αυτού μπορούν να επηρεάσουν το εγχειρόμενο συναίσθημα;

Από την υφιστάμενη επιστημονική βιβλιογραφία, δεν δίνεται κάποια απάντηση στα ερωτήματα αυτά. Για παράδειγμα, δεν αναφέρεται κάπου κάποια ειδοποιός διαφορά μεταξύ των τεχνικών χαρακτηριστικών του ηχητικού γεγονότος ικανή να εγείρει στον ακροατή αντίθετες

και ακραίες συναισθηματικές καταστάσεις. Μία απάντηση σε αυτό το ερώτημα, δηλαδή στο “Ποια είναι η ειδοποιός διαφορά στα τεχνικά χαρακτηριστικά μεταξύ ακραίων συναισθηματικών καταστάσεων;” θα μπορούσε να αποτυπώσει την επήρεια ενός τεχνικού χαρακτηριστικού ικανού να μεταβάλλει ισχυρά την συναισθηματική κατάσταση του ακροατή. Δεδομένης της γενικότητας των ηχητικών γεγονότων, σε αντιπαράβολή με την ομιλία και την μουσική, αυτό το τεχνικό χαρακτηριστικό θα μπορούσε να θεωρηθεί ως ένας γενικός παράγοντας στην συναισθηματική επήρεια του ήχου πάνω στον άνθρωπο. Ακόμα, και αναφορικά με το ερώτημα “Πώς τα τεχνικά χαρακτηριστικά ενός ηχητικού γεγονότος ή/και της πηγής αυτού μπορούν να επηρεάσουν το εγειρόμενο συναίσθημα;”, έχουν γίνει ήδη κάποιες προσπάθειες ανίχνευσης της επήρειας της χωρικής τοποθέτησης της πηγής στο συναίσθημα του ακροατή αλλά χωρίς να υπάρχει κάποιο αναλυτικό μοντέλο της επήρειας αυτής. Δηλαδή, πως μεταβάλλεται η επήρεια στο συναίσθημα για διαφορετικές θέσεις της πηγής γύρω από τον ακροατή; Τέλος, ενώ υπάρχει ένα υφιστάμενο πλαίσιο μελέτης της επήρειας του ηχητικού περιβάλλοντος στον άνθρωπο, αυτό δεν λαμβάνει υπόψη την συναισθηματική συνιστώσα. Αποδεδειγμένης της συναισθηματικής επήρειας του ήχου, εν γένει, πάνω στον άνθρωπο, εύλογα δημιουργείται το ερώτημα με ποιο τρόπο το πλαίσιο αυτό θα μπορούσε να επαυξηθεί έτσι ώστε η συναισθηματική επήρεια να λαμβάνεται υπόψη.

Απαντήσεις στα παραπάνω ερωτήματα δίνει η παρούσα διατριβή. Με αυτό τον τρόπο συμβάλει στην προαγωγή της έρευνας πάνω στην συναισθηματική επήρεια του ήχου στον άνθρωπο και οι απαντήσεις σε αυτά τα ερωτήματα συνιστούν την πρωτοτυπία της.

3.2 Θεωρητική προσέγγιση

Για την θεωρητική προσέγγιση με στόχο την διερεύνηση των παραπάνω ερωτημάτων ακολουθήθηκε μία σειρά βημάτων. Έτσι, αρχικά έγινε η επιλογή του τύπου του συναισθηματικού μοντέλου που θα χρησιμοποιόταν στα επόμενα στάδια της έρευνας. Αυτή συνίσταται στην επιλογή συναισθηματικών μοντέλων, διερεύνηση σχέσεων μεταξύ των στοιχείων του μοντέλου που επιλέχθηκε, επιλογή των στοιχείων των ηχητικών γεγονότων που θα αναλυθούν για την αναγνώριση συναισθηματικής πληροφορίας και σχεδιασμό των πειραμάτων εξόρυξης αυτής της πληροφορίας. Έπειτα, αναζητήθηκαν τα κεντρικά τεχνικά χαρακτηριστικά τα οποία καταστούν ήχους ικανούς να εγείρουν ακραίες (από άποψη Διέγερσης και Ευχαρίστησης)

συναισθηματικές καταστάσεις στους ακροατές. Για αυτό το σκοπό χρησιμοποιήθηκαν φωνητικές εντολές οι οποίες εκφέρθηκαν με αντίστοιχη συναισθηματική φόρτιση. Αν και αυτό εμπίπτει στην αναγνώριση συναισθήματος από ομιλία, η χρήση φωνητικών εντολών και όχι ελεύθερης ομιλίας και τα χρησιμοποιούμενα τεχνικά μεγέθη έδωσαν την ευκαιρία για πρώτη τριβή με την εξόρυξη συναισθηματικής πληροφορίας, την αρχική συσχέτιση τεχνικών χαρακτηριστικών με το εγείρομενο συναίσθημα και την ανάδειξη των τεχνικών χαρακτηριστικών εκείνων που μπορούν να εγείρουν ακραίες τιμές συναισθηματικών καταστάσεων στο μοντέλο που επιλέχθηκε. Έχοντας το συναισθηματικό μοντέλο και μία πρώτη εκτίμηση της επίρειας των τεχνικών χαρακτηριστικών στο εγείρομενο συναίσθημα, ακολούθησε η αναζήτηση της θέσης της συναισθηματικής επίρειας από γενικούς ήχους σε ένα ευρύτερο πλαίσιο ή/και μοντέλο. Προς αυτό το σκοπό προτάθηκε ένα νέο μοντέλο Ακουστικής Οικολογίας το οποίο είναι μία επαύξηση του υφιστάμενου μοντέλου. Τέλος, δεδομένου του θεωρητικού ορισμού του πλαισίου της συναισθηματικής διάστασης του ηχητικού περιβάλλοντος του ανθρώπου, έγινε προσπάθεια ορισμού των στοιχείων εκείνων που μετέχουν στην συναισθηματική επίρεια του ανθρώπου από τα ηχητικά γεγονότα που των περιβάλλουν.

3.2.1 Συναισθηματικά μοντέλα

Κατόπιν δευτερογενούς έρευνας στην υφιστάμενη βιβλιογραφία και της οποίας τα αποτελέσματα μπορούν να φανούν στην Ενότητα 2.2, επιλέχθηκε ένα διαστατικό μοντέλο δύο διαστάσεων, ήτοι: α) Διέγερση, και β) Ευχαρίστηση. Αν και θα μπορούσε να εισαχθεί και τρίτη διάσταση, αυτή θα συνεπαγόταν και παραπάνω πολυπλοκότητα [154]. Έτσι, λόγω του ήδη μικρού ερευνητικού υπόβαθρου στον τομέα της εξόρυξης συναισθηματικής πληροφορίας από ηχητικά γεγονότα, επιλέχθηκε να χρησιμοποιηθούν οι δύο προαναφερθείσες και καθιερωμένες διαστάσεις.

Πιο συγκεκριμένα, η θεωρητική προσέγγιση στα συναισθηματικά μοντέλα έγινε με το μοντέλο των βασικών συναισθημάτων. Από την μία, το μοντέλο αυτό μπορεί να προσφέρει ιδιαίτερη ευκολία σε εργασίες που σχετίζονται με την σχέση ανθρώπινων φυσιολογικών λειτουργιών αλλά, από την άλλη, δεν επιτρέπει αφενός μία αναλυτική προσέγγιση της μεταβολής του συναισθήματος κατά την ακρόαση διαφορετικών ήχων και αφετέρου την χρήση αποτελεσμάτων από άλλες έρευνες. Το πρώτο οφείλεται στον διακριτό χαρακτήρα των λεκτικών περιγραφών

των συναισθημάτων που χρησιμοποιούνται για το μοντέλο. Έτσι, οι περιγραφές αυτές αν και μπορούν να εκληφθούν ως μία ομαδοποίηση των περιοχών ενός διαστατικού μοντέλου, η υποβόσκουσα θεωρία του μοντέλου των βασικών συναισθημάτων έχειδειχθεί ότι δεν μπορεί να ισχύσει σε όλες τις περιπτώσεις [92]. Το δεύτερο, δηλαδή η δυσκολία ενσωμάτωσης αποτελεσμάτων προηγούμενων ερευνών λόγω της χρήσης συγκεκριμένων λεκτικών περιγραφών συναισθημάτων, έχει ήδη περιγραφεί και καταγραφεί ως πρόβλημα στην διεθνή επιστημονική βιβλιογραφία. Λόγω της διαφορετικής γλώσσας, δηλαδή Ελληνικά και Αγγλικά, το πρόβλημα αυτό μεγαλώνει. Αναλυτικότερα, η δυσκολία της χρήσης αποτελεσμάτων άλλων ερευνών που έχουν χρησιμοποιήσει κάποιο διακριτό μοντέλο έγκειται στις καθαυτές λεκτικές περιγραφές. Εφόσον δεν υπάρχει κάποιο καθιερωμένο, κοινά αποδεκτό και ευρέως χρησιμοποιούμενο σετ τέτοιων περιγραφών, η επιλογή των λέξεων για τα συναισθήματα που θα χρησιμοποιηθούν επαφίεται στον εκάστοτε ερευνητή. Έτσι, έχουν ήδη αναφερθεί προβλήματα μεταξύ χρήσης συνωνύμων λέξεων σε εργασίες που είναι γραμμένες στην αγγλική γλώσσα [14]. Το πρόβλημα αυτό θεωρήθηκε σχεδόν σίγουρο ότι θα κλιμακωνόταν εφόσον θα έπρεπε είτε να χρησιμοποιηθεί μία μετάφραση κάποιων κοινών και συχνότερα χρησιμοποιούμενων λέξεων είτε να χρησιμοποιηθούν λέξεις στα αγγλικά. Στην μεν πρώτη περίπτωση δημιουργούνται οι δυσκολίες της έτερο-αξιολόγησης ερευνών που χρησιμοποιήσαν άλλο σετ λέξεων καθώς και της διατήρησης του ακριβούς νοήματος των λέξεων κατόπιν της απαραίτητης μετάφρασής τους. Στην δε δεύτερη, δημιουργείται το πρόβλημα της συνέχειας και ορθής αντίληψης του νοήματος από ανθρώπους με μητρική γλώσσα άλλη από την αγγλική. Συνεπώς, το μοντέλο αυτό αν και προσέφερε μία καταρχήν βάση για την περαιτέρω έρευνα δεν υιοθετήθηκε.

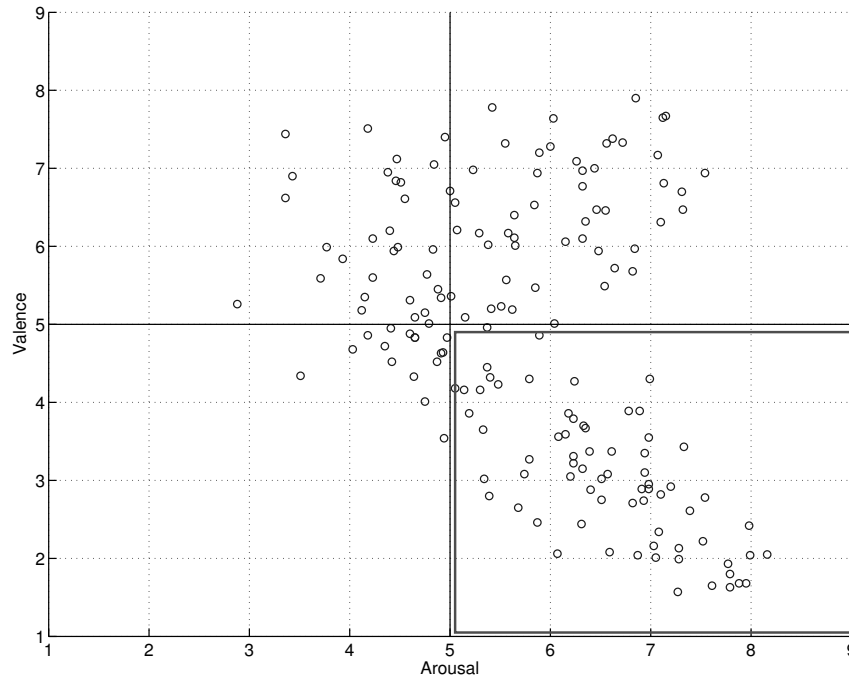
Τα παραπάνω συντέλεσαν στο να υπάρχουν δύο τελικές επιλογές για το μοντέλο που θα επιλεγόταν στην μετέπειτα έρευνα. Η πρώτη είναι ο συνδυασμός διαστατικού με διακριτού και η δεύτερη το αμιγώς διαστατικό. Όμως, και όπως μπορεί να φανεί από την Ενότητα 2.2.6, στην μικρή και υφιστάμενη επιστημονική βιβλιογραφία πάνω στον τομέα της εξόρυξης συναισθηματικής πληροφορίας από ηχητικά γεγονότα, το ευρέως χρησιμοποιούμενο μοντέλο είναι διαστατικό. Επιπροσθέτως, από την Ενότητα 2.2.2 μπορεί να φανεί η μετακύλιση της προτίμησης των εργασιών σε διαστατικά μοντέλα. Συνεπώς, το διαστατικό μοντέλο δύο διαστάσεων, Διέγερσης και Ευχαρίστησης, επιλέχθηκε για την παρούσα διατριβή.

3.2.2 Συσχέτιση υψηλής διέγερσης και χαμηλής ευχαρίστησης με τεχνικά χαρακτηριστικά

Έχοντας ως δεδομένο το μοντέλο Διέγερσης – Ευχαρίστησης, επιχειρήθηκε να απαντηθεί το ερώτημα σχετικά με την ύπαρξη ή μη και αναγνώριση (εάν υπάρχει) της κύριας διαφοράς μεταξύ ακραίων καταστάσεων στο μοντέλο αυτό. Κάτι τέτοιο θα μπορούσε να οριοθετήσει ποιοτικά και ποσοτικά τα τεχνικά χαρακτηριστικά που έχουν πρωτεύοντα ρόλο στην επήρεια της συναισθηματικής κατάστασης του ανθρώπου. Επίσης, με την υιοθέτηση (όπως γράφτηκε παραπάνω) των φωνητικών εντολών ως ηχητικό ερέθισμα, αυτή η προσέγγιση θα φανέρωνε πιθανή ύπαρξη καθολικών χαρακτηριστικών (δηλαδή, με ισχύ και στις τρεις κατηγορίες ηχητικών ερεθισμάτων) ικανά να επηρεάσουν συναισθηματικά τον ακροατή. Για αυτό το σκοπό έγινε μία ομαδοποίηση της συναισθηματικής κατάστασης η οποία αναφέρεται ως αλληλένδετη στην διεθνή βιβλιογραφία [103], δηλαδή της μεγάλης διέγερσης και ελάχιστης ευχαρίστησης. Αναφορικά με ηχητικά γεγονότα, αυτή η κατάσταση μπορεί να φανεί και στο **Σχήμα 2.9** όπου αναπαρίσταται από το τέταρτο τεταρτημόριο του σχήματος που ορίζουν οι άξονες και έχει ως κέντρο του το σημείο (5, 5).

Πιο συγκεκριμένα, η συναισθηματική κατάσταση αυτή μοντελοποιήθηκε και αντιστοιχήθηκε στην βίωση φόβου και θυμού. Ο συνδυασμός αυτών των δύο συναισθημάτων, και για λόγους καλύτερου προσδιορισμού στους συμμετέχοντες των πειραμάτων, αντιστοιχήθηκε στην κατάσταση του stress. Μία γραφική απεικόνιση της αντιστοίχησης της κατάστασης του stress με αποτυπωμένες τις συναισθηματικές επισημάνσεις των ήχων IADS είναι στο **Σχήμα 3.1**. Για την διαδικασία αναγνώρισης συναισθήματος, σχεδιάστηκε ένα πείραμα εξόρυξης συναισθηματικής πληροφορίας από φωνητικές εντολές με συμμετοχή μη επαγγελματιών ηθοποιών. Θα πρέπει να σημειωθεί όμως πως οι συμμετέχοντες είχαν σχέση με τον καλλιτεχνικό χώρο ως φοιτητές του Τμήματος Τεχνών Ήχου & Εικόνας, κάτι που μπορεί ίσως να παρέχει ένα εχέγγυο για την αναπαράσταση συγκεκριμένων συναισθηματικών καταστάσεων κατά βούληση.

Όπως μπορεί να φανεί από το **Σχήμα 3.1**, τα ηχητικά γεγονότα που ανήκουν στους IADS και είναι στο τεταρτημόριο που αντιστοιχήθηκε το stress δείχνουν μία περισσότερη συγκέντρωση όσο αφορά την απόστασή τους από την διαγώνιο του τεταρτημορίου αυτού αλλά και έχουν τις μεγαλύτερες τιμές διέγερσης και τις ελάχιστες ευχαρίστησης. Το πρώτο γεγο-



Σχήμα 3.1: Απεικόνιση της μοντελοποίησης της κατάστασης του stress σε σχέση με το μοντέλο Διέγερσης (Arousal) - Ευχαρίστησης (Valence) και των επισημάνσεων των ήχων IADS

νός δείχνει πως η αύξηση της διέγερσης και η ελάττωση της ευχαρίστησης έχουν κάποια αναλογία και το δεύτερο ότι είναι μία κατάλληλη περιοχή στο μοντέλο Διέγερσης – Ευχαρίστησης για να χρησιμοποιηθεί ως αυτόνομη ακραία κατάσταση. Ο λόγος είναι πως το τρίτο τεταρτημόριο (χαμηλή ευχαρίστηση) δεν έχει καθόλου σχεδόν ηχητικά γεγονότα ενώ το πρώτο τεταρτημόριο (υψηλή διέγερση) έχει μεγάλη διασπορά στις επισημάνσεις των ηχητικών γεγονότων. Τέλος, το δεύτερο τεταρτημόριο έχει ελάχιστες καταχωρήσεις. Άρα, για τους παραπάνω λόγους επιλέχθηκε η μοντελοποίηση της συναισθηματικής κατάστασης του stress η οποία και αντιστοιχήθηκε στο τρίτο τεταρτημόριο. Αυτό συντέλεσε σε δύο συναισθηματικές καταστάσεις, ήτοι: α) stress, και β) όχι stress.

Ως τεχνικά χαρακτηριστικά επιλέχθηκαν τέτοια που να μπορούν να αντιστοιχηθούν σε ηχητικές αλλά και μουσικές ιδιότητες. Έτσι, επιλέχθηκε ένα σετ χαρακτηριστικών που μπορούν να προσεγγίσουν την προσωδία αλλά και να είναι χρήσιμα σε εξόρυξη συναισθηματικής πληροφορίας από ηχητικά γεγονότα ή/και μουσική. Αυτά τα μεγέθη απαριθμούνται στον **Πίνακα 3.1**. Η επιλογή τους έγινε βάσει συναισθηματικών ηχητικών προφίλ της ομιλίας για την κατάσταση της χαμηλής Ευχαρίστησης και υψηλής Διέγερσης από δημοσιευμένες εργασίες [155, 147].

Πίνακας 3.1: Τα τεχνικά χαρακτηριστικά που επιλέχθηκαν για εξαγωγή από τα ηχητικά δεδομένα φωνητικών εντολών

A/A	Σύμβολο	Ονομασία
1	F_0	Θεμελιώδης συχνότητα
2	S_R	Ρυθμός ομιλίας (speech rate)
3	E	Ενέργεια

3.2.3 Ένταξη συναισθηματικής συνιστώσας στο μοντέλο ηχητικού περιβάλλοντος του ανθρώπου

Τα ηχητικά γεγονότα, ως το σύνολο των ήχων που μπορεί να ακούσει ένας άνθρωπος μία δεδομένη χρονική στιγμή, αποτελούν το ηχητικό του περιβάλλον [3]. Με αυτό μπορεί να αλληλεπιδράσει αλλά και αντλήσει πληροφορίες από αυτό, όπως κίνηση, μέγεθος της πηγής αλλά και σχετική της θέση (π.χ. μπροστά, πίσω, πλάγια). Την σχέση του ηχητικού περιβάλλοντος με τον άνθρωπο μελετάει το αντικείμενο της Ακουστικής Οικολογίας [156, 157]. Μέσα στο ηχητικό περιβάλλον και από τις κατηγορίες ηχητικών ερεθισμάτων, ήτοι: α) ομιλία, β) μουσική, και γ) ηχητικά γεγονότα, η τελευταία είναι αυτή στην οποία ανήκουν κατά πλειοψηφία τα ερεθίσματα που λαμβάνει ο άνθρωπος [3]. Συνδυάζοντας από την μία το γεγονός ότι κατά κύριο λόγο ο άνθρωπος αντιλαμβάνεται ηχητικά γεγονότα από το περιβάλλον του και, από την άλλη, ότι αυτά φαίνεται ότι μπορούν να του εγείρουν συναισθήματα, γεννάται το ερώτημα σχετικά με το ποια είναι η θέση των ηχητικών αυτών ερεθισμάτων μέσα στο ακουστικό περιβάλλον και πως ορίζονται. Σε αυτό το ερώτημα, ανάμεσα και στα άλλα που παρουσιάστηκαν παραπάνω, εστίασε, υπό ένα θεωρητικό πρίσμα, η ανά χειράς διατριβή.

λόγω της διαφαινόμενης συναισθηματικής επίδρασης στον άνθρωπο που έχουν τα ηχητικά γεγονότα, αναζητήθηκε ένα αρχικό πλαίσιο ορισμού αυτών έτσι ώστε να μπορούν να ενταχθούν στην ευρύτερη έννοια της ακουστικής οικολογίας. Το πλαίσιο αυτό θα πρέπει: α) να λαμβάνει υπόψη του τα εγγενή χαρακτηριστικά ενός ηχητικού γεγονότος, β) να προσδιορίζει την σημαντικότητα του σημασιολογικού περιεχομένου των γενικών ήχων στην συναισθηματική αλληλεπίδραση ανθρώπου/αχροατή και ηχητικών γεγονότων, και γ) να εισάγει και την συναισθηματική επήρεια του ανθρώπου από το ηχητικό του περιβάλλον. Για την διερεύνηση προς ένα τέτοιο πλαίσιο χρησιμοποιήθηκε το μόνο, κατά τον καιρό που έλαβε χώρα αυτή

Πίνακας 3.2: Τα στοιχεία του ηχητικού γεγονότος, όπως αυτό ορίζεται από την
Συναισθηματική Ακουστική Οικολογία

A/A Στοιχείο ηχητικού γεγονότος

- 1 Κυματομορφή (sound waveform)
Εκδήλωση της στιγμιαίας θέσης της πηγής σε σχέση με τον ακροατή
- 2 (manifestation of source's instantaneous spatial position, relative to receiver's)
- 3 Διάρκεια (duration)
- 4 Ένδειξη της διαδικασίας παραγωγής ήχου (indication of the sound's creation procedure)
- 5 Ένδειξη της φύσης των ταλαντευμένων σωμάτων, δηλαδή στερεά, υγρά, αέρια (evidence of vibrating objects' nature state)
- 6 Σημασιολογικό περιεχόμενο (semantic context)

η διερεύνηση, διαθέσιμο σετ ηχητικών γεγονότων, ήτοι το IADS, εντός του πλαισίου εργασίας που παρουσιάστηκε από τον Gaver [5] σχετικά με τα χαρακτηριστικά των ηχητικών γεγονότων.

Έτσι, προτείνεται το πλαίσιο της Συναισθηματικής Ακουστικής Οικολογίας (Affective Acoustic Ecology) [3]. Μέσα σε αυτό κυρίαρχο δομικό λόγο έχει το ηχητικό γεγονός το οποίο ορίζεται από τα στοιχεία που φαίνονται στον Πίνακα 3.2. Επίσης, προσδιορίζεται ότι πλέον το ηχητικό περιβάλλον του ανθρώπου μπορεί να επιδράσει πάνω στον ίδιο και συναισθηματικά και πως αυτή η σχέση μπορεί να έχει πρωτεύοντα ρόλο ιδιαίτερα κατά τον σχεδιασμό αστικών περιοχών [158]. Δηλαδή, η χρήση των στοιχείων της αστικής περιοχής μπορεί να οδηγήσει σε συγκεκριμένα ηχητικά πεδία τα οποία, σύμφωνα με την Συναισθηματική Ακουστική Οικολογία, μπορούν να επιδράσουν πάνω στους κατοίκους. Έτσι, μία κατάλληλη σχεδίαση του περιβάλλοντος της αστικής περιοχής μπορεί να επαυξήσει την ποιότητα ζωής των κατοίκων μειώνοντας αρνητικές συναισθηματικές αλληλεπιδράσεις των τελευταίων με το περιβάλλον τους.

Πιο συγκεκριμένα, σύμφωνα με το πλαίσιο που παρουσιάζεται από τον Gaver [5], διάφορα υλικά ή σώματα μπορούν να αλληλεπιδράσουν για την δημιουργία ήχου. Αναλόγως της φύσης των σωμάτων αλλά και του τρόπου αλληλεπίδρασης, το τελικώς παραγόμενο ηχητικό

γεγονός έχει τέτοια τεχνικά χαρακτηριστικά που επιτρέπουν την αναγνώριση της διαδικασίας παραγωγής ήχου αλλά και της φύσης των σωμάτων που αλληλεπέδρασαν. Τέτοια μπορεί να είναι η ενέργεια που περιέχεται στις υψηλές συχνότητες, η διαμόρφωση του πλάτους κ.α. Επιπροσθέτως, η σχετική, ως προς τον ακροατή, στιγμιαία θέση της πηγής αλλά και η ποιοτική φύση του περιβάλλοντος της πηγής (π.χ. μεγάλος χώρος, ανακλαστικός χώρος κτλ) είναι χαρακτηριστικά του ηχητικού γεγονότος που επίσης αντιλαμβάνεται ο άνθρωπος. Ακόμα, όπως μπορεί να φανεί από τους ήχους IADS, ένα ηχητικό γεγονός δύναται να εγείρει συναισθήματα είτε έχοντας ένα ισχυρό σημασιολογικό περιεχόμενο, π.χ. κλάμα μωρού, ήχος από άνθρωπο που πληγώνεται, χειροκρότημα κτλ., είτε χωρίς, π.χ. περπάτημα, ήχος γραφομηχανής, σφύριγμα, ήχος αέρα, ήχος από αφρίζουσα συσκευή (jip fizz) κτλ. Ο συνδυασμός των ανωτέρω με τα χαρακτηριστικά του ηχητικού γεγονότος, όπως αυτά φαίνονται στον **Πίνακα 3.2** και αναλύθηκαν στην Ενότητα 2.2.4, οδήγησε στην πρόταση του επαυξημένου μοντέλου της Ακουστικής Οικολογίας με σκοπό την εισαγωγή και μοντελοποίηση και της συναισθηματικής συνιστώσας στο ακουστικό περιβάλλον του ανθρώπου. Τέλος, με την προαναφερθείσα μοντελοποίηση δόθηκε η ευκαιρία για εστιασμένη διερεύνηση στα χαρακτηριστικά του ηχητικού γεγονότος, έτσι όπως αυτό ορίστηκε, και την επίρειά τους στην συναισθηματική κατάσταση του ακροατή.

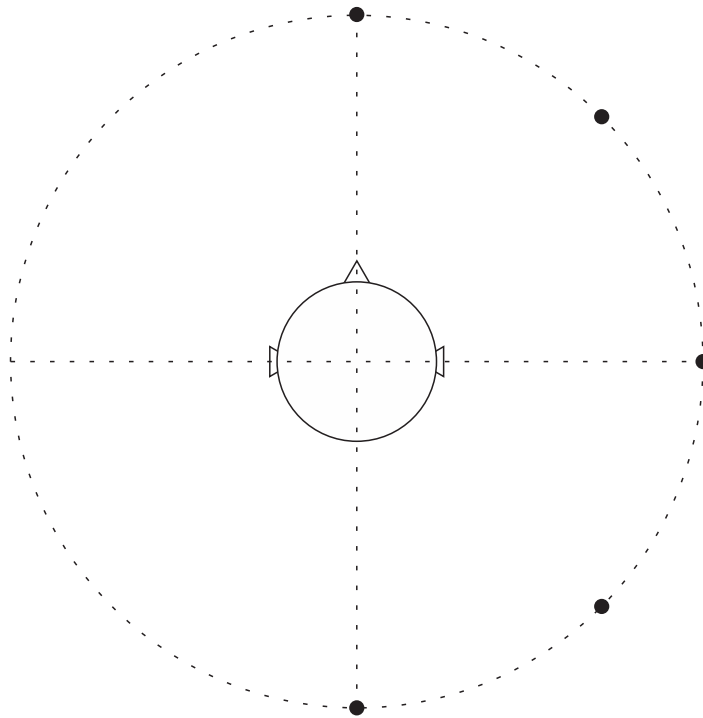
3.2.4 Επίρεια χωρικής τοποθέτησης της πηγής στο συναίσθημα

Ένα από τα χαρακτηριστικά που έχει το ηχητικό γεγονός είναι και η θέση της πηγής του, σχετικά με τον ακροατή. Αν και, όπως αναφέρθηκε προηγουμένως, έχουν αναφερθεί στην διεθνή βιβλιογραφία εργασίες που μελετάνε την επίρεια της θέσης της πηγής, αυτές αφορούν ένα συγκεκριμένο συναίσθημα (“Φόβος”) και περιορισμένες θέσεις της πηγής, ήτοι ακριβώς μπροστά και ακριβώς πίσω από τον ακροατή [37, 152]. Αναλυτική παρουσίαση των ερευνών αυτών έχει δοθεί προηγουμένως στην Ενότητα 2.2.6. Με βάση αυτά, δημιουργήθηκε το ερώτημα σχετικά με το πως μπορεί να επηρεάσει η θέση της πηγής την συναισθηματική κατάσταση του ακροατή. Υπάρχουν διαφορές στο εγειρόμενο συναίσθημα, αναλόγως της θέσης της πηγής; Δηλαδή, δεδομένου ενός ηχητικού γεγονότος που εγείρει συγκεκριμένη συναισθηματική κατάσταση, ο ακροατής θα αισθάνεται το ίδιο ακόμα και εάν η πηγή μετακινηθεί σχετικά με αυτόν; Π.χ., εάν είναι σε 45° , σε 90° ή και σε 135° η πηγή, θα έχει

διαφορετική επήρεια στην συναισθηματική κατάσταση του ακροατή; Εάν υπάρχει διαφορά, αυτή πως αποτυπώνεται σε ένα διαστατικό μοντέλο; Επίσης, θα πρέπει να σημειωθεί πως δεν υπάρχει κάποιο σετ ηχητικών γεγονότων με συναισθηματική επισήμανση που να παρέχει και χωρική τοποθέτηση αυτών.

Τα παραπάνω ερωτήματα ερευνήθηκαν από την παρούσα διατριβή. Η θεωρητική προσέγγιση σε αυτά έγινε με την επιλογή κατάλληλου πλαισίου για την χωρική τοποθέτηση της πηγής, επιλογή συναισθηματικού μοντέλου και τρόπου συναισθηματικής επισήμανσης και πρωταρχικού ηχητικού υλικού που να διαθέτει τέτοια επισήμανση. Αναλυτικότερα, οι μέχρι τώρα δημοσιευμένες εργασίες δεν είχαν μελετήσει αναλυτικά, τόσο χωρικά όσο και σε σχέση με το συναίσθημα, την επήρεια της θέσης της πηγής. Η μία [37] ερευνά αποκλειστικά μόνο δύο θέσεις της πηγής, συγκεκριμένα όταν η πηγή είναι μπροστά και πίσω από τον ακροατή, και ένα μόνο συναίσθημα, τον “Φόβο”. Τα αποτελέσματα της έρευνας αυτής έδειξαν ότι η θέση της πηγής έχει επήρεια στο συναίσθημα του ακροατή. Από την άλλη, η δεύτερη εργασία [152] χρησιμοποίησε ένα υπολογιστικό μοντέλο ενός χώρου και μέσα σε αυτόν τοποθέτησε, εικονικά, την πηγή και τον δέκτη. Τα ηχητικά ερεθίσματα ου χρησιμοποιήθηκαν για την αναγνώριση συναισθήματος επεξεργάστηκαν ψηφιακά έτσι ώστε το τελικό ηχητικό αποτέλεσμα να αντιπροσωπεύει το πραγματικό για τις δεδομένες θέσεις πηγής και ακροατή μέσα στο μοντέλο του χώρου. Με αυτά τα τελικά ηχητικά ερεθίσματα διεξήχθησαν πειράματα αναγνώρισης συναισθήματος, επικεντρωμένα στον “Φόβο”, τα οποία έδειξαν ότι όσο πιο κοντά στον ακροατή είναι η πηγή και όσο πιο μεγάλος είναι ο χώρος μέσα στον οποίο είναι η πηγή τόσο μεγαλώνει ο φόβος που νιώθει ο ακροατής. Οι θέσεις που είχε η πηγή ήταν πάντα μπροστά από τον ακροατή. Παρόλα αυτά, παρέμεινε το ερώτημα σχετικά με άλλες θέσεις της πηγής αλλά και με τις συναισθηματικές καταστάσεις, αλλά και η ανάγκη για την ύπαρξη ενός σετ ηχητικών γεγονότων με συναισθηματική επισήμανση τα οποία να έχουν και χωρική τοποθέτηση. Το ερώτημα αυτό ενισχύεται από το γεγονός ότι όλα τα διαθέσιμα σετ συναισθηματικά επισημασμένων ηχητικών γεγονότων περιέχουν μονοφωνικά δεδομένα χωρίς χωρική τοποθέτηση και οι υφιστάμενες μελέτες εξόρυξης συναισθηματικής πληροφορίας από ηχητικά γεγονότα δεν λαμβάνουν την χωρική τοποθέτηση υπόψη. Ακόμα, αν και η απόσταση φαίνεται ότι επηρεάζει το συναίσθημα του ακροατή, δεν μελετήθηκε λόγω της εστίασης που έγινε στην έλλειψη ερευνητικών αποτελεσμάτων και ενός αναλυτικού μοντέλου για την επήρεια της χωρικής τοποθέτησης της πηγής αναφορικά με τον ακροατή.

Για να μελετηθεί η επίρεια της χωρικής τοποθέτησης της πηγής πρέπει να χρησιμοποιηθεί τεχνολογία κατάλληλη να μπορεί να αναπαραστήσει το ηχητικό ερέθισμα μπροστά, σε πλευρικές θέσεις αλλά και πίσω από τον ακροατή. Μία τυπική και ώριμη τεχνολογία είναι η αμφιωτική επεξεργασία. Με αυτή, το ηχητικό ερέθισμα μπορεί να αναπαρασταθεί γύρω από τον ακροατή και σε εύρος γωνιών από 0 έως και 360 μοίρες με μόνο περιορισμό την χρήση ακουστικών από τον ακροατή, αντί για ηχείων. Η τεχνολογία αυτή μοντελοποιεί το ακουστικό όργανο του ανθρώπου και τον τρόπο που επιδρά αυτό πάνω στα ηχητικά κύματα που έρχονται από διάφορες διευθύνσεις μέσω ενός πλήθους συναρτήσεων μεταφοράς, διαφορετικές για κάθε αυτί. Κάθε μία από αυτές τις συναρτήσεις ανταποκρίνεται σε διαφορετική γωνία μεταξύ της πηγής και του ακροατή. Έτσι, για να αναπαρασταθεί ένα ηχητικό ερέθισμα σε μία συγκεκριμένη γωνία γύρω από τον ακροατή, τα ηχητικά δεδομένα του ερεθίσματος φιλτράρονται με τις συναρτήσεις μεταφοράς για την γωνία αυτή και για το δεξί και αριστερό αυτί. Η διαδικασία αυτή έχει ως αποτέλεσμα την παραγωγή δύο καναλιών ήχου, ένα για το αριστερό και ένα για το δεξί αυτί. Τέλος, με την αναπαραγωγή των δύο καναλιών αυτών μέσω ακουστικών επιτυγχάνεται η χωρική τοποθέτηση του ηχητικού ερεθίσματος [159]. Έτσι, για μελέτη την επίρειας της θέσης της πηγής χρησιμοποιήθηκε η αμφιωτική τεχνολογία και το σύνολο ήχων IADS. Βάσει της συμμετρικής ιδιότητας των συναρτήσεων μεταφοράς που χρησιμοποιούνται κατά την αμφιωτική επεξεργασία, έγινε χωρική τοποθέτηση της πηγής σε 5 γωνίες, από εντελώς μπροστά στον ακροατή μέχρι και εντελώς πίσω του. Για την μοντελοποίηση των συναισθημάτων χρησιμοποιήθηκε το μοντέλο Διέγερσης – Ευχαρίστησης και η αποτύπωση έγινε με την μέθοδο SAM, που αναλύεται στην Ενότητα 2.1.6, ενώ παράχθηκε ένα πρωτότυπο, μοναδικό και το πρώτο παγκοσμίως σετ ηχητικών γεγονότων με συναισθηματική επισήμανση και χωρική τοποθέτηση, το οποίο είναι και ένα από τα 3 διαθέσιμα σετ με συναισθηματικά επισημασμένα ηχητικά γεγονότα. Η αναλυτική περιγραφή των πειραμάτων καθώς και τα αποτελέσματα είναι σε επόμενες ενότητες. Στο **Σχήμα 3.2** είναι μία απεικόνιση της χωρικής τοποθέτησης της πηγής όπως προσεγγίστηκε κατά τον σχεδιασμό των πειραμάτων.



Σχήμα 3.2: Η χωρική τοποθέτηση της πηγής όπως επιλέχθηκε για την μελέτη της επίρρειάς της στην συναισθηματική κατάσταση του ακροατή

3.2.5 Συσχέτιση τεχνικών χαρακτηριστικών και συναισθήματος

Μετά από τα παραπάνω, έγινε η διερεύνηση της σχέσης των τεχνικών χαρακτηριστικών ενός ηχητικού γεγονότος και του εγείρομένου συναισθήματος. Αυτή η προσέγγιση έγινε για την προσπάθεια απάντησης στο ερώτημα εάν υπάρχουν κάποια συγκεκριμένα χαρακτηριστικά των ηχητικών γεγονότων που μπορούν να επηρεάσουν την συναισθηματική κατάσταση του ακροατή και, εάν υπάρχουν, ποια είναι αυτά. Εφόσον τα ηχητικά γεγονότα μπορούν να παραχθούν από όλες τις πιθανές πηγές ήχου, ακόμα και τον άνθρωπο, και εφόσον είναι ικανά να μεταφέρουν και σημασιολογικό περιεχόμενο, μία απάντηση σε αυτό το ερώτημα θα μπορούσε να αποκαλύψει ίσως μία εγγενή επικοινωνία του ανθρώπου με τη φύση μέσω της αυθόρμητης αναγνώρισης από τον πρώτο κάποιων συγκεκριμένων χαρακτηριστικών του ήχου. Δηλαδή, εάν κάποια χαρακτηριστικά των ηχητικών γεγονότων είναι ικανά να προκαλέσουν συγκεκριμένα συναισθήματα στον άνθρωπο αυτό ενδέχεται να σημαίνει πως στην πορεία εξέλιξης του ανθρώπου έχει δημιουργηθεί και έχει κληρονομηθεί η γνώση για την αναγνώριση συγκεκριμένων τεχνικών χαρακτηριστικών του ηχητικού περιβάλλοντος τα οποία μπορούν να

επιδράσουν πάνω του.

Για αυτό το σκοπό πρώτα επιλέχθηκαν τα τεχνικά χαρακτηριστικά που θα εξαχθούν από τα ηχητικά γεγονότα, έπειτα το σετ ηχητικών γεγονότων που θα χρησιμοποιούταν και τέλος έγιναν πειράματα εξόρυξης συναισθηματικής πληροφορίας. Η διεξαγωγή των τελευταίων ήταν σύμφωνη με την διαδικασία που εικονίζεται στο Σχήμα 2.1. Έτσι, επιλέχθηκαν τεχνικά χαρακτηριστικά από όλες τις κατηγορίες που εμφανίζονται στον Πίνακα 2.2. Ως σετ ήχων επιλέχθηκαν οι ήχοι IADS. Από αυτούς εξάχθηκαν όλα τα χαρακτηριστικά που επιλέχθηκαν και με αυτά πραγματοποιήθηκαν πειράματα κατηγοριοποίησης. Τα τελευταία άρχισαν με την πρώτη προσέγγιση η οποία χρησιμοποίησε ηχητικά γεγονότα, ακολουθώντας το μοντέλο της Συναισθηματικής Ακουστικής Οικολογίας και συνεχίστηκε με την διερεύνηση της επίρειας του ρυθμού στην διέγερση και την ευχαρίστηση.

Κατά την διαδικασία αυτή χρησιμοποιήθηκαν αρκετοί αλγόριθμοι και χαρακτηριστικά, που παρατίθενται σε επόμενη ενότητα. Ο λόγος για αυτό είναι πως στην βιβλιογραφία δεν αναφέρεται κάτι παρόμοιο και έτσι έγινε η προσπάθεια να ερευνηθούν αρκετά χαρακτηριστικά και αλγόριθμοι μηχανικής μάθησης. Αναλυτική περιγραφή των πειραμάτων δίνεται στις επόμενες ενότητες.

3.3 Πειραματική προσέγγιση

Κάθε στάδιο της θεωρητικής προσέγγισης υποστηρίχθηκε και τεκμηριώθηκε από παράλληλη πειραματική μελέτη. Η τελευταία επέτρεψε την πειραματική και πρακτική επιβεβαίωση της θεωρητικής προσέγγισης και τα αποτελέσματα φανέρωσαν απαραίτητες αλλαγές, συμπληρώσεις ή επεκτάσεις που έπρεπε να γίνουν. Σε αυτή την ενότητα παρουσιάζεται η ακριβής οργάνωση και οι τεχνικές λεπτομέρειες κάθε πειράματος.

3.3.1 Αναγνώριση stress από φωνητικές εντολές

Για την διερεύνηση των τεχνικών χαρακτηριστικών που έχουν πρωτεύοντα ρόλο σε μία κατάσταση υψηλής διέγερσης και χαμηλής ευχαρίστησης, όπως αυτή περιγράφηκε στην Ενότητα 3.2.2, έγιναν πειράματα ανίχνευσης stress από φωνητικές εντολές. Ως stress μοντε-

Πίνακας 3.3: Η μεταβολή της τιμής των τεχνικών χαρακτηριστικών που επιλέχθηκαν για την αναγνώριση stress σε σχέση με τα βασικά συναισθήματα. Με F_0 είναι η θεμελιώδης συχνότητα, ως S_R συμβολίζεται ο ρυθμός ομιλίας (speech rate) και E είναι η ενέργεια

Συναίσθημα	Τεχνικό χαρακτηριστικό		
	F_0	S_R	E
Θυμός	Αύξηση	Αύξηση	Αύξηση
Χαρά	Αύξηση	Μείωση	Αύξηση
Λύπη	Μείωση	Μείωση	Μείωση
Φόβος	Αύξηση	Αύξηση	Αύξηση

λοποιήθηκε η συναισθηματική κατάσταση χαμηλής Ευχαρίστησης και υψηλής Διέγερσης, δηλαδή, και σύμφωνα με τις αντιστοιχίσεις διακριτών συναισθημάτων που αναφέρονται σε δημοσιευμένες έρευνες, η κατάσταση ταυτόχρονης βίωσης Φόβου και Θυμού, ενώ ως τεχνικά χαρακτηριστικά επιλέχθηκαν εκείνα που αφενός αναφέρονται στην βιβλιογραφία ως χρήσιμα για την ανίχνευση Φόβου και Θυμού και, αφετέρου, μπορούν να μοντελοποιήσουν την προσωδία της ανθρώπινης φωνής [147, 160]. Η τελευταία είναι τα μουσικά χαρακτηριστικά της εκφοράς λέξεων, δηλαδή η διακύμανση του ρυθμού, της ενέργειας και της θεμελιώδους συχνότητας [161], και κρίθηκε χρήσιμο να χρησιμοποιηθεί λόγω της εγγενούς, και προφανούς, σχέσης της με την μουσική. Η σχέση της μεταβολής της τιμής των τεχνικών χαρακτηριστικών που επιλέχθηκαν, και αναφέρονται στην Ενότητα 3.2.2, με τα βασικά συναισθήματα στα οποία ανήκει ο Φόβος και ο Θυμός εμφανίζεται στον **Πίνακα 3.3**, σύμφωνα με δημοσιευμένες εργασίες [147, 160]. Στον Πίνακα αυτόν παρατίθενται οι μεταβολές των τεχνικών μεγεθών για τα αντίστοιχα συναισθήματα. Έτσι, κατά την βίωση θυμού παρατηρείται μία αύξηση στην θεμελιώδη συχνότητα της ομιλίας, μία αύξηση του ρυθμού ομιλίας καθώς και μία αύξηση στην ενέργεια που περιέχεται στο σήμα ομιλίας. Για την υλοποίηση του πειράματος ηχογραφήθηκαν 5 φωνητικές εντολές οι οποίες παρουσιάζονται στον **Πίνακα 3.4**.

Όπως μπορεί να παρατηρηθεί, το σημασιολογικό περιεχόμενο των λέξεων διαφέρει. Οι δύο πρώτες εκφέρονται συνήθως σε κατάσταση stress ενώ οι δύο τελευταίες όχι. Η τρίτη, και σε ένα οικιακό περιβάλλον, θα μπορούσε να εκφερθεί και με stresss, όπως στην άτυχη περίπτωση που ξεσπάσει φωτιά, ή και χωρίς, όπως στην περίπτωση που κάποιος ζητάει να του δώσουνε αναπτήρα, να ειδοποιήσει για φωτιά στο τζάκι κτλ. Η επιλογή των λέξεων βασίστηκε σε δύο

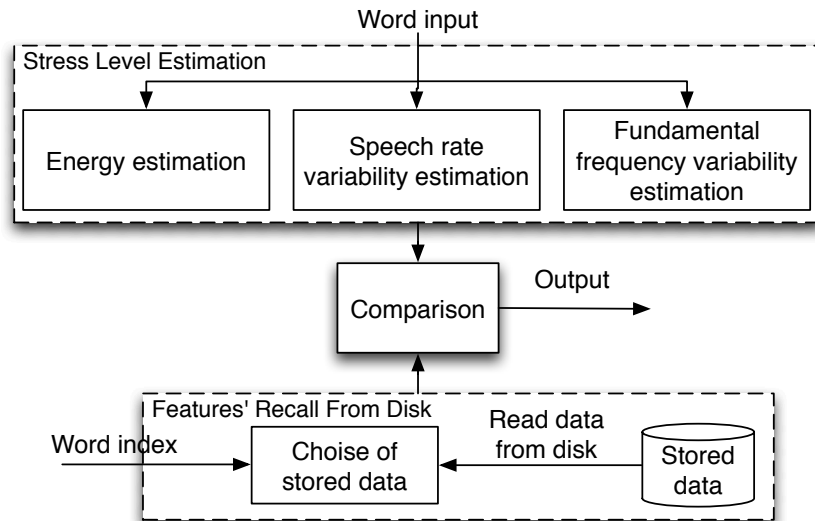
Πίνακας 3.4: Οι φωνητικές εντολές που χρησιμοποιήθηκαν στο σύστημα αναγνώρισης stress

A/A	Φωνητική εντολή	A/A	Φωνητική εντολή
1	Κλέφτης	4	Καλά
2	Σεισμός	5	Καλημέρα
3	Φωτιά		

παράγοντες. Ο πρώτος είναι η προαναφερθείσα διαφορά της ποιότητας του σημασιολογικού περιεχομένου τους. Με αυτή την επιλογή λέξεων και αντίστοιχου σημασιολογικού περιεχομένου ενδέχεται να αποκαλυφθεί η όποια εξάρτηση υπάρχει στην διαδικασία αναγνώρισης stress από το σημασιολογικό περιεχόμενο. Ο δεύτερος λόγος είναι το πλαίσιο στο οποίο υλοποιήθηκε το συγκεκριμένο πείραμα, το οποίο είναι η υλοποίηση ενός συστήματος αυτόματης αναγνώρισης φωνητικών εντολών για οικιακό περιβάλλον και επιλογής αυτοματοποιημένης δράσης αναλόγως την εντολή. Έτσι, στην περίπτωση που κάποιος χρησιμοποιήσει αυτό το σύστημα και θέλει να ανάψει το τζάκι του, θα πει απλώς την λέξη φωτιά. Εάν θέλει να παρθούν μέτρα, όπως ειδοποίηση της πυροσβεστικής υπηρεσίας, για κάποια φωτιά που έχει ξεσπάσει στο σπίτι τότε, και δεδομένου ότι θα πει την λέξη “φωτιά” με το ανάλογο stress, θα γίνει αναγνώριση από το σύστημα της λέξης και του επιπέδου stress και θα παρθούν τα αντίστοιχα μέτρα.

Η διαδικασία εξόρυξης συναισθηματικής πληροφορίας υλοποιήθηκε με την χρήση ενός συστήματος αναγνώρισης stress που αναπτύχθηκε συγκεκριμένα για αυτό το σκοπό. Αποτελείται από δύο υπομονάδες, α) μία για την εξαγωγή των τεχνικών χαρακτηριστικών και την υλοποίηση της αναγνώρισης του επιπέδου stress, και β) μία για την σύγκριση των τιμών που υπολογίστηκαν και των τιμών που είχαν καταγραφεί κατά την εκπαίδευση του συστήματος. Μία σχηματική αναπαράσταση φαίνεται στο **Σχήμα 3.3**.

Η πρώτη υπομονάδα φαίνεται στο Σχήμα με το διακριτικό “Stress Level Estimator” ενώ η δεύτερη με το “Comparison”. Όπως μπορεί να φανεί, από κάθε λέξη εξαγόntonταν τα τεχνικά χαρακτηριστικά που φαίνονται στον **Πίνακα 3.3** και στη συνέχεια οι τιμές τους χρησιμοποιούντουσαν ως είσοδο στην υπομονάδα σύγκρισης. Ταυτόχρονα, και βάσει της λέξης που εξυπηρετούσε ως αρχική είσοδο στο σύστημα, γινόταν ανάκληση των αντίστοιχων τιμών για την ίδια λέξη από την μνήμη που είχε στη διάθεσή του το σύστημα και είχαν



Σχήμα 3.3: Το σύστημα αναγνώρισης stress από φωνητικές εντολές

καταγραφεί κατά την διαδικασία εκπαίδευσης.

Για την διαδικασία εξαγωγής των τεχνικών χαρακτηριστικών, και κατά την διάρκεια αναγνώρισης και κατά την εκπαίδευση, δεν γινόταν κατάτμηση του σήματος. Η ενέργεια υπολογίστηκε σύμφωνα με την Εξίσωση:

$$E = \frac{1}{N} \sum_{i=1}^N S_i^2 \quad (3.1)$$

όπου E είναι η ενέργεια, N το πλήθος των δειγμάτων στο σήμα και S_i το δείγμα του σήματος. Η θεμελιώδης συχνότητα, F_0 , υπολογίστηκε βάσει της:

$$F_0 = \underset{\min(f)}{\operatorname{argmax}}(|S[f]|), \quad f = 1, 2, \dots, N_f \quad (3.2)$$

όπου S_f είναι ο συντελεστής του γρήγορου διακριτού μετασχηματισμού Fourier (FFT) για την ζώνη συχνοτήτων f και N_f το πλήθος των συντελεστών αυτών. Τέλος ο ρυθμός ομιλίας υπολογίστηκε από:

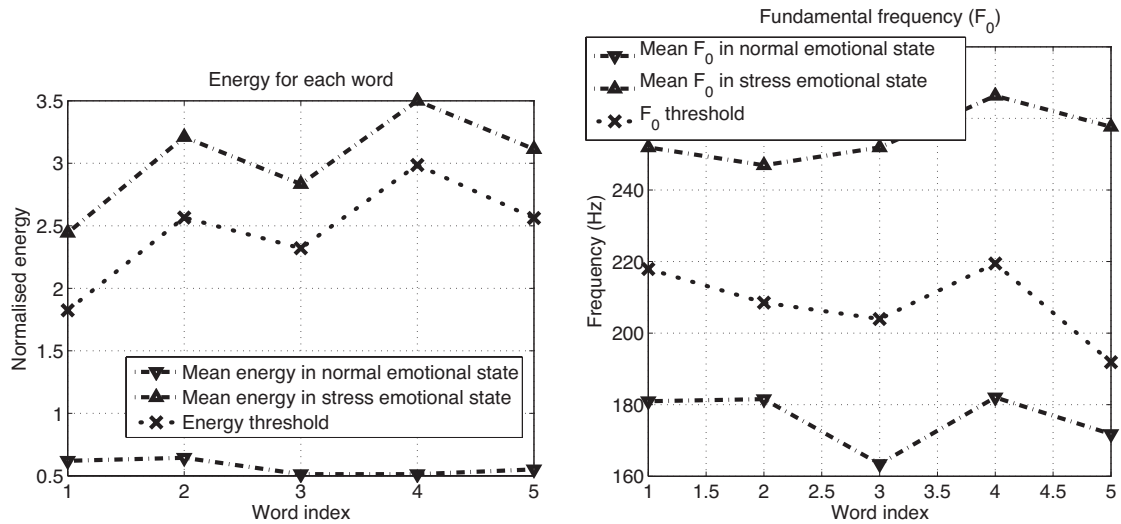
$$S_R = \frac{E(S[Sp])}{E(S[N])} \quad (3.3)$$

όπου S_R είναι ο ρυθμός ομιλίας, $E(S[Sp])$ είναι το πλήθος των δειγμάτων του σήματος που

έχουν δεδομένα ομιλίας και $E(S[N])$ είναι το πλήθος των δειγμάτων του σήματος που δεν έχουν ομιλία. Η εκπαίδευση του συστήματος αναγνώρισης έγινε με 10 συμμετέχοντες, με τους μισούς να είναι γένους θηλυκού, όπου κάθε ένας πρόφερε τις λέξεις με την προσομοίωση βίωσης stress και χωρίς. Σε κάθε συμμετέχων δόθηκαν οδηγίες σχετικά με την διαδικασία εκπαίδευσης πριν εκείνη αρχίσει ενώ κάθε λέξη ηχογραφήθηκε με ψηφιακά μέσα και κάθε ηχογράφηση είχε διάρκεια 2 δευτερολέπτων. Η διάρκεια αυτή επιλέχθηκε λόγω ότι θα γινόταν ηχογράφηση μίας λέξης και έτσι έπρεπε να υπάρχει ένα σταθερό χρονικό πλαίσιο εκφοράς της για να μπορέσει να υπολογιστεί ο ρυθμός ομιλίας. Επίσης, 2 δευτερόλεπτα αρκούν για την εκφορά μίας λέξης από αυτές που χρησιμοποιήθηκαν. Στη συνέχεια, έτεροι 10 διαφορετικοί ακροατές αξιολόγησαν την εκφορά των λέξεων σχετικά με το εάν μεταφέρεται αξιόπιστα το συναίσθημα του stress. Όσες ηχογραφήσεις παρουσίασαν διαφωνία μεταξύ του συναισθήματος που ήθελε να μεταφέρει ο ομιλητής και της αξιολόγησης και η διαφορά αυτή ήταν παραπάνω από 20% δεν χρησιμοποιήθηκαν για την καθαυτή εκπαίδευση του συστήματος, έτσι ώστε να ελαχιστοποιηθεί η υποκειμενική κρίση στην αξιολόγηση του εκφερόμενου συναισθήματος.

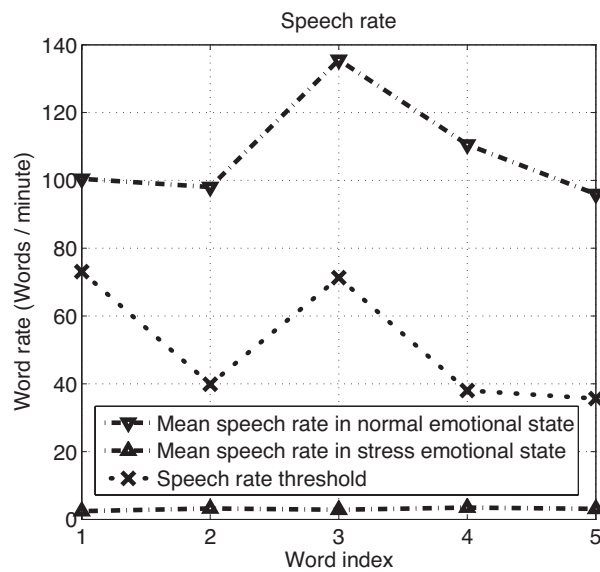
Από κάθε ηχογράφηση που χρησιμοποιήθηκε στην εκπαίδευση του συστήματος έγινε η εξαγωγή της ενέργειας, της θεμελιώδους συχνότητας και του ρυθμού ομιλίας, βάσει των εξισώσεων που δόθηκαν προηγουμένως. Για κάθε λέξη, έγινε ο υπολογισμός του μέσου όρου των προηγούμενων μεγεθών από όλες τις ηχογραφήσεις για αυτή τη λέξη και για κάθε συναισθηματική κατάσταση ξεχωριστά. Έτσι, για κάθε λέξη και για έκαστη συναισθηματική κατάσταση (με stress και χωρίς) υπήρχαν τρία μεγέθη, ήτοι: α) μέση ενέργεια, β) μέση θεμελιώδη συχνότητα, και γ) μέσος ρυθμός ομιλίας. Με την σύγκριση αυτών των τιμών μεταξύ των διαφορετικών συναισθηματικών καταστάσεων θεσπίστηκαν ποσοτικά όρια όπου οι τιμές πάνω από αυτά σηματοδοτούσαν την ύπαρξη stress. Τα όρια αυτά καθώς και οι μέσες τιμές για stress και χωρίς για την ενέργεια, θεμελιώδη συχνότητα και τον ρυθμό ομιλίας φαίνονται στα **Σχήματα 3.4α', 3.4β' και 3.4γ'** αντιστοίχως.

Θα πρέπει να σημειωθεί πως αν και στην βιβλιογραφία αναφέρεται μία αύξηση του ρυθμού ομιλίας κατά την βίωση Φόβου και Θυμού [160, 147, 162], στο Σχήμα 3.4γ' παρατηρείται μία μείωση. Εκτιμάται ότι, ο λόγος για αυτό είναι το γεγονός πως κατά την εκφορά φωνητικών εντολών ή μεμονωμένων λέξεων, ο ομιλητής στοχεύει στο να γίνει ακουστός και κατανοητός και έτσι επιμηχύνει την διάρκεια της λέξης [2]. Τα αποτελέσματα του παραπάνω πειράματος



(α') Το όριο και οι μέσες τιμές με stress και χωρίς για την ενέργεια

(β') Το όριο και οι μέσες τιμές με stress και χωρίς για την θεμελιώδη συχνότητα



(γ') Το όριο και οι μέσες τιμές με stress και χωρίς για τον ρυθμό ομιλίας

Σχήμα 3.4: Τα όρια και οι μέσες τιμές για stress και χωρίς για την ενέργεια, θεμελιώδη συχνότητα και ρυθμό ομιλίας

παρουσιάζονται αναλυτικά σε επόμενη ενότητα.

3.3.2 Αναγνώριση διέγερσης και ευχαρίστησης από ηχητικά γεγονότα

Κατά την πρόταση του νέου πλαισίου της Συναισθηματικής Ακουστικής Οικολογίας έγινε και ένα πείραμα εξόρυξης συναισθηματικής πληροφορίας από ηχητικά γεγονότα. Σε αυτό χρησιμοποιήθηκαν οι ήχοι IADS, 13 τεχνικά χαρακτηριστικά, διαστατικό μοντέλο Διέγερσης – Ευχαρίστησης με 4 διάκριτες περιοχές και επιβλεπόμενοι αλγόριθμοι κατηγοριοποίησης.

Πιο συγκεκριμένα, το πείραμα αυτό βασίστηκε στην τυπική διαδικασία εξόρυξης πληροφορίας από ηχητικά δεδομένα που φαίνεται στο Σχήμα 2.1. Ως ηχητικά και συναισθηματικής επισήμανσης δεδομένα χρησιμοποιήθηκαν αυτά που παρέχονται στο σετ IADS. Κάθε ήχος εξ αυτών προσφέρθηκε ως είσοδο στην διαδικασία εξαγωγής τεχνικών χαρακτηριστικών. Στην τελευταία, κάθε σήμα εισόδου της χωριζόταν σε μη επικαλυπτόμενα χρονικά πλαίσια μήκους 0.01 δευτερολέπτου. Από κάθε τέτοιο πλαίσιο εξήχθησαν τα τεχνικά χαρακτηριστικά που φαίνονται στον Πίνακα 3.5. Ο λόγος για την επιλογή αυτών των χαρακτηριστικών είναι διττός. Από την μία, δεν υπήρχαν, κατά την στιγμή της επιλογής των χαρακτηριστικών αυτών, δημοσιευμένες εργασίες πάνω στην αναγνώριση συναισθημάτων που να παρουσιάζουν ερευνητικά αποτελέσματα και έτσι να μπορεί να γίνει μία πρώτη εκτίμηση για την επιλογή των χαρακτηριστικών. Από την άλλη, τα χαρακτηριστικά αυτά τυγχάνουν ευρείας χρήσης στην αναγνώριση συναισθήματος από ομιλία και μουσική. Έτσι, αφενός επιλέχθηκαν ως μία πρώτη προσπάθεια προσέγγισης της διαδικασίας αναγνώρισης συναισθήματος από ηχητικά γεγονότα και αφετέρου επειδή έχουν χρησιμοποιηθεί στο παρελθόν σε συγγενικά πεδία και για παρόμοιες διαδικασίες, όπως φαίνεται και στο Κεφάλαιο 2.

Επίσης, για την έρευνα σχετικά με την επήρεια της μεταβολής των παραπάνω μεγεθών στο συναίσθημα του ακροατή, υπολογίστηκαν στατιστικά και μαθηματικά μεγέθη τα οποία παρατίθενται στον Πίνακα 3.6.

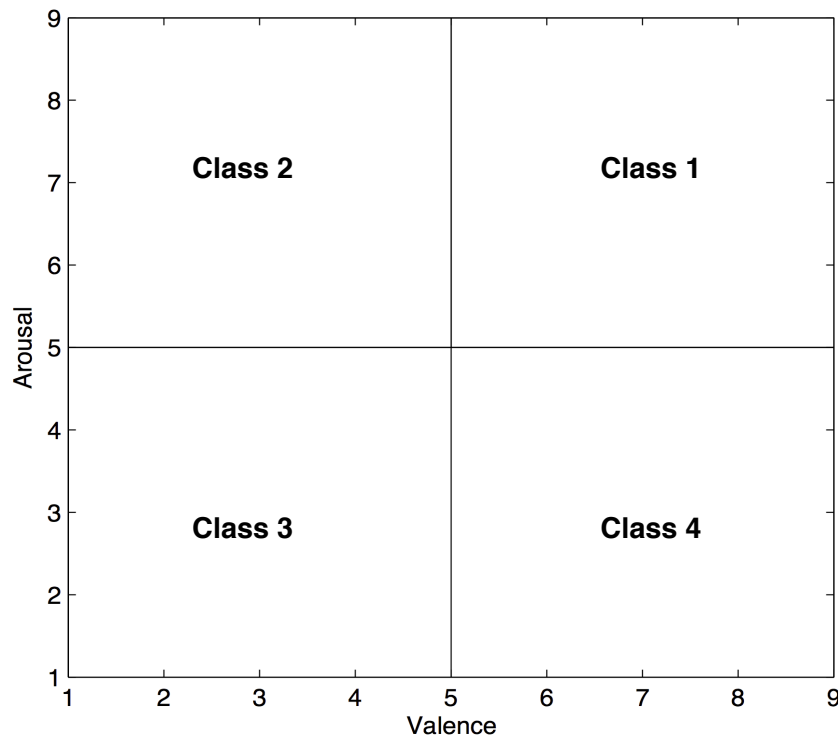
Η παραπάνω διαδικασία οδήγησε σε έναν χώρο 101 διαστάσεων για κάθε ηχητικό γεγονός. Όπου δεδομένου ότι οι MFCC ήταν 13, η ενέργεια ανά ζώνη συχνοτήτων αφορούσε ένα

Πίνακας 3.5: Τα τεχνικά χαρακτηριστικά που χρησιμοποιήθηκαν στο πείραμα εξόρυξης συναισθηματικής πληροφορίας από τους ήχους IADS

A/A	Τεχνικό χαρακτηριστικό	A/A	Τεχνικό χαρακτηριστικό
1	Ενέργεια ανά ζώνη συχνοτήτων	8	Spectral centroid
2	Μέση ενέργεια του σήματος	9	Spectral rolloff
3	Low energy	10	Spectral flux
4	Zero crossing rate	11	Spectral variability
5	MFCC	12	Compactness
6	Beat strength	13	Θεμελιώδης συχνότητα
7	Beat spectrum		

Πίνακας 3.6: Τα στατιστικά και μαθηματικά μεγέθη που χρησιμοποιήθηκαν στην εξόρυξη συναισθηματικής πληροφορίας από τους ήχους IADS

A/A	Μέγεθος
1	Πρώτη παράγωγος
2	Τυπική απόκλιση
3	Πρώτη παράγωγος της τυπικής απόκλισης



Σχήμα 3.5: Γραφική απεικόνιση της αντιστοίχησης τεταρτημορίων και κλάσεων συναισθηματικής κατάστασης που χρησιμοποιήθηκε

πλήθος τέτοιων ζωνών, για το spectral variability δεν υπολογίστηκαν στατιστικά μεγέθη (εφόσον αφορούσε όλο το σήμα) και για κάθε άλλο μέγεθος υπολογίστηκαν τρία στατιστικά μεγέθη, σύμφωνα με τον **Πίνακα ;;**, σημαίνει πως υπήρχαν $99/3 = 33$ διαφορετικά μεγέθη για κάθε σήμα, ένα το spectral variability και ένα ακόμα η κλάση του ηχητικού γεγονότος. Η εξαγωγή των τεχνικών χαρακτηριστικών και οι υπολογισμοί των στατιστικών και μαθηματικών μεγεθών έγιναν με την βοήθεια του λογισμικού JAudio¹⁰. Η τελευταία απαριθμούνταν από τέσσερις περιοχές οι οποίες ήταν αντιστοιχισμένες στα ισάριθμα τεταρτημόρια του χώρου του συναισθηματικού μοντέλου. Δηλαδή, η κλάση 1 στο πρώτο τεταρτημόριο, η κλάση 2 στο δεύτερο τεταρτημόριο κ.ο.κ. Μία γραφική απεικόνιση αυτής της αντιστοίχησης είναι στο **Σχήμα 3.5**.

Τέλος, την κατηγοριοποίηση υλοποίησαν δύο επιβλεπόμενοι αλγόριθμοι, ήτοι: α) Artificial Neural Networks (ANN), και β) Support Vector Machines (SVM). Η χρήση των αλγορίθμων και η εξέταση των παραγόμενων μοντέλων έγινε με την χρήση του περιβάλλοντος μηχανικής μάθησης WEKA¹¹. Επειδή δεν υπήρχε κάποιο ξεχωριστό σετ δεδομένων για την

¹⁰<http://jaudio.sourceforge.net>

¹¹<http://www.cs.waikato.ac.nz/ml/weka/>

εκπαίδευση των αλγορίθμων, χρησιμοποιήθηκε η τεχνική stratified cross fold validation με πλήθος αναδιπλώσεων 10. Τα αποτελέσματα αυτού του πειράματος παρουσιάζονται σε επόμενη ενότητα.

3.3.3 Διερεύνηση της επίρειας των ρυθμικών χαρακτηριστικών ηχητικών γεγονότων στην διέγερση

Σε αυτό το πείραμα επιχειρήθηκε η διερεύνηση της συσχέτισης μίας κοινής παρατήρησης, δηλαδή ότι ο ρυθμός του ήχου μπορεί να επηρεάσει την διέγερση του ακροατή. Αναφορικά με την εξόρυξη συναισθηματικής πληροφορίας από μουσική ή/και ομιλία, η παραπάνω εμπειρική γνώση έχει ήδη επαληθευθεί εκτενώς [38, 140]. Η απόδειξη της ισχύος αυτού του γεγονότος στους γενικούς ήχους μπορεί να δείξει, αφενός, πως ο ρυθμός είναι ένα καθολικό χαρακτηριστικό του ήχου που μπορεί να επηρεάσει την συναισθηματική κατάσταση του ακροατή. Αφετέρου, μπορεί επίσης να αναδείξει τα χαρακτηριστικά εκείνα τα οποία μπορούν να χρησιμοποιηθούν για καθολική αναγνώριση συναισθηματικής πληροφορίας από τα ρυθμικά χαρακτηριστικά του ηχητικού σήματος.

Προς αυτό το σκοπό διεξήχθη πείραμα εξόρυξης συναισθηματικής πληροφορίας από ηχητικά γεγονότα χρησιμοποιώντας μόνο ρυθμικά χαρακτηριστικά και ένα διαστατικό μοντέλο μίας διάστασης, της Διέγερσης. Ως συναισθηματικά επισημασμένα δεδομένα χρησιμοποιήθηκαν οι ήχοι IADS και από την συναισθηματική τους επισήμανση μόνο αυτή της διέγερσης. Τα τεχνικά χαρακτηριστικά που εξήχθησαν συνοψίζονται στον **Πίνακα 3.7**.

Τα χαρακτηριστικά αυτά εξήχθησαν από μικρά χρονικά πλαίσια του σήματος για έκαστο ηχητικό γεγονός. Λόγω της μη ύπαρξης πρότερης μελέτης σχετικά με την συσχέτιση της διέγερσης και των ρυθμικών χαρακτηριστικών από ηχητικά γεγονότα αλλά και της περιορισμένης βιβλιογραφίας σχετικά με την εν γένει εξόρυξη συναισθηματικής πληροφορίας από ηχητικά γεγονότα, δεν υπήρχε κάποιο προτεινόμενο μήκος για τα χρονικά πλαίσια του σήματος. Έτσι, χρησιμοποιήθηκε ένα σύνολο τέτοιων μηκών τα οποία δίνονται στον **Πίνακα 3.8** ενώ η επικάλυψη των πλαισίων είχε οριστεί σταθερή και ίση προς 20%. Αντί για τις τιμές που υπολογίστηκαν για κάθε τεχνικό χαρακτηριστικό χρησιμοποιήθηκαν στατιστικά μεγέθη αυτών που επίσης φαίνονται στον **Πίνακα 3.6**. Με αυτό το τρόπο διατηρήθηκε χαμηλός

Πίνακας 3.7: Τα τεχνικά χαρακτηριστικά και τα στατιστικά μεγέθη που χρησιμοποιήθηκαν για την συσχέτιση του ρυθμού και της διέγερσης

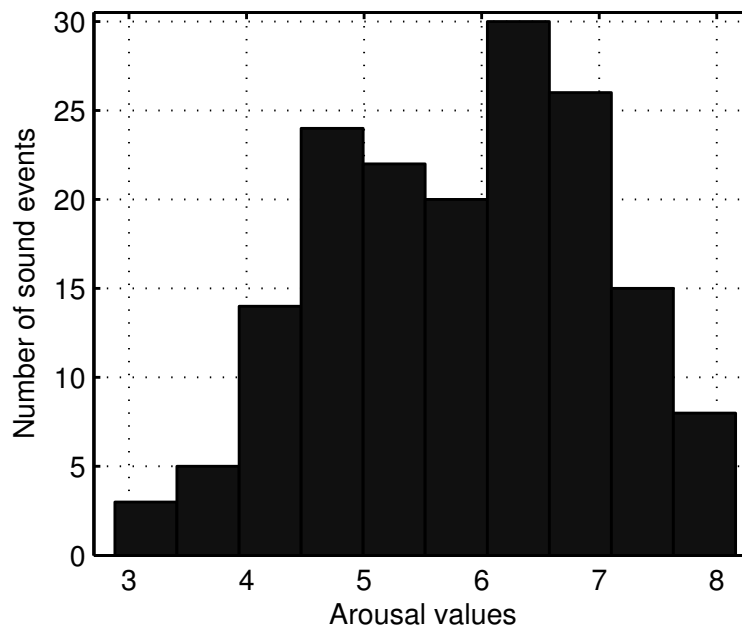
A/A	Τεχνικό χαρακτηριστικό	Στατιστικό μέγεθος
1	Beat spectrum	Μέση τιμή
2	Onsets	Τυπική απόκλιση
3	Tempo	Κλίση
4	Fluctuation	Κύρτωση
5	Event density	Ασσυμετρία
6	Pulse clarity	

Πίνακας 3.8: Η χρονική διάρκεια των πλαισίων που χρησιμοποιήθηκαν για την συσχέτιση των ρυθμικών χαρακτηριστικών με την διέγερση

A/A	Χρονική διάρκεια (sec)
1	0.8
2	1.0
3	1.2
4	1.4
5	1.6
6	1.8
7	2.0

ο αριθμός των χαρακτηριστικών για κάθε ηχητικό γεγονός ενώ παράλληλα έγινε εφικτή η αποτύπωση της διακύμανσης ενός μεγέθους κατά την εξέλιξη του ηχητικού γεγονότος. Με την διατήρηση λίγων χαρακτηριστικών για την διαδικασία εκπαίδευσης του συστήματος, μειώνεται ο ενδεχόμενος κίνδυνος υπερεξειδίκευσης του αλγόριθμου Μηχανικής Μάθησης στο πρόβλημα που καλείται να λύσει.

Ακόμα, στο σετ ήχων IADS υπήρξε μία κατηγοριοποίηση των ήχων αναφορικά με την τιμή διέγερσης. Συγκεκριμένα, σχηματίστηκαν δύο κλάσεις με την πρώτη να περιέχει τους ήχους που είχαν τιμή διέγερσης κάτω από από την μέση τιμή των συναισθηματικών επισημάνσεων, ήτοι 5, και με την δεύτερη να έχει τους υπόλοιπους. Στο **Σχήμα 3.6** απεικονίζεται το ιστόγραμμα των τιμών της διέγερσης για το σύνολο των ήχων του σετ IADS.

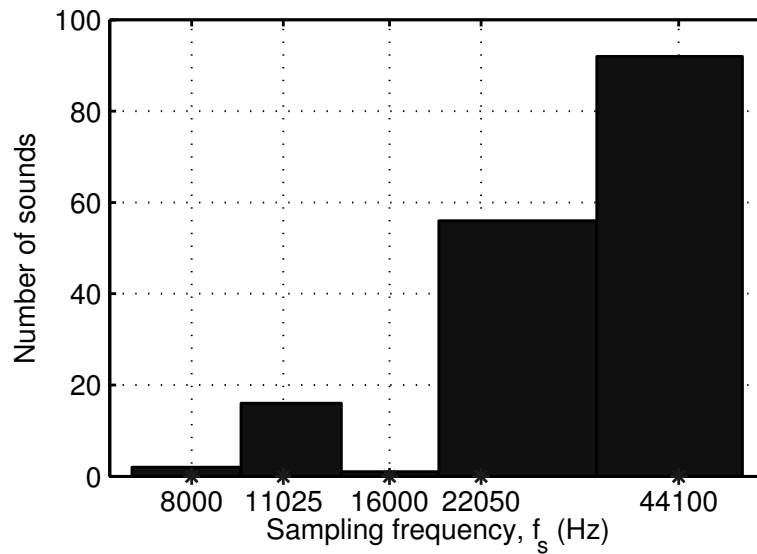


Σχήμα 3.6: Το ιστόγραμμα των τιμών της Διέγερσης (Arousal) για τους ήχους IADS.

Ως “Number of sound events” είναι το πλήθος των ηχητικών γεγονότων

Αναλυτικότερα, για κάθε ένα ήχο που περιέχεται στο IADS έγιναν 6 αντίγραφα και έτσι υπήρχαν σύνολο 7 διαφορετικά αρχεία για το ίδιο ηχητικό γεγονός με αποτέλεσμα ένα σύνολο ηχητικών γεγονότων ίσο με $167 \times 7 = 1169$. Αν και η συχνότητα δειγματοληψίας διαφέρει στους ήχους IADS, επιλέχτηκε να μην γίνει κάποια επεξεργασία (υπέρ-δειγματοληψία ή υπό-δειγματοληψία) έτσι ώστε να μην αλλοιωθούν, κατά οποιοδήποτε τρόπο, τα αρχικά δεδομένα. Στο **Σχήμα 3.7** είναι ένα ιστόγραμμα των τιμών της συχνότητας δειγματοληψίας των εν λόγω ήχων.

Από το σύνολο αυτό των ηχητικών γεγονότων δημιουργήθηκαν τελικά 7 διαφορετικές ομάδες όπου κάθε μία εμπεριείχε ένα μοναδικό αντίγραφο για κάθε ήχο. Για κάθε ομάδα χρησιμοποιήθηκε η αντίστοιχη χρονική διάρκεια των πλαισίων που φαίνεται στον **Πίνακα 3.8**. Από κάθε ομάδα και για κάθε χρονικό πλαίσιο έκαστου ηχητικού γεγονότος εξήχθησαν τα τεχνικά χαρακτηριστικά που φαίνονται στον **Πίνακα 3.7**. Έπειτα, τα στατιστικά μεγέθη που φαίνονται στον ίδιο Πίνακα υπολογίστηκαν από το σύνολο των τιμών κάθε χαρακτηριστικού (μία τιμή για κάθε χρονικό πλαίσιο ενός ηχητικού γεγονότος). Οι τιμές των στατιστικών μεγεθών για κάθε τεχνικό χαρακτηριστικό χρησιμοποιήθηκαν ως τα χαρακτηριστικά των ηχητικών γεγονότων βάσει των οποίων θα γινόταν η κατηγοριοποίηση. Επίσης, σε αυτά προστέθηκε η κλάση έκαστου ηχητικού γεγονότος, όπως προαναφέρθηκε.



Σχήμα 3.7: Το ιστόγραμμα των συχνοτήτων δειγματοληψίας (Sampling frequency) για τους ήχους IADS. Ως “Number of sounds” είναι το πλήθος των ηχητικών γεγονότων

Πίνακας 3.9: Τα τεχνικά χαρακτηριστικά που χρησιμοποιήθηκαν στο πείραμα συσχέτισης ρυθμικών χαρακτηριστικών με την διέγερση και τα αντίστοιχα γράμματα για την γρήγορη δεικτοδότησή τους

Τεχνικό χαρακτηριστικό	Γράμμα	Τεχνικό χαρακτηριστικό	Γράμμα
beatspectrumstd	A	onsetskurtosis	N
eventdensitystd	B	beatspectrumskewness	O
eventdensityskewness	C	pulseclaritygradient	P
onsetsgradient	D	beatspectrumkurtosis	Q
fluctuationkurtosis	E	pulseclaritykurtosis	R
beatspectrumgradient	F	eventdensitykurtosis	S
eventdensitymean	G	beatspectrummean	T
tempomean	H	eventdensitygradient	U
pulseclaritystd	I	pulseclaritymean	V
fluctuationmean	J	onsetsmean	W
fluctuationstd	K	pulseclarityskewness	X
fluctuationskewness	L	onsetsstd	Y
onsetsskewness	M	fluctuationgradient	Z

Αυτή η διαδικασία είχε σαν αποτέλεσμα ένα σύνολο χαρακτηριστικών του οποίου το πλήθος των στοιχείων ήταν ίσο με 27, όπου το ένα ήταν η κλάση της διέγερσης. Για την αξιοποίηση των χαρακτηριστικών αυτών, και προτού διεξαχθεί η διαδικασία δημιουργίας και επαλήθευσης του μοντέλου, έγινε μία αξιολόγηση της πληροφορίας που περιέχουν. Για την εύκολη αναφορά έκαστου χαρακτηριστικού, έγινε η δεικτοδότησή τους με γράμματα ακολουθώντας το σχήμα αντιστοίχισης που φαίνεται στον **Πίνακα 3.9**. Έτσι, εξετάστηκε για κάθε χαρακτηριστικό η συσχέτισή του με τα άλλα και έγινε επίσης και ιεράρχηση των χαρακτηριστικών αναφορικά με την ποσότητα πληροφορίας που εμπεριέχουν. Τα αποτελέσματα της εξέτασης της συσχέτισης των χαρακτηριστικών συνοψίζονται στον **Πίνακα 3.10**.

Όπως μπορεί να φανεί, τα τεχνικά χαρακτηριστικά που χρησιμοποιήθηκαν είναι σε μεγάλο βαθμό ασυσχέτιστα και έτσι δεν αποκλείστηκε κάποιο για την μετέπειτα διαδικασία. Για την ιεράρχηση χρησιμοποιήθηκαν οι αλγόριθμοι Information gain attribute evaluation και SVM attribute evaluation [45]. Ο πρώτος εξετάζει την σημαντικότητα κάθε χαρακτηριστικού υπολογίζοντας το κέρδος πληροφορίας αναφορικά με τις κλάσεις του προβλήματος με την χρήση μετρικών εντροπίας. Ο δεύτερος είναι μία τεχνική support vector machines (SVM) που εξετάζει την αποδοτικότητα κάθε χαρακτηριστικού για το πρόβλημα κατηγοριοποίησης στις δοθέντες κλάσεις. Η ιεράρχηση των χαρακτηριστικών σε σχέση με τον αλγόριθμο αλλά και την χρονική διάρκεια του παραθύρου που χρησιμοποιήθηκε δίνεται στην αντίστοιχη ενότητα των αποτελεσμάτων.

Τα χαρακτηριστικά αυτά προσφέρθηκαν ως είσοδο σε 3 διαφορετικούς αλγόριθμους κατηγοριοποίησης, ήτοι: α) ANN, που χρησιμοποιούνται σε εργασίες ανάλυσης σημασιολογικού περιεχομένου και παρέχουν αυξημένα αποτελέσματα κατηγοριοποίησης [163, 164], β) Logistic regression, που στηρίζονται σε μία στατιστική τεχνική που παράγει ένα μη-γραμμικό μοντέλο παλινδρόμησης το οποίο συσχετίζει το αποτέλεσμα της κατηγοριοποίησης στην πιθανότητα εξόδου του αποτελέσματος αυτού [163, 165], και γ) k-NN, ο οποίος είναι ένα ευρέως γνωστός ευριστικός αλγόριθμος για άμεση κατηγοριοποίηση, συγκρίνοντας τα χαρακτηριστικά ενός δεδομένου με αυτά από άλλα, ήδη κατηγοριοποιημένα, δεδομένα και προσδίδει στο προς κατάταξη δεδομένο την κλάση των k κατηγοριοποιημένων δεδομένων που απέχουν λιγότερο. Επιπροσθέτως, ως ANN χρησιμοποιήθηκε ένα νευρωνικό δίκτυο με δύο σιγμοειδή κρυφά επίπεδα και ένα γραμμικό επίπεδο εξόδου. Για την διαδικασία επαλήθευσης του μοντέλου που παρήγαγε έκαστος αλγόριθμος χρησιμοποιήθηκε η τεχνική stratified cross validation [45]

Πίνακας 3.10: Συσχέτιση τεχνικών χαρακτηριστικών κατά την αναγνώριση διέγερσης από ρυθμικά τεχνικά χαρακτηριστικά.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	1.0	0.3	-0.1	0.0	0.1	-0.7	0.0	0.1	0.3	0.2	0.1	0.2	0.0	-0.1	0.1	-0.1	-0.3	-0.1	-0.1	-0.6	-0.1	0.1	-0.1	0.0	0.2	0.2
B	0.3	1.0	-0.1	0.1	-0.1	-0.2	0.6	-0.1	0.2	0.0	0.0	-0.1	-0.3	-0.3	0.0	-0.1	-0.1	-0.2	-0.3	-0.1	0.0	0.3	0.2	0.1	0.3	0.0
C	-0.1	-0.1	1.0	-0.1	0.1	0.0	-0.4	-0.2	0.1	-0.2	-0.3	-0.1	0.3	0.2	0.2	0.1	-0.2	0.1	0.5	0.0	-0.1	-0.1	-0.3	0.1	0.1	-0.3
D	0.0	0.1	-0.1	1.0	-0.1	0.1	0.1	0.1	-0.2	0.0	0.1	0.0	-0.2	-0.1	-0.1	0.1	0.1	0.1	0.0	0.0	0.5	0.0	0.1	0.1	0.0	-0.1
E	0.1	-0.1	0.1	-0.1	1.0	0.0	-0.2	0.0	0.1	-0.1	0.0	0.8	0.2	0.2	0.0	-0.1	0.0	0.1	0.1	0.0	-0.1	-0.3	-0.2	0.0	0.1	-0.1
F	-0.7	-0.2	0.0	-0.1	0.0	1.0	0.0	-0.1	-0.2	0.0	0.1	0.0	0.0	0.1	0.1	0.0	0.2	0.1	0.1	0.5	0.0	-0.1	0.0	0.0	-0.1	0.0
G	0.0	0.6	-0.4	0.1	-0.2	0.0	1.0	0.1	-0.1	0.0	0.1	-0.3	-0.6	-0.4	-0.1	0.0	0.0	-0.2	-0.2	0.1	0.1	0.5	0.7	0.0	-0.3	0.1
H	0.1	-0.1	-0.2	0.1	0.0	-0.1	0.1	1.0	-0.1	0.0	0.0	-0.1	0.1	0.0	-0.1	0.1	0.0	0.1	-0.1	0.0	0.0	0.2	0.0	-0.2	-0.2	0.0
I	0.3	0.2	0.1	-0.2	0.1	-0.2	-0.1	1.0	0.0	-0.2	0.1	0.2	0.1	0.2	0.1	0.2	-0.1	-0.1	-0.3	0.1	-0.2	-0.1	0.0	-0.3	-0.1	0.2
J	0.2	0.0	-0.2	0.0	-0.1	0.0	0.0	0.0	0.0	1.0	0.9	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	-0.4	-0.1	0.0	0.0	-0.1	-0.1	0.7
K	0.1	0.0	-0.3	0.1	0.0	0.1	0.1	0.0	-0.2	0.9	1.0	0.2	-0.1	0.0	-0.1	0.0	0.1	0.1	0.0	-0.2	-0.1	0.1	0.1	0.1	-0.1	-0.1
L	0.2	-0.1	-0.1	0.0	0.8	0.0	-0.3	-0.1	0.1	0.1	0.2	1.0	0.3	0.2	0.0	-0.1	0.0	0.1	0.0	-0.1	-0.2	-0.3	-0.4	0.0	0.2	0.1
M	0.0	-0.3	0.3	-0.2	0.2	0.0	-0.6	0.1	0.2	0.1	-0.1	0.3	1.0	0.6	0.2	-0.1	-0.1	0.0	0.1	-0.2	-0.2	-0.3	-0.9	0.1	-0.1	0.0
N	-0.1	-0.3	0.2	-0.1	0.2	0.1	-0.4	0.0	0.1	0.1	0.0	0.2	0.6	1.0	0.0	0.0	0.1	0.1	0.1	0.0	-0.1	-0.2	-0.5	0.1	-0.4	0.0
O	0.1	0.0	0.2	-0.1	0.0	0.1	-0.1	-0.1	0.2	0.0	-0.1	0.0	0.2	0.0	1.0	-0.1	-0.5	-0.1	0.1	-0.4	-0.1	0.0	-0.2	0.1	0.1	-0.1
P	-0.1	-0.1	0.1	0.1	0.0	0.0	0.1	0.0	-0.1	0.0	0.0	-0.1	-0.1	0.0	-0.1	1.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1	-0.1	-0.1	0.0
Q	-0.3	-0.1	-0.2	0.1	0.0	0.2	0.0	0.0	-0.1	0.0	0.1	0.0	-0.1	0.1	-0.5	0.1	1.0	0.1	0.1	0.4	0.0	0.0	0.1	-0.2	-0.1	0.1
R	-0.1	-0.2	0.1	0.1	0.1	-0.2	0.1	-0.3	0.0	0.1	-0.3	0.1	0.1	0.0	-0.1	0.0	0.1	1.0	0.0	0.1	0.1	-0.2	0.0	0.1	0.0	0.0
S	-0.1	-0.3	0.5	0.0	0.1	0.1	-0.2	-0.1	0.1	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.0	1.0	0.1	0.0	-0.2	-0.1	-0.1	0.0	0.0
T	-0.6	-0.1	0.0	0.0	0.5	0.1	0.0	-0.2	-0.4	-0.2	-0.4	-0.2	-0.1	-0.2	0.0	-0.4	0.0	0.4	0.1	1.0	0.0	0.1	0.2	0.0	-0.1	-0.2
U	-0.1	0.0	-0.1	0.5	-0.1	0.0	0.1	0.0	-0.1	0.0	-0.1	-0.2	-0.2	-0.1	-0.1	0.1	0.0	0.1	0.0	0.0	1.0	-0.1	0.1	0.1	0.0	0.0
V	0.1	0.3	-0.1	0.0	-0.3	-0.1	0.5	0.2	0.0	0.0	0.1	-0.3	-0.3	-0.2	0.0	0.0	0.0	-0.2	-0.2	0.1	-0.1	1.0	0.4	-0.3	-0.2	0.1
W	-0.1	0.2	-0.3	0.1	-0.2	0.0	0.7	0.0	-0.3	0.0	0.1	-0.4	-0.9	-0.5	-0.2	0.1	0.1	0.0	-0.1	0.2	0.1	0.4	1.0	-0.1	-0.3	0.1
X	0.0	0.1	0.1	0.1	0.0	0.0	0.0	-0.2	-0.1	-0.1	-0.1	0.0	0.1	0.1	0.1	-0.1	-0.2	0.1	-0.1	0.0	0.1	-0.3	-0.1	1.0	0.1	-0.2
Y	0.2	0.3	0.1	0.0	0.1	-0.1	-0.3	-0.2	0.2	-0.1	-0.1	0.2	-0.1	-0.4	0.1	-0.1	-0.1	0.0	0.0	-0.1	0.0	-0.2	-0.3	0.1	1.0	0.0
Z	0.2	0.0	-0.3	-0.1	-0.1	0.0	0.1	0.0	0.0	0.7	0.7	0.1	0.0	0.0	-0.1	0.0	0.1	0.0	0.0	-0.2	0.0	0.1	-0.2	0.0	0.1	0.0

όπου το πλήθος των αναδιπλώσεων ήταν ίσο με 3, 8 και 24. Ο λόγος για την επιλογή αυτών των αριθμών ήταν για να διατηρηθεί ένας ισομερισμός του διαθέσιμου σετ ήχων, όπου το πλήθος τους, για κάθε σύνολο από τα 7 που δημιουργήθηκαν, ήταν ίσο με 167. Επίσης, χρησιμοποιήθηκε και η τεχνική leave-one-out [45], όπου στην παρούσα περίπτωση ισοδυναμεί με την stratified cross validation και πλήθος αναδιπλώσεων ίσο με 167. Τα αποτελέσματα αυτού του πειράματος παρουσιάζονται σε επόμενη ενότητα. Σημειώνεται επίσης ότι, τα αποτελέσματα αυτά αποτελούν την πρώτη δημοσίευση στην διεθνή βιβλιογραφία που μελετά την συσχέτιση των ρυθμικών χαρακτηριστικών των ηχητικών γεγονότων με την διέγερση του ακροατή.

3.3.4 Διερεύνηση της επίρειας της χωρικής τοποθέτησης της πηγής στο συναίσθημα

Ακολουθώντας το μοντέλο της Συναισθηματικής Ακουστικής Οικολογίας και τον ορισμό που δόθηκε για το ηχητικό γεγονός μέσα στα πλαίσια αυτού του μοντέλου, στην παρούσα ενότητα παρουσιάζεται η διερεύνηση της επίρειας της χωρικής τοποθέτησης της πηγής στο συναίσθημα. Η χωρική τοποθέτηση είναι ένα εγγενές χαρακτηριστικό κάθε πηγής, δεδομένου ότι κάθε πηγή έχει φυσική απόσταση και, ως εκ τούτου, μία χωρική τοποθέτηση σε σχέση με τον ακροατή. Άρα, αυτή η θέση της πηγής αποτελεί αναπόσπαστο μέρος των χαρακτηριστικών της το οποίο μπορεί να αντιληφθεί και ο ακροατής.

Η αξιολόγηση της επίρειας της θέσης της πηγής στην συναισθηματική κατάσταση του ακροατή έγινε με την διεξαγωγή πειραμάτων ακρόασης ηχητικών γεγονότων. Όπως αναφέρεται παραπάνω, στην διεθνή βιβλιογραφία υπάρχει παρόμοια διερεύνηση αλλά μόνο για δύο θέσεις, ήτοι: α) μπροστά, και β) πίσω από τον ακροατή, και ένα συναίσθημα, τον “Φόβο”. Μέχρι την δημοσίευση του πειράματος που περιγράφεται παρακάτω, καθώς και του σετ ηχητικών γεγονότων με συναισθηματική επισήμανση και χωρική τοποθέτηση που παράχθηκε, δεν είχε δημοσιευθεί άλλη έρευνα με ισοδύναμη αναλυτική προσέγγιση της επίρειας της χωρικής τοποθέτησης της πηγής στο συναίσθημα του ακροατή. Όπως αναφέρθηκε συνοπτικά και προηγουμένως, για την μελέτη αυτή χρησιμοποιήθηκε το σετ ήχων IADS και η αμφιωτική τεχνολογία. Με αυτά τα δύο εργαλεία δημιουργήθηκε ένα σετ ήχων με χωρική τοποθέτηση το οποίο χρησιμοποιήθηκε σε πειράματα αξιολόγησης συναισθηματικής κατάστασης με το μο-

ντέλο SAM. Τα πειράματα έγιναν με χρήση μίας διαδικτυακής πλατφόρμας που αναπτύχθηκε για τον συγκεκριμένο σκοπό [166] όπου η ακρόαση των ήχων γινόταν με ακουστικά (λόγω της χρήσης της αμφιωτικής τεχνολογίας). Επίσης, έγινε χρήση διασταστικού μοντέλου συναισθημάτων Διέγερσης – Ευχαρίστησης.

Αναλυτικότερα, το σετ ήχων IADS χρησίμευσε ως το αρχικό ηχητικό υλικό για την αμφιωτική επεξεργασία. Για την τελευταία επιλέχθηκαν 5 γωνίες $\theta(k) = \{0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ\}$, με $k \in [1, 5]$. Εδώ θα πρέπει να αναφερθεί πως αν και ο λογισμός της συμμετρίας των HRTF, που αναφέρθηκε στην Ενότητα 3.2.4, ο οποίος οδήγησε στο να μην χρησιμοποιηθούν γωνίες μεγαλύτερες από 180° είναι ορθός τεχνικά, στην βιβλιογραφία αναφέρονται πιθανές διαφοροποιήσεις στην αντίληψη του ανθρώπου όταν το ηχητικό ερέθισμα προέρχεται από δεξιά και από αριστερά [167]. Έτσι, για την πλήρη επαλήθευση των αποτελεσμάτων και για τις περιοχές που είναι στα αριστερά του ακροατή ενδέχεται να χρειάζεται μία διεξοδικότερη μελέτη. Συνεχίζοντας, για κάθε μία από τις γωνίες $\theta(k)$ δημιουργήθηκε ένα σετ αμφιωτικών ηχητικών γεγονότων τα οποία σταθμίστηκαν στο να έχουν όλα στάθμη αναπαραγωγής $-4.44dBFS$ για την αποφυγή ψαλιδισμού κατά την μετέπειτα διαδικασία επεξεργασίας. Έτσι δημιουργήθηκαν τα αμφιωτικά σετ ηχητικών γεγονότων $S_a(k) = \{s_b(1, k), s_b(2, k), \dots, s_b(167, k)\}$, όπου $s_b(i, k)$ είναι το αμφιωτικό ηχητικό γεγονός με αύξων αριθμό i , στο σετ IADS, και για την γωνία με αύξων αριθμό k . Τα τελευταία μαζί με έναν μετρητή, c και ο οποίος θα χρησιμοποιηθεί στην συνέχεια για την ορθή υλοποίηση της πειραματικής διαδικασίας, για την καταγραφή των φορών αναπαραγωγής έκαστου ηχητικού γεγονότος αποθηκεύτηκαν σε μία βάση δεδομένων και σε πλειάδα (tuple) της μορφής:

$$t_a(k) = (s_b(i, k), c(i, k)) \quad (3.4)$$

Με αυτά τα δεδομένα διεξήχθη ένα διαδικτυακό πείραμα συναισθηματικής επισήμανσης. Για να συμπεριληφθούν όσο το δυνατό περισσότεροι συμμετέχοντες στο πείραμα αλλά και να αποφευχθούν διάφοροι επηρεασμοί που μπορεί να ανακύψουν σε ένα περιβάλλον εργαστηρίου, π.χ. συναισθηματική αποστασιοποίηση λόγω του μη οικείου χώρου, απόσπαση προσοχής κτλ [168], χρησιμοποιήθηκε μία διαδικτυακή προσέγγιση στο πείραμα συναισθηματικής επισήμανσης. Επίσης, αυτή η προσέγγιση πρόσφερε πολλά οφέλη:

- Η παράλληλη πρόσβαση στο πείραμα από διαφορετικούς χρήστες δύναται να μειώσει

Πίνακας 3.11: Η χώρα καταγωγής των συμμετεχόντων ανά ποσοστό του συνολικού αριθμού συμμετεχόντων

Χώρα	(%)	Χώρα	(%)
Ελλάδα	18.92%	Γαλλία	5.85%
Ολλανδία	3.28%	Φινλανδία	5.85%
Αυστραλία	3.28%	Νέα Ζηλανδία	3.28%
Ηνωμένο Βασίλειο	12.27%	Δανία	4.56%
Καναδάς	4.56%	Γερμανία	3.28%
Άγνωστη (.com και .org καταλήξεις email)		34.87%	

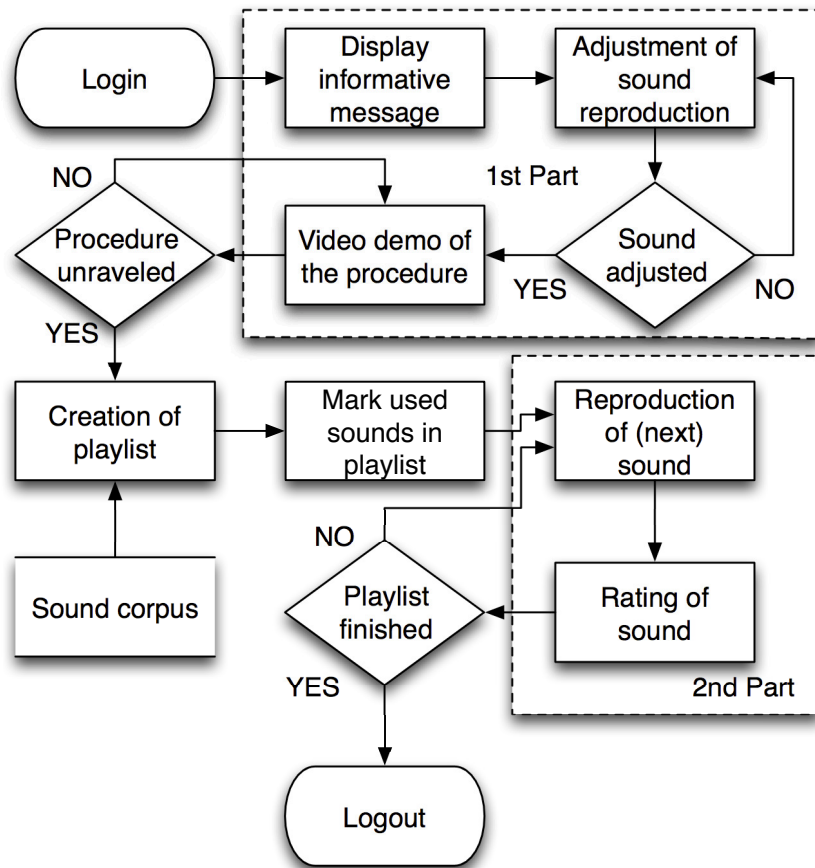
τον απαιτούμενο χρόνο για την διεξαγωγή του πειράματος

- Κάθε συμμετέχων μπορεί να πραγματοποιήσει το πείραμα κατά βούληση και σε χρόνο και τόπο που επιλέγει εκείνος. Έτσι είναι δυνατή η ελαχιστοποίηση της κούρασης του συμμετέχοντα και ο ψυχολογικός επηρεασμός από ένα μη γνώριμο περιβάλλον, κάτι που είναι γνωστό και με τον όρο *user fatigue* [168]
- Λόγω της διαδικτυακής πρόσβασης, οι συμμετέχοντες μπορούν να είναι από οποιοδήποτε σημείο του πλανήτη. Έτσι διασφαλίζεται η ευρύτητα, ως προς τις πολιτισμικές, πολιτιστικές και κοινωνικές διαφορές, των αποτελεσμάτων. Για την διασφάλιση αυτού του χαρακτηριστικού, οι οδηγίες του πειράματος που προσφέρονταν από την πλατφόρμα ήταν γραμμένες σε δύο γλώσσες: ελληνικά και αγγλικά, και υιοθετήθηκε ένας μηχανισμός ελέγχου εισόδου στην πλατφόρμα με e-mail. Στον **Πίνακα 3.11** είναι μία σύνοψη του ποσοστού ανά χώρα για τους συμμετέχοντες στο σύνολο και για το εν λόγω πείραμα σύμφωνα με τις καταλήξεις από τις διευθύνσεις e-mail τις οποίες δήλωναν οι χρήστες κατά την εγγραφή τους στην πλατφόρμα

Από την άλλη, ο απομακρυσμένος χαρακτήρας του πειράματος εμπεριέχει ορισμένους κινδύνους για την ακρίβεια των παραγόμενων αποτελεσμάτων. Ο πρώτος εξ αυτών ήταν η εν γένει χρήση ακουστικών. Τόσο η καθαυτή χρήση τους, όσο και η ρύθμιση της στάθμης αναπαραγωγής τους. Για παράδειγμα, θα μπορούσε κάποιος συμμετέχων στο πείραμα να μην φορέσει ακουστικά. Αυτό το γεγονός θα ακύρωνε την χωρική τοποθέτηση της πηγής, σύμφωνα με τα δεδομένα του πειράματος. Για αυτό το λόγο υπήρχε ενημερωτικό μήνυμα με

υποχρεωτική κατάδειξη της ανάγνωσής του, το οποίο αφενός ενημέρωνε για την αναγκαία χρήση ακουστικών όσο για μία αναγκαστική διαδικασία ρύθμισης της στάθμης αναπαραγωγής. Η τελευταία γινόταν για κάθε συμμετέχοντα και περιλάμβανε την αναπαραγωγή ενός τόνου $1kHz$ και την ρύθμιση της στάθμης αναπαραγωγής των ακουστικών σε επίπεδο που ο χρήστης ένιωθε άνετα. Βέβαια, πάντα παραμένει ο κίνδυνος ο χρήστης να μην συμμορφωθεί με τις οδηγίες και, για παράδειγμα να μην φορέσει ακουστικά, σημαίνοντας έτσι την ακύρωση του βασικού παράγοντα του πειράματος· δηλαδή της χωρικής τοποθέτησης. Επίσης, λόγω της απουσίας εξειδικευμένου προσωπικού, σε αντίθεση με το περιβάλλον ενός εργαστηρίου, αλλά και της πιθανής ύπαρξης εξωτερικών ηχητικών και μη ενοχλήσεων, ενδέχεται η συναισθηματική επισήμανση από τον συμμετέχοντα να μην είναι ακριβής. Αυτοί όμως οι κίνδυνοι εντάσσονται στην αξιοπιστία, ακεραιότητα και εμπιστοσύνη στον συμμετέχοντα και υπάρχουν, είτε αυτούσιοι είτε με ισοδύναμη μορφή, σε κάθε πείραμα υποκειμενικής αξιολόγησης. Δηλαδή, ακόμα και σε ένα καλά ελεγχόμενο εργαστηριακό περιβάλλον, ένας κακόβουλος και μη υπεύθυνος συμμετέχων σε ένα πείραμα μπορεί να απαντήσει με ψεύδη σε μία διαδικασία υποκειμενικής αξιολόγησης. Ακόμα, ο μηχανισμός καταγραφής των e-mail χρησίμευσε και ως ελεγκτικός μηχανισμός για μην τέλεση του πειράματος από τον ίδιο συμμετέχοντα δύο φορές. Επειδή όμως ελλόχευε ο κίνδυνος για την δημιουργία διαφορετικού λογαριασμού από τον ίδιο χρήστη, στην προ-επεξεργασία των αποτελεσμάτων διαγράφηκαν όσες καταχωρήσεις είχαν πολύ όμοιο προφίλ χρήστη. Έτσι, ο τελικός αριθμός χρηστών με έγκυρες καταχωρήσεις ανήλθε στους 215 διαφορετικούς χρήστες/διευθύνσεις e-mail που αντιστοιχεί σε ένα σύνολο 3,905 συναισθηματικών επισημάνσεων.

Το καθαυτό τεστ ακρόασης και συναισθηματικής επισήμανσης ήταν οργανωμένο σε δύο στάδια. Τόσο αυτά όσο και η εν γένει διαδικασία του τεστ ακρόασης περιγράφονται γραφικά στο **Σχήμα 3.8**. Στην πρώτη φάση, παρουσιαζόταν στον χρήστη ένα ενημερωτικό μήνυμα σχετικά με την επιβεβλημένη χρήση ακουστικών και την διαδικασία ρύθμισης της στάθμης αναπαραγωγής τους. Έπειτα, και κατόπιν της κατάδειξης από τον χρήστη ότι έλαβε γνώση του μηνύματος, γινόταν η αναπαραγωγή ενός τόνου $1kHz$ και στάθμης $0dBFS$, όπου κατά την διάρκεια αυτής ο χρήστης έπρεπε να προσαρμόσει την στάθμη των ακουστικών του σε ένα επίπεδο που ο ίδιος ένιωθε άνετα. Τέλος, γινόταν αναπαραγωγή ενός ενημερωτικού video σχετικά με την χρήση της πλατφόρμας και την ακόλουθη διαδικασία συναισθηματικής επισήμανσης. Σε αυτό το video αναδεικνυόταν στους χρήστες ότι δεν υπάρχει κάποια σωστή ή λάθος απάντηση, καθώς και τους γινόταν μία επίδειξη της χρήσης του μοντέλου SAM.

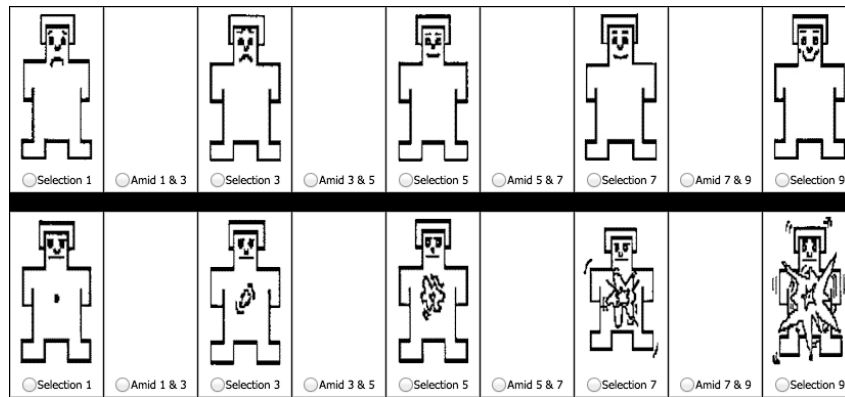


Σχήμα 3.8: Η διαδικασία των τεστ ακρόασης κατά την μελέτη της χωρικής τοποθέτησης της πηγής στο συναίσθημα του ακροατή

Στο **Σχήμα 3.9** είναι μία απεικόνιση των επιλογών συναισθηματικής επισήμανσης που ήταν διαθέσιμες στους συμμετέχοντες. Με το πέρας αυτού του σταδίου καταρτιζόταν αυτόματα μία ακολουθία ήχων που θα αναπαραγότουσαν για τον εκάστοτε συμμετέχοντα βάσει της σχέσης:

$$I(k) = \min_{c(i,k)}(t_a(k)) \quad (3.5)$$

όπου $I(k)$ είναι μία πλειάδα της μορφής $t_a(k)$. Με την κατάρτιση της $I(k)$, και για όλα τα ηχητικά γεγονότα $s_b(i, k)$ που ήταν σε αυτή, ο μετρητής $c(i, k)$ αυξανόταν κατά ένα έτσι ώστε να υπάρχει η ένδειξη του πλήθους χρήσεων των ηχητικών γεγονότων αυτών σε μία λίστα αναπαραγωγής. Έπειτα, γινόταν η εκκίνηση του δεύτερου σταδίου που ήταν η καθεαυτού τέλεση των συναισθηματικών επισημάνσεων. Σε αυτές παρατηρήθηκε το γεγονός ότι, πιθανώς λόγω του απομακρυσμένου χαρακτήρα, ορισμένοι χρήστες δεν ολοκλήρωσαν την



Σχήμα 3.9: Οι διαθέσιμες επιλογές για την συναισθηματική επισήμανση των ηχητικών γεγονότων κατά την μελέτη της χωρικής τοποθέτησης της πηγής στο συναίσθημα του ακροατή

διαδικασία επισήμανσης. Έτσι, οι επισημάνσεις που δημιουργήθηκαν από τους χρήστες αυτούς αφαιρέθηκαν συνολικά. Το γεγονός αυτό οδήγησε αρχικά σε ένα σύνολο 32 ηχητικών γεγονότων s_{init} (i'_{init}) με χωρική τοποθέτηση και συναισθηματική επισήμανση, όπως ορίζεται παρακάτω. Το σύνολο αυτό ονομάστηκε Binaural Emotionally Annotated Digital Sounds (BEADS) το οποίο είναι το πρώτο παγκοσμίως σετ συναισθηματικά επισημασμένων ηχητικών γεγονότων και ένα από τα τρία διαθέσιμα σετ ηχητικών γεγονότων με συναισθηματική επισήμανση [86]¹². Τα ηχητικά γεγονότα, το σημασιολογικό τους περιεχόμενο και η σχέση του i'_{init} με το i είναι στον **Πίνακα 3.12** και περιγράφεται στην Εξίσωση:

$$i = i'_{init} + f(i) \quad (3.6)$$

Πίνακας 3.12: Η αντιστοιχία του δείκτη i' με το i καθώς και το σημασιολογικό περιεχόμενο των αντίστοιχων ηχητικών γεγονότων για το σετ ηχητικών γεγονότων BEADS

Δείκτης i'_{init}/i	Σημασιολογικό περιεχόμενο	Δείκτης i'_{init}/i	Σημασιολογικό περιεχόμενο
1/5	Σκύλος	17/99	Βροχή
2/16	Κότες	18/106	Αντίστροφη μέτρηση

Πίνακας 3.12 – Συνέχεια σε επόμενη σελίδα

¹²Το σύνολο BEADS δημοσιεύθηκε και διατίθεται ελεύθερα από τον ιστότοπο του Εργαστηρίου Επεξεργασίας Οπτικοακουστικών Σημάτων του Τμήματος Τεχνών Ήχου & Εικόνας, Ιόνιο Πανεπιστήμιο

Πίνακας 3.12 – Συνέχεια από προηγούμενη σελίδα

Δείκτης i'_{init}/i	Σημασιολογικό περιεχόμενο	Δείκτης i'_{init}/i	Σημασιολογικό περιεχόμενο
3/18	Κροταλίας	19/107	Κόρνες αμαξιών
4/20	Κοκκινολαίμης	20/112	Αέρας
5/21	Τροπικό δάσος	21/113	Συντριβή αεροπλάνου
6/26	Γυναικείοι ήχοι συνουσίας	22/114	Βλάβη μηχανής
7/35	Άντρας που γελάει	23/115	Ατύχημα δικύκλου
8/38	Γέλια	24/122	Έκρηξη
9/42	Ζευγάρι που φταρνίζεται	25/130	Τηλέφωνο
10/50	Έμετος	26/133	Ρολόι
11/54	Σφύριγμα	27/135	Κούκος
12/62	Γυναίκα που κλαίει	28/142	Κουλοχέρης (μηχανή)
13/66	Θύμα	29/146	Περπάτημα
14/80	Γραφομηχανή	30/157	Άρπα
15/85	Γράψιμο	31/159	Απόσπασμα έργου του Bach
16/96	Νεροχύτης	32/160	Χορωδία

όπου, $f(i)$ είναι ένας ακέραιος αριθμός που αναπαριστά το σύνολο των ήχων IADS που δεν περιλήφθηκαν στο τελικό σύνολο μέχρι τον ήχο i . Στην συνέχεια, το σύνολο των ηχητικών γεγονότων που χρησιμοποιήθηκε επαυξήθηκε με την επανάληψη του πειράματος ακρόασης. Το τελικό σύνολο αποτελείται 50 πλήρως επισημασμένων ηχητικών γεγονότων $s(i')$ ή, ισοδύναμα, σε $5 \times 50 = 250$ συναισθηματικά επισημασμένους ήχους $s_b(i', k)$. Κάτι που οδήγησε σε 2,946 συναισθηματικές επισημάνσεις στο τελικό σύνολο. Ο δείκτης i' είναι ο αύξων αριθμός των ηχητικών γεγονότων που είχαν πλήρη επισημάνση και σχετίζεται με τον αύξων αριθμό τους στο σύνολο IADS μέσω της σχέσης:

$$i = i' + f(i) \quad (3.7)$$

όπου όλα τα σύμβολα είναι όπως ορίστηκαν προηγουμένως. Η σχέση του i' με το i για το τελικό σετ δεδομένων καθώς και το σημασιολογικό περιεχόμενο των ηχητικών γεγονότων φαίνονται στον Πίνακα 3.13. Τα νέα σύνολα ηχητικών γεγονότων ονομάζονται $S'_a(k)$ και περιέχουν τους ήχους $s_b(i', k)$ με $i' \in [1, 150]$ και ανταποκρίνονται στην ίδια γωνία $\theta(k)$ με τους $s_b(i, k)$.

Πίνακας 3.13: Η αντιστοιχία του i' με το i καθώς και το σημασιολογικό περιεχόμενο των τελικών ηχητικών γεγονότων

Δείκτης i'_{init}/i	Σημασιολογικό περιεχόμενο	Δείκτης i'_{init}/i	Σημασιολογικό περιεχόμενο
1/1	Γάτα	26/85	Γράψιμο
2/5	Σκύλος	27/88	Ιπποδρομία
3/7	Μωρό	28/95	Βάψιμο
4/9	Παιδιά 1	29/96	Νεροχύτης
5/12	Μέλισσες	30/99	Βροχή 1
6/16	Κότες	31/104	Ελικόπτερο 1
7/18	Κροταλίας	32/106	Αντίστροφη μέτρηση
8/20	Κοκκινολαίμης	33/107	Κόρνες αμαξιών
9/21	Τροπικό δάσος	34/112	Άνεμος
10/26	Γυναικείοι ερωτικοί ήχοι 1	35/113	Συντριβή αεροπλάνου
11/27	Γυναικείοι ερωτικοί ήχοι 2	36/114	Βλάβη μηχανής
12/34	Αγόρι που γελάει	37/115	Ατύχημα δικύκλου
13/35	Άντρας που γελάει	38/117	Καταιγίδα
14/38	Γέλια	39/122	Έκρηξη
15/40	Άντρας που βήχει	40/130	Τηλέφωνο 1

Πίνακας 3.13 – Συνέχεια σε επόμενη σελίδα

Πίνακας 3.13 – Συνέχεια από προηγούμενη σελίδα

Δείκτης i'_{init}/i	Σημασιολογικό περιεχόμενο	Δείκτης i'_{init}/i	Σημασιολογικό περιεχόμενο
16/42	Ζευγάρι που φταρνίζεται	41/133	Ρολόι
17/43	Άντρας που ασθμαίνει	42/135	Κούκος
18/46	Άντρας που φταρνίζεται	43/140	Ξυπνητήρι
19/50	Έμετος	44/142	Κουλοχέρης (παιχνίδι) 2
20/54	Σφύριγμα	45/146	Περπάτημα
21/62	Μάχη 2	46/150	Άνοιγμα και σερβίρισμα ποτού
22/63	Μάχη 3	47/157	Άρπα
23/66	Θύμα	48/159	Απόσπασμα έργου του Bach
24/67	Ερπυσμός	49/160	Χορωδία
25/80	Γραφομηχανή	50/167	Ηλεκτρισμός

Τα αποτελέσματα του πειράματος [99] της επίρειας της χωρικής θέσης της πηγής παρατίθενται σε επόμενη ενότητα.

3.3.5 Διερεύνηση της επίρειας των ρυθμικών χαρακτηριστικών ηχητικών γεγονότων στην ευχαρίστηση

Με την ολοκλήρωση του πειράματος για την συσχέτιση των ρυθμικών χαρακτηριστικών ενός ηχητικού γεγονότος και της διέγερσης, πραγματοποιήθηκε πείραμα για την συσχέτιση των ίδιων χαρακτηριστικών και της ευχαρίστησης. Με την διεξαγωγή του πειράματος αυτού και την επεξεργασία των αποτελεσμάτων έγινε δυνατή η πλήρης θεώρηση της επίρειας των ρυθμικών χαρακτηριστικών ενός ηχητικού γεγονότος στην συναισθηματική κατάσταση

του ακροατή επειδή αυτή, στο μοντέλο αναπαράστασης συναισθήματος που έχει επιλεγεί, ολοκληρώνει το σχετικό χώρος αναπαράστασης.

Η διαδικασία δεν διέφερε από αυτή που ακολουθήθηκε στην διερεύνηση της σχέσης των ρυθμικών χαρακτηριστικών με την διέγερση. Έτσι, από το σύνολο συναισθηματικά επισημασμένων ήχων που επιλέχθηκε, έγινε εξαγωγή τεχνικών χαρακτηριστικών και διεξάχθηκαν πειράματα κατηγοριοποίησης. Αναλυτικότερα, ως αρχικό σετ δεδομένων επιλέχθηκαν οι ήχοι IADS και ως μοντέλο συναισθημάτων ένα διαστατικό με μία διάσταση, αυτή της Ευχαρίστησης. Για να διατηρηθεί μία συνέχεια σε σχέση με την έρευνα πάνω στα ρυθμικά χαρακτηριστικά και την διέγερση επιλέχθηκε η ίδια προσέγγιση στην εξαγωγή χαρακτηριστικών [30]. Έτσι, δημιουργήθηκαν 6 επιπλέον αντίγραφα του σετ IADS. Σε κάθε ένα από τα συνολικά 7 σετ εφαρμόστηκε αρχικά μία τμηματοποίηση των ηχητικών γεγονότων σε χρονικά πλαίσια και έπειτα, για κάθε πλαίσιο κάθε ηχητικού γεγονότος, έγινε η εξαγωγή των τεχνικών χαρακτηριστικών. Τα χρονικά πλαίσια και η διάρκειά τους φαίνεται στον Πίνακα 3.8 ενώ η χρονική επικάλυψη ήταν 20%. Έτσι, στο πρώτο σετ οι ήχοι τμηματοποιήθηκαν σε χρονικά πλαίσια διάρκειας 0.8 δευτερολέπτων, στο δεύτερο σε πλαίσια διάρκειας 1.0 δευτερολέπτου κ.ο.κ. Για κάθε ηχητικό γεγονός, και από τα επιμέρους τμήματά του, έγινε εξαγωγή των τεχνικών χαρακτηριστικών που φαίνονται στον **Πίνακα 3.14** καθώς και κανονικοποίηση στην στάθμη $0dBFS$. Αυτή η διαδικασία οδήγησε στην λήψη αρκετών τιμών, δεδομένου ότι κάθε ήχος IADS είχε διάρκεια 6.0 δευτερολέπτων. Για γίνει η εισαγωγή των δεδομένων στον αλγόριθμο μηχανικής μάθησης αλλά και να μην χρησιμοποιηθούν πολλές διαστάσεις του χώρου των χαρακτηριστικών κατά την κατηγοριοποίηση, επιλέχθηκε να χρησιμοποιηθούν στατιστικά μεγέθη των τιμών των τεχνικών χαρακτηριστικών. Επίσης, αυτή η επιλογή επιτρέπει και την αποτύπωση της διακύμανσης ενός χαρακτηριστικού κατά την εξέλιξη του ηχητικού γεγονότος. Τα στατιστικά μεγέθη φαίνονται επίσης στον **Πίνακα 3.14**.

Έτσι, εάν $t_{wl}[i]$ είναι η χρονική διάρκεια έκαστου διαφορετικού χρονικού πλαισίου i με $i \in [1, 7]$ τότε τα διαφορετικά σετ ήχων ορίζονται ως:

$$S_{i,x'} = f(S_x[n]), x \in [1, 167], i \in [1, 7] \quad (3.8)$$

όπου $S_x[n]$ είναι ήχος στο IADS σετ και f είναι η συνάρτηση κανονικοποίησης σε $0dBFS$. Για κάθε τμήμα κάθε ήχου στο $S_{i,x'}$ εφαρμόστηκε το παράθυρο hamming με πλήθος

Πίνακας 3.14: Τεχνικά χαρακτηριστικά και στατιστικά μεγέθη που χρησιμοποιήθηκαν κατά την συσχέτιση ρυθμικών χαρακτηριστικών και ευχαρίστησης

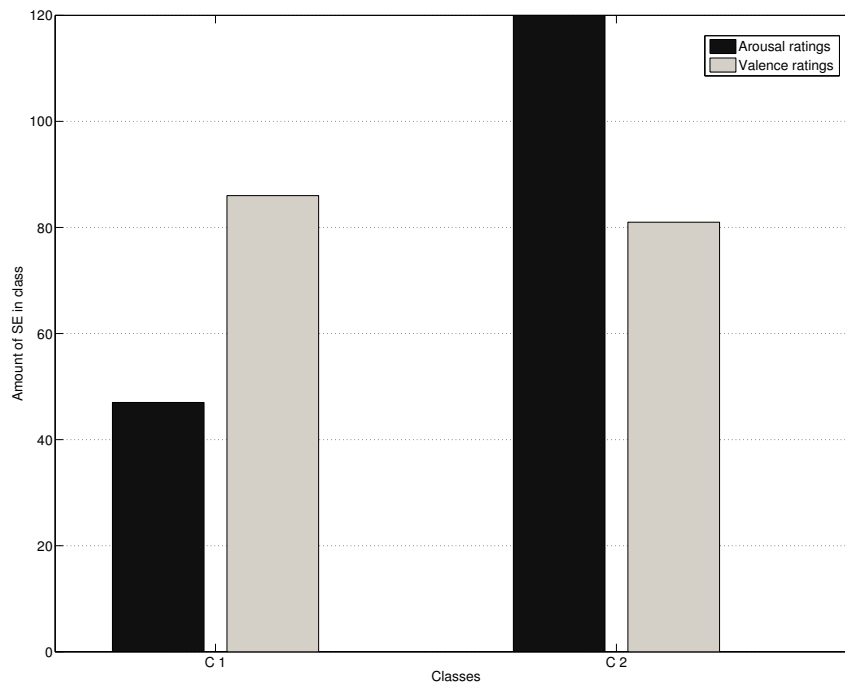
A/A	Τεχνικό χαρακτηριστικό	Στατιστικό μέγεθος
1	Beat spectrum	Μέση τιμή (M)
2	Onsets	Τυπική απόκλιση (STD)
3	Τέμπο	Μέση τιμή κλίσης (gradient mean)
4	Fluctuation	Τυπική απόκλιση κλίσης (gradient STD)
5	Event density	Ασσυμετρία
6	Pulse clarity	Κύρτωση

δειγμάτων:

$$n_{wh} = t_{wt}[i] \times Fs_{x'}, x' \in [1, 167], i \in [1, 7] \quad (3.9)$$

όπου $Fs_{x'}$ είναι η συχνότητα δειγματοληψίας έκαστου ηχητικού γεγονότος, και διαφορετική σε κάθε ένα από αυτά, x' . Ακόμα, χρησιμοποιήθηκε και η συναισθηματική επισήμανση της Διέγερσης και της Ευχαρίστησης ως επιπλέον χαρακτηριστικά. Το δεύτερο χρησιμοποιήθηκε για την επαλήθευση του μοντέλου και κβαντίστηκε σε δύο τιμές που αντιπροσωπεύουν δύο κλάσεις ευχαρίστησης, ήτοι: α) C_1 με ευχαρίστηση $V[x'] \in [1, 5)$, και β) C_2 με $V[x'] \in [5, 9]$, όπου $V[x']$ είναι η επισήμανση για την ευχαρίστηση του ηχητικού γεγονότος x' . Η ίδια κβάντιση τηρήθηκε και στην διέγερση ενώ η κατανομή των ηχητικών γεγονότων στις δύο κλάσεις για την Ευχαρίστηση και την Διέγερση είναι στο **Σχήμα 3.10**.

Η διέγερση χρησιμοποιήθηκε λόγω της σχέσης που αναφέρεται στο [103], δηλαδή είναι μη πιθανό ένας ακροατής να νιώθει δυσαρεστημένος από ένα ηχητικό γεγονός και ταυτόχρονα μη διεγερμένος. Όμως, για την πλήρη διερεύνηση της σχέσης των ρυθμικών χαρακτηριστικών σχηματιστήκαν δύο τελικά σετ ήχων: ένα με την διέγερση ως επιπρόσθετο χαρακτηριστικό και ένα χωρίς και είναι τα:



Σχήμα 3.10: Η κατανομή σε κλάσεις για την ευχαρίστηση και την διέγερση

$$R_{i,x'} = \{S_{i,x'}, V[x']\}, i \in [1, 7], x' \in [1, 167] \quad (3.10)$$

$$R_{i,x'} = \{S_{i,x'}, V[x'], A[x']\}, i \in [1, 7], x' \in [1, 167]$$

Για την διαδικασία κατηγοριοποίησης χρησιμοποιήθηκαν διάφοροι αλγόριθμοι, τόσο αυτόνομοι όσο και μετά-μάθησης. Συγκεκριμένα χρησιμοποιήθηκαν οι:

- Decision stump [169], ο οποίος χρησιμοποιεί μόνο ένα χαρακτηριστικό για να κατηγοριοποιήσει το σετ δεδομένων
- RepTree [170], που είναι ένας γρήγορος αλγόριθμος κατηγοριοποίησης χρησιμοποιώντας δένδρα απόφασης βάσει του κέρδους πληροφορίας
- C4.5 [171], που είναι ένας τυπικός αλγόριθμος απόφασης με χρήση δένδρων
- Support Vector Machines [68], ένας μη-πιθανολογικός δυαδικός και γραμμικός ταξινομητής
- AdaBoost [74], ένας αλγόριθμος μετά-μάθησης στον οποίο χρησιμοποιήθηκε ο Se-

quential Minimal Optimization ως κεντρικός

- Bagging [75], ένας αλγόριθμος μετά-μάθησης στο οποίο χρησιμοποιήθηκε ο Sequential Minimal Optimization ως κεντρικός

Η επαλήθευση των μοντέλων έγινε με την τεχνική stratified cross validation χρησιμοποιώντας το τυπικό πλήθος αναδιπλώσεων 10. Τα αποτελέσματα αυτού του πειράματος παρουσιάζονται στο επόμενο Κεφάλαιο.

ΚΕΦΑΛΑΙΟ 4

ΑΠΟΤΕΛΕΣΜΑΤΑ

Σε αυτό το κεφάλαιο παρατίθενται τα αποτελέσματα που ελήφθησαν κατά την πειραματική διαδικασία που περιγράφηκε αναλυτικά στο προηγούμενο Κεφάλαιο. Η σειρά παράθεσής τους θα ακολουθήσει αυτή της εμφάνισης στο προαναφερθέν Κεφάλαιο. Έτσι, πρώτα θα παρατεθούν τα αποτελέσματα που λήφθηκαν κατά το πείραμα της αναγνώρισης stress από φωνητικές εντολές. Θα ακολουθήσουν τα αποτελέσματα από την διαδικασία εξόρυξης συναισθηματικής πληροφορίας στα πλαίσια της πρότασης του πλαισίου της Συναισθηματικής Ακουστικής Οικολογίας. Μετά είναι τα αποτελέσματα από την συσχέτιση του ρυθμού και της Διέγερσης και από την επήρεια της θέσης της πηγής στην συναισθηματική κατάσταση του ακροατή. Τέλος, παρουσιάζεται το σετ ηχητικών δεδομένων BEADS και ακολουθούν τα αποτελέσματα από την συσχέτιση του ρυθμού και της Ευχαρίστησης, τα οποία οδηγούν στην ολοκλήρωση της παρουσίασης των αποτελεσμάτων. Η συζήτηση και ανάλυση των αποτελεσμάτων βρίσκεται στο επόμενο Κεφάλαιο.

4.1 Αναγνώριση stress από φωνητικές εντολές

Τα αποτελέσματα αναγνώρισης stress από φωνητικές εντολές αφορούν το πείραμα που εκτελέστηκε, και περιγράφεται στο προηγούμενο Κεφάλαιο, για την διερεύνηση της ανίχνευσης ακραίων συναισθηματικών καταστάσεων από τεχνικά χαρακτηριστικά. Διενεργήθηκε στα πλαίσια υλοποίησης ενός συστήματος έξυπνου μικροφώνου το οποίο μπορεί να αντιλαμβάνεται ένα σετ φωνητικών εντολών και την συναισθηματική κατάσταση του χρήστη και να εκτελεί, αναλόγως, διάφορες προγραμματισμένες διαδικασίες.

Για την εξαγωγή των αποτελεσμάτων από το σύστημα αναγνώρισης stress από φωνητικές εντολές έγινε μία παράθεση της ευστοχίας του εν λόγω συστήματος, με είσοδο λέξεις με stress και χωρίς από 16 συμμετέχοντες, διαφορετικούς από αυτούς του σταδίου της εκπαίδευσης. Αναλυτικότερα, το σύστημα αναγνώρισης stress χρησιμοποιήθηκε ως υποσύστημα ενός μεγαλύτερου το οποίο οδηγούσε ένα έξυπνο μικρόφωνο τεχνολογίας MEMS. Οι λειτουργίες του μικροφώνου αυτού περιελάμβαναν και την αναγνώριση φωνητικών εντολών, όπου με ένα σετ τέτοιων ο χρήστης μπορούσε να χρησιμοποιήσει το έξυπνο μικρόφωνο. Για παράδειγμα, μπορούσε να ανάψει φωτιά στο τζάκι ή να καλέσει, άμεσα και αυτόματα, σε βοήθεια. Η κλήση σε βοήθεια πραγματοποιούνταν από το μικρόφωνο αυτόματα, με την αναγνώριση της φωνητικής εντολής και του stress του ομιλητή. Στο σύστημα αναγνώρισης stress δόθηκαν ως είσοδοι φωνητικές εντολές που μπορεί να αναγνωρίσει το γενικότερο σύστημα του έξυπνου μικροφώνου και οι οποίες είχαν εκφερθεί με stress και χωρίς. Από την έξοδο του συστήματος, ήτοι αναγνώριση stress ή όχι, παράχθηκαν τα αποτελέσματα ανά λέξη και παρουσιάζονται στον **Πίνακα 4.1**. Επίσης, στον ίδιο Πίνακα φαίνονται και οι μέσες τιμές ευστοχίας ανά είσοδο. Η στήλη A/A είναι ο αύξων αριθμός λέξης όπως σημειώνεται στον **Πίνακα 3.4**. Στον **Πίνακα 4.1** ως ευστοχία νοείται ο λόγος των ορθών αναγνωρίσεων συναισθηματικής κατάστασης προς το συνολικό πλήθος αυτών, για κάθε λέξη.

Όπως μπορεί να φανεί από τον **Πίνακα 4.1**, η ευστοχία του συστήματος αναγνώρισης stress από φωνητικές εντολές είναι της τάξης του 75% ενώ η μέγιστη τιμή απαντάται στην έξη 3 για τις λέξεις με stress και στην λέξη 4 για τις λέξεις χωρίς stress. Από τα ποσοστά ευστοχίας μπορεί να φανεί ότι το σημασιολογικό περιεχόμενο δεν έχει μεγάλη βαρύτητα στην αναγνώριση stress αλλά και ότι μπορεί να γίνει αναγνώριση ακραίων συναισθηματικών καταστάσεων από ηχητικά δεδομένα με τεχνικά χαρακτηριστικά. Η πλήρης ανάλυση και

Πίνακας 4.1: Η ευστοχία αναγνώρισης stress, δεδομένης μίας λέξης, με είσοδο λέξεις με και χωρίς stress

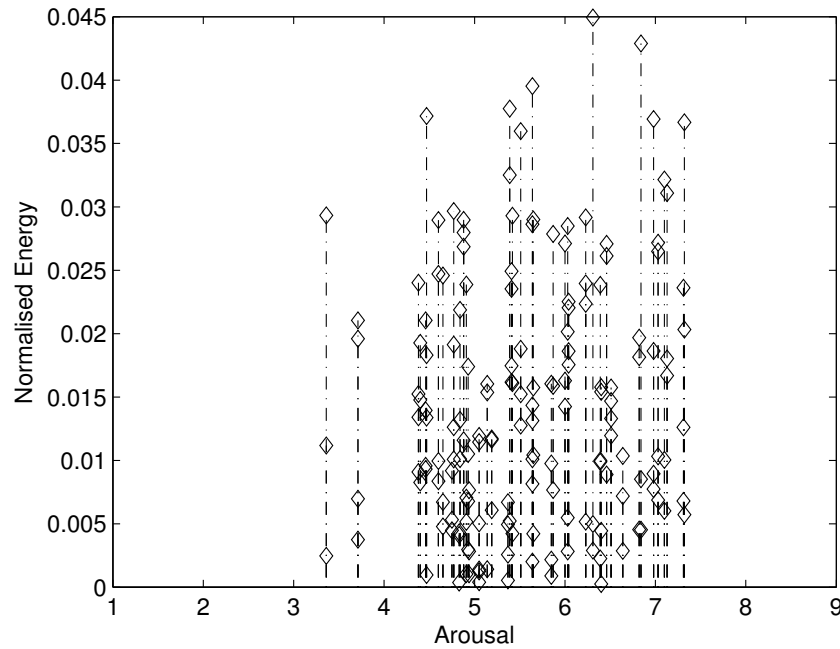
A/A λέξης	Ευστοχία σε είσοδο με stress	Ευστοχία σε είσοδο χωρίς stress
1	78.05%	70.11%
2	71.59%	76.60%
3	80.85%	69.66%
4	79.54%	80.85%
5	69.51%	71.59%
Μέση τιμή	75.90%	73.76%

συζήτηση των αποτελεσμάτων δίνεται στο επόμενο Κεφάλαιο.

4.2 Πρώτη προσπάθεια αναγνώρισης συναισθήματος στα πλαίσια της Συναισθηματικής Ακουστικής Οικολογίας

Έπειτα από το πείραμα συσχέτισης ακραίων συναισθηματικών καταστάσεων με τεχνικά χαρακτηριστικά, έγινε μία πρώτη διερεύνηση αναγνώρισης συναισθήματος από ηχητικά γεγονότα. Χρησιμοποιήθηκαν 4 συναισθηματικές κλάσεις και ένα πλήθος 100 τεχνικών μεγεθών για κάθε ηχητικό γεγονός. Τα τεχνικά μεγέθη αυτά χρησιμοποιήθηκαν ως είσοδο σε αλγόριθμους κατηγοριοποίησης. Πριν την εξαγωγή των αποτελεσμάτων από τους αλγόριθμους κατηγοριοποίησης, έγινε μία διερεύνηση σχετικά με την ενέργεια σε ένα ηχητικό γεγονός και την συναισθηματική επισήμανση Διέγερσης που έχει.

Στο **Σχήμα 4.1** είναι μία αναπαράσταση της μέσης ενέργειας ανά ηχητικό γεγονός του σετ IADS σε σχέση με την τιμή διέγερσής του. Στον **Πίνακα 4.3** είναι μία αύξουσα λίστα, σε σχέση με την τιμή Διέγερσης, με 7 ηχητικά γεγονότα, επίσης από το σετ IADS, και την αντίστοιχη ενέργεια σε αυτά. Τα ηχητικά γεγονότα αυτά επιλέχθηκαν βάσει του σημασιολογικού τους περιεχομένου, της ενέργειας που υπολογίστηκε και της επισήμανσης



Σχήμα 4.1: Αποτύπωση της κανονικοποιημένης ενέργειας (Normalized energy) ενός ηχητικού γεγονότος σε σχέση με την επισήμανση Διέγερσης (Arousal) στους ήχους IADS της Διέγερσης που συνοδεύει κάθε ένα από αυτά.

Πίνακας 4.2: Αποτελέσματα ευστοχίας αναγνώρισης συναισθηματικής πληροφορίας

Αλγόριθμος	Ποσοστό ευστοχίας
SVM	43.7%
ANN	36.5%

Στον **Πίνακα 4.2** είναι τα αποτελέσματα που παρήχθησαν κατά την διαδικασία εξόρυξης συναισθηματικής πληροφορίας από ηχητικά γεγονότα, στα πλαίσια του πειράματος που περιγράφηκε στην Ενότητα 3.3.2.

Όπως μπορεί να φανεί στον **Πίνακα 4.2**, τα ποσοστά ευστοχίας της διαδικασίας αναγνώρισης συναισθήματος, από τους αλγόριθμους που χρησιμοποιήθηκαν και είναι οι Artificial Neural Networks (ANN) και Support Vector Machines (SVM), είναι μη αξιοποιήσιμα. Από την άλλη όμως, από τον **Πίνακα 4.3** μπορεί να εξαχθεί ένα χρήσιμο συμπέρασμα σχετικά με την συσχέτιση της ενέργειας, της Διέγερσης αλλά και του σημασιολογικού περιεχομένου. Τέλος, ένα ακόμα συμπέρασμα από αυτή την Διερεύνηση είναι και το γεγονός ότι οι τέσσερις κλάσεις συναισθήματος που λήφθηκαν υπόψη είτε δεν ήταν ο βέλτιστος αριθμός, ίσως μία μεγαλύτερη κβάντιση να επέφερε καλύτερα αποτελέσματα, ή αυτές οι 4 κλάσεις δεν χρησι-

Πίνακας 4.3: Λίστα κανονικοποιημένης ενέργειας, E , 7 ενδεικτικών ηχητικών γεγονότων με αύξουσα τιμή διέγερσης

Σημσιολογικό περιεχόμενο	E	Διέγερση
Συναγερμός αεροπορικής επιδρομής	0.0293	3.36
Αντρικές κραυγές	0.0210	3.71
Γυναικείες κραυγές	0.0210	4.46
Μουσικό όργανο Bongos	0.0372	4.47
Soda Fizz	0.0009	5.85
Αντρικοί ήχοι συνουσίας	0.0029	6.64
Γραφομηχανή	0.0057	7.32

μοποιήθηκαν με τρόπο που να ωφελεί την διαδικασία αναγνώρισης συναισθήματος. Ίσως μία ιεραρχική προσέγγιση να ήταν προτιμότερη. Περισσότερη ανάλυση των αποτελεσμάτων από αυτό το πείραμα θα γίνει στο επόμενο Κεφάλαιο.

4.3 Συσχέτιση ρυθμού και Διέγερσης για ηχητικά γεγονότα

Το πρώτο πείραμα αναγνώρισης συναισθήματος από ηχητικά γεγονότα ακολούθησε ένα ακόμα, με την διαφορά ότι το δεύτερο εστίασε αποκλειστικά σε ένα τεχνικό χαρακτηριστικό του ήχου, αυτό του ρυθμού, λήφθηκε υπόψη μόνο η συναισθηματική κατάσταση της Διέγερσης και τα ηχητικά γεγονότα είχαν ομαδοποιηθεί σε 2 συναισθηματικές κλάσεις. Ως σετ δεδομένων χρησιμοποιήθηκε πάλι το IADS, τα ηχητικά δεδομένα υπέστησαν μία προεπεξεργασία (παραθυροποίηση) ενώ ως αλγόριθμοι κατηγοριοποίησης χρησιμοποιήθηκαν οι: α) K Nearest neighbours, β) Artificial Neural Networks, και γ) Logistic Regression. Για την επαλήθευση του μοντέλου χρησιμοποιήθηκε η τεχνική Stratified Cross Validation και η Leave-One-Out (LOO). Η πλήρης θεωρητική προσέγγιση αλλά και η λεπτομέρειες της πειραματικής διαδικασίας αποτυπώνονται στο προηγούμενο Κεφάλαιο.

Σε αυτό το πείραμα παρήχθησαν δύο αποτελέσματα. Ένα σχετικά με την ιεράρχηση των τεχνικών χαρακτηριστικών, βάσει της διαδικασίας που περιγράφηκε στην Ενότητα 3.3.3, και

Πίνακας 4.4: Ιεράρχηση των τεχνικών χαρακτηριστικών βάσει της χρονικής διάρκειας του πλαισίου (w) και του αλγόριθμου ιεράρχησης που χρησιμοποιήθηκε. Η σειρά εμφάνισης είναι και η σειρά ιεράρχησης (πρώτα είναι τα σημαντικότερα)

	$w = 0.8s$	$w = 1.0s$	$w = 1.2s$	$w = 1.4s$	$w = 1.6s$	$w = 1.8s$	$w = 2.0s$							
InfoGain SVM1	InfoGain SVM1	InfoGain SVM1	InfoGain SVM1	InfoGain SVM1	InfoGain SVM1	InfoGain SVM1	InfoGain SVM1							
D	Z	R	K	M	K	S	J	Q	J	N	K	A	K	
F	J	I	M	X	J	F	K	K	G	K	I	Z	W	J
C	K	D	J	L	F	Z	B	V	L	X	J	Q	E	
N	L	A	W	G	W	C	G	H	E	S	E	Y	L	
E	I	B	E	P	I	H	E	X	W	M	L	N	G	
M	G	S	L	N	A	W	L	M	N	H	Y	P	B	
K	B	J	F	F	E	B	I	L	Y	J	W	E	X	
Z	F	K	I	K	X	J	F	J	G	K	I	K	W	
J	C	M	D	J	L	K	Z	K	V	Z	X	J	Q	
G	D	L	R	A	M	L	S	N	Q	Y	N	B	A	
B	E	F	B	E	P	I	H	Y	X	W	M	X	N	
L	M	W	S	W	N	G	W	E	M	E	H	L	P	
I	N	E	A	I	G	E	C	W	H	L	S	G	Y	
R	O	H	P	Y	B	O	V	R	S	U	T	U	T	
T	P	Z	N	Z	R	D	X	D	A	R	F	O	H	
Q	W	Y	C	C	D	R	Y	T	F	V	Q	D	M	
A	X	Q	G	U	Q	P	A	I	Z	B	O	Z	I	
S	Y	X	U	O	V	T	N	C	O	G	P	R	S	
H	V	T	O	S	T	Q	M	P	B	C	A	C	F	
W	U	C	V	D	H	Y	U	F	U	Q	D	M	V	
O	T	P	Z	B	Z	V	D	S	D	T	R	T	O	
P	Q	N	Y	R	C	X	R	A	T	F	V	H	D	
V	R	O	H	T	Y	M	O	B	R	A	U	F	U	
U	S	V	X	H	O	U	T	U	C	D	G	V	R	
X	H	G	T	Q	S	A	Q	Z	P	O	C	I	C	
Y	A	U	Q	V	U	N	P	O	I	P	B	S	Z	

ένα για την ευστοχία του μοντέλου. Το πρώτο είναι στον **Πίνακα 4.4** ενώ το δεύτερο στον **4.5**. Η παράθεση της ιεράρχησης των τεχνικών χαρακτηριστικών τηρεί την δεικτοδότηση των τελευταίων σύμφωνα με τον **Πίνακα 3.9**.

Πίνακας 4.5: Η ευστοχία των παραγόμενων μοντέλων για την συσχέτιση των ρυθμικών χαρακτηριστικών και της διέγερσης αναλόγως του χρονικού πλαισίου (διάρκειας w), του αλγόριθμου που χρησιμοποιήθηκε και της τεχνικής επαλήθευσης

	Αλγόριθμος	3-fold	8-fold	24-fold	LOO
$w = 0.8s$	ANN	76.65%	77.84%	79.04%	80.84%
	LR	83.83%	85.03%	85.63%	85.63%
	KNN	82.05%	84.43%	85.03%	85.03%
$w = 1.0s$	ANN	78.44%	79.04%	79.64%	81.02%
	LR	81.44%	86.23%	87.21%	88.37%
	KNN	83.23%	83.23%	83.23%	84.43%
$w = 1.2s$	ANN	76.65%	77.96%	77.96%	80.02%
	LR	85.63%	85.63%	85.63%	85.72%
	KNN	84.43%	85.03%	85.63%	85.63%
$w = 1.4s$	ANN	71.26%	76.05%	77.32%	78.25%
	LR	85.63%	85.63%	85.63%	85.63%
	KNN	84.43%	83.83%	83.83%	83.83%
$w = 1.6s$	ANN	75.45%	77.84%	80.24%	81.32%
	LR	85.63%	85.63%	85.63%	85.63%
	KNN	85.63%	85.03%	85.63%	85.63%
$w = 1.8s$	ANN	77.25%	80.24%	82.63%	84.57%
	LR	85.63%	85.63%	86.45%	87.71%
	KNN	85.03%	85.03%	85.03%	85.03%
$w = 2.0s$	ANN	77.84%	76.65%	78.25%	79.32%
	LR	85.63%	85.63%	85.63%	86.04%
	KNN	83.83%	85.03%	85.03%	85.03%

Όπως μπορεί να φανεί από τα αποτελέσματα, η ακρίβεια/ευστοχία της διαδικασίας κατη-

γοριοποίησης είναι της τάξης του 87% και αποτελεί μία από τις πιο υψηλά αναφερόμενες, στην διεθνή επιστημονική βιβλιογραφία, σχετικά με την αναγνώριση συναισθημάτων από ηχητικά γεγονότα. Αυτό το αποτέλεσμα ενισχύει την κοινή πεποίθηση ότι ο ρυθμός επηρεάζει την Διέγερση, καθιστώντας έτσι το τεχνικό χαρακτηριστικό αυτό ένα σημαντικό και γενικό χαρακτηριστικό για την επήρεια του ήχου στον άνθρωπο. Η πλήρης ανάλυση των αποτελεσμάτων βρίσκεται στο επόμενο Κεφάλαιο.

4.4 Επήρεια χωρικής τοποθέτησης της πηγής στο συναίσθημα του ακροατή

Μετά από τα πειράματα των οποίων τα αποτελέσματα παρατίθενται παραπάνω, ακολούθησε η διερεύνηση της επήρειας της χωρικής τοποθέτησης της πηγής στο συναίσθημα του ακροατή. Για αυτό το σκοπό, διενεργήθηκε ένα πείραμα ακρόασης (listening test) υποκειμενικής αξιολόγησης με αμφιωτικά επεξεργασμένους ήχους, προερχόμενους από το σετ IADS. Η χωρική τοποθέτηση, μέσω της αμφιωτικής τεχνολογίας, που χρησιμοποιήθηκε είναι για 0, 45, 90, 135 και 180 μοίρες, με 0° να είναι ακριβώς μπροστά από τον ακροατή. Η αναλυτική περιγραφή της θεωρητικής προσέγγισης και της πειραματικής διαδικασίας παρατίθεται στο προηγούμενο Κεφάλαιο.

Στον **Πίνακα 4.6** είναι η μέση τιμή και η τυπική απόκλιση του πλήθους των συναισθηματικών επισημάνσεων ανά γωνία και συνολικά για κάθε $s_b(i', k)$. Επίσης, έγινε σύγκριση των τιμών της συναισθηματικής επισήμανσης που με αυτές από τους αρχικούς IADS ήχους. Πιο συγκεκριμένα, εάν $\overline{A}_a[x]$ και $\overline{A}_v[x]$ είναι οι μέσες τιμές συναισθηματικής επισήμανσης Διέγερσης και Ευχαρίστησης αντιστοίχως και για ένα σήμα x βάσει των:

$$\overline{A}_a[x] = \frac{1}{M} \sum_{p=1}^M A_a[x, p] \quad (4.1)$$

$$\overline{A}_v[x] = \frac{1}{M} \sum_{p=1}^M A_v[x, p] \quad (4.2)$$

τότε η προαναφερθείσα σύγκριση έγινε υπολογίζοντας την απόσταση $D_a(i')$ και $D_v(i')$ αντιστοίχως, βάσει των εξισώσεων:

Πίνακας 4.6: Η μέση τιμή και η τυπική απόκλιση (σ) για το πλήθος των επισημάνσεων που έλαβε κάθε ήχος ανά γωνία και συνολικά

Πλήθος επισημάνσεων	Γωνίες				
	0°	45°	90°	135°	180°
Μέση τιμή	9.54	10.60	9.84	10.36	9.92
σ	3.11	3.72	3.84	3.02	3.90
	Μέσος όρος μέσης τιμής				10.05
	Μέσος όρος σ				3.52

$$D_a(i') = \overline{A}_a[s(i' + f(i))] - \overline{A}_a[s_b(i', 1)] \quad (4.3)$$

$$D_v(i') = \overline{A}_v[s(i' + f(i))] - \overline{A}_v[s_b(i', 1)]$$

Πίνακας 4.7: Η διαφορά των επισημάνσεων των συναισθηματικών επισημάνσεων από το πείραμα της επήρειας της χωρικής τοποθέτησης και των αρχικών ήχων IADS

i'	$D_a(i')$	$D_v(i')$	i'	$D_a(i')$	$D_v(i')$
1	-2.09	-0.75	26	0.58	0.23
2	2.25	1.67	27	0.14	1.55
3	-0.54	1.07	28	-0.72	1.21
4	-1.04	0.84	29	0.52	1.31
5	0.70	-0.97	30	-0.07	1.02
6	0.77	0.35	31	-1.99	1.75
7	2.53	-0.81	32	0.22	1.46
8	-1.24	0.55	33	0.22	-0.37
9	1.65	-1.63	34	0.26	0.46
10	1.18	-0.17	35	1.37	-0.59
11	0.05	0.14	36	1.99	-0.68
12	-0.11	0.95	37	1.11	-0.54
13	0.19	0.13	38	-2.83	1.79
14	0.88	0.01	39	-0.33	0.31

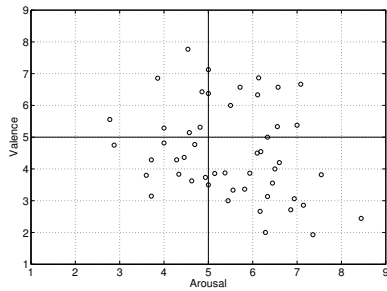
Πίνακας 4.7 – Συνέχεια σε επόμενη σελίδα

Πίνακας 4.7 – Συνέχεια από προηγούμενη σελίδα

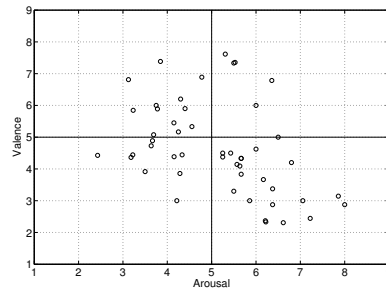
i'	$D_a(i')$	$D_v(i')$	i'	$D_a(i')$	$D_v(i')$
15	0.43	-0.54	40	0.72	2.13
16	1.48	0.72	41	-1.42	0.61
17	-0.83	-0.42	42	-0.20	0.71
18	-0.06	0.04	43	-1.45	1.86
19	0.30	0.08	44	0.00	1.99
20	-0.58	0.79	45	2.09	0.08
21	0.70	-1.08	46	-0.49	-0.30
22	1.58	-0.58	47	0.58	1.88
23	0.52	-0.25	48	-0.05	1.03
24	0.72	-1.79	49	-1.14	1.76
25	0.10	0.24	50	0.25	-0.01
Μέση διέγερση:		0.18	Μέση ευχαρίστηση:		0.38
σ διέγερσης:		1.12	σ ευχαρίστησης:		0.98
Απόλυτη μέση διέγερση:		0.87	Απόλυτη μέση ευχαρίστηση:		0.84
Απόλυτο σ διέγερσης:		0.73	Απόλυτο σ ευχαρίστησης:		0.62

Οι τιμές των $D_a(i')$ και $D_v(i')$ καθώς και οι μέσες τιμές αυτών αλλά και η τυπική απόκλιση (σ) παρατίθενται στον **Πίνακα 4.7**. Στα **Σχήματα 4.2α'** έως **4.2ε'** παρατίθενται τα διαγράμματα διασποράς των συναισθηματικών επισημάνσεων για κάθε $S'_a(k)$.

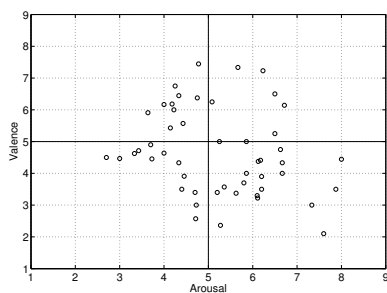
Ιδιαίτερο ενδιαφέρον έχει η μεταβολή των \vec{v}_a κατά την μετακίνηση της πηγής. Αυτό μπορεί να μοντελοποιηθεί με την θεώρηση ενός διανυσματικού χώρου, Διέγερσης – Ευχαρίστησης, όπου για κάθε μετακίνηση της πηγής από μία γωνία σε μία άλλη, συνεχόμενη, σχηματίζεται ένα διάνυσμα. Το τελευταίο έχει ως αρχή την συνισταμένη των τιμών Διέγερσης – Ευχαρίστησης κατά την γωνία εκκίνησης και ως τέλος την συνισταμένη των ίδιων μεγεθών αλλά για την γωνία τερματισμού. Έτσι, κάθε σημείο σε αυτά μπορεί να περιγραφεί από την καρτεσιανές του συντεταγμένες: $(a(i', k), v(i', k))$, ισοδύναμες με τις μέσες τιμές επισημάνσεως που έλαβε το εκάστοτε ηχητικό ερέθισμα για συγκεκριμένη γωνία $\theta(k)$ και υπολογιζόμενες ως εξής:



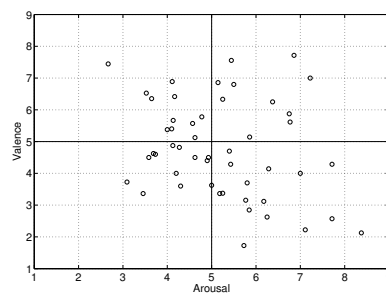
(α') Μέσες τιμές Διέγερσης και Ευχαρίστησης για το $S'_a(1)$ ($\theta(k)=0^\circ$)



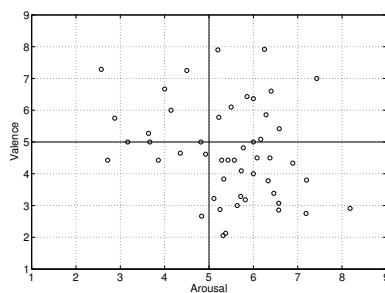
(β') Μέσες τιμές Διέγερσης και Ευχαρίστησης για το $S'_a(2)$ ($\theta(k)=45^\circ$)



(γ') Μέσες τιμές Διέγερσης και Ευχαρίστησης για το $S'_a(3)$ ($\theta(k)=90^\circ$)



(δ') Μέσες τιμές Διέγερσης και Ευχαρίστησης για το $S'_a(4)$ ($\theta(k)=135^\circ$)



(ε') Μέσες τιμές Διέγερσης και Ευχαρίστησης για το $S'_a(5)$ ($\theta(k)=180^\circ$)

Σχήμα 4.2: Μέσες τιμές Διέγερσης και Ευχαρίστησης για τα $S'_a(k)$, με $k \in [1, 5]$

$$a(i', k) = \overline{A}_a[s_b(i', k)] \quad (4.4)$$

$$v(i', k) = \overline{A}_v[s_b(i', k)] \quad (4.5)$$

Έτσι, ένα ακόμα μέγεθος που λήφθηκε ως αποτέλεσμα είναι η διακύμανση των $\overline{A}_a[s_b(i', k)]$ και $\overline{A}_v[s_b(i', k)]$ για διαφορετικές γωνίες $\theta(k)$. Το σετ τιμών που αναφέρθηκε, και μπορεί να ονομαστεί ως διανύσματα $\vec{v}\vec{a}$ (Διέγερσης – Ευχαρίστησης), για συνεχόμενες γωνίες μπορεί να εκφραστεί ως:

$$\begin{aligned} \vec{v}\vec{a}(i', k) &= (a(i', k) - a(i', k-1)) \vec{a} \\ &+ (v(i', k) - v(i', k-1)) \vec{v} \end{aligned} \quad (4.6)$$

ή, ισοδύναμα:

$$\vec{v}\vec{a}(i', k) = (\overline{A}_a^d[s_b(ik)]) \vec{a} + (\overline{A}_v^d[s_b(ik)]) \vec{v} \quad (4.7)$$

όπου \vec{a} και \vec{v} είναι τα μοναδιαία διανύσματα της Διέγερσης και Ευχαρίστησης τα οποία ορίζουν τον διανυσματικό χώρο, αντιστοίχως, $k \in [2, 5]$ και τα $\overline{A}_a^d[s_b(ik)]$ και $\overline{A}_v^d[s_b(ik)]$ ορίζονται ως:

$$\overline{A}_a^d[s_b(ik)] = \overline{A}_a[s_b(i', k)] - \overline{A}_a[s_b(i', k-1)] \quad (4.8)$$

$$\overline{A}_v^d[s_b(ik)] = \overline{A}_v[s_b(i', k)] - \overline{A}_v[s_b(i', k-1)] \quad (4.9)$$

Η γωνίες των παραπάνω διανυσμάτων, αντίθετα με την φορά του ρολογιού, καθώς και το μέτρο τους υπολογίζονται βάσει των εξισώσεων:

$$|\vec{v}\vec{a}(i', k)| = \sqrt{\left(\overline{A}_a^d[s_b(i', k)]\right)^2 + \left(\overline{A}_v^d[s_b(i', k)]\right)^2} \quad (4.10)$$

$$\phi(i', k) = \arctan\left(\frac{\overline{A}_a^d[s_b(ik)]}{\overline{A}_v^d[s_b(ik)]}\right) \quad (4.11)$$

Οι τελικές τιμές των μέτρων των διανυσμάτων και των γωνιών $\phi(i', k)$ εμφανίζονται στον Πίνακα 4.8.

Πίνακας 4.8: Τα μέτρα και οι γωνίες των διανυσμάτων $|\vec{v}\hat{a}|$ αναλόγως τις γωνιακές μεταβάσεις

i'	Γωνιακές μεταβάσεις							
	$0^\circ \rightarrow 45^\circ$		$45^\circ \rightarrow 90^\circ$		$90^\circ \rightarrow 135^\circ$		$135^\circ \rightarrow 180^\circ$	
	$\phi(i', 2)$	$ \vec{v}\hat{a}(i', 2) $	$\phi(i', 3)$	$ \vec{v}\hat{a}(i', 3) $	$\phi(i', 4)$	$ \vec{v}\hat{a}(i', 4) $	$\phi(i', 5)$	$ \vec{v}\hat{a}(i', 5) $
1	-150.89	1.80	-171.35	1.11	69.78	0.84	47.54	0.88
2	85.14	2.10	-79.94	1.27	82.54	0.77	164.05	1.27
3	144.58	1.31	-7.99	0.74	-137.54	1.33	91.83	1.57
4	88.75	1.35	-112.02	1.19	39.54	2.30	-112.31	1.20
5	126.05	1.18	-18.71	0.60	14.29	1.56	176.19	2.15
6	-132.75	1.15	173.93	0.53	0.00	1.92	141.99	1.25
7	136.90	1.04	-126.03	0.45	49.14	0.87	-17.44	1.86
8	174.68	2.60	-36.43	1.09	142.79	0.59	141.51	1.22
9	-78.24	1.44	82.52	0.55	-104.93	0.34	5.06	2.16
10	179.06	1.36	26.57	0.99	135.00	0.31	24.15	0.88
11	-148.39	1.27	45.00	0.71	34.70	0.88	-111.97	1.71
12	61.46	0.51	-60.95	0.74	-82.41	0.27	58.90	1.31
13	-115.69	1.40	93.84	1.02	95.71	0.71	-15.54	1.96
14	-114.02	1.94	56.31	0.90	-146.63	0.72	1.96	2.21
15	143.51	1.44	-71.57	1.36	56.31	0.97	-132.27	0.21
16	105.48	0.80	78.50	1.00	-106.26	0.42	-14.62	0.28
17	-150.36	1.06	133.78	1.48	26.57	0.67	-87.99	0.52
18	48.37	1.51	-136.70	1.79	106.77	1.48	-42.88	1.45
19	121.15	1.52	-103.64	0.96	-54.46	0.78	141.67	0.52
20	175.36	0.26	118.07	0.27	0.00	0.14	-66.64	0.62
21	-60.95	1.14	150.86	1.44	-93.22	0.55	-5.87	0.81
22	-23.20	1.90	75.47	1.16	-125.42	1.42	106.70	0.46
23	159.01	1.23	29.36	1.29	-105.95	0.81	81.76	0.53
24	-145.95	0.64	22.75	0.60	-55.30	0.28	-82.57	0.37
25	-144.46	0.66	-57.69	0.56	136.23	1.00	-6.30	1.56
26	175.96	1.11	-35.33	1.49	122.35	1.33	93.18	0.38
27	55.01	0.55	180.00	0.64	34.01	1.10	-141.34	0.99
28	104.04	0.52	90.00	0.62	-120.26	0.69	-19.98	1.17
29	14.38	0.64	169.77	1.02	-35.80	1.07	-15.22	2.34
30	168.02	0.34	91.70	1.02	-64.71	1.14	172.57	0.97

Πίνακας 4.8 – Συνέχεια σε επόμενη σελίδα

Πίνακας 4.8 – Συνέχεια από προηγούμενη σελίδα

i'	$0^\circ \rightarrow 45^\circ$		$45^\circ \rightarrow 90^\circ$		$90^\circ \rightarrow 135^\circ$		$135^\circ \rightarrow 180^\circ$	
	$\phi(i', 2)$	$ \vec{v}\hat{a}(i', 2) $	$\phi(i', 3)$	$ \vec{v}\hat{a}(i', 3) $	$\phi(i', 4)$	$ \vec{v}\hat{a}(i', 4) $	$\phi(i', 5)$	$ \vec{v}\hat{a}(i', 5) $
31	152.88	0.84	107.65	0.58	99.46	1.52	-71.39	2.59
32	-135.00	0.94	-60.26	0.38	121.26	2.08	-38.66	2.05
33	-120.74	0.47	134.88	1.78	171.47	0.36	-75.96	1.55
34	80.54	0.65	-71.57	1.19	-25.02	0.61	129.92	1.95
35	-28.07	1.89	142.68	1.41	-77.47	0.69	148.57	0.72
36	162.65	2.00	-31.86	2.71	164.05	1.32	55.01	1.57
37	6.48	1.85	169.58	1.92	-22.09	1.73	172.98	2.48
38	171.87	0.94	2.73	2.34	174.39	2.61	-61.75	0.69
39	-176.69	1.08	-27.31	1.96	152.41	2.73	-20.03	1.48
40	1.17	0.56	144.46	0.22	134.47	1.10	18.12	0.69
41	-134.44	1.03	129.62	1.90	-33.69	1.56	-60.26	1.07
42	160.35	0.83	48.61	0.77	175.96	1.25	25.23	1.37
43	151.78	2.58	-5.57	1.72	-70.02	1.46	103.84	0.81
44	171.23	3.36	28.49	1.25	-170.54	0.17	-93.27	0.42
45	-1.71	0.76	-71.57	0.29	-131.19	0.97	118.24	0.80
46	-127.12	1.16	-101.51	0.79	55.01	1.74	148.57	0.75
47	59.71	2.12	3.67	0.93	180.00	2.11	-30.26	1.54
48	76.07	1.28	-114.21	1.36	29.54	0.86	-90.00	0.70
49	102.76	0.78	-17.20	2.20	-170.54	0.65	-120.96	0.83
50	-20.62	2.06	135.00	1.68	-45.00	0.47	108.43	0.35

Τα αποτελέσματα δείχνουν ότι η θέση της πηγής έχει επίρεια στην συναίσθηματική κατάσταση του ακροατή. Πιο συγκεκριμένα, φαίνονται δύο κλάσεις ηχητικών γεγονότων: μία που καθώς η πηγή μετακινείται προς τα πίσω στον ακροατή, η διέγερσή του αυξάνει ενώ μειώνεται η ευχαρίστησή του, και μία που καθώς η πηγή μετακινείται προς τα πίσω μειώνεται η Διέγερση και αυξάνεται η Ευχαρίστηση του ακροατή. Η πλήρης ανάλυση των αποτελεσμάτων είναι στο επόμενο Κεφάλαιο.

4.5 Συσχέτιση ρυθμού και Ευχαρίστησης

Τέλος, διενεργήθηκε ένα πείραμα διερεύνησης της συσχέτισης του ρυθμού με την Ευχαρίστηση. Με αυτό το πείραμα ολοκληρώνεται η διερεύνηση της επίρειας του ρυθμού πάνω στον ακροατή. Για την πειραματική διαδικασία, χρησιμοποιήθηκε το σετ IADS και ένα πλήθος τεχνικών χαρακτηριστικών που αναλύονται στο προηγούμενο Κεφάλαιο. Τα ηχητικά γεγονότα διαχωρίστηκαν σε δύο κλάσεις. Μία με τα δεδομένα που είχαν επισημανθεί με Ευχαρίστηση κάτω από την μέση τιμή, δηλαδή το 5, και μία με τα υπόλοιπα. Επίσης, στην διαδικασία κατηγοριοποίησης χρησιμοποιήθηκαν δύο σετ δεδομένων εισόδων στους αλγόριθμους κατηγοριοποίησης. Ένα με τα τεχνικά χαρακτηριστικά και με την επισήμανση της Ευχαρίστησης και ένα δεύτερο με τα τεχνικά χαρακτηριστικά, την επισήμανση της Ευχαρίστησης και την επισήμανση της Διέγερσης. Το πρώτο ονομάστηκε $R'_{i,x'}$ και το δεύτερο $R''_{i,x'}$.

Για κάθε αλγόριθμο, σετ δεδομένων, κλάσης (C_1 και C_2) και χρονική διάρκεια τμήματος λήφθηκαν η ακρίβεια (precision), η ανάκληση (recall), η ευστοχία (accuracy) και το μέσο λάθος (mean square error). Τα σχετικά αποτελέσματα εμφανίζονται στους **Πίνακες 4.9** έως **4.11** αντιστοίχως. Με γκρι σκίαση επισημαίνονται τα αποτελέσματα εκείνα που έχουν τις μεγαλύτερες τιμές σε κάθε σειρά του εκάστοτε Πίνακα, δηλαδή είτε ακρίβειας και ανάκλησης είτε ευστοχίας. Συγκριμένα, στον **Πίνακα 4.9** είναι οι τιμές της ανάκλησης και της ακρίβειας για όλους τους αλγόριθμους και για το σετ ήχων $R'_{i,x'}$, στον **Πίνακα 4.10** είναι οι τιμές της ανάκλησης και της ακρίβειας για όλους τους αλγόριθμους και για το σετ ήχων $R''_{i,x'}$ ενώ στον **Πίνακα 4.11** είναι οι τιμές της ευστοχίας και του μέσου λάθους για όλους τους αλγόριθμους και όλα τα σετ ήχων.

παραθύρου

Τα αποτελέσματα δείχνουν ότι η επίρεια του ρυθμού στην Ευχαρίστηση δεν είναι τόσο μεγάλη όπως στην περίπτωση της Διέγερσης. Από την άλλη, με την χρήση του ρυθμού ως ένα επιπλέον χαρακτηριστικό, η ακρίβεια των παραγόμενων μοντέλων αυξάνει, κάτι το οποίο δείχνει αφενός την σημαντικότητα του ρυθμού στην επίρεια της Διέγερσης και αφετέρου ότι με την χρήση της Διέγερσης ως ένα επιπλέον χαρακτηριστικό, η αναγνώριση της Ευχαρίστησης μόνο από τεχνικά μεγέθη συσχετιζόμενα με τον ρυθμό, είναι εφικτή ως ένα σημείο. Η πλήρης ανάλυση των αποτελεσμάτων είναι στο επόμενο Κεφάλαιο.

Πίνακας 4.9: Οι τιμές της ανάληψης (R) και της ακρίβειας (P) για κάθε λάσση (C_1 και C_2), για όλους τους αλγόριθμους και για το σετ ήλων $R'_{i,gr}$. Με γκρι σκίαση είναι τα καλά με τις μέγιστες τιμές ανά λάσση και ως "D. Stump" είναι ο "Decision Stump".

	Ada Boost		Bagging		D. Stump		C4.5		SVM		RepTree		SMO	
	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2
$i = 1$ (μήκος παραθύρου = 0.8 seconds)														
R	0.558	0.568	0.616	0.556	0.512	0.556	0.512	0.494	0.535	0.519	0.709	0.284	0.523	0.593
P	0.578	0.548	0.596	0.577	0.550	0.517	0.518	0.488	0.541	0.512	0.513	0.479	0.577	0.539
$i = 2$ (μήκος παραθύρου = 1.0 seconds)														
R	0.640	0.543	0.616	0.469	0.593	0.420	0.802	0.235	0.628	0.556	0.791	0.321	0.640	0.519
P	0.598	0.587	0.552	0.535	0.520	0.493	0.527	0.528	0.600	0.584	0.553	0.591	0.585	0.575
$i = 3$ (μήκος παραθύρου = 1.2 seconds)														
R	0.628	0.457	0.628	0.457	0.605	0.494	0.488	0.531	0.616	0.506	0.651	0.395	0.640	0.469
P	0.551	0.536	0.551	0.536	0.559	0.541	0.525	0.494	0.570	0.554	0.533	0.516	0.561	0.551
$i = 4$ (μήκος παραθύρου = 1.4 seconds)														
R	0.581	0.494	0.547	0.519	0.512	0.444	0.674	0.395	0.535	0.506	0.709	0.358	0.523	0.519
P	0.549	0.526	0.547	0.519	0.494	0.462	0.542	0.533	0.535	0.506	0.540	0.537	0.536	0.506
$i = 5$ (μήκος παραθύρου = 1.6 seconds)														
R	0.593	0.617	0.616	0.556	0.570	0.741	0.547	0.654	0.581	0.531	0.686	0.556	0.605	0.580
P	0.622	0.588	0.596	0.577	0.700	0.619	0.627	0.576	0.568	0.544	0.621	0.625	0.605	0.580
$i = 6$ (μήκος παραθύρου = 1.8 seconds)														
R	0.628	0.617	0.558	0.531	0.488	0.802	0.430	0.667	0.593	0.556	0.535	0.691	0.628	0.617
P	0.635	0.610	0.558	0.531	0.724	0.596	0.578	0.524	0.586	0.563	0.648	0.583	0.635	0.610
$i = 7$ (μήκος παραθύρου = 2.0 seconds)														
R	0.593	0.543	0.616	0.531	0.279	0.765	0.523	0.519	0.581	0.568	0.640	0.444	0.558	0.568
P	0.580	0.557	0.582	0.566	0.558	0.500	0.536	0.506	0.588	0.561	0.550	0.537	0.578	0.548

Πίνακας 4.10: Οι τιμές της ανάκλησης (R) και της ακρίβειας (P) για κάθε κλάση (C_1 και C_2), για όλους τους αλγόριθμους και για το σετ ήχων $R'_{i,gr}$. Με γκρι σκίαση είναι τα κελιά με τις μέγιστες τιμές ανά κλάση και ως “D. Stump” είναι ο “Decision Stump”.

Ada Boost		Bagging		D. Stump		C4.5		SVM		RepTree		SMO		
C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	
$i = 1$ (μήκος παραθύρου = 0.8 seconds)														
R	0.628	0.593	0.674	0.593	0.512	0.556	0.744	0.506	0.558	0.556	0.651	0.346	0.640	0.593
P	0.621	0.600	0.637	0.632	0.550	0.517	0.615	0.651	0.571	0.542	0.514	0.483	0.6250	0.608
$i = 2$ (μήκος παραθύρου = 1.0 seconds)														
R	0.651	0.543	0.698	0.593	0.593	0.420	0.698	0.531	0.581	0.580	0.756	0.370	0.674	0.556
P	0.602	0.595	0.645	0.649	0.520	0.493	0.612	0.623	0.595	0.566	0.560	0.588	0.617	0.616
$i = 3$ (μήκος παραθύρου = 1.2 seconds)														
R	0.616	0.519	0.663	0.519	0.605	0.494	0.733	0.395	0.593	0.519	0.663	0.383	0.674	0.556
P	0.576	0.560	0.594	0.592	0.559	0.541	0.563	0.582	0.567	0.545	0.533	0.517	0.617	0.616
$i = 4$ (μήκος παραθύρου = 1.4 seconds)														
R	0.593	0.605	0.593	0.580	0.512	0.444	0.535	0.667	0.570	0.543	0.744	0.346	0.616	0.580
P	0.614	0.583	0.600	0.573	0.494	0.462	0.630	0.574	0.570	0.543	0.547	0.560	0.609	0.588
$i = 5$ (μήκος παραθύρου = 1.6 seconds)														
R	0.640	0.580	0.628	0.543	0.570	0.741	0.663	0.593	0.581	0.580	0.721	0.531	0.640	0.531
P	0.618	0.603	0.593	0.579	0.700	0.619	0.633	0.623	0.595	0.566	0.620	0.642	0.591	0.581
$i = 6$ (μήκος παραθύρου = 1.8 seconds)														
R	0.663	0.593	0.663	0.531	0.488	0.802	0.547	0.593	0.616	0.630	0.535	0.691	0.663	0.556
P	0.633	0.623	0.600	0.597	0.724	0.596	0.588	0.552	0.639	0.607	0.648	0.583	0.613	0.608
$i = 7$ (μήκος παραθύρου = 2.0 seconds)														
R	0.628	0.605	0.628	0.605	0.279	0.765	0.558	0.556	0.628	0.556	0.640	0.444	0.605	0.642
P	0.628	0.605	0.628	0.605	0.558	0.500	0.571	0.542	0.600	0.584	0.550	0.537	0.642	0.605

Πίνακας 4.11: Ευστοχία (A) και μέσο λάθος (E) για όλους τους αλγόριθμους και τα σετ $R'_{i,x'}$ και $R''_{i,x'}$. Με γκρι σκίαση είναι τα καλύτερα με τις μέγιστες τιμές ανά σετ ήχων.

	Ada Boost	Bagging	D. Stump	C4.5	SVM	RepTree	SMO							
	$R'_{i,x'}$	$R''_{i,x'}$	$R'_{i,x'}$	$R''_{i,x'}$	$R'_{i,x'}$	$R''_{i,x'}$	$R'_{i,x'}$	$R''_{i,x'}$						
$i = 1$ (μήκος παραθύρου = 0.8 seconds)														
A	56.29	61.08	58.68	63.47	53.29	53.30	50.30	62.87	52.70	55.69	50.30	50.30	55.69	61.68
E	0.543	0.531	0.554	0.528	0.514	0.514	0.606	0.566	0.563	0.562	0.554	0.548	0.666	0.619
$i = 2$ (μήκος παραθύρου = 1.0 seconds)														
A	59.28	59.88	54.49	64.67	50.90	50.90	52.70	61.68	59.28	58.08	56.29	56.89	58.08	61.68
E	0.537	0.515	0.570	0.530	0.515	0.515	0.558	0.601	0.534	0.527	0.517	0.519	0.647	0.619
$i = 3$ (μήκος παραθύρου = 1.2 seconds)														
A	54.49	56.89	54.49	59.28	55.09	55.09	50.90	56.89	56.29	55.69	52.70	52.70	55.69	61.68
E	0.548	0.524	0.568	0.555	0.503	0.503	0.592	0.591	0.554	0.551	0.530	0.534	0.666	0.619
$i = 4$ (μήκος παραθύρου = 1.4 seconds)														
A	53.89	59.88	53.29	58.68	47.90	47.90	53.89	59.88	52.10	55.69	53.89	55.09	52.10	59.88
E	0.539	0.517	0.573	0.555	0.517	0.517	0.555	0.587	0.554	0.546	0.554	0.541	0.692	0.633
$i = 5$ (μήκος παραθύρου = 1.6 seconds)														
A	60.48	61.08	58.68	58.68	65.27	65.27	59.88	62.87	55.69	58.08	62.28	62.87	59.28	58.68
E	0.523	0.526	0.558	0.538	0.477	0.477	0.579	0.591	0.532	0.524	0.496	0.492	0.638	0.642
$i = 6$ (μήκος παραθύρου = 1.8 seconds)														
A	62.28	62.87	54.49	59.89	64.07	64.07	54.49	56.89	57.49	62.28	61.08	61.08	62.28	61.08
E	0.520	0.494	0.565	0.543	0.480	0.480	0.617	0.624	0.527	0.510	0.502	0.502	0.614	0.624
$i = 7$ (μήκος παραθύρου = 2.0 seconds)														
A	56.89	61.68	57.49	61.68	51.50	51.50	52.10	55.69	57.49	59.28	54.49	54.49	56.29	62.28
E	0.551	0.522	0.549	0.543	0.519	0.520	0.649	0.631	0.548	0.533	0.529	0.529	0.661	0.614

ΚΕΦΑΛΑΙΟ 5

ΣΥΖΗΤΗΣΗ ΚΑΙ ΑΝΑΛΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Στο παρόν κεφάλαιο θα γίνει η ανάλυση και η συζήτηση των αποτελεσμάτων που παρουσιάστηκαν προηγουμένως. Προς ευκολία του αναγνώστη, η οργάνωση της συζήτησης θα ακολουθήσει την ίδια δομή με αυτή του Κεφαλαίου 4. Έτσι, στην Ενότητα 5.1 αναπτύσσεται η συζήτηση των αποτελεσμάτων κατά την αναγνώριση stress από φωνητικές εντολές. Στην Ενότητα 5.2 αναλύονται τα αποτελέσματα που λήφθηκαν κατά την πρώτη διερεύνηση αναγνώρισης συναισθημάτων από ηχητικά γεγονότα και εντός του πλαισίου της Συναισθηματικής Ακουστικής Οικολογίας ενώ στην Ενότητα 5.3 γίνεται η ποιοτική ανάλυση των αποτελεσμάτων της συσχέτισης των ρυθμικών χαρακτηριστικών ενός ηχητικού γεγονότος με την διέγερση. Τέλος, στις Ενότητες 5.4 και 5.5 αναλύονται τα αποτελέσματα από την έρευνα σχετικά με την επήρεια της γωνιακής θέσης της πηγής στην συναισθηματική κατάσταση του ακροατή και των ρυθμικών χαρακτηριστικών στην ευχαρίστηση, αντιστοίχως.

5.1 Συσχέτιση υψηλής Διέγερσης και χαμηλής Ευχαρίστησης με τεχνικά χαρακτηριστικά του ήχου

Στον **Πίνακα 4.1** παρατίθενται οι τιμές της ευστοχίας αναγνώρισης λέξεων με stress και χωρίς. Οι τιμές αυτές κυμαίνονται από 69.51%, για την λέξη “Καλημέρα”, έως και 80.85%, για τις λέξεις “Φωτιά” και “Καλά”. Η αναγνώριση φωνητικών εντολών είναι αρκετά πάνω από το εμπειρικό ποσοστό του τυχαίου, δηλαδή 50%. Επιπροσθέτως, η μέση απόδοση φαίνεται ότι είναι μεγαλύτερη όταν ως είσοδος υπάρχουν λέξεις που δεν έχουν εκφερθεί με stress ενώ για αυτές με stress η απόδοση μειώνεται κατά 2%. Πιο συγκεκριμένα, κατά την είσοδο στο σύστημα αναγνώρισης λέξεων με stress η μέση απόδοση είναι 75.9% ενώ με λέξεις χωρίς stress είναι 73.76%. Όμως, δεδομένου, από την μία, του διαφορετικού σημασιολογικού περιεχομένου έκαστης λέξης και, από την άλλη, το γεγονός ότι δεν υπάρχει κάποια προϋπάρχουσα βάση δεδομένων ομιλίας με συναισθηματική επισήμανση, αναδεικνύεται το ενδιαφέρον για την αναγνώριση έκαστης λέξης. Αυτό οφείλεται στο γεγονός ότι μία αναλυτικότερη προσέγγιση της αναγνώρισης έκαστης λέξης θα μπορέσει να φανερώσει την ύπαρξη ή μη πιθανών εξαρτήσεων από το σημασιολογικό περιεχόμενο κατά την αναγνώριση της ακραίας συναισθηματικής κατάστασης από τεχνικά χαρακτηριστικά.

Έτσι, αναλύοντας τα επί μέρους αποτελέσματα για έκαστη λέξη, φαίνεται ότι υπάρχει ένα συνολικό μέγιστο και ελάχιστο της απόδοσης και για τους δύο τύπους λέξεων, δηλαδή με stress και χωρίς. Το πρώτο είναι ακριβώς στο 80.75% ενώ το δεύτερο περίπου στο 69%. Ακόμα, και σύμφωνα με τις τιμές ευστοχίας που παρατίθενται στον **Πίνακα 4.1**, υπάρχουν λέξεις που η αναγνώριση ήταν καλύτερη όταν είχαν εκφερθεί με stress και λέξεις με ακριβώς αντίθετη συμπεριφορά, δηλαδή παρουσιάστηκε καλύτερη αναγνώριση όταν εκφέρθηκαν χωρίς stress. Πιο συγκεκριμένα, στην πρώτη περίπτωση ανήκουν οι λέξεις:

- “Κλέφτης”, λέξη με A/A 1 και ποσοστό ευστοχίας 78.05% και 70.11% κατά την εκφορά της με stress και χωρίς, αντιστοίχως
- “Καλά”, λέξη με A/A 3 και ποσοστό ευστοχίας 80.85% και 69.66% κατά την εκφορά της με stress και χωρίς, αντιστοίχως

Στην περίπτωση αυτή λοιπόν, παρατηρείται πως δύο λέξεις με σημασιολογικό περιεχόμενο εντελώς διαφορετικό μεταξύ τους και αντίθετες χρήσεις, δηλαδή η μία (“Κλέφτης”) να έχει εύλογη εκφορά με stress ενώ η άλλη (“Καλημέρα”) όχι, παρουσιάζουν παρόμοια συμπεριφορά ως προς την αναγνώριση και συγκεκριμένα κατά την εκφορά τους με stress. Στην δεύτερη περίπτωση λέξεων, δηλαδή αυτών που παρουσίασαν καλύτερο ποσοστό αναγνώρισης όταν εκφράστηκαν χωρίς stress ανήκουν οι εξής:

- “Σεισμός”, λέξη με A/A 2 και ποσοστό ευστοχίας 71.59% και 76.60% κατά την εκφορά της με stress και χωρίς, αντιστοίχως
- “Φωτιά”, λέξη με A/A 4 και ποσοστό ευστοχίας 79.54% και 80.85% κατά την εκφορά της με stress και χωρίς, αντιστοίχως
- “Καλημέρα”, λέξη με A/A 5 και ποσοστό ευστοχίας 69.51% και 71.59% κατά την εκφορά της με stress και χωρίς, αντιστοίχως

Πρόκειται πάλι για λέξεις που υπό φυσιολογικές συνθήκες η έκφρασή τους αντιβαίνει αυτή που έπρεπε να γίνει στα πλαίσια του πειράματος. Δηλαδή, υπάρχουν λέξεις που μπορούν να εκφερθούν χωρίς stress κάτω από φυσιολογικές συνθήκες, π.χ. “Καλημέρα”, και λέξεις που η εκφορά τους χωρίς ακραία συναισθηματική κατάσταση δεν είναι τόσο συνηθισμένη, π.χ. “Σεισμός”. Δεδομένου του σημασιολογικού περιεχομένου των λέξεων στις δύο παραπάνω λίστες αλλά και της περίπτωσης που αυτές παρουσίασαν το μέγιστο ποσοστό αναγνώρισης, δεν φαίνεται κάποια συσχέτιση του σημασιολογικού περιεχομένου με την απόδοση της παραπάνω διαδικασίας. Ακόμα, λόγω της αντιστοίχισης του stress με την συμβολή των συναισθημάτων “Φόβος” και “Θυμός” και της αντίστοιχης μεταβολής των κατάλληλων τεχνικών χαρακτηριστικών (όπως αναφέρθηκε στην Ενότητα 3.2.2) φαίνεται αφενός ότι η αντιστοιχία αυτή μπορεί να επαληθευτεί πειραματικά και αφετέρου πως τα τεχνικά χαρακτηριστικά που επιλέχθηκαν, δηλαδή η θεμελιώδης συχνότητα, ο ρυθμός ομιλίας και η ενέργεια, μπορούν, ως ένα μέτρο, να συντελέσουν σε μία αναγνώριση της ακραίας συναισθηματικής κατάστασης με αυξημένη διέγερση και μειωμένη ευχαρίστηση.

Αναφορικά με τα όρια που παρουσιάζονται στο **Σχήμα 3.4** και αναφέρονται στις τιμές των χαρακτηριστικών που χρησιμοποιήθηκαν οι οποίες σηματοδοτούν την ύπαρξη stress (δηλαδή, μία τιμή πάνω από το όριο δείχνει την ύπαρξη stress), μπορεί να φανεί ότι στην ενέργεια το όριο που χρησιμοποιήθηκε ήταν πολύ πιο κοντά στην περίπτωση των λέξεων με stress

σε σχέση με τον ρυθμό ομιλίας και την θεμελιώδη συχνότητα. Αν και αυτό το γεγονός σχετίζεται περισσότερο με την αναγνώριση συναισθήματος από ομιλία, και ως εκ τούτου είναι εκτός του πεδίου της ανά χείρας διατριβής, θα μπορούσε να γραφεί πως παρατηρείται μία μεγαλύτερη συγκέντρωση των τιμών για τον ρυθμό ομιλίας και την θεμελιώδη συχνότητα από ό,τι για την ενέργεια. Δηλαδή, παρουσιάζεται μεγάλη διακύμανση των τιμών της ενέργειας κατά την εκφορά λέξεων χωρίς stress. Αυτή η διακύμανση όμως δεν παρατηρείται στα άλλα μεγέθη, δηλαδή στο ρυθμό ομιλίας και στην θεμελιώδη συχνότητα.

Ανακεφαλαιώνοντας την προηγούμενη ανάλυση, γίνεται φανερό πως η ακραία συναισθηματική κατάσταση που μπορεί να μεταδοθεί μέσω ηχητικών ερεθισμάτων μπορεί να αναγνωρισθεί με κάποια σχετική ακρίβεια χρησιμοποιώντας τεχνικά χαρακτηριστικά. Αναφορικά με την αναγνώριση, αν και τα ποσοστά δεν είναι αρκετά υψηλά, παρόλα αυτά, θα πρέπει να αναφερθεί πως το πλήθος των τεχνικών χαρακτηριστικών είναι ίσως το μικρότερο που αναφέρεται στην διεθνή βιβλιογραφία σχετικά με την αναγνώριση συναισθημάτων από ηχητικά δεδομένα. Επιπροσθέτως, την ισχυρή συσχέτιση τεχνικών χαρακτηριστικών και ακραίας συναισθηματικής κατάστασης τονίζει η χρήση λέξεων που κάτω από φυσιολογικές συνθήκες δεν συνάδει η εκφορά τους με stress. Αυτό το γεγονός δείχνει πως ενδεχομένως υπάρχει μία εγγενής αντίληψη στον άνθρωπο σχετικά με τα τεχνικά χαρακτηριστικά που θα μπορούσε να έχει ένα ηχητικό ερέθισμα το οποίο θα τους εγείρει τέτοια συναισθηματική κατάσταση. Λόγω της εξελικτικής πορείας του ανθρώπου ως μέρος της Φύσης, αυτό μπορεί να σημαίνει ότι αυτή η αντίληψη πιθανόν να προέρχεται από τα χαρακτηριστικά του περιβάλλοντος του ανθρώπου. Δηλαδή, η εγγενής γνώση για τα τεχνικά χαρακτηριστικά που ο άνθρωπος αντιλαμβάνεται ότι σηματοδοτούν ακραία συναισθηματική κατάσταση είναι προϊόν μάθησης από την αλληλεπίδραση του με το περιβάλλον. Έτσι, ενδέχεται σε τέτοιες καταστάσεις το σημασιολογικό περιεχόμενο των ήχων να είναι ήσσονος σημασίας.

5.2 Αναγνώριση συναισθήματος από ηχητικά γεγονότα στα πλαίσια της Συναισθηματικής Ακουστικής Οικολογίας

Το πείραμα για την διερεύνηση της συσχέτισης υψηλής Διέγερσης και χαμηλής Ευχαρίστησης με τα τεχνικά χαρακτηριστικά του ήχου ακολούθησε μία πρώτη προσπάθεια αναγνώρισης συναισθήματος από ηχητικά γεγονότα στα πλαίσια του μοντέλου της Συναισθηματικής Ακουστικής Οικολογίας. Προς αυτό το σκοπό, έγιναν πειράματα Μηχανικής Μάθησης με ηχητικά γεγονότα από το σετ IADS. Από τα ηχητικά αυτά γεγονότα, εξάχθηκε ένα πλήθος τεχνικών χαρακτηριστικών που, μαζί με την συναισθηματική επισήμανση των ηχητικών γεγονότων, χρησίμευσε ως είσοδος σε αλγόριθμους κατηγοριοποίησης. Ακόμα, η συναισθηματική επισήμανση των ηχητικών γεγονότων είχε κβαντιστεί σε τέσσερις κλάσεις. Μία για τα ηχητικά γεγονότα που είχαν υψηλή Διέγερση και υψηλή Ευχαρίστηση, μία δεύτερη για αυτά που είχαν υψηλή Διέγερση και χαμηλή Ευχαρίστηση, μία τρίτη για αυτά που είχαν χαμηλή Διέγερση και υψηλή Ευχαρίστηση και μία τελευταία για όσα είχαν χαμηλή Διέγερση και χαμηλή Ευχαρίστηση. Ως υψηλή ηχητικά γεγονότα με Ευχαρίστηση ή Διέγερση χαρακτηρίστηκαν εκείνα που είχαν τιμή συναισθηματικής επισήμανσης για την Ευχαρίστηση ή την Διέγερση, αντιστοίχως, πάνω από 5. Η τιμή 5 επιλέχθηκε λόγω του ότι είναι η μέση τιμή της κλίμακας επισήμανσης, όπως φαίνεται σε προηγούμενο Κεφάλαιο.

Σύμφωνα με τον **Πίνακα 4.2**, η ευστοχία (accuracy) όλων των αλγορίθμων που χρησιμοποιήθηκαν είναι κάτω από το όριο της τύχης, δηλαδή μικρότερη από 50%. Συγκεκριμένα, ο αλγόριθμος SVM είχε ευστοχία αναγνώρισης 43.7% ενώ τα τεχνικά νευρωνικά δίκτυα 36.5%. Έτσι παρατηρείται μία αδυναμία κατηγοριοποίησης των ηχητικών γεγονότων, ως προς την ορθή τους συναισθηματική κλάση, με αυτή την επιλογή τεχνικών χαρακτηριστικών. Αυτό το γεγονός μπορεί να έχει δύο αιτιολογήσεις: είτε τα τεχνικά χαρακτηριστικά που χρησιμοποιήθηκαν δεν ήταν κατάλληλα, με αποτέλεσμα να αποκλείονται χρήσιμα δεδομένα για την διαδικασία κατηγοριοποίησης, είτε το σημασιολογικό περιεχόμενο των ήχων έχει πολύ ισχυρή επίδραση πάνω στην συναισθηματική επισήμανση των ηχητικών γεγονότων, κάτι που έρχεται σε αντίθεση με προηγούμενα ευρήματα, όπως αυτά στην αμέσως προηγούμενη ενότητα. Δηλαδή, στην αμέσως προηγούμενη ενότητα γίνεται φανερό ότι σε ακραίες συναισθηματικές καταστάσεις το σημασιολογικό περιεχόμενο δεν έχει ισχυρή επίδραση πάνω

στον εγειρόμενο συναίσθημα του ακροατή. Το τελευταίο όμως μπορεί να στηριχθεί και από τη σχέση ενέργειας και Διέγερσης που φαίνεται στο **Σχήμα 4.1**, καθώς και στον **Πίνακα 4.3**. Συγκεκριμένα, στο προαναφερθέν σχήμα φαίνεται με γραφική απεικόνιση ότι αφενός οι μέγιστες τιμές τις ενέργειας αντιστοιχούν σε τιμές Διέγερσης ανάμεσα στο 6 και το 8, οι οποίες δεν είναι υψηλές, ενώ στην περιοχή Διέγερσης με τιμές στο διάστημα (6, 7) παρατηρείται μία πτώση της ενέργειας. Ακόμα, στον **Πίνακα 4.3** φαίνεται ότι η ενέργεια δεν ακολουθεί την ίδια αυξητική τάση όπως η διέγερση.

Όμως, η παραπάνω και, αρχικά, εμφανής σχέση σημασιολογικού περιεχομένου η οποία δείχνει να υπερτερεί της επίρειας της ενέργειας στην διέγερση φαίνεται να είναι σε αντίθεση με ευρήματα από προηγούμενες και ήδη δημοσιευμένες εργασίες, αλλά και με την εμπειρική και κοινή γνώση ότι η ενέργεια, δηλαδή το πόσο δυνατά ακούγεται ένα ηχητικό ερέθισμα, μπορεί να επηρεάσει την διέγερση [28]. Αυτό γεγονός καταδεικνύει εμφανώς ότι ναι μεν η ενέργεια μπορεί να έχει ισχυρή επίρεια στην διέγερση, αλλά υπάρχει ένα συγκεκριμένο χρονικό πλαίσιο μέσα στο οποίο θα πρέπει να αναζητηθεί η τιμή της. Έτσι, η εύρεση αυτής της βέλτιστης, εάν υπάρχει, χρονικής διάρκειας των πλαισίων θα μπορούσε να επαυξήσει την διαδικασία εξόρυξης συναισθηματικής πληροφορίας από ηχητικά γεγονότα.

Ακόμα, θα πρέπει να σημειωθεί πως τα τεχνικά χαρακτηριστικά που χρησιμοποιήθηκαν στην παρούσα εργασία δεν καλύπτουν μεγάλο φάσμα των ιδιοτήτων ενός ηχητικού γεγονότος. Έτσι, αν και υπάρχουν τεχνικά χαρακτηριστικά από όλες τις κατηγορίες (όπως αυτά παρουσιάστηκαν στον **Πίνακα 2.2**), αυτά φαίνεται ότι δεν είναι ικανά να μπορούν να περιγράψουν κατάλληλα και πλήρως, βάσει της διαδικασίας αναγνώρισης συναισθήματος, το ηχητικό γεγονός. Συγκεκριμένα, και σύμφωνα με τον **Πίνακα 3.5**, χρησιμοποιήθηκαν μόνο 2 χαρακτηριστικά σχετικά με τον ρυθμό, τα beat spectrum και beat strength. Τα τελευταία, αν και ανήκουν στα χαρακτηριστικά που σχετίζονται με τον ρυθμό ενός ηχητικού σήματος, δεν μπορούν να παρέχουν ευρεία γκάμα πληροφοριών σχετικά με αυτόν γιατί εστιάζουν στην περιοδικότητα και την ένταση του ρυθμού. Έτσι, δεν υπάρχει πληροφορία για το τέμπο, την διακύμανση, την καθαρότητα παλμών κτλ. Ακόμα, δεν έχουν συμπεριληφθεί στοιχεία σχετικά με την τονικότητα ή μεγέθη όπως η τραχύτητα.

Επιπροσθέτως, σύμφωνα με το **Σχήμα 3.5**, υπήρξε ένας διαχωρισμός και αντιστοίχιση των περιοχών του επιπέδου Διέγερσης – Ευχαρίστησης στις τέσσερις κλάσεις που αναφέρθηκαν στην αρχή αυτής της ενότητας. Έτσι, ενδέχεται το μικρό ποσοστό ευστοχίας των αλγόριθ-

μων να επηρεάστηκε από την κατάτμηση σε συναισθηματικές καταστάσεις και την επιλογή των τεχνικών χαρακτηριστικών. Έτσι, αν και για τις τέσσερις κλάσεις του συγκεκριμένου προβλήματος κατηγοριοποίησης, δηλαδή την αναγνώριση σε ποια από τις τέσσερις κλάσεις ανήκει ένα ηχητικό γεγονός, υπήρξε χαμηλή δυνατότητα κατηγοριοποίησης, ενδέχεται με τα ίδια τεχνικά χαρακτηριστικά αλλά με λιγότερες κλάσεις να επιτυγχάνονταν αυξημένα ποσοστά ευστοχίας.

Εν κατακλείδι, αν και τα ποσοστά ευστοχίας στην προαναφερθείσα μελέτη ήταν αρκετά χαμηλά, τα αποτελέσματα εν γένει αποκάλυψαν πτυχές του προβλήματος για μετέπειτα έρευνα. Από αυτές, οι πιο αξιοσημείωτες είναι:

1. Περαιτέρω διερεύνηση της πιθανής επίρειας του σημασιολογικού περιεχομένου στο συναίσθημα. Μία πρώτη ματιά στα παραπάνω δεδομένα και ανάλυση παρέχει κάποιες ενδείξεις ότι το σημασιολογικό περιεχόμενο μπορεί να έχει κάποια επίρεια πάνω στο εγείρομενο συναίσθημα του ακροατή. Έτσι, και σε συνδυασμό με τα ευρήματα και την ανάλυση της Ενότητας 5.1, θα πρέπει να ερευνηθεί εάν το καθαυτό σημασιολογικό περιεχόμενο είναι η βασική αιτία που προκαλεί συναισθήματα στον ακροατή ή εάν, λόγω της σύνδεσης συγκεκριμένων τεχνικών χαρακτηριστικών με πληροφορίες που μπορούν να εγείρουν συναισθήματα, έχει γίνει φυσική επιλογή για την ηχητική παρουσίαση αντίστοιχου σημασιολογικού περιεχομένου με ηχητικά ερεθίσματα που να έχουν τα προαναφερθέντα τεχνικά χαρακτηριστικά. Δηλαδή, επειδή κάποια κάποια τεχνικά χαρακτηριστικά εγείρουν συγκεκριμένα συναισθήματα, τότε έχει γίνει χρήση αυτής της αντιστοίχισης για την χρήση παρόμοιων ήχων με σημασιολογικό περιεχόμενο που να ταιριάζει με το συναίσθημα που εγείρουν αυτοί οι ήχοι. Για παράδειγμα, το κλάμα ή η έντονη και επιθετική φωνή έχουν τιμές τεχνικών χαρακτηριστικών τέτοιες ώστε να εγείρουν τα επιθυμητά συναισθήματα σε έκαστη περίπτωση.
2. Προσδιορισμός της βέλτιστης χρονικής διάρκειας πλαισίου για την τεχνική ανάλυση των ηχητικών δεδομένων. Όπως έγινε φανερό από την παραπάνω ανάλυση, αρκετά από τα ευρήματα μπορούν να οφείλονται σε μη κατάλληλη επιλογή χρονικής διάρκειας του πλαισίου που χρησιμοποιήθηκε για την κατάτμηση των ηχητικών γεγονότων. Έτσι, θα πρέπει να γίνει περαιτέρω διερεύνηση σχετικά με το ποια χρονική διάρκεια πλαισίου θα πρέπει να χρησιμοποιηθεί έτσι ώστε να αποδοθεί κατάλληλα η διακύμανση των τεχνικών χαρακτηριστικών κατά την χρονική εξέλιξη ενός ηχητικού γεγονότος.

3. Επιλογή κατάλληλων τεχνικών χαρακτηριστικών για την ποσοτική περιγραφή του ηχητικού υλικού. Σύμφωνα με τα παραπάνω, υπήρχε πληροφορία στο σήμα η οποία δεν είχε αποτυπωθεί πλήρως βάσει των τεχνικών χαρακτηριστικών που επιλέχθηκαν να εξαχθούν από τα ηχητικά δεδομένα. Έτσι, ενδέχεται μία διαφορετική τεχνική προσέγγιση να έχει καλύτερα αποτελέσματα. Δηλαδή, είτε ένα μεγαλύτερο σετ τεχνικών χαρακτηριστικών ή περισσότερη εστίαση σε συγκεκριμένες ποιότητες του σήματος, όπως ο ρυθμός.

5.3 Συσχέτιση ρυθμικών χαρακτηριστικών με διέγερση

Με εφιαλτήριο τις πτυχές για μετέπειτα έρευνα που αποκαλύφθηκαν από την ανάλυση της προηγούμενης ενότητας, το επόμενο πείραμα εστίασε στην συσχέτιση των ρυθμικών χαρακτηριστικών μόνο με την Διέγερση. Έτσι, διενεργήθηκε ένα πείραμα Μηχανικής Μάθησης με το ίδιο σετ ηχητικών δεδομένων του προηγούμενου πειράματος, ήτοι το σετ IADS, από το οποίο εξήχθησαν τεχνικά χαρακτηριστικά που σχετίζονται αποκλειστικά με τον ρυθμό. Η συναισθηματική επισήμανση των δεδομένων κατηγοριοποιήθηκε σε δύο κλάσεις, όπου η μία περιείχε τα ηχητικά δεδομένα με επισήμανση Διέγερσης κάτω από 5 και η άλλη αυτά με επισήμανση μεγαλύτερη ή ίση του 5. Τα τεχνικά χαρακτηριστικά μαζί με την κλάση χρησίμευσαν ως είσοδο σε αλγόριθμους κατηγοριοποίησης με επιβλεπόμενη μάθηση των οποίων οι επιδόσεις παρατέθηκαν στον προηγούμενο Κεφάλαιο.

Η ανάλυση και συζήτηση των αποτελεσμάτων αυτού του πειράματος έχει δύο μέρη, ήτοι: α) την αξιολόγηση των τεχνικών χαρακτηριστικών ως προς την πληροφορία που περιέχουν για το συγκεκριμένο πρόβλημα Μηχανικής Μάθησης (δηλαδή, την αναγνώριση Διέγερσης βάσει του ρυθμού), και β) την αξιολόγηση των αποτελεσμάτων ευστοχίας των αλγορίθμων. Σχετικά με το πρώτο, μπορεί να παρατηρηθεί από τον **Πίνακα 3.10** ότι τα τεχνικά χαρακτηριστικά δεν παρουσιάζουν συσχέτιση μεταξύ τους και έτσι θεωρήθηκε χρήσιμο να συμμετάσχουν όλα τα τεχνικά χαρακτηριστικά στην διαδικασία εξόρυξης συναισθηματικής πληροφορίας. Έτσι, ένα εύλογο ερώτημα θα μπορούσε να ήταν κατά πόσο υπάρχουν χαρακτηριστικά τα οποία συνεισφέρουν περισσότερο στην επήρεια της Διέγερσης και, εάν υπάρχουν, ποια είναι αυτά;

Πίνακας 5.1: Τα πρώτα δεκατρία χαρακτηριστικά κατά την αξιολόγησή τους στο πείραμα συσχέτισης τεχνικών ρυθμικών χαρακτηριστικών και διέγερσης

Δείκτης	Όνομα	Δείκτης	Όνομα
A	beatspectrumstd	K	fluctuationstd
B	eventdensitystd	L	fluctuationskewness
D	onsetsgradient	M	onsetskewness
E	fluctuationkurtosis	R	pulseclaritykurtosis
F	beatspectrumgradient	S	eventdensitykurtosis
I	pulseclaritystd	W	onsetsmean
J	fluctuationmean		

Σε αυτό το ερώτημα θα μπορούσε να απαντήσει μία λεπτομερής ανάλυση του **Πίνακα 4.4**, η οποία αποκαλύπτει πως τα τεχνικά χαρακτηριστικά χωρίζονται σε δύο ομάδες οι οποίες παραμένουν οι ίδιες ανεξαρτήτως των άλλων παραμέτρων (π.χ. ο αλγόριθμος αξιολόγησης). Η μία ομάδα που περιέχει τα πρώτα δεκατρία χαρακτηριστικά και μία ακόμα που περιέχει όλα τα υπόλοιπα. Τα τεχνικά χαρακτηριστικά που συμμετέχουν στις ομάδες δεν αλλάζουν, ανεξαρτήτως της μεθόδου αξιολόγησης της πληροφορίας τους αλλά και την διάρκεια του χρονικού πλαισίου που χρησιμοποιήθηκε κατά την εξαγωγή τους. Έτσι, σε όλες τις περιπτώσεις αξιολόγησης της πληροφορίας των τεχνικών χαρακτηριστικών, ως πρώτα δεκατρία είναι πάντα τα ίδια χαρακτηριστικά. Αυτά φαίνονται στον **Πίνακα 5.1**.

Σύμφωνα με την ενότητα που περιγράφει το συγκεκριμένο πείραμα, στην διαδικασία κατηγοριοποίησης δεν χρησιμοποιήθηκαν οι τιμές των τεχνικών χαρακτηριστικών καθαυτές, οι οποίες είχαν εξαχθεί από τα χρονικά πλαίσια του αντίστοιχου σήματος, αλλά στατικά μεγέθη αυτών των τιμών αυτών. Έτσι, η μέση τιμή ενός χαρακτηριστικού αφορά την μέση από τις τιμές του χαρακτηριστικού αυτού για όλα τα χρονικά πλαίσια του σήματος από το οποίο εξήχθη. Στον **Πίνακα 5.1** μπορεί να φανεί πως από τα συνολικά 13 χαρακτηριστικά, μόνο τα 2 αφορούν την μέση τιμή (mean) τεχνικού χαρακτηριστικού. Τα υπόλοιπα 11 αφορούν είτε την τυπική απόκλιση (std), είτε την κλίση (gradient) είτε τις δύο στατιστικές ροπές που χρησιμοποιήθηκαν (κύρτωση και ασυμμετρία). Αυτό το γεγονός φανερώνει σε ένα καθαρά τεχνικό επίπεδο, πως περισσότερη πληροφορία σχετικά με την αναγνώριση Διέγερσης από ρυθμικά χαρακτηριστικά μπορεί να δώσει η εξέλιξη και διακύμανση των τιμών τεχνικών

χαρακτηριστικών και όχι μία συνολική, μέση τιμή. Πιο συγκεκριμένα, παρατηρείται ότι 4 χαρακτηριστικά από τον **Πίνακα 5.1**, αυτά με δείκτη A , B , I και K , αφορούν την τυπική απόκλιση του μεγέθους, δίνοντας έτσι την ένδειξη ότι σε ένα μέγεθος η ευρύτητα της απόκλισης από μία κεντρική τιμή μπορεί να δώσει πολύτιμη πληροφορία στην εξόρυξη συναισθηματικής πληροφορίας. Ακόμα, η κύρτωση, δηλαδή η σχέση των τιμών του μεγέθους στα χρονικά άκρα (αρχή και τέλος) και στην χρονική μέση του ηχητικού γεγονότος, φαίνεται ότι μπορεί να προσφέρει σημαντική πληροφορία σχετικά με την αναγνώριση Διέγερσης, λόγω της σχετικά μεγαλύτερης εμφάνισής της, από τα υπόλοιπα στατιστικά μεγέθη, στον **Πίνακα 5.1** όπου εμφανίζεται 3 φορές και είναι τα μεγέθη με δείκτη E , $\bullet R$ και S . Τέλος, στον προαναφερθέν Πίνακα υπάρχουν δύο καταχωρήσεις σχετικά με την κλίση και μία σχετικά με την ασυμμετρία. Αυτές με δείκτη D και F για την κλίση και L για την ασυμμετρία. Αυτό αναδεικνύει πως η μείωση ή η αύξηση ενός μεγέθους αλλά και το πότε, χρονικά, εμφανίστηκε η μέγιστη τιμή μπορούν να επηρεάσουν επίσης την διαδικασία αναγνώρισης συναισθήματος από ηχητικά γεγονότα.

Από μία μη τεχνική σκοπιά, το προαναφερθέν γεγονός αναδεικνύει πως ο ακροατής επηρεάζεται από διακυμάνσεις των χαρακτηριστικών του ηχητικού του περιβάλλοντος και όχι από σταθερές τιμές. Έτσι, δεν είναι οι καθαυτές τιμές των τεχνικών χαρακτηριστικών οι οποίες γίνονται αντιληπτές από τον άνθρωπο και του επικοινωνούν συναισθηματική πληροφορία αλλά η διακύμανσή των τιμών αυτών. Αυτό δείχνει καθαρά πως μπορεί δύο ηχητικά γεγονότα να έχουν διαφορετικές τιμές για τα τεχνικά χαρακτηριστικά αλλά, λόγω της εξέλιξης του ηχητικού γεγονότος, να έχουν την ίδια συναισθηματική επίδραση στον ακροατή. Το τελευταίο αποτέλεσμα είναι πολύ σημαντικό μιας και δείχνει ότι η συναισθηματική επίδραση ή η συναισθηματική αλληλεπίδραση του ανθρώπου με το ηχητικό του περιβάλλον εξαρτάται από την χρονική εξέλιξη του δεύτερου και όχι από την αρχική ή την σταθερή του κατάσταση. Η ισχύς των παραπάνω συμπερασμάτων επαυξάνεται δεδομένου ότι οι ήχοι IADS έχουν σημασιολογικό περιεχόμενο. Έτσι, φαίνεται πως αφενός υπάρχουν τεχνικά χαρακτηριστικά τα οποία μπορούν να υπερβούν τις διαφορές στο σημασιολογικό περιεχόμενο κατά την διαδικασία αναγνώρισης Διέγερσης και αφετέρου, πως ο ρυθμός ενός ηχητικού ερεθίσματος με την διέγερση συνδέονται πιο ισχυρά με τα τεχνικά χαρακτηριστικά του πρώτου παρά με το σημασιολογικό του περιεχόμενο.

Σχετικά με τα αποτελέσματα ευστοχίας των αλγόριθμων κατηγοριοποίησης μπορεί να πα-

ρατηρηθεί πως όλα ήταν αρκετά πάνω από το όριο της τύχης, ήτοι 50%. Συγκεκριμένα, το ελάχιστο ήταν 71.26% ενώ το μέγιστο 88.37%. Το δεύτερο αποτέλεσμα αναδεικνύει την ισχύ της πεποίθησης ότι ο ρυθμός ενός ήχου μπορεί να επηρεάσει την συναισθηματική κατάσταση του ακροατή, και συγκεκριμένα την διέγερση, ενώ ισχυροποιεί την υπόθεση για την ύπαρξη καθολικών τεχνικών χαρακτηριστικών που σχετίζονται με τον ρυθμό και μπορούν να παρέχουν υψηλά ποσοστά ευστοχίας στην αναγνώριση Διέγερσης από ηχητικά δεδομένα. Πιο αναλυτικά, φαίνεται ότι ο αλγόριθμος *LR* παρέχει την μεγαλύτερη ευστοχία ανεξαρτήτως της χρονικής διάρκειας των πλαισίων που χρησιμοποιήθηκαν και του πλήθους αναδιπλώσεων. Η ελάχιστη ευστοχία με αυτό τον αλγόριθμο ήταν 81.44%, η οποία επιτεύχθηκε με χρονική διάρκεια πλαισίου 1 δευτερόλεπτο και πλήθος αναδιπλώσεων 3, και η μέγιστη 88.37%, με χρονική διάρκεια πλαισίου επίσης 1 δευτερόλεπτο και χρησιμοποιώντας την *LOO* τεχνική. Ο δεύτερος σε κατάταξη αλγόριθμος είναι ο *KNN* ενώ ο τρίτος ο *ANS*. Σχετικά με τον αλγόριθμο *KNN*, η ελάχιστη ευστοχία του φαίνεται ότι είναι 82.05%, για χρονική διάρκεια πλαισίου 0.8 δευτερόλεπτα και με πλήθος αναδιπλώσεων 3, ενώ η μέγιστη 85.63%, με διάφορες χρονικές διάρκειες πλαισίων και πλήθος αναδιπλώσεων. Τέλος, ο *ANS* είχε ελάχιστη ευστοχία 71.26%, με χρονική διάρκεια πλαισίου 1.4 δευτερόλεπτα και πλήθος αναδιπλώσεων 3, και μέγιστη 84.57%, με χρονική διάρκεια πλαισίου 1.8 δευτερόλεπτα και με την τεχνική *LOO*. Αν και δεν έχει δημοσιευθεί εργασία που να εστιάζει αποκλειστικά στην συσχέτιση των ρυθμικών χαρακτηριστικών ενός μουσικού κομματιού και την Διέγερση, τα παραπάνω αποτελέσματα ευστοχίας αναγνώρισης συναισθήματος από ηχητικά γεγονότα, συγκριτικά με αυτά που αναφέρονται στην διεθνή βιβλιογραφία για την αναγνώριση συναισθήματος από μουσική και παρατίθενται στο Κεφάλαιο 2, είναι αρκετά υψηλά. Έτσι, στο Κεφάλαιο 2 παρουσιάζονται μέγιστα ποσοστά ευστοχίας αναγνώρισης συναισθήματος από μουσική της τάξης του 88%, σχεδόν ίσα με αυτά του συγκεκριμένου πειράματος.

5.4 Χωρική θέση ηχητικών πηγών και συναισθηματική κατάσταση του ακροατή

Σύμφωνα με τον ορισμό του ηχητικού γεγονότος εντός του μοντέλου της Συναισθηματικής Ακουστικής Οικολογίας, το ηχητικό γεγονός μπορεί να μεταφέρει, εντός των άλλων, πληροφορία σχετικά με χαρακτηριστικά της πηγής. Η χωρική τοποθέτηση της πηγής αποτελεί ένα

τέτοιο χαρακτηριστικό, μιας και εγγενώς η ακουστική πηγή έχει μία φυσική υπόσταση και, ως εκ τούτου, μία χωρική τοποθέτηση σχετικά με τον ακροατή. Από υφιστάμενες εργασίες, σύμφωνα με το Κεφάλαιο 2, έχει δειχθεί μία σχέση της θέσης της πηγής και του συναισθήματος του φόβου. Όμως, στις εργασίες που εστίασαν σε αυτή την σχέση, δηλαδή θέσης πηγής και συναισθήματος, δεν αναλύεται και δεν αναφέρεται κάποια αναλυτική συσχέτιση της θέσης της πηγής και του συναισθήματος του ακροατή [99]. Σε όλες τις εργασίες γίνεται είτε εστίαση μόνο σε δύο συγκεκριμένες θέσεις (π.χ. μπροστά και πίσω από τον ακροατή) ή/και μόνο σε ένα συγκεκριμένο συναίσθημα, π.χ. “Φόβος” [99]. Στην παρούσα διατριβή ερευνήθηκε η επίρεια της γωνιακής θέσης της πηγής σχετικά με τον ακροατή. Για το σκοπό αυτό, έγινε το πείραμα που αναλύθηκε και παρουσιάστηκε στο Κεφάλαιο 3, όπου ήταν η διενέργεια πειράματος ακρόασης με υποκειμενική αξιολόγηση του συναισθήματος για χωρικά τοποθετημένα ηχητικά γεγονότα. Ως δεδομένα χρησιμοποιήθηκαν πέντε σετ αμφιωτικά επεξεργασμένων ηχητικών γεγονότων, όπου ως πρωτογενές υλικό (δηλαδή, τα δεδομένα που επεξεργάστηκαν με την αμφιωτική τεχνολογία) ήταν το σετ IADS. Τα πέντε σετ αυτά ανταποκρινόντουσαν στις εξής γωνίες γύρω από τον ακροατή: α) 0 μοίρες, β) 45 μοίρες, γ) 90 μοίρες, δ) 135 μοίρες, και ε) 180 μοίρες. Τα ηχητικά γεγονότα για κάθε σετ επισημάνθηκαν για την Διέγερση και την Ευχαρίστηση που εγείρανε στον ακροατή και τα αποτελέσματα παρουσιάστηκαν στο Κεφάλαιο 4.

Μία περαιτέρω ανάλυση και παρατήρηση των τιμών στον **Πίνακα 4.7** μπορεί να αποκαλύψει ότι η μέγιστη διαφορά των τιμών της επισήμανσης της Διέγερσης μεταξύ των ήχων IADS και τον αμφιωτικών ηχητικών γεγονότων είναι 2.47 ενώ για την Ευχαρίστηση η αντίστοιχη διαφορά είναι 2.07. Από την άλλη όμως, η μέγιστη απόλυτη διαφορά των τιμών της επισήμανσης για την Διέγερση και για την Ευχαρίστηση είναι 0.89 και 0.85 αντιστοίχως. Το γεγονός αυτό δείχνει ότι η μέγιστη διαφορά της επισήμανσης που είχαν οι πρωταρχικοί ήχοι IADS και οι ήχοι με χωρική τοποθέτηση 0° είναι λιγότερο από μία θέση στην μέθοδο SAM. Δεδομένου ότι το σετ IADS περιείχε μονοφωνικά ηχητικά γεγονότα και, άρα, η χωρική του τοποθέτηση ήταν ακριβώς μπροστά από τον ακροατή, οι προαναφερθείσες διαφορές στις τιμές των επισημάνσεων αναδεικνύουν ότι η αμφιωτική επεξεργασία δεν αλλοίωσε τα ηχητικά γεγονότα με τέτοιο τρόπο που να μεταβάλλει την συναισθηματική τους επιρροή στον ακροατή. Επίσης, εστιάζοντας στις μέσες τιμές που φαίνονται στον προαναφερθέντα Πίνακα, φαίνεται ότι οι ήχοι με χωρική τοποθέτηση εγείρουν μεγαλύτερη διέγερση και μικρότερη ευχαρίστηση στον ακροατή, χωρίς να υπάρχει μεγάλη απόκλιση των τιμών των επισημάνσεων, για την

Διέγερση και την Ευχαρίστηση, κατά μέσο όρο. Όμως, παρά την μικρή διαφορά μεταξύ των παρουσιαζόμενων αποτελεσμάτων και των ήχων IADS αναφορικά με τις μέσες τιμές και από τον **Πίνακα 4.7** μπορεί να παρατηρηθεί πως υπάρχουν ηχητικά γεγονότα τα οποία παρουσιάζουν διαφορά επισήμανσης πάνω από 1 βαθμίδα στο μοντέλο SAM. Δεδομένης της απουσίας εγγενούς (από τα ηχητικά δεδομένα του σετ IADS) χωρικής τοποθέτησης καθώς η συναισθηματική επισήμανση των ήχων IADS έγινε με την χρήση μονοφωνικών αρχείων τα οποία αναπαραγότουσαν από ηχεία, το προαναφερθέν γεγονός δείχνει πως οι νέες τιμές των επισημάνσεων με την χωρική τοποθέτηση των γεγονότων είναι διορθωμένες σε σχέση με αυτών των IADS. Δηλαδή, η αναπαραγωγή των IADS έγινε με την χρήση ηχείων, τα οποία προσδίδουν μία χωρική τοποθέτηση στην πηγή, ήτοι του σημείο που είναι το ηχείο. Αναλόγως την θέση του ακροατή σε σχέση με το ηχείο, η αντίληψη της θέσης της πηγής μπορεί να αλλάξει. Δηλαδή, ένας ακροατής που βρίσκεται μπροστά ακριβώς από το ηχείο, αντιλαμβάνεται την πηγή μπροστά του ενώ, απεναντίας, κάποιος ακροατής που βρίσκεται σε μία πλάγια θέση σε σχέση με το ηχείο, αντιλαμβάνεται την χωρική τοποθέτηση της πηγής στις πλάγιες, προς αυτόν, θέσεις. Με την υλοποίηση των πειραμάτων στα πλαίσια της παρούσας διατριβής και την σταθερή και συγκεκριμένη χωρική τοποθέτηση της πηγής για όλους τους ακροατές, μέσω της χρήσης της αμφιωτικής τεχνολογίας και των ακουστικών, φαίνεται ότι ενώ οι τιμές επισήμανσης κατά μέσο όρο δεν έχουν μεγάλες αποκλίσεις, μερικά ηχητικά γεγονότα παρουσιάζουν διαφορά στην επισήμανση πάνω από μία κλίμακα (ήτοι, μία φιγούρα) στο μοντέλο SAM. Όπου, αφενός βάσει της τεκμηρίωσης του μοντέλου SAM [102] και, αφετέρου, της χρήσης του για την συναισθηματική επισήμανση των ήχων στο σετ IADS [103], η διαφορά μίας θέσης καταδεικνύει μία αντιληπτή, από τον ακροατή, διαφορά στην συναισθηματική κατάσταση. Έτσι, ενώ εν γένει δεν υπάρχει μεγάλη διαφοροποίηση των τιμών, μερικά ηχητικά γεγονότα έχουν αναπροσαρμοσμένες τιμές επισήμανσης. Άρα, δεδομένης της χωρικής τοποθέτησης που χρησιμοποιήθηκε και στις δύο διαδικασίες επισήμανσης (της διατριβής και των IADS), για τους ήχους που έχουν $|D_a(i')| > 1$ και $|D_v(i')| > 1$ μπορεί να θεωρηθεί ότι οι νέες τιμές ανταποκρίνονται καλύτερα στην πραγματικότητα μιας και σε αυτές υπάρχει και η επίρεια από την χωρική τοποθέτηση της πηγής. Αυτός ο λόγος ήταν και η παραγωγή του σετ ήχων BEADS, το οποίο πλέον μπορεί να χρησιμεύσει, πέρα από ένα ακόμα σετ συναισθηματικά επισημασμένων ηχητικών γεγονότων όπου συμβάλλει στην αύξηση του πλήθους των διαθέσιμων σετ συναισθηματικά επισημασμένων ηχητικών γεγονότων στα 3, ως ένα σετ ηχητικών γεγονότων των οποίων οι συναισθηματικές επισημάνσεις

περιλαμβάνουν και την επίρεια της χωρικής τοποθέτησης της πηγής.

Όμως, υπάρχει το ερώτημα κατά πόσο οι τιμές αυτές είναι όντως αντιπροσωπευτικές και δεν εμπίπτουν στο τυχαίο. Δηλαδή, δεδομένου του πλήθους συμμετεχόντων στο πείραμα που διεξήχθη με την χρήση του σετ δεδομένων BEADS και του πλήθους συμμετεχόντων στην επισήμανση των ήχων IADS, υπάρχει η πιθανότητα τα δύο διαφορετικά σετ επισημάνσεων να συγκλίνουν. Αυτό θα σήμαινε πως οι προαναφερθείσες διαφορές στις τιμές οφείλονται απλά στο διαφορετικό πλήθος συμμετεχόντων. Το ερώτημα αυτό εξετάστηκε με την διεξαγωγή στατιστικών τεστ τα οποία είχαν σκοπό να εξετάσουν την στατιστική σημαντικότητα της διαφοράς των τιμών συναισθηματικής επισήμανσης που ελήφθησαν. Δηλαδή, να εξετάσουν κατά πόσο είναι σημαντικά διαφορετικές οι τιμές της επισήμανσης που λήφθηκαν από τα τεστ ακρόασης που έγιναν για την δημιουργία του BEADS με τις αντίστοιχες τιμές που έχουν οι ήχοι IADS. Εάν υπάρχει σημαντική διαφορά τότε αυτό σημαίνει πως οι τιμές συναισθηματικής επισήμανσης του σετ BEADS είναι στην πραγματικότητα διαφορετικές από αυτές στο σετ IADS και δεν θα συγκλίνουν με αυτές με την διεξαγωγή του πειράματος από επιπλέον συμμετέχοντες. Το μέγεθος που δείχνει την σημαντικότητα της διαφοράς λέγεται στατιστική σημαντικότητα (statistical significance), συμβολίζεται με το γράμμα p και δείχνει την πιθανότητα να παρατηρηθούν όμοιες τιμές (μεταξύ των δύο συνόλων) όταν μία υπόθεση (π.χ. οι δύο τιμές έχουν ίση μέση τιμή) είναι αληθινή. Το αντίστοιχο τεστ που επιλέχθηκε ονομάζεται εξέταση υπόθεσης (hypothesis testing).

Η βασική υπόθεση ορίστηκε ως εξής: είναι η μέση τιμή των συναισθηματικών επισημάνσεων που λήφθηκαν από τα διαδικτυακά τεστ ακρόασης ίδια με την αντίστοιχη τιμή επισήμανσης από το σετ IADS;. Για αυτό το σκοπό, και επειδή δεν είναι διαθέσιμο όλο το σύνολο των επισημάνσεων από τους IADS αλλά μόνο οι μέσες τιμές και η τυπική απόκλιση, έγιναν μονόπλευρα (one-sided) z-tests. Σε αυτά, ελήφθησαν υπόψη οι τιμές από τα ηχητικά γεγονότα για τα οποία βρέθηκε ότι $|D_a(i')| > 1$ και $|D_v(i')| > 1$. Τα αποτελέσματα του z-test μαζί με τις τιμές των $|D_a(i')|$ και $|D_v(i')|$ φαίνονται στον **Πίνακα 5.2**.

Πίνακας 5.2 – Συνέχεια σε επόμενη σελίδα

Πίνακας 5.2 – Συνέχεια από προηγούμενη σελίδα

i'	Σημαιολογικό περιεχόμενο	$D_a(i')$	p	i'	Σημαιολογικό περιεχόμενο	$D_v(i')$	p
$ D_a(i') > 1$				$ D_v(i') > 1$			

Πίνακας 5.2: Σημαιολογικό περιεχόμενο, τιμές των $|D_a(i')|$ και $|D_v(i')|$ και του p από τα z-tests για τους ήχους με $|D_a(i')| > 1$ και $|D_v(i')| > 1$

i'	Σημαιολογικό περιεχόμενο	$D_a(i')$	p	i'	Σημαιολογικό περιεχόμενο	$D_v(i')$	p
$ D_a(i') > 1$				$ D_v(i') > 1$			
1	Γάτα	-2.09	0.00	2	Σκύλος	1.67	0.02
2	Σκύλος	2.25	0.00	3	Μωρό	1.07	0.18
4	Παιδιά 1	-1.04	0.23	9	Τροπικό δάσος	-1.63	0.06
7	Κροαταλίας	2.53	0.00	21	Μάχη 2	-1.08	0.26
8	Κοκκινολαίμης	-1.24	0.15	24	Ερπυσμός	-1.79	0.00
9	Τροπικό δάσος	1.65	0.05	27	Ιπποδρομία	1.55	0.01
10	Γυναικείοι ήχοι συνουσίας 1	1.18	0.01	28	Βάψιμο	1.21	0.00
16	Ζευγάρι που φταρνίζεται	1.48	0.06	29	Νεροχύτης	1.31	0.01
22	Μάχη 3	1.58	0.01	30	Βροχή 1	1.02	0.05
31	Ελικόπτερο 1	-1.99	0.00	31	Ελικόπτερο 1	1.75	0.00
35	Συντριβή αεροπλάνου	1.37	0.03	32	Αντίστροφη μέτρηση	1.46	0.01
36	Βλάβη μηχανής	1.99	0.01	38	Καταιγίδα	1.79	0.07
37	Ατύχημα δικύκλου	1.11	0.15	40	Τηλέφωνο 1	2.13	0.00
38	Τηνδεροστομ	-2.83	0.00	43	Ξυπνητήρι	1.86	0.03
41	Ρολόι	-1.42	0.01	44	Κουλοχέρης (παιχνίδι)	1.99	0.00
43	Ξυπνητήρι	-1.45	0.02	47	Άρπα	1.88	0.00
45	Περπάτημα	2.09	0.00	48	Απόσπασμα έργου του Bach	1.03	0.07

Πίνακας 5.2 – Συνέχεια σε επόμενη σελίδα

Πίνακας 5.2 – Συνέχεια από προηγούμενη σελίδα

i'	Σημαιολογικό περιεχόμενο	$D_a(i')$	p	i'	Σημαιολογικό περιεχόμενο	$D_v(i')$	p
		$ D_a(i') > 1$				$ D_v(i') > 1$	
49	Χορωδία	-1.14	0.24	49	Χορωδία	1.76	0.01
	Πλήθος ηχητικών γεγονότων				Πλήθος ηχητικών γεγονότων		
			18				18

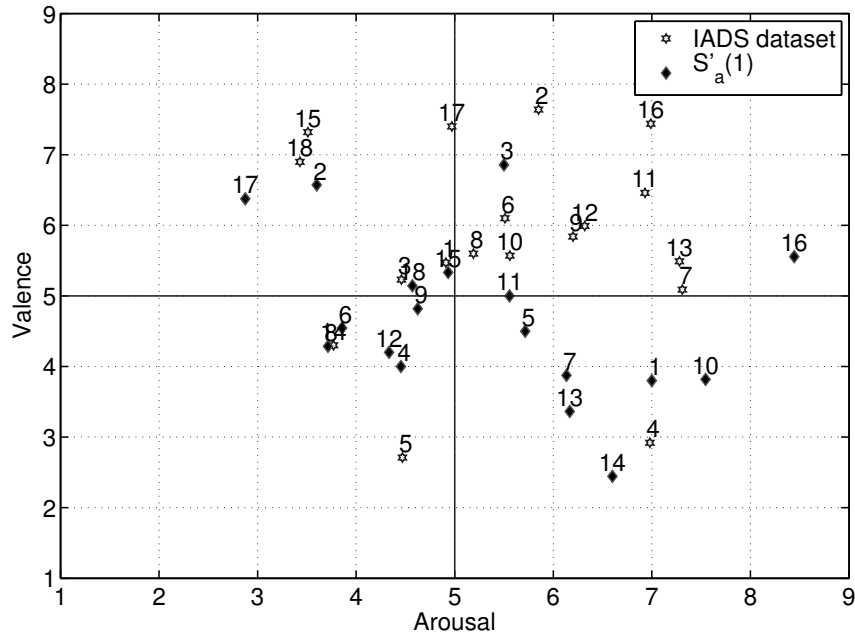
Όπως μπορεί να φανεί από τον **Πίνακα 5.2**, το 55.55% (10 από τα 18 συνολικά) ηχητικά γεγονότα έχουν στατιστικά σημαντικές διαφορές στις τιμές επισήμανσης για κάθε συναισθηματική διάσταση. Συγκεκριμένα, η τιμή του p για αυτά τα ηχητικά γεγονότα είναι κατά το μέγιστο 0.01. Επίσης, το 72.22% (13 από 18) έχει τιμή του $p \leq 0.05$, ενώ τιμή του $p \leq 0.1$ παρατηρείται για το 77.77% (14 από 18) και 83.33% (15 από 18) των ηχητικών γεγονότων. Έτσι, μπορεί να συμπεράνει κανείς με βεβαιότητα ότι οι τιμές για την συναισθηματική επισήμανση που λήφθηκαν από τα πειράματα είναι τελικά διαφορετικές από αυτές των ήχων IADS.

Σχετικά με τις καθαυτές τιμές της Διέγερσης, μπορεί να παρατηρηθεί πως μόνο 4 ηχητικά γεγονότα, αυτά με $i' = 1, 2, 7$ και 38, έχουν τιμή που σηματοδοτεί επόμενη ή προηγούμενη φιγούρα στο μοντέλο SAM. Υπενθυμίζοντας πως το μοντέλο SAM έχει εκτός από τις φιγούρες και ενδιάμεσες θέσεις, η διαφορά αυτή σημαίνει δύο θέσεις, πριν ή μετά, στο μοντέλο SAM. Τα υπόλοιπα ηχητικά γεγονότα διαφέρουν κατά το πολύ μία θέση Διέγερσης, και όχι φιγούρα, στο μοντέλο αυτό. Επιπροσθέτως, το σηματολογικό περιεχόμενο των ηχητικών γεγονότων στον **Πίνακα 5.2** δεν παρουσιάζει κάποια συστηματικότητα, εφόσον αυτό διαφέρει άρδην μεταξύ των ηχητικών γεγονότων και περιλαμβάνει μία γκάμα από ήχους ζώνων μέχρι μηχανικούς ήχους και μουσική. Έτσι, οι προαναφερθείσες διαφορές δεν μπορούν να αιτιολογηθούν από πιθανή επήρεια του σηματολογικού περιεχομένου. Συνεπώς, η μόνη επήρεια στην διέγερση φαίνεται να είναι η χωρική τοποθέτηση της πηγής. Εστιάζοντας στην ευχαρίστηση, οι μέσες τιμές των $D_v(i')$ και $|D_v(i')|$ είναι 0.38 και 0.84 αντιστοίχως. 18 ηχητικά γεγονότα παρουσιάζουν τιμή για το $|D_v(i')| \geq 1$ ενώ μόνο 1 έχει τιμή $|D_v(i')| \geq 2$. Έτσι, μόνο ένα εξ αυτών έχει επισημανθεί με διαφορά τριών θέσεων στο μοντέλο SAM και 17 με διαφορά μίας θέσης. Σημειώνεται δε ότι και στην περίπτωση αυτή, το σηματολογικό περιεχόμενο των ήχων παρουσιάζει κάποια συστηματικότητα.

Συνοψίζοντας την παραπάνω ανάλυση, οι τιμές της συναισθηματικής επισήμανσης για την διέγερση και την ευχαρίστηση που λήφθηκαν κατά το πείραμα, παρουσιάζουν μία σχετική συμφωνία με αυτές των IADS για την γωνία $\theta(1) = 0^\circ$. Η εξέταση της υπόθεσης για το εάν η διαφορά των τιμών που ελήφθησαν ανήκει απλά στην διαφορά του πλήθους των συμμετεχόντων για τα τεστ ακρόασης έδειξε ότι η τελευταία δεν οφείλεται σε στατιστικές αποκλίσεις λόγω του μεγέθους του δείγματος. Έτσι, εφόσον η μόνη αλλαγή στα ηχητικά δεδομένα ήταν η εφαρμογή της χωρικής θέσης της πηγής, τότε τα παραπάνω δείχνουν καθαρά πως οι νέες τιμές οφείλονται αποκλειστικά στην χωρική τοποθέτηση της πηγής στον χώρο μπροστά από τον ακροατή. Με μία πρώτη ματιά, αυτό το γεγονός δείχνει πως οι επισημάνσεις των ήχων IADS μπορούν να αναπροσαρμοστούν, έτσι ώστε να ανταποκρίνονται ορθότερα σε μία πραγματική κατάσταση, διότι όλες οι πηγές έχουν μία θέση στο φυσικό κόσμο. Αυτή την θέση την αντιλαμβάνεται ο ακροατής. Όμως, με μία δεύτερη ματιά μπορεί να φανεί ότι μερικές πηγές στον **Πίνακα 5.2** μπορεί να θεωρηθούν ως σημειακές, αλλά άλλες όχι. Για παράδειγμα, τα ηχητικά γεγονότα με $i' = 9, 30$ και 38 δεν μπορούν να θεωρηθούν ότι προέρχονται από σημειακές πηγές, καθώς το σημασιολογικό τους περιεχόμενο είναι αυτό του ηχοτοπίου, π.χ. “Τροπικό δάσος” για το ηχητικό γεγονός με $i' = 9$, “Βροχή” για τον ήχο με $i' = 30$ και “Καταιγίδα” για τον ήχο με $i' = 38$. Ως εκ τούτου, μία χωρική τοποθέτηση των ηχητικών γεγονότων αυτών αντιβαίνει σε μία πραγματική κατάσταση. Συνεπώς, για αυτά τα ηχητικά γεγονότα, τα οποία δεν αντιστοιχούν σε πηγές με συγκεκριμένη χωρική τοποθέτηση, η παραπάνω θεώρηση της αναπροσαρμογής των επισημάνσεων των ήχων IADS δεν ισχύει.

Όμως, οι παραπάνω διαφορές μπορούν να καταλήξουν σε εντελώς διαφορετικές συναισθηματικές καταστάσεις που περιγράφονται από τις επισημάνσεις; Η απάντηση σε αυτή την ερώτηση είναι πολύ σημαντική μιας και σηματοδοτεί το επίπεδο της διαφορετικότητας των τιμών για $\theta(1) = 0^\circ$. Για να δοθεί η παραπάνω απάντηση έγινε μία αποτύπωση των τιμών των συναισθηματικών επισημάνσεων στο επίπεδο Διέγερσης – Ευχαρίστησης, η οποία φαίνεται στο **Σχήμα 5.1** έτσι ώστε να μελετηθούν καλύτερα οι τιμές των συναισθηματικών επισημάνσεων που δόθηκαν προηγουμένως στον **Πίνακα 5.2**.

Από το **Σχήμα 5.1** μπορεί να φανεί ότι μόνο 4 ηχητικά γεγονότα, με δείκτη i' ίσον με 1, 2, 28 και 38, αλλάζουν τεταρτημόριο και στις δύο συναισθηματικές διαστάσεις. Η αλλαγή αυτή είναι σημαντική μιας και λόγω της έλλειψης μιας ποσοτικής αντιστοιχίας των τιμών



Σχήμα 5.1: Η αποτύπωση των συναισθηματικών επισημάνσεων, στο επίπεδο Διέγερσης (Arousal) – Ευχαρίστησης (Valence), για τα ηχητικά γεγονότα με χωρική τοποθέτηση και $\theta(1) = 0^\circ$

της Διέγερσης και της Ευχαρίστησης με διάκριτα συναισθήματα, και βάσει της υφιστάμενης βιβλιογραφίας έτσι όπως παρουσιάζεται στο αντίστοιχο κεφάλαιο, σηματοδοτεί σημαντική αλλαγή στην συναισθηματική κατάσταση. Για παράδειγμα, όλες οι δημοσιευμένες εργασίες που έχουν αντιστοιχίσει διακριτά συναισθήματα στο συνεχές επίπεδο Διέγερσης – Ευχαρίστησης, έχουν διαφορετική λεκτική περιγραφή συναισθήματος για διαφορετικό τεταρτημόριο του επιπέδου αυτού. Επιπροσθέτως, 7 ηχητικά γεγονότα ($i' = 4, 7, 8, 9, 16, 22$ και 36) έχουν τιμές Διέγερσης σε διαφορετικό τεταρτημόριο και 5 ($i' = 27, 29, 30, 31$ και 40) τιμές Ευχαρίστησης σε άλλο τεταρτημόριο. Συνεπώς, από συνολικά 30 διαφορετικές συναισθηματικές επισημάνσεις ηχητικών γεγονότων που εμφανίζονται στον **Πίνακα 5.2** μόνο 16 αντιστοιχούν σε εμφανή αλλαγή της συναισθηματικής κατάστασης του ακροατή. Έτσι, το γεγονός της εκδήλωσης της χωρικής τοποθέτησης δείχνει να αλλάζει, ως ένα βαθμό, το αντιλαμβανόμενο συναίσθημα από τον ακροατή αλλά, παράλληλα, προσδίδει στις λαμβανόμενες συναισθηματικές επισημάνσεις μία πιο αληθινή υπόσταση λόγω του ότι επικοινωνείται η, φυσική τους, χωρική τοποθέτηση σε σχέση με τον ακροατή.

Όμως, η παραπάνω ανάλυση είναι για μία συγκεκριμένη θέση της πηγής στον χώρο και σχετικά με τον ακροατή, ήτοι ακριβώς μπροστά του. Σε ένα πραγματικό σενάριο, η πηγή είναι

αρκετά πιθανό είτε να μετακινείται ή να είναι σταθερή αλλά σε μία άλλη θέση, δηλαδή είτε στα πλάγια του ακροατή είτε, ακόμα, όπισθεν αυτού. Συνεχίζοντας και εστιάζοντας στις τιμές που λήφθηκαν για τις υπόλοιπες γωνίες, μία λεπτομερής εξέταση του **Πίνακα 4.8** μπορεί να αποκαλύψει ότι υπάρχει μία συστηματικότητα στην διακύμανση των τιμών. Συγκεκριμένα, υπάρχουν τρεις κλάσεις ηχητικών γεγονότων αναφορικά με την αύξηση του k , δηλαδή με την αύξηση της γωνίας από την ακριβώς μπροστινή θέση σε σχέση με τον ακροατή:

1. Κλάση 1 (C_1), στην οποία ανήκουν τα ηχητικά γεγονότα τα οποία έχουν επισημανθεί με μεγαλύτερη διέγερση και μικρότερη ευχαρίστηση κατά την μετακίνηση της πηγής περιφερειακά του ακροατή και από μπροστά προς τα πίσω του
2. Κλάση 2 (C_2), η οποία περιέχει τα ηχητικά γεγονότα τα οποία έχουν επισημανθεί με μικρότερη διέγερση και μεγαλύτερη ευχαρίστηση κατά την ίδια κίνηση της πηγής με την κλάση C_1
3. Κλάση 3 (C_3), στην οποία περιέχονται τα ηχητικά γεγονότα που δεν ανήκουν στις δύο παραπάνω κλάσεις

Στον **Πίνακα 5.3** είναι τα ηχητικά γεγονότα των δύο πρώτων κλάσεων μαζί με το σημασιολογικό τους περιεχόμενο, την γωνία $\phi(i', k)$, που σχηματίζει το διάνυσμα \vec{v}_a και το μέτρο των διανυσμάτων $|\vec{v}_a(i', k)|$, τα οποία έχουν ως αρχικό σημείο την συνισταμένη των επισημάνσεων για την Διέγερση και την Ευχαρίστηση στην γωνία $k - 1$ και ως τελικό σημείο την συνισταμένη των συναισθηματικών επισημάνσεων για την γωνία k . Προφανώς, υπάρχουν ηχητικά γεγονότα τα οποία δεν μπορούν να αποδοθούν σε μία από τις C_1 και C_2 , διότι η συμπεριφορά τους, αναφορικά με την μεταβολή των τιμών επισημάνσεως, μεταβάλλεται και δεν έχει την συστηματικότητα που απαιτείται για την ένταξη σε αυτές τις δύο κλάσεις. Στο **Σχήμα 5.2** απεικονίζεται το ποσοστό των ηχητικών γεγονότων που ανήκουν σε μία εκ των C_1 και C_2 , καθώς επίσης και στην ένωση των C_1 και C_2 , σε σχέση με το σύνολο των ηχητικών γεγονότων στο πείραμα (ήτοι, 50), καθώς η θέση της πηγής μετατοπίζεται προς την πίσω μεριά του ακροατή.

Πίνακας 5.3: Τα ηχητικά γεγονότα στις κλάσεις C_1 και C_2 για όλες τις γωνίες $\theta(k)$

i'	Σημαιολογικό περιεχόμενο	$\phi(i', k)$	$ \vec{v}\hat{a}(i', k) $	i'	Σημαιολογικό περιεχόμενο	$\phi(i', k)$	$ \vec{v}\hat{a}(i', k) $
Κλάση 1				Κλάση 2			
k=2 (γωνιακή μετάβαση από 0° σε 45°)							
9	Τροπικό δάσος	-78.24	1.44	3	Μωρό	144.58	1.31
21	Μάχη 2	-60.95	1.14	5	Μέλισσες	126.05	1.18
22	Μάχη 3	-23.20	1.90	7	Κροταλίας	136.90	1.04
35	Συντριβή αεροπλά- νου	-28.07	1.89	8	Κοκκινολαίμης	174.68	2.60
45	Περπάτημα	-1.71	0.76	10	Γυναικείοι ερωτικοί ήχοι 1	179.06	1.36
50	Ηλεκτρισμός	-20.62	2.06	15	Άντρας που βήχει	143.51	1.44
				16	Ζευγάρι που φταρνί- ζεται	105.48	0.80
				19	Έμετος	121.15	1.52
				20	Σφύριγμα	175.36	0.26
				23	Θύμα	159.01	1.23
				26	Γράψιμο	175.96	1.11
				28	Βάψιμο	104.04	0.52
				30	Βροχή 1	168.02	0.34
				31	Ελικόπτερο 1	152.88	0.84
				36	Βλάβη μηχανής	162.65	2.00
				38	Καταιγίδα	171.87	0.94
				42	Κούκος	160.35	0.83
				43	Ξυπνητήρι	151.78	2.58
				44	Κουλοχέρης (παι- χνίδι) 2	171.23	3.36
				49	Χορωδία	102.76	0.78
k=3 (γωνιακή μετάβαση από 45° σε 90°)							
2	Σκύλος	-79.94	1.27	6	Κότες	173.93	0.53
3	Μωρό	-7.99	0.74	13	Άντρας που γελάει	93.84	1.02
5	Μέλισσες	-18.71	0.60	17	Άντρας που ασθμαί- νει	133.78	1.48
8	Κοκκινολαίμης	-36.43	1.09	20	Σφύριγμα	118.07	0.27
12	Αγόρι που γελάει	-60.95	0.74	21	Μάχη 2	150.86	1.44

Πίνακας 5.3 – Συνέχεια σε επόμενη σελίδα

Πίνακας 5.3 – Συνέχεια από προηγούμενη σελίδα

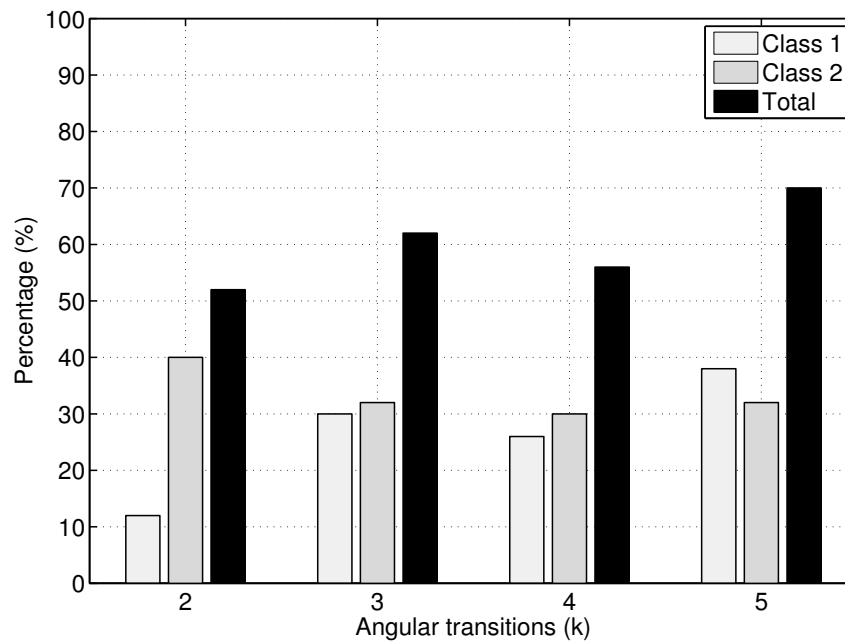
i'	Σημαιολογικό περιεχόμενο	$\phi(i', k)$	$ \vec{v}a(i', k) $	i'	Σημαιολογικό περιεχόμενο	$\phi(i', k)$	$ \vec{v}a(i', k) $
Κλάση 1				Κλάση 2			
15	Άντρας που βήχει	-71.57	1.36	27	Ιπποδρομία	180.00	0.64
25	Γραφομηχανή	-57.69	0.56	28	Βάψιμο	90.00	0.63
26	Γράψιμο	-35.33	1.49	29	Νεροχύτης	169.77	1.02
32	Αντίστροφη μέτρηση	-60.26	0.38	30	Βροχή 1	91.70	1.02
34	Άνεμος	-71.57	1.19	31	Ελικόπτερο 1	107.65	0.58
36	Βλάβη μηχανής	-31.86	2.71	33	Κόρνες αμαξιών	134.88	1.78
39	Έκρηξη	-27.31	1.96	35	Συντριβή αεροπλάνου	142.68	1.41
43	Ξυπνητήρι	-5.57	1.72	37	Ατύχημα δικύκλου	169.58	1.92
45	Περπάτημα	-71.57	0.29	40	Τηλέφωνο 1	144.46	0.22
49	Χορωδία	-17.20	2.20	41	Ρολόι	129.62	1.90
				50	Ηλεκτρισμός	135.00	1.68
k=4 (γωνιακή μετάβαση από 90° σε 135°)							
6	Κότες	0.00	1.92	8	Κοκκινολάιμης	142.79	0.59
12	Αγόρι που γελάει	-82.41	0.27	10	Γυναικείοι ερωτικοί ήχοι 1	135.00	0.31
19	Έμετος	-54.46	0.78	13	Άντρας που γελάει	95.71	0.71
20	Σφύριγμα	0.00	0.14	18	Άντρας που φταρνίζεται	106.77	1.48
24	Ερπυσμός	-55.30	0.28	25	Γραφομηχανή	136.23	1.00
29	Νεροχύτης	-35.80	1.07	26	Γράψιμο	122.35	1.33
30	Βροχή 1	-64.71	1.14	31	Ελικόπτερο 1	99.46	1.52
34	Άνεμος	-25.02	0.61	32	Αντίστροφη μέτρηση	121.26	2.08
35	Συντριβή αεροπλάνου	-77.47	0.69	33	Κόρνες αμαξιών	171.47	0.36
37	Ατύχημα δικύκλου	-22.09	1.73	36	Βλάβη μηχανής	164.05	1.32
41	Ρολόι	-33.69	1.56	38	Καταιγίδα	174.39	2.61
43	Ξυπνητήρι	-70.02	1.46	39	Έκρηξη	152.41	2.73
50	Ηλεκτρισμός	-45.00	0.47	40	Τηλέφωνο 1	134.47	1.10
				42	Κούκος	175.96	1.25
				47	Άρπα	180.00	2.11

Πίνακας 5.3 – Συνέχεια σε επόμενη σελίδα

Πίνακας 5.3 – Συνέχεια από προηγούμενη σελίδα

i'	Σημαιολογικό περιεχόμενο	$\phi(i', k)$	$ \vec{v}\hat{a}(i', k) $	i'	Σημαιολογικό περιεχόμενο	$\phi(i', k)$	$ \vec{v}\hat{a}(i', k) $
Κλάση 1				Κλάση 2			
k=5 (γωνιακή μετάβαση από 135° σε 180°)							
7	Κροταλίας	-17.44	1.86	2	Σκύλος	164.05	1.27
13	Άντρας που γελάει	-15.54	1.96	3	Μωρό	91.83	1.57
16	Ζευγάρι που φταρνίζεται	-14.62	0.28	5	Μέλισσες	176.19	2.15
17	Άντρας που ασθμαίνει	-87.99	0.52	6	Κότες	141.99	1.25
18	Άντρας που φταρνίζεται	-42.88	1.45	8	Κοκκινολαΐμης	141.51	1.22
20	Σφύριγμα	-66.64	0.62	19	Έμετος	141.67	0.52
21	Μάχη 2	-5.87	0.81	22	Μάχη 3	106.70	0.46
24	Ερπυσμός	-82.57	0.37	26	Γράψιμο	93.18	0.38
25	Γραφομηχανή	-6.30	1.56	30	Βροχή 1	172.57	0.97
28	Βάψιμο	-19.98	1.17	34	Άνεμος	129.92	1.95
29	Νεροχύτης	-15.22	2.34	35	Συντριβή αεροπλάνου	148.57	0.72
31	Ελικόπτερο 1	-71.39	2.59	37	Ατύχημα δικύκλου	172.98	2.48
32	Αντίστροφη μέτρηση	-38.66	2.05	43	Ξυπνητήρι	103.84	0.81
33	Κόρνες αμαξιών	-75.96	1.55	45	Περπάτημα	118.24	0.80
38	Καταιγίδα	-61.75	0.69	46	Άνοιγμα και σερβίρισμα ποτού	148.57	0.75
39	Έκρηξη	-20.03	1.48	50	Ηλεκτρισμός	108.43	0.35
41	Ρολόι	-60.26	1.07				
47	Άρπα	-30.26	1.54				
48	Απόσπασμα έργου του Bach	-90.00	0.70				

Όπως μπορεί να φανεί από το **Σχήμα 5.2** υπάρχει μία καθαρή τάση για αύξηση του πλήθους των γεγονότων που ανήκουν είτε στην κλάση C_1 , είτε στην C_2 , καθώς η πηγή μετατοπίζεται προς τα πίσω. Επιπροσθέτως, βάσει των αποτελεσμάτων του **Πίνακα 5.3**, το σημαιολογικό περιεχόμενο των ηχητικών γεγονότων δεν φαίνεται να έχει κάποια επίδραση σε αυτή



Σχήμα 5.2: Το ποσοστό των ηχητικών γεγονότων που ανήκουν στις κλάσεις C_1 , C_2 και $C_1 \cup C_2$

την τάση, μιας και είναι αρκετά μη συστηματικό και διαφέρει πολύ ανάμεσα στα ηχητικά γεγονότα. Στον **Πίνακα 5.4** απεικονίζονται τα στατιστικά στοιχεία (μέγιστο, ελάχιστο, μέση τιμή και τυπική απόκλιση) για την γωνία $\phi(i', k)$ και το $|\vec{v}\hat{a}(i', k)|$ αναφορικά με τις γωνιακές μεταβάσεις.

Πίνακας 5.4: Στατιστικά στοιχεία για την γωνία $|\phi(i', k)|$ και το μέτρο $|\vec{v}\hat{a}(i', k)|$ για τα ηχητικά γεγονότα που είναι στις C_1 και C_2 . Οι τιμές για την γωνία $|\phi(i', k)|$ είναι σε μοίρες ενώ για το μέτρο των διανυσμάτων σε θέσεις στο μοντέλο SAM

	Κλάση 1		Κλάση 2	
	$ \phi(i', k) $	$ \vec{v}\hat{a}(i', k) $	$ \phi(i', k) $	$ \vec{v}\hat{a}(i', k) $
k = 2 (γωνιακή μετάβαση 0° → 45°)				
Max	78.24	2.06	179.06	3.36
Min	1.71	0.76	102.76	0.26
Mean	35.46	1.53	149.37	1.30
σ	28.44	0.51	25.50	0.79
k = 3 (γωνιακή μετάβαση 45° → 90°)				
Max	79.94	2.71	180.00	1.92

Πίνακας 5.4 – Συνέχεια σε επόμενη σελίδα

Πίνακας 5.4 – Συνέχεια από προηγούμενη σελίδα

	Κλάση 1		Κλάση 2	
	$ \phi(i', \mathbf{k}) $	$ \vec{v}\hat{a}(i', \mathbf{k}) $	$ \phi(i', \mathbf{k}) $	$ \vec{v}\hat{a}(i', \mathbf{k}) $
Min	5.57	0.29	90.00	0.22
Mean	43.60	1.22	135.36	1.10
σ	25.35	0.70	29.44	0.58
k = 4 (γωνιακή μετάβαση 90° → 135°)				
Max	82.41	1.92	180.00	2.73
Min	0.00	0.14	95.71	0.31
Mean	43.54	0.93	140.82	1.37
σ	27.11	0.59	28.33	0.75
k = 5 (γωνιακή μετάβαση 135° → 180°)				
Max	90.00	2.59	176.19	2.48
Min	5.87	0.28	91.83	0.35
Mean	43.33	1.30	135.01	1.10
σ	29.67	0.68	28.63	0.65

Κατά την μετακίνηση της πηγής γύρω από τον ακροατή, για τα περισσότερα ηχητικά γεγονότα φαίνεται ότι οι τιμές συναισθηματικής επισήμανσης για την Διέγερση και την Ευχαρίστηση είναι αντιστρόφως ανάλογες. Δηλαδή, μείωση στην μία εκ των δύο συνοδεύεται από αύξηση στην άλλη. Αυτή η ένδειξη ισχυροποιείται περισσότερο κατά την μετακίνηση της πηγής στα όρια του οπτικού πεδίου του ακροατή και προς τα έξω από αυτό, δηλαδή για $k \geq 3$, και αντίστοιχες τιμές $\theta(k) \geq 90$. Για $k = 5$, δηλαδή για $\theta(k) = 180^\circ$, το ποσοστό των ηχητικών γεγονότων που ανήκει σε μία από τις κλάσεις 1 και 2 φτάνει το 70%. Όμως, γιατί τελικά είναι σημαντική αυτή η κατανομή σε κλάσεις; Βάσει του αρχικού διαστατικού μοντέλου που προτάθηκε από τους Posner, Russel και Peterson [93] αλλά και πλείστες δημοσιευμένες εργασίες που αναφέρθηκαν στο Κεφάλαιο 2, στο επίπεδο Διέγερσης – Ευχαρίστησης μπορούν να αντιστοιχηθούν κάποιες λεκτικές περιγραφές συναισθημάτων. Έτσι, η μεταβολή των τιμών των $|\vec{v}\hat{a}(i', k)|$ και $\phi(i', k)$, και συνεπώς των τιμών Διέγερσης και Ευχαρίστησης των ηχητικών γεγονότων, σημαίνει την αλλαγή και στην λεκτική περιγραφή που μπορεί να χρησιμοποιηθεί για να περιγράψει το εγείρομενο συναίσθημα. Στην περίπτωση των ηχητικών γεγονότων που

ανήκουν στις κλάσεις 1 και 2 αυτό σημαίνει μία προβλέψιμη συμπεριφορά της αλλαγής των συναισθημάτων. Επιπροσθέτως, θα πρέπει να συνυπολογιστεί ότι, και λαμβάνοντας υπόψη το **Σχήμα 2.9**, το γεγονός πως είναι σχετικά δύσκολο ο ακροατής να νιώθει δυσαρέσκεια κατά την ακρόαση ενός γενικού ήχου και ταυτόχρονα να μην νιώθει διέγερση.

Έτσι, βάσει της προηγούμενη ανάλυσης, καταλήγουμε ότι οι συναισθηματικές επισημάνσεις για τα ηχητικά γεγονότα των κλάσεων 1 και 2 έχουν την καθαρή τάση να μετακινούνται κατά την αύξηση της $\theta(k)$ προς συγκεκριμένες περιοχές του επιπέδου Διέγερσης – Ευχαρίστησης. Συγκεκριμένα, τα γεγονότα στην κλάση 1 έχουν την τάση να μετακινούνται προς το δεύτερο τεταρτημόριο ή να εγείρουν συναισθήματα που βρίσκονται σε θέση βρισκόμενη σε φορά αντίθετη προς την κατεύθυνση του ρολογιού. Από την άλλη, τα ηχητικά γεγονότα στην κλάση 2 έχουν την τάση να μετακινούνται προς συναισθήματα που τοποθετούνται στο τέταρτο τεταρτημόριο ή σε θέση βρισκόμενη σε φορά όπως του ρολογιού, όταν η $\theta(k)$ αυξάνεται.

Αναλυτικότερα και εστιάζοντας στην κλάση 1 και για όλες τις γωνιακές μεταβάσεις, μία αύξηση στην διέγερση συνοδεύεται από μία μείωση στην ευχαρίστηση. Καθώς αυξάνεται το k , άρα μεγαλώνει η γωνία $\theta(k)$, οι μέγιστες και ελάχιστες απόλυτες τιμές των γωνιών $\phi(i', k)$ αυξάνονται επίσης. Κατά την κίνηση της πηγής προς τις πλευρικές θέσεις, σε σχέση με τον ακροατή, η Ευχαρίστηση εμφανίζει μία αυξημένη επιρροή από την αλλαγή στην συναισθηματική κατάσταση έως και την μετάβαση από $45^\circ \rightarrow 90^\circ$. Από το σημείο αυτό και για μεγαλύτερες τιμές του k , η επίδραση της αλλαγής στην συναισθηματική κατάσταση εστιάζεται περισσότερο στην Διέγερση. Λόγω της έλλειψης μίας αναλυτικής και ποσοτικής αντιστοιχίας τιμών στο πεδίο Διέγερσης – Ευχαρίστησης με λεκτικές περιγραφές συναισθημάτων, δεν μπορεί να γίνει και κάποια περαιτέρω συσχέτιση των λαμβανόμενων τιμών επισήμανσης για την κλάση 1 και λεκτικών περιγραφών συναισθημάτων. Παρόλα αυτά, με την χρήση του μοντέλου SAM γίνεται δυνατή η αντιστοίχιση των μεταβολών των συναισθηματικών καταστάσεων στις επιλογές του μοντέλου αυτού. Βάσει των τιμών $\phi(i', k)$ και $|\vec{v}\vec{a}(i', k)|$, στον **Πίνακα 5.5** είναι οι μέσες τιμές των $|\vec{v}\vec{a}(i', k)|$ για τα γεγονότα που ανήκουν στην C_1 . Συνοψίζοντας την ανάλυση για τα γεγονότα στην κλάση αυτή, παρατηρείται ότι όταν η πηγή είναι εκτός των ορίων του οπτικού πεδίου του ακροατή επηρεάζεται η διέγερση του τελευταίου περισσότερο από την ευχαρίστηση. Επιπροσθέτως, οι όποιες αλλαγές στην συναισθηματική κατάσταση του ακροατή είναι προσανατολισμένες σε υψηλότερη διέγερση και χαμηλότερη ευχαρίστηση. Αυτό το γεγονός αντικατοπτρίζει συναισθήματα που τοποθετούνται στο δεύτερο τεταρτη-

Πίνακας 5.5: Μέσες τιμές των στοιχείων των διανυσμάτων ΔA για τις κλάσεις 1 και 2

	Διέγερση Ευχαρίστηση		Διέγερση Ευχαρίστηση	
	C_1		C_2	
$k = 2$	1.25	0.89	-1.12	0.66
$k = 3$	0.88	0.84	-0.78	0.77
$k = 4$	0.67	0.64	-1.06	0.87
$k = 5$	0.95	0.89	-0.78	0.78

μόριο του επίπεδου Διέγερσης – Ευχαρίστησης, ή κοντά στον άξονα της Ευχαρίστησης (για γεγονότα που βρίσκονται στο τρίτο ή τέταρτο τεταρτημόριο), είτε κοντά στις υψηλές περιοχές του άξονα της Διέγερσης (για γεγονότα που βρίσκονται στο πρώτο τεταρτημόριο). Έτσι, οι γενικοί ήχοι που ανήκουν στην κλάση 1 τείνουν να διεγείρουν τον ακροατή και να τον κάνουν να νιώθει λιγότερο ευχάριστα καθώς αυτοί μετακινούνται στο οπτικό του πεδίο προς τα άκρα του ή εκτός αυτού.

Τα ηχητικά γεγονότα που ανήκουν στην κλάση 2 παρουσιάζουν μία παρόμοια αλλά αντίθετη συμπεριφορά στο εγείρομενο συναίσθημα σε σχέση με αυτά στην κλάση 1. Συγκεκριμένα, τα ηχητικά γεγονότα στην κλάση 2 προσανατολίζονται στις ακριβώς αντίθετες περιοχές του επίπεδου Διέγερσης – Ευχαρίστησης. Στον **Πίνακα 5.5** φαίνονται επίσης οι μέσες τιμές των στοιχείων των διανυσμάτων \vec{v}_A κατά αντίστοιχο τρόπο με πριν. Από αυτόν τον Πίνακα φαίνεται καθαρά ότι η επίρεια των ηχητικών γεγονότων στην κλάση 2 προσανατολίζεται ελαφρά στην διέγερση. Αναλυτικότερα, τα διανύσματα \vec{v}_A για τα ηχητικά γεγονότα στην κλάση 2 παρουσιάζουν ένα μέγιστο της γωνίας $\phi(i', k)$ σχεδόν ίσο με 135° . Εν γένει όμως, δεν παρουσιάζουν κάποια αξιοσημείωτη επίρεια πάνω στην Διέγερση ή την Ευχαρίστηση. Αντιθέτως, η επίδραση στις δύο αυτές συναισθηματικές συνιστώσες φαίνεται ισοδύναμη, λόγω των μέσων τιμών που παρατίθενται στον **Πίνακα 5.5**, και παρουσιάζει μείωση καθώς η πηγή μετακινείται εκτός του οπτικού πεδίου του ακροατή. Επιπροσθέτως, τα ηχητικά γεγονότα στην κλάση C_2 τυγχάνουν επισημάνσεων που τα τοποθετούν σε κατάσταση υψηλής Ευχαρίστησης και χαμηλής Διέγερσης ή περιοχές που: α) είναι κοντά στον άξονα της Ευχαρίστησης όταν η αρχική επισήμανση, δηλαδή για $k = 1$, ήταν στο πρώτο τεταρτημόριο, και β) είναι κοντά στις χαμηλές τιμές του άξονα της Διέγερσης όταν η αρχική, $k = 1$, συναισθηματική επισήμανση βρισκόταν στο τρίτο τεταρτημόριο. Όμως, η αδυναμία εύρεσης αναλυτικής

και ποσοτικής συσχέτισης των τιμών στο επίπεδο Διέγερσης – Ευχαρίστησης δεν επιτρέπεται πάλι μία περαιτέρω ανάλυση. Όμως, αυτό που μπορεί να συναχθεί τελικά είναι ότι τα ηχητικά γεγονότα στην κλάση 2 αυξάνουν την ευχαρίστηση και μειώνουν την διέγερση του ακροατή.

Ανακεφαλαιώνοντας την παραπάνω ανάλυση, φαίνεται καθαρά ότι υπάρχουν ηχητικά γεγονότα που παρουσιάζουν μία συστημικότητα αναφορικά με την επήρεια της συναισθηματικής κατάστασης του ακροατή καθώς η πηγή μετακινείται από μπροστά του προς τα άκρα ή και εκτός του οπτικού πεδίου του. Φαίνονται δύο είδη της επήρειας αυτής: στο πρώτο είναι η αύξηση της Διέγερσης και η μείωση της Ευχαρίστησης και στο δεύτερο η μείωση της Διέγερσης και η αύξηση της Ευχαρίστησης. Το πλήθος των ηχητικών γεγονότων που υπακούουν είτε στο ένα είτε στο άλλο είδος επήρειας αυξάνεται καθώς η πηγή μετακινείται προς τις πλευρικές θέσεις του ακροατή ή/και από πίσω του. Επιπροσθέτως, όταν η πηγή μετακινείται ακριβώς πίσω από τον ακροατή το πλήθος των ηχητικών γεγονότων που προξενούν μεγαλύτερη διέγερση και μικρότερη ευχαρίστηση αυξάνεται και είναι μεγαλύτερο από αυτό των ηχητικών γεγονότων με την αντίθετη συμπεριφορά. Η μεταβολή των τιμών των διανυσμάτων \vec{v}_A φαίνεται σχετικά συστηματική, δεδομένου ότι η τυπική απόκλιση για την γωνία $|\phi(i', k)|$ (δηλαδή, την γωνία μεταξύ των \vec{v}_A και του άξονα της Διέγερσης) έχει ένα μέγιστο με όριο την τιμή 29.67° . Για την αποσαφήνιση του τελευταίου, θα πρέπει να αναφερθεί ότι μικρές τιμές της τυπικής απόκλισης για τη γωνία $|\phi(i', k)|$ δείχνουν ότι οι αλλαγές της Διέγερσης και της Ευχαρίστησης και για τις δύο κλάσεις προσανατολίζονται σε μία συγκεντρωμένη περιοχή. Για την πρώτη κλάση (C_1) στο δεύτερο τεταρτημόριο και για την δεύτερη (C_2) στο τέταρτο τεταρτημόριο του χώρου Διέγερσης – Ευχαρίστησης.

Όμως, ανάμεσα στα 50 ηχητικά γεγονότα που έχουν μελετηθεί, είναι και κάποια που παρουσιάζουν ένα άλλο ενδιαφέρον στοιχείο, κυρίως σε σχέση με το σημασιολογικό τους περιεχόμενο. Πρόκειται για τα ηχητικά γεγονότα με καθαρά μουσικό σημασιολογικό περιεχόμενο (είναι αυτά με i' 47, 48, 49). Οι μέσες τιμές των επισημάνσεων ανά γωνία $\theta(k)$ για την Διέγερση και την Ευχαρίστηση, για αυτά τα ηχητικά γεγονότα, φαίνονται στον **Πίνακα 5.6**. Όπως μπορεί να παρατηρηθεί, υπάρχει μία αξιοσημείωτη επήρεια πάνω στην συνιστώσα της Ευχαρίστησης, ειδικά όταν η πηγή μετακινείται προς τις πλευρικές θέσεις. Άρα, ο ακροατής φαίνεται να νιώθει πιο ευχαριστημένος όταν η πηγή φεύγει από μπροστά του και μετακινείται προς τα πλάγια.

Πίνακας 5.6: Οι μέσες τιμές επισημάνσεων για την Διέγερση, $\bar{a}(i', k)$, και την Ευχαρίστηση, $\bar{v}(i', k)$, για τα ηχητικά γεγονότα με μουσικό σημασιολογικό περιεχόμενο

Σημασιολογικό περιεχόμενο	i'	k				
		1	2	3	4	5
$\bar{a}(i', k)$						
Άρπα	47	2.78	3.85	4.78	2.67	4.00
Απόσπασμα έργου του Bach	48	5.00	5.31	4.75	5.50	5.86
Χορωδία	49	4.57	4.40	6.50	5.86	5.43
$\bar{v}(i', k)$						
Άρπα	47	5.56	7.39	7.45	7.45	6.67
Απόσπασμα έργου του Bach	48	6.38	7.62	6.37	6.80	6.10
Χορωδία	49	5.14	5.90	5.25	5.14	4.43

Η παρατήρηση αυτή παρουσιάζει ιδιαίτερο ενδιαφέρον. Πιο συγκεκριμένα, το μέγιστο της συνιστώσας της Ευχαρίστησης για όλες τις περιπτώσεις ηχητικών γεγονότων με μουσικό σημασιολογικό περιεχόμενο παρατηρείται στην γωνία των 45° . Αυτό το γεγονός είναι αρκετά πιθανό να συνδέεται με την ευρεία εφαρμοσμένη τεχνική αναπαραγωγής ήχου κατά την οποία προσδίδεται χωρική απόσταση μεταξύ των πηγών αναπαραγωγής μουσικής για την επαύξηση της εμπύθισης του ακροατή. Μία τεχνική που είναι ευρέως γνωστή ως στερεοφωνική αναπαραγωγή. Όταν όμως η πηγή μετακινείται στις 90° και πλέον (δηλαδή, $k \geq 3$), μία εφαρμοζόμενη τοποθέτηση πηγών αναπαραγωγής μουσικής, τα ηχητικά γεγονότα δεν παρουσιάζουν όλα την ίδια επήρεια στην συναισθηματική κατάσταση του ακροατή.

Επιπροσθέτως, για τιμές του k ίσες με 3 και 4, δηλαδή για γωνίες 90 και 135 μοιρών, τα ηχητικά γεγονότα που παρουσιάζουν μία αυξητική επίδραση στην ευχαρίστηση του ακροατή είναι αυτά με σημασιολογικό περιεχόμενο “Άρπα” και “Απόσπασμα έργου του Bach”. Την ίδια συμπεριφορά δεν έχει το ηχητικό γεγονός με σημασιολογικό περιεχόμενο “Χορωδία”. Αυτό το γεγονός είναι επίσης πιθανό να βασίζεται σε συνήθεις καταστάσεις ακρόασης μουσικής όπου σε μία τυπική διαρρύθμιση οι ανθρώπινες φωνές (δηλαδή το τραγούδι ή/και η χορωδία) βρίσκονται στις μπροστινές θέσεις σε σχέση με τον ακροατή. Έτσι, η τοποθέτηση του ηχητικού γεγονότος με $i' = 49$ σε πλευρικές θέσεις έχει μειωτική επίδραση στην Ευ-

χαρίστησή που διεγείρει στον ακροατή. Αυτή η πιθανή συσχέτιση με τις τυπικές συνθήκες ακρόασης μουσικής φαίνεται ότι ισχύει για όλα τα ηχητικά γεγονότα που φαίνονται στον **Πίνακα 5.6** και για όλες τις θέσεις εκτός για ακριβώς πίσω από αυτόν. Τέλος, η διέγερση δεν δείχνει κάποια συστηματική αλλαγή με την μετακίνηση της πηγής και για τα ηχητικά γεγονότα με σημασιολογικό περιεχόμενο μουσικής. Αυτό το γεγονός δείχνει πως η Διέγερση που εγείρεται μέσω της μουσικής δεν επηρεάζεται από την τοποθέτηση των πηγών αλλά από τεχνικά χαρακτηριστικά τα οποία αφορούν την κυματομορφή, σύμφωνα με τον ορισμό του ηχητικού γεγονότος εντός του πλαισίου της Συναισθηματικής Ακουστικής Οικολογίας.

5.5 Συσχέτισης ρυθμικών χαρακτηριστικών με ευχαρίστηση

Το τελευταίο πείραμα που διεξήχθη στην παρούσα διατριβή είναι η συσχέτιση των ρυθμικών χαρακτηριστικών ενός ηχητικού γεγονότος με την Ευχαρίστηση. Έχοντας ως δεδομένο την συσχέτιση του ρυθμού με την Διέγερση του ακροατή, η έρευνα πάνω στον εάν σχετίζεται και με την Ευχαρίστηση θα αποκαλύψει την πλήρη επίρεια του ρυθμού ενός ήχου πάνω στον ακροατή. Έτσι, διεξήχθη ένα πείραμα Μηχανικής Μάθησης χρησιμοποιώντας τα ηχητικά γεγονότα του σετ IADS. Από τα ηχητικά γεγονότα αυτά εξήχθησαν τεχνικά χαρακτηριστικά που σχετίζονται με τον ρυθμό, και είχαν χρησιμοποιηθεί και στο προηγούμενο πείραμα συσχέτισης του ρυθμού με την Διέγερση. Η διαδικασία εξαγωγής τεχνικών χαρακτηριστικών ακολούθησε την αντίστοιχη διαδικασία από το πείραμα συσχέτισης ρυθμικών χαρακτηριστικών και Διέγερσης. Τα τεχνικά χαρακτηριστικά αυτά μαζί με την συναισθηματική επισήμανση της Ευχαρίστησης και της Διέγερσης χρησιμοποιήθηκαν ως είσοδος σε αλγόριθμους κατηγοριοποίησης, οι οποίοι, ονομαστικά, ήταν οι: α) Decision Stump, β) C4.5, γ) SVM, δ) RepTree, ε) SMO, στ) AdaBoost, και η) Bagging. Τα αποτελέσματα είχαν ένα μέγιστο για το μέγεθος της ακρίβειας ίσο με 0.651, για την ανάκληση η μέγιστη τιμή ήταν 0.756 και για την ευστοχία 64.6%.

Η ανάλυση των αποτελεσμάτων της έρευνας σχετικά με την συσχέτιση των ρυθμικών χαρακτηριστικών με την συνιστώσα της Ευχαρίστησης θα κινηθεί σε δύο άξονες. Ο πρώτος αφορά τον καθαυτό στόχο της προαναφερθείσας μελέτης. Την συσχέτιση των ρυθμικών χα-

ρακτηριστικών με την Ευχαρίστηση. Ο δεύτερος αφορά την επίδοση και καταλληλότητα των τεχνικών και αλγορίθμων μηχανικής μάθησης που χρησιμοποιήθηκαν για τον συγκεκριμένο στόχο. Μία προσέγγιση αρκετά κοινή στις μελέτες εξόρυξης συναισθηματικής πληροφορίας από ηχητικά δεδομένα γιατί προσφέρει στοιχεία των τεχνικών και των αλγορίθμων έτσι ώστε αυτοί να μπορούν να αξιοποιηθούν σε μελλοντικές έρευνες. Εστιάζοντας στον δεύτερο άξονα, βάσει των αποτελεσμάτων που λήφθηκαν από τα πειράματα που έγιναν με τους αλγόριθμους, που αναφέρονται παραπάνω αλλά και στην Ενότητα 3.3.5, φαίνεται από τις τιμές των αποτελεσμάτων της ακρίβειας, της ανάκλησης, της ευστοχίας και του μέσου λάθους μία καθαρή υπεροχή των αλγορίθμων μετά-μάθησης σε σχέση με τους υπόλοιπους. Η αυξημένη επίδοση του αλγόριθμου μετά-μάθησης AdaBoost σε σχέση με τους υπόλοιπους, και δη με τον πιο εξελιγμένο από αυτούς, τον SMO, είναι εμφανή από τους **Πίνακες 4.9** και **4.10**, όπου για παραπλήσιες επιδόσεις ανάκλησης, ακρίβειας και ευστοχίας, ο AdaBoost έχει μικρότερο μέσο λάθος και ίσο με 0.515. Ανάμεσα στους υπόλοιπους αλγόριθμους, δηλαδή όχι τους μετά-μάθησης, ο SMO και ο RepTree υλοποιούν την διαδικασία αναγνώρισης της Ευχαρίστησης με μεγαλύτερη επιτυχία, φτάνοντας για την ακρίβεια, την ανάκληση και την ευστοχία σε τιμές ίσες με 0.642, 0.674 και 62.28 ο πρώτος και 0.648, 0.756 και 62.28 ο δεύτερος, αντιστοίχως. Η ιδανική χρονική διάρκεια παραθύρου φαίνεται ότι είναι ίση με 1.6 δευτερόλεπτα. Η διάρκεια αυτή, ενώ είναι λίγο λιγότερο από το ένα τρίτο της συνολικής διάρκειας των ηχητικών γεγονότων, φαίνεται ότι μπορεί να δώσει την κατάλληλη πληροφορία για την αναγνώριση Ευχαρίστησης από ρυθμικά χαρακτηριστικά. Επίσης, από τα αποτελέσματα που φαίνονται στους **Πίνακες 4.9** και **4.10** μπορεί να παρατηρηθεί ότι η κλάση C_1 , δηλαδή η κλάση με τα ηχητικά γεγονότα που έχουν τιμές συναισθηματικής επισήμανσης πάνω από 5, είναι δυσκολότερο να αναγνωριστεί από την κλάση C_2 , δηλαδή από αυτή που περιέχει ηχητικά γεγονότα που έχουν επισημανθεί με τιμή για την Ευχαρίστηση κάτω από 5. Έτσι, στους Πίνακες αυτούς φαίνεται ότι συνεχώς και για τους περισσότερους αλγόριθμους οι τιμές για τα μεγέθη της ακρίβειας, της ανάκλησης και της ευστοχίας η κλάση C_2 λαμβάνει μεγαλύτερες τιμές. Αυτό το γεγονός, δηλαδή η δυσκολία αναγνώρισης υψηλής Ευχαρίστησης σε σχέση με την χαμηλή, έχει παρατηρηθεί και από προηγούμενες δημοσιευμένες εργασίες [172]. Σε κάθε περίπτωση, σύμφωνα με τον **Πίνακα 4.10**, φαίνεται ότι η διαδικασία εξόρυξης της συναισθηματικής πληροφορίας επωφελείται σημαντικά από την χρήση της Διέγερσης ως χαρακτηριστικό, αποδεικνύοντας την έντονη σχέση μεταξύ της τελευταίας και της Ευχαρίστησης. Αυτό είναι ένα στοιχείο που έχει αναφερθεί και από τους δημιουργούς

του σετ ήχων IADS [103]. Τα αποτελέσματα με τον αλγόριθμο AdaBoost ξεπερνούν το 65% για την ανάκληση και την ακρίβεια για την κλάση C_1 , ενώ για την C_2 τα αντίστοιχα ποσοστά είναι πάνω από 60% και 63% αντιστοίχως. Παρόλα αυτά, παρατηρούνται ακόμα μεγαλύτερα ποσοστά αλλά με τον αλγόριθμο Decision Stump. Ένα γεγονός που οφείλεται σε πιθανές, και τυχαίες, συνθήκες που ευνοούν τα αποτελέσματα από αυτόν τον απλό αλγόριθμο. Έτσι, είναι σχεδόν απίθανο ότι παρόμοια επίδοση θα εμφανιστεί με τον Decision stump σε παρόμοια πειράματα αλλά διαφορετικά αριθμητικά δεδομένα.

Πιο συγκεκριμένα, μία πρώτη ματιά στους **Πίνακες 4.9, 4.10 και 4.11** μπορεί να αποκαλύψει ότι ο αλγόριθμος Decision stump παρουσιάζει την μέγιστη τιμή ευστοχίας και την ελάχιστη τιμή μέσου λάθους ενώ, από την άλλη, όλοι οι αλγόριθμοι παρουσιάζουν τις βέλτιστες τιμές τους στο εύρος χρονικής διάρκειας παραθύρου από 1.6 έως 1.8 δευτερόλεπτα. Συνυπολογίζοντας όμως την απλή φύση του αλγόριθμου αυτού, μία τέτοια επίδοση και δη καλύτερη από τους υπόλοιπους, πιο εξελιγμένους και πολύπλοκους αλγόριθμους μοιάζει αρκετά παράξενη. Επίσης, φαίνεται καθαρά πως οι εν γένει επιδόσεις των αλγορίθμων αυξάνουν με την ταυτόχρονη χρήση της Διέγερσης ως χαρακτηριστικό, βάσει των αποτελεσμάτων για το σετ ήχων $R''_{i,x}$ (το σετ ήχων όπου έγινε ταυτόχρονη χρήση και της Διέγερσης) και για όλα τις μετρικές αξιολόγησης απόδοσης, ήτοι ανάκληση, ακρίβεια, ευστοχία και μέσο λάθος. Αυτό δεν παρατηρείται για τον Decision stump, όπου η μέγιστη και ελάχιστη τιμή για την ευστοχία και το μέσο λάθος, αντίστοιχα είναι 65.27% και 0.477. Αυτό δικαιολογείται από το γεγονός ότι ο συγκεκριμένος αλγόριθμος χρησιμοποιεί έναν κανόνα/χαρακτηριστικό που επιλέγει και βάσει αυτού κατηγοριοποιεί τα δεδομένα. Έτσι, δεν παρατηρείται κάποια αλλαγή κατά την ταυτόχρονη θεώρηση και της Διέγερσης ως χαρακτηριστικό. Αν, και βάσει αυτού του αλγόριθμου, θα μπορούσε να βγει το συμπέρασμα ότι η κατηγοριοποίηση των ηχητικών γεγονότων εξαρτάται μόνο από ένα χαρακτηριστικό με αυτό να μην είναι η Διέγερση, μία παρατήρηση στις τιμές των μετρικών από τους άλλους, περισσότερο εξελιγμένους, αλγορίθμους θα δώσει διαφορετικό αποτέλεσμα. Επίσης, σε αυτό συνηγορεί και η φύση του αλγόριθμου αυτού που αναδεικνύει ότι τα αποτελέσματά του είναι σχεδόν σίγουρο ότι οφείλονται σε τυχαίες εξαρτήσεις των τιμών των τεχνικών χαρακτηριστικών που βοηθάνε στην τελική διαμόρφωση της επίδοσής του.

Έτσι, εστιάζοντας στους υπόλοιπους αλγόριθμους μπορεί να παρατηρηθεί ότι οι βέλτιστες τιμές των μετρικών αξιολόγησης απόδοσης λαμβάνονται όταν το χρονικό παράθυρο είναι με-

Πίνακας 5.7: Ακρίβεια και ανάκληση για τον αλγόριθμο RepTree και για χρονικό παράθυρο διάρκειας 1.6 δευτερολέπτων

	Κλάση C_1	Κλάση C_2
	Σετ ήχων $R'_{i,x'}$	
Ανάκληση	0.686	0.556
Ακρίβεια	0.621	0.625
	Σετ ήχων $R''_{i,x'}$	
Ανάκληση	0.721	0.531
Ακρίβεια	0.620	0.642

γέθους 1.6 και 1.8 δευτερολέπτων. Δηλαδή, για τα χρονικά παράθυρα με δείκτη $i = 5$ και 6. Για $i = 7$, δηλαδή για χρονικό παράθυρο μεγέθους 2 δευτερολέπτων, παρατηρείται μία πτώση στις επιδόσεις όλων των αλγορίθμων. Έτσι, ως βέλτιστος χρόνος, $t_{optimal}$, για τον συγκεκριμένο στόχο, δηλαδή αυτόν της αναγνώρισης Ευχαρίστησης από ρυθμικά τεχνικά χαρακτηριστικά, είναι στο χρονικό διάστημα δευτερολέπτων [1.4, 2]. Για το $t_{optimal}$ μπορεί να φανεί ότι οι αλγόριθμοι AdaBoost και RepTree έχουν τις καλύτερες επιδόσεις. Πιο συγκεκριμένα, για $i = 5$ ο αλγόριθμος RepTree έχει την μεγαλύτερη ευστοχία και το ελάχιστο μέσο λάθος και στα δύο σετ ήχων, δηλαδή $R'_{i,x}$ και $R''_{i,x}$, και από όλους τους άλλους αλγορίθμους με εξαιρούμενο τον Decision stump. Οι τιμές των μετρικών αυτών είναι 62.28% και 62.87% για την ευστοχία, για τα σετ $R'_{i,x}$ και $R''_{i,x}$, και για το μέσο λάθος 0.496 και 0.492, αντιστοίχως. Επίσης, για αυτό τον αλγόριθμο παρατηρείται ότι δεν υπήρξε αξιοσημείωτη βελτίωση με την θεώρηση της Διέγερσης ως χαρακτηριστικό, εφόσον η βελτίωση για το μέσο λάθος ήταν 0.007 και για την ευστοχία 0.59%. Αν και μία πρώτη άποψη θα μπορούσε να ήταν πως το παραπάνω γεγονός του μη επηρεασμού από την διέγερση έρχεται σε αντίθεση με τα ευρήματα από προηγούμενες έρευνες, δηλαδή ότι υπάρχει ισχυρή σύνδεση μεταξύ Διέγερσης και Ευχαρίστησης ιδιαίτερα κατά την ακρόαση ήχων που προκαλούν χαμηλή ευχαρίστηση [103], μία δεύτερη ματιά στους **Πίνακες 4.9** και **4.10** μπορεί να αποκαλύψει ότι οι τιμές της ακρίβειας και της ανάκλησης βελτιώθηκαν με την χρήση της Διέγερσης. Αυτές οι τιμές για χρονικό παράθυρο διάρκειας 1.6 δευτερολέπτων (δηλαδή, αυτό με δείκτη $i = 5$) συνοψίζονται στον **Πίνακα 5.7**.

Ακόμα, αναφορικά με το σετ $R'_{i,x}$, μπορεί να παρατηρηθεί ότι το ποσοστό των ηχητικών

γεγονότων της κλάσης C_2 που αναγνωρίζονται εσφαλμένα ότι δεν ανήκουν σε αυτή, είναι μεγαλύτερο από το ποσοστό των ηχητικών γεγονότων της κλάσης C_1 που επίσης δεν αναγνωρίζονται ως να ανήκουν στην C_1 . Δηλαδή, εάν ως FN_A συμβολίζεται το ποσοστό των ηχητικών γεγονότων που δεν αναγνωρίζονται ότι ανήκουν στην κλάση τους A , τότε ισχύει ότι $FN_{C_2} > FN_{C_1}$. Παρόλα αυτά, και για τις δύο κλάσεις σχεδόν το ίδιο ποσοστό, με διαφορά 0.004 στην ακρίβεια, γεγονότων έχει αναγνωριστεί από τον RepTree ότι ανήκει στη σωστή κλάση. Αναλύοντας την χρήση της Διέγερσης ως χαρακτηριστικό σε σχέση με την επίδοση του ίδιου αλγόριθμου, παρατηρείται ότι αυξάνεται η ανάκληση για την κλάση C_1 και μειώνεται για την C_2 . Δηλαδή μειώνεται το ποσοστό των ηχητικών γεγονότων που, ενώ ανήκουν στην κλάση C_1 , κατηγοριοποιούνται ότι ανήκουν στην κλάση C_2 και αυξάνεται το ποσοστό των ηχητικών γεγονότων που, ενώ ανήκουν στην κλάση C_2 , κατηγοριοποιούνται ότι ανήκουν στην C_1 . Ακόμα, με την χρήση της Διέγερσης ως χαρακτηριστικό μειώνεται η ακρίβεια για την κλάση C_2 . Συνεπώς, για $i = 5$ (χρονικό παράθυρο διάρκειας .16 δευτερολέπτων) η χρήση της Διέγερσης βελτιώνει την αναγνώριση των ηχητικών γεγονότων που έχουν χαμηλή Ευχαρίστηση (C_1) αλλά, παράλληλα, μειώνει την αναγνώριση των ηχητικών γεγονότων που ανήκουν στην κλάση της υψηλής Ευχαρίστησης (C_2). Έτσι, και για χρονική διάρκεια πλαισίου 1.6 δευτερολέπτων, υπάρχει μία καθαρή ένδειξη ότι τα ρυθμικά χαρακτηριστικά ενός ηχητικού γεγονότος μπορούν να συντελέσουν σε μία ορθή αναγνώριση της χαμηλής Ευχαρίστησης, η οποία μπορεί να επαυξηθεί από την ταυτόχρονη θεώρηση της Διέγερσης ως επιπλέον χαρακτηριστικό. Αυτό είναι σύμφωνα με την προαναφερθείσα σχέση χαμηλής Ευχαρίστησης και αυξημένης Διέγερσης που έχει αναφερθεί σε δημοσιευμένες εργασίες [103]. Παρόλα αυτά, η αναγνώριση των ηχητικών γεγονότων με επισήμανση χαμηλής Ευχαρίστησης παρουσιάζει δυσκολία με την αποκλειστική χρήση ρυθμικών τεχνικών χαρακτηριστικών.

Μία παρόμοια συμπεριφορά στην επίδοση της αναγνώρισης Ευχαρίστησης μπορεί να παρατηρηθεί και για $i = 6$ (χρονικό παράθυρο διάρκειας 1.8 δευτερολέπτων) αλλά για τους αλγόριθμους AdaBoost και SMO. Για αυτή την χρονική διάρκεια πλαισίου, δηλαδή 1.8 δευτερόλεπτα, οι δύο αυτοί αλγόριθμοι παρουσιάζουν τις βέλτιστες τιμές σε σχέση με τους υπόλοιπους, εξαιρουμένου του Decision stump. Οι τιμές αυτές για αυτούς τους δύο αλγόριθμους φαίνονται στον **Πίνακα 5.8**.

Αμφότεροι οι προαναφερθέντες αλγόριθμοι παρουσιάζουν τιμές ανάκλησης και ακρίβειας πά-

Πίνακας 5.8: Ακρίβεια και ανάκληση των αλγορίθμων AdaBoost και SMO για χρονικό παράθυρο διάρκειας 1.8 δευτερολέπτων ($i = 6$)

	Κλάση C_1		Κλάση C_2	
	AdaBoost	SMO	AdaBoost	SMO
Σετ ήχων $R'_{i,x'}$				
Ανάκληση	0.628	0.628	0.617	0.617
Ακρίβεια	0.635	0.635	0.610	0.610
Σετ ήχων $R''_{i,x'}$				
Ανάκληση	0.663	0.663	0.593	0.556
Ακρίβεια	0.633	0.613	0.623	0.608

νω από 60% και για τις δύο κλάσεις στο $R'_{i,x'}$. Αναλυτικότερα, ο AdaBoost επιτυγχάνει το υψηλότερο ποσοστό ανάκλησης, δηλαδή δεδομένης μίας κλάσης κατηγοριοποιεί το μεγαλύτερο ποσοστό ηχητικών γεγονότων αυτής της κλάσης ως να ανήκουν στην κλάση που είναι. Η θεώρηση της Διέγερσης ως επιπλέον χαρακτηριστικό οδηγεί σε μία αύξηση της ανάκλησης για την κλάση C_1 , αλλά σε μείωση αντίστοιχη για την κλάση C_2 , ενδυναμώνοντας την συσχέτιση Διέγερσης και Ευχαρίστησης. Όμως, κανένας από τους δύο αυτούς αλγόριθμους δεν υποσκελίζει τις επιδόσεις του αλγόριθμου RepTree με $i = 5$. Πιο συγκεκριμένα, αν και ο AdaBoost παρουσιάζει ισοδύναμη ευστοχία με τον RepTree, ο τελευταίος έχει καλύτερα αποτελέσματα σχετικά με τις τιμές μέσου λάθους. Έτσι, αναδεικνύεται το γεγονός ότι αλγόριθμοι που κάνουν χρήση δένδρων τείνουν να αποδίδουν καλύτερα από αυτούς που βασίζονται σε συναρτήσεις και αλγόριθμους μετά-μάθησης που χρησιμοποιούν βασικούς κατηγοριοποιητές με συναρτήσεις.

Συνοψίζοντας την παραπάνω ανάλυση, εστιάζοντας στην πιθανή συσχέτιση ρυθμού και Ευχαρίστησης και συνυπολογίζοντας ότι τα αποτελέσματα που παρατίθενται παραπάνω είναι τα ακριβέστερα που έχουν αναφερθεί στην διεθνή επιστημονική βιβλιογραφία για την αναγνώριση Ευχαρίστησης από ηχητικά γεγονότα, είναι φανερό ότι: α) η αναγνώριση της Ευχαρίστησης που διεγείρουν τα ηχητικά γεγονότα χρησιμοποιώντας αποκλειστικά ρυθμικά χαρακτηριστικά είναι εφικτή έως ένα σημείο και, ως εκ τούτου, η συναισθηματική κατάσταση της Ευχαρίστησης του ακροατή δείχνει επηρεάζεται και από τα ρυθμικά χαρακτηριστικά, σε αντίθετη περίπτωση τα αποτελέσματα από τους αλγόριθμους θα κυμαίνονταν στα όρια της

τύχης, β) η φαινόμενη βέλτιστη χρονική διάρκεια του χρονικού πλαισίου για την αλγοριθμική ανάλυση κατά την αναγνώριση Ευχαρίστησης από ρυθμικά χαρακτηριστικά είναι 1.6 δευτερόλεπτα. Τα παραπάνω καθίστανται περισσότερο εμφανή με τον συνυπολογισμό των τιμών της ανάκλησης, της ακρίβειας, της ευστοχίας και του μέσου λάθους για τους αλγόριθμους που χρησιμοποιήθηκαν. Όπως μπορεί να φανεί από την προηγούμενη διεξοδική ανάλυση, τα χαρακτηριστικά των ηχητικών γεγονότων φαίνεται να έχουν μία περιορισμένη επίδραση πάνω στην εγειρόμενη ευχαρίστηση η οποία αντανακλάται κιόλας στις, σχετικά, χαμηλές τιμές των μετρικών (ανάκληση, ακρίβεια κτλ). Αυτό το συμπέρασμα ισχυροποιείται βάσει προηγούμενων μελετών, όπως αυτή που παρατίθεται παραπάνω και στην Ενότητα 3.3.3, σχετικά με την σχέση ρυθμικών χαρακτηριστικών και Διέγερσης όπου χρησιμοποιήθηκαν τα ίδια τεχνικά χαρακτηριστικά, όμοιοι αλγόριθμοι αλλά με καλύτερα αποτελέσματα [30].

Επιπροσθέτως, η ταυτόχρονη θεώρηση της Διέγερσης ως επιπλέον χαρακτηριστικό μπορεί να βελτιώσει τα αποτελέσματα της αναγνώρισης, αλλά μόνο για τα ηχητικά γεγονότα που εγείρουν χαμηλή ευχαρίστηση. Σε αυτά που είναι στον αντίποδα, δηλαδή εγείρουν υψηλή ευχαρίστηση, η χρήση της Διέγερσης φαίνεται ότι δεν δημιουργεί κάποια ουσιαστική βελτίωση. Αυτό το γεγονός συνδυαζόμενο με την παραπάνω ανάλυση, αναδεικνύει με ισχυρό τρόπο ότι η διαδικασία αναγνώρισης Ευχαρίστησης θα μπορούσε να επωφεληθεί από μία ιεραρχική προσέγγιση κατηγοριοποίησης, αλλά και με την χρήση επιπρόσθετων χαρακτηριστικών, όπως αυτό της Διέγερσης. Ακόμα, παρόλο που οι τιμές των μετρικών αξιολόγησης απόδοσης μπορούν να θεωρηθούν χαμηλές σε σχέση με τις εν γένει μελέτες εξόρυξης συναισθηματικής πληροφορίας από ηχητικά δεδομένα, τα αποτελέσματα της παρούσας μελέτης παρέχουν χρήσιμες ενδείξεις και αποδείξεις έτσι ώστε η σχέση Ευχαρίστησης και ρυθμικών χαρακτηριστικών των ηχητικών γεγονότων να αποβεί κερδοφόρα για την βελτιστοποίηση της εν γένει εξόρυξης συναισθηματικής πληροφορίας από ηχητικά γεγονότα. Τέλος, υπάρχει μία δυσκολία στην αναγνώριση των ηχητικών γεγονότων με υψηλή ευχαρίστηση. Μία πιθανή εξήγηση μπορεί να είναι ο δυαδικός ορισμός των κλάσεων της Ευχαρίστησης: υψηλή και χαμηλή. Ίσως περισσότερες κλάσεις, π.χ. χαμηλή, μέση και υψηλή, να αυξήσουν τα αποτελέσματα αναγνώρισης. Για το παρόν πρόβλημα κατηγοριοποίησης, δηλαδή σε χαμηλή και υψηλή κλάση, φαίνεται ότι μία ιεραρχική προσέγγιση θα μπορούσε να αποφέρει υψηλότερες τιμές στα μετρικά αξιολόγησης απόδοσης.

ΚΕΦΑΛΑΙΟ 6

ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΠΡΟΕΚΤΑΣΕΙΣ

Στην ανά χείρας διατριβή εξετάστηκε η σχέση συναισθημάτων και γενικών ήχων. Οι γενικοί ήχοι, ή ηχητικά γεγονότα, είναι μη-λεκτικοί (non verbal) και μη-μουσικοί (non-musical) ήχοι και, ως εκ τούτου, αποτελούν μία γενική μορφή ήχου, σε αντίθεση με την ομιλία και την μουσική όπου υπάρχουν συγκεκριμένες δομές οργάνωσης του ήχου. Για παράδειγμα, στην μουσική υπάρχει η αρμονία και ο ρυθμός ενώ στην ομιλία υπάρχουν οι φθόγγοι και τα φωνήματα, όπου μπορούν να θεωρηθούν ως κατανομές ενέργειας ανά συχνότητα, ή ζώνη συχνοτήτων, ή ως κατανομή ενέργειας στο πεδίο του χρόνου (π.χ. η περίπτωση του ρυθμού της μουσικής). Η σχέση συναισθήματος και ομιλίας ή μουσικής είναι εμπειρικά γνωστή. Υπάρχει η κοινή γνώση ότι η ομιλία και η μουσική μπορούν να εγείρουν συναισθήματα. Πέρα από τις λέξεις που μπορεί να εκφέρει κανείς ή τους στίχους ενός τραγουδιού που μπορούν να επηρεάσουν συναισθηματικά με το νόημά τους, ο ήχος καθαυτός μπορεί να επηρεάσει συναισθηματικά τον ακροατή. Αυτή την εμπειρική και κοινή γνώση συσχέτισης ομιλίας ή/και μουσικής με το συναίσθημα του ακροατή τους, πλαισιώνουν, πλέον, επιστημονικές εργασίες που αποδεικνύουν πως το ηχητικό ερέθισμα μεταφέρει εγγενώς πληροφορία η οποία

σχετίζεται με την έγερση συναισθημάτων στον ακροατή. Για την ακρίβεια, υπάρχουν τα επιστημονικά πεδία της Αναγνώρισης Συναισθήματος από Ομιλία (Speech Emotion Recognition, SER), της Εξόρυξης Μουσικής Πληροφορίας (Music Information Retrieval, MIR) και της Αναγνώρισης Συναισθήματος από Μουσική (Music Emotion Recognition, MER) τα οποία εστιάζουν στην διερεύνηση της προαναφερθείσας συσχέτισης, δηλαδή στην συσχέτιση ομιλίας ή/και μουσικής με το συναίσθημα του ακροατή. Μία ενδελεχής επισκόπηση των πρόσφατων αποτελεσμάτων στα προαναφερθέντα πεδία έχει ήδη δοθεί στο Κεφάλαιο 2.

Η αναγνώριση συναισθήματος από μουσική ή/και ομιλία είναι εδραιωμένο πεδίο έρευνας και υπάρχει μία πληθώρα δημοσιευμένων εργασιών πάνω σε αυτόν τον τομέα. Στην αναγνώριση συναισθήματος από μουσική, η έρευνα εστιάζει στην χρήση ενός ανθρωποκεντρικού και εστιασμένου στο περιεχόμενο παράγοντα για την κατηγοριοποίηση της μουσικής πληροφορίας. Έτσι, αντί για τυπικά σχήματα κατηγοριοποίησης, όπως (π.χ.) το “Καλλιτέχνης, Έτος, Όνομα Άλμπουμ”, η μουσική μπορεί να οργανωθεί αναλόγως με το συναίσθημα που εγείρει στον ακροατή της. Για παράδειγμα, σε “Στεναχωρημένη”, “Χαρούμενη” κτλ. Αυτό σημαίνει μία πιο εύκολη αναζήτηση από τον άνθρωπο μιας και αφενός μπορεί να βρει μουσικά δεδομένα χωρίς να γνωρίζει τον καλλιτέχνη τους (αλλά να τον μάθει μετά, αφού τα βρει) και, αφετέρου, εστιασμένα σε αυτό που θέλει. Δηλαδή, στο συναίσθημα που θα ήθελε την συγκεκριμένη ώρα να εγείρει η μουσική. Τα αποτελέσματα από τα πεδία MIR και MER δείχνουν ευστοχία στην αναγνώριση συναισθήματος μέχρι και 88%. Από την άλλη, στην αναγνώριση συναισθήματος από ομιλία, η έρευνα εστιάζει σε τρόπους για την ορθή διαχείριση της αλληλεπίδρασης με τον άνθρωπο. Δηλαδή, με την δυνατότητα αναγνώρισης του συναισθήματος του ομιλητή συνδυαζόμενη με την αναγνώριση των όσων είπε, να γίνεται μία ορθότερη και πιο στοχευμένη αξιοποίηση των όσων ειπώθηκαν.

Όμως, η αναγνώριση συναισθήματος από ηχητικά γεγονότα είναι σχετικά νέο πεδίο έρευνας, με λίγες δημοσιευμένες εργασίες να εστιάζουν αποκλειστικά σε αυτή την διαδικασία και, ως εκ τούτου, τα αποτελέσματά τους να μην μπορούν να έχουν μία συνεκτικότητα, κυρίως λόγω της μη ύπαρξης ενός ικανού πλήθους δημοσιευμένων αποτελεσμάτων που θα μπορούσαν να προσφέρουν ένα στιβαρό υπόβαθρο για το πεδίο αυτό. Από την άλλη όμως, λόγω της γενικότητας των ηχητικών γεγονότων αλλά και του προφανούς μεγαλύτερου πλήθους ηχητικών ερεθισμάτων που μπορούν να καταταγούν σε αυτή την κατηγορία, τα αποτελέσματα από την έρευνα στην αναγνώριση συναισθήματος από ηχητικά γεγονότα μπορούν να αναδείξουν τα

εξής ερωτήματα αλλά και συντελέσουν στα:

- Κατανόηση για την εν γένει σχέση των τεχνικών χαρακτηριστικών του ήχου με το εγειρόμενο συναίσθημα του ακροατή. Δηλαδή, πως επηρεάζει ένα τεχνικό χαρακτηριστικό του ήχου (π.χ. ρυθμός) το πώς νιώθει ο ακροατής, γενικά; Λόγω της γενικότητας των ηχητικών γεγονότων, οποιοδήποτε αποτέλεσμα των επιστημονικών εργασιών που εστιάζουν στην διαδικασία εξόρυξης συναισθηματικής πληροφορίας από ηχητικά γεγονότα και το οποίο συσχετίζει ένα ή περισσότερα τεχνικά χαρακτηριστικά με το συναίσθημα του ακροατή αφενός θα αναδείξει την εν γένει επήρεια αυτού του τεχνικού χαρακτηριστικού στο συναίσθημα του ανθρώπου και, αφετέρου, θα προσφέρει γνώση για περαιτέρω διερεύνηση για την συσχέτιση αυτού του χαρακτηριστικού με το συναίσθημα του ακροατή στις πιο ειδικές κατηγορίες ήχου, ήτοι της μουσικής και της ομιλίας.
- Ανάδειξη της επήρειας των εγγενών χαρακτηριστικών της πηγής, τα οποία μπορούν να επικοινωνηθούν με το ηχητικό ερέθισμα, στο συναίσθημα του ακροατή. Ένα ηχητικό γεγονός, πέρα από την κυματομορφή του, έχει μία πληθώρα άλλων στοιχείων που μεταφέρει και ο ακροατής μπορεί να τα αντιληφθεί, τα οποία αναλύθηκαν εκτενώς στο Κεφάλαιο 3. Μέσα σε αυτά είναι η χωρική τοποθέτηση της ηχητικής πηγής σε σχέση με τον ακροατή. Κάθε πηγή ήχου είναι ένα φυσικό σώμα, δηλαδή έχει φυσικές διαστάσεις. Συνεπώς, έχει μία φυσική υπόσταση η οποία έχει μία χωρική συσχέτιση με τον ακροατή. Δηλαδή, μπορεί να είναι μπροστά του, στις πλευρικές θέσεις του ακροατή ή και πίσω του. Αυτή η χωρική συσχέτιση γίνεται αντιληπτή από τον ακροατή, όπου ο τελευταίος μπορεί να γνωρίζει, και μόνο μέσω του ηχητικού ερεθίσματος, που βρίσκεται η πηγή σχετικά με αυτόν. Επηρεάζει αυτή η γνώση το πώς θα νιώσει ο ακροατής από ένα ηχητικό γεγονός; Υπάρχει διαφορά στο πώς νιώθει ο άνθρωπος όταν αντιλαμβάνεται ένα ηχητικό ερέθισμα όταν η πηγή αυτού είναι, π.χ., εκτός του οπτικού του πεδίου από όταν η ηχητική αυτή πηγή είναι μπροστά του; Η απάντηση σε τέτοιου είδους ερωτήματα, δηλαδή στη συσχέτιση των εγγενών χαρακτηριστικών της πηγής, τα οποία μπορούν να γίνουν αντιληπτά από το ηχητικό γεγονός, με το συναίσθημα του ακροατή μπορεί να χρησιμοποιηθεί σε ψυχαγωγικές εφαρμογές, σε σχεδιασμούς αστικών περιοχών αλλά και στην παραγωγή καλλιτεχνικού αποτελέσματος. Για παράδειγμα, εάν υπάρχει συσχέτιση της χωρικής τοποθέτησης της πηγής με το συναίσθημα

του ακροατή τότε σε ένα πλήρως εικονικό περιβάλλον, π.χ. σοβαρό παιχνίδι (serious game), ή σε ένα επαυξημένο (augmented reality) η χωρική θέση της πηγής θα μπορούσε να χρησιμοποιηθεί για να μεταβάλλει κατάλληλα την συναισθηματική κατάσταση του ακροατή. Ακόμα, σε περιπτώσεις καλλιτεχνικού έργου όπου υπάρχει η ευχέρεια και η δυνατότητα για τοποθέτηση των ηχητικών πηγών σε θέσεις πίσω ή στα πλάγια των θεατών ή του ακροατηρίου (π.χ. θέατρο, καλλιτεχνικές εγκαταστάσεις, κινηματογράφος), η πιθανή συσχέτιση της χωρικής τοποθέτησης της πηγής θα προσέφερε μία ακόμα παράμετρο συναισθηματικής αλληλεπίδρασης του καλλιτέχνη, μέσω του έργου του, με τους θεατές λόγω της επίρειας της συναισθηματικής κατάστασης των τελευταίων και από την χωρική τοποθέτηση της πηγής.

- Εν γένει κατανόηση την συναισθηματικής αλληλεπίδρασης του ανθρώπου με το ηχητικό του περιβάλλον. Μέσω της έρευνας για την αναγνώριση συναισθηματικής πληροφορίας από ηχητικά γεγονότα, θα υπάρξει η δυνατότητα για την κατανόηση της συναισθηματικής επίρειας του ηχητικού περιβάλλοντος στον άνθρωπο. Ήδη, η σχέση του ανθρώπου με το ηχητικό του περιβάλλον μοντελοποιείται μέσω της Ακουστικής Οικολογίας. Εφόσον τα ηχητικά γεγονότα είναι μία γενική μορφή ήχου, η γνώση για την επίρεια αυτών στο συναίσθημα του ακροατή θα βελτιώσει τον τρόπο με τον οποίο μοντελοποιείται η σχέση του ανθρώπου με το ηχητικό του περιβάλλον. Επιπροσθέτως, θα πρέπει να βρεθεί και ένα νέο μοντέλο Ακουστικής Οικολογίας το οποίο θα μπορεί να περιλαμβάνει και την προαναφερθείσα συναισθηματική αλληλεπίδραση.

Στα παραπάνω εστιάζει η παρούσα διατριβή. Στην διεθνή επιστημονική βιβλιογραφία, δεν αναφέρονται παρόμοιες εργασίες που να εστιάζουν στα ίδια ερωτήματα και προβλήματα στο πεδίο αναγνώρισης συναισθημάτων από ηχητικά γεγονότα. Επιπροσθέτως, ακόμα και σε εργασίες που αντιμετωπίζουν παρόμοια προβλήματα αλλά με ηχητικά δεδομένα μουσικής, η ανά χείρας διατριβή έχει να επιδείξει αποτελέσματα που είναι αρκετά υψηλά, τα υψηλότερα κατά το καλύτερο της γνώσης του συγγραφέα.

Σε αυτό το Κεφάλαιο θα παρατεθούν τα συμπεράσματα που προκύπτουν από την διατριβή αυτή, εστιάζοντας σε δύο βασικά στοιχεία ενός ηχητικού γεγονότος: α) την κυματομορφή, και άρα τα τεχνικά χαρακτηριστικά του καθαυτού σήματος, και β) τα χαρακτηριστικά της πηγής που μπορεί να μεταφέρει ως πληροφορία το ηχητικό γεγονός. Συνεπώς, στην αμέσως επόμενη ενότητα δίνονται τα συμπεράσματα για την αναγνώριση συναισθήματος α-

πό τεχνικά χαρακτηριστικά ηχητικών δεδομένων, ακολουθούμενα, σε άλλη ενότητα, από τα συμπεράσματα για την αναγνώριση συναισθήματος από χαρακτηριστικά της πηγής που μεταφέρει το ηχητικό γεγονός. Τέλος, παρουσιάζεται η πρωτοτυπία της διατριβής και έπειτα παρατίθεται μία σύνοψη όλων των συμπερασμάτων ταυτόχρονα με προτάσεις για μελλοντικές εφαρμογές.

6.1 Επήρεια συναισθημάτων από τεχνικά χαρακτηριστικά ηχητικών γεγονότων

Όπως δόθηκε σε προηγούμενο Κεφάλαιο, τα τεχνικά χαρακτηριστικά του ήχου είναι ο ρυθμός, η χροιά, η ενέργεια και το τονικό ύψος. Αυτά, χρησιμοποιούνται στην διαδικασία αναγνώρισης συναισθήματος από ομιλία και μουσική μέσω της εξαγωγής από το ηχητικό σήμα των αντίστοιχων τεχνικών μεγεθών. Οι τιμές των τεχνικών μεγεθών, μαζί με τις τιμές των συναισθηματικών επισημάνσεων των αντίστοιχων ηχητικών δεδομένων, χρησιμοποιούνται ως είσοδος σε συστήματα που εκτελούν διαδικασίες Μηχανικής Μάθησης. Η έξοδος από τα συστήματα αυτά είναι ένα μοντέλο κατηγοριοποίησης, κατάλληλο για ναμαντέψει την συναισθηματική επισήμανση για άλλα και αντίστοιχα (δηλαδή, εάν χρησιμοποιήθηκε ομιλία ως ηχητικό σήμα, τότε θα πρέπει να χρησιμοποιηθεί ομιλία ως άλλο) ηχητικά δεδομένα.

Η επιστημονική βιβλιογραφία που υπάρχει στα πεδία αναγνώρισης συναισθήματος από ομιλία και μουσική, παρέχει μία πληθώρα αποτελεσμάτων αναφορικά με την χρήση διαφόρων τεχνικών χαρακτηριστικών, και αντίστοιχων μεγεθών, και αλγορίθμων κατηγοριοποίησης. Συγκεκριμένα για την αναγνώριση συναισθήματος από ομιλία, υπάρχουν ήδη συναισθηματικά προφίλ ομιλίας. Δηλαδή, συσχέτιση τιμών τεχνικών χαρακτηριστικών με το εκλαμβανόμενο ή εκφερόμενο συναίσθημα. Για παράδειγμα, εάν κανείς μιλήσει πιο δυνατά, με μεγαλύτερη θεμελιώδη συχνότητα, με μεγαλύτερο ρυθμό ομιλίας και με καθαρή φωνή τότε είναι πολύ πιθανό να νιώθει θυμό. Στην αναγνώριση συναισθήματος από μουσική δεν έχει αναφερθεί κάποια παρόμοια αναλυτική συσχέτιση. Παρόλα αυτά, υπάρχουν ενδείξεις για την επήρεια τεχνικών χαρακτηριστικών στο συναίσθημα του ακροατή. Παραδείγματος χάρη, η ενέργεια έχει βρεθεί ότι σχετίζεται σε υψηλό βαθμό με την Διέγερση του ακροατή. Χρησιμοποιείται στις περισσότερες δημοσιευμένες εργασίες ως ένα από τα κυρίαρχα τεχνικά χαρακτηρι-

κά για την αναγνώριση της Διέγερσης, ενώ σε άλλες ως το μόνο. Ο ρυθμός είναι επίσης ένα τεχνικό χαρακτηριστικό που έχει συσχετιστεί ισχυρά με το συναίσθημα του ακροατή και συγκεκριμένα με την Διέγερση. Επίσης χρησιμοποιείται σε δημοσιευμένες έρευνες ως χαρακτηριστικό για την ανίχνευση της Διέγερσης. Ποτέ όμως ως το μόνο τεχνικό χαρακτηριστικό για την αναγνώριση της Διέγερσης, όπως γίνεται με την ενέργεια.

Εστιάζοντας όμως στα ηχητικά γεγονότα, ένα υπόβαθρο έρευνας όπως αυτό που αναφέρθηκε για την ομιλία και την μουσική, δεν υπάρχει. Δεν υπάρχει καν μία αρχική γνώση για την συσχέτιση των τεχνικών χαρακτηριστικών του ήχου με τα συναισθήματα. Όμως, εάν βρισκόταν μία τέτοια συσχέτιση για ένα τεχνικό χαρακτηριστικό, θα μπορούσε να αποκαλύψει την καθολικότητα της επίρειας αυτού του χαρακτηριστικού στο συναίσθημα του ακροατή. Δηλαδή, εάν η ενέργεια ενός ηχητικού γεγονότος έχει μείζονα ρόλο στην εγειρόμενη Διέγερση του ακροατή, τότε, με τον συνδυασμό των ευρημάτων από τα πεδία αναγνώρισης συναισθήματος από ομιλία και μουσική, θα φαινόταν καθαρά ότι η ενέργεια του ήχου επηρεάζει τον ακροατή ασχέτως του τύπου του ηχητικού ερεθίσματος.

Στην παρούσα διατριβή δείχθηκε ότι η μέση ενέργεια ενός ηχητικού γεγονότος δεν έχει σχέση με την Διέγερση του ακροατή, όταν αυτή δεν είναι σε μία ακραία τιμή της. Δηλαδή, όταν στο μοντέλο συναισθημάτων που επιλέχθηκε στην εκάστοτε εργασία, η Διέγερση του ηχητικού γεγονότος δεν βρίσκεται στις μέγιστες τιμές του μοντέλου. Στις μη συναισθηματικά ακραίες καταστάσεις, δείχθηκε ότι η μέση ενέργεια δεν μπορεί να επηρεάσει την Διέγερση του ακροατή. Συγκεκριμένα, μέσω ενός σετ από συναισθηματικά επισημασμένα ηχητικά γεγονότα, ήτοι το IADS, φαίνεται πως ηχητικά γεγονότα με μία μέση τιμή ενέργειας, εγείρουν μεγαλύτερη διέγερση στον ακροατή από άλλα με μικρότερη μέση τιμή ενέργειας. Ένας παράγοντας που θα μπορούσε να επηρεάσει την σχέση ενέργειας ηχητικού γεγονότος και Διέγερσης, είναι το σημασιολογικό περιεχόμενο. Όμως, επίσης μέσω του ίδιου σετ ηχητικών γεγονότων δείχθηκε πως αυτό δεν συμβαίνει. Συνεπώς, η ενέργεια του ηχητικού ερεθίσματος επηρεάζει σημαντικά την Διέγερση του ακροατή μόνο στις περιπτώσεις που το ηχητικό ερέθισμα είναι είτε ομιλία, είτε μουσική.

Από την άλλη, στην παρούσα διατριβή δείχθηκε ότι ο ρυθμός έχει σημαντική επίρεια στην Διέγερση του ακροατή και όχι τόσο σημαντική στην Ευχαρίστηση. Έτσι, την εμπειρική γνώση ότι ο ρυθμός του ήχου μπορεί να διεγείρει τον ακροατή, επιβεβαιώνει η ανά χείρας διατριβή. Συγκεκριμένα, με το σετ IADS, έγινε μία σειρά πειραμάτων Μηχανικής Μάθησης.

Σε αυτά χρησιμοποιήθηκε μόνο η συναισθηματική επισήμανση της Διέγερσης και από τα ηχητικά δεδομένα εξήχθησαν τεχνικά μεγέθη που αφορούσαν αποκλειστικά τον ρυθμό. Η ευστοχία των παραγόμενων μοντέλων έφτασε ως και 88%, σχεδόν στην ίδια τιμή με την μέγιστη τιμή ευστοχίας αναγνώρισης συναισθήματος από μουσική, όπως αναφέρεται στην διεθνή βιβλιογραφία. Αυτό το γεγονός αναδεικνύει την σημαντικότητα του ρυθμού ως τεχνικό χαρακτηριστικό, στην συναισθηματική επήρεια του ακροατή. Με τον συνδυασμό των ευρημάτων από τα πεδία αναγνώρισης ομιλίας και μουσικής, όπως αυτά αναφέρονται συνοπτικά παραπάνω και αναλυτικότερα σε προηγούμενο Κεφάλαιο, φαίνεται καθαρά ότι ο ρυθμός έχει μία καθολική ισχύ στην επήρεια της Διέγερσης του ανθρώπου από ηχητικά ερεθίσματα.

Με μία επόμενη σειρά πειραμάτων Μηχανικής Μάθησης, που διενεργήθηκε στα πλαίσια της παρούσας διατριβής, δείχθηκε η καθολικότητα της επήρειας του ρυθμού ενός ηχητικού ερεθίσματος στην συνολική συναισθηματική κατάσταση του ακροατή. Στα πειράματα αυτά, χρησιμοποιήθηκε το ίδιο σετ ηχητικών γεγονότων αλλά και τεχνικών μεγεθών με αυτό της διερεύνησης της επήρειας του ρυθμού ενός ηχητικού γεγονότος στην Διέγερση. Η διαφορά είναι ότι στην νέα σειρά πειραμάτων, έγινε εστίαση στην Ευχαρίστηση. Η μέγιστη τιμή ευστοχίας αναγνώρισης ήταν της τάξης του 65%, μία από τις μεγαλύτερες τιμές που έχει αναφερθεί για την αναγνώριση Ευχαρίστησης από ηχητικά δεδομένα εν γένει και η μεγαλύτερη τιμή, μέχρι στιγμής, για την αναγνώριση Ευχαρίστησης από ηχητικά γεγονότα. Παράλληλα, αυτή η μέγιστη τιμή ευστοχίας δείχνει ότι υπάρχουν και άλλοι παράγοντες, δηλαδή τεχνικά χαρακτηριστικά, που επηρεάζουν την Ευχαρίστηση και θα πρέπει να ληφθούν υπόψη. Συνεπώς, ο ρυθμός ενός ηχητικού δεδομένου μπορεί να επηρεάσει την Διέγερση του ακροατή με ισχυρό τρόπο, ενώ την Ευχαρίστηση μέχρι ένα ποσοστό.

Συμπερασματικά, με την παρούσα διατριβή αναδεικνύεται η συσχέτιση της συναισθηματικής επήρειας του ακροατή από δύο τεχνικά χαρακτηριστικά ηχητικών δεδομένων, εν γένει, αλλά και ηχητικών γεγονότων ιδιαίτερος. Φαίνεται καθαρά πως η ενέργεια δεν έχει μείζονα ρόλο στην Διέγερση του ακροατή για όλων των τύπων ηχητικά ερεθίσματα. Τέλος, δείχνεται ότι ο ρυθμός μπορεί να επηρεάσει ισχυρά την Διέγερση του ακροατή και έως ένα σημείο την Ευχαρίστησή του.

6.2 Επήρεια συναισθημάτων από χαρακτηριστικά της πηγής

Βάσει των αποτελεσμάτων της παρούσας διατριβής, είναι φανερό ότι ο άνθρωπος αλληλεπιδρά και συναισθηματικά με το ηχητικό του περιβάλλον. Η εν γένει αυτή αλληλεπίδραση, χωρίς όμως την συναισθηματική διάσταση, είναι αντικείμενο της Ακουστικής Οικολογίας. Έτσι, προτάθηκε από την παρούσα διατριβή η επαύξηση αυτού του πεδίου με το να συμπεριλάβει και την αποδεδειγμένη συναισθηματική αλληλεπίδραση. Ως βασικό στοιχείο του επαυξημένου ηχητικού περιβάλλοντος θα είναι το ηχητικό γεγονός το οποίο θα περιγράφεται από ένα πλήθος στοιχείων ικανά να χρησιμοποιηθούν για την εξόρυξη, διαχείριση και επικοινωνία όλων των ειδών της πληροφορίας. Εκτός από τα τεχνικά χαρακτηριστικά του ήχου, τα οποία ενυπάρχουν στην καθαυτή κυματομορφή του ηχητικού γεγονότος και τα αντιλαμβάνεται ο άνθρωπος, υπάρχουν διάφορα χαρακτηριστικά της πηγής που μεταφέρονται ως πληροφορία από ένα ηχητικό ερέθισμα. Αυτά μοντελοποιούνται από το ηχητικό γεγονός, όπως αυτό ορίστηκε στα πλαίσια της Συναισθηματικής Ακουστικής Οικολογίας, και είναι τα εξής:

- εκδήλωση της στιγμιαίας χωρικής τοποθέτησης της πηγής
- ένδειξη του τρόπου παραγωγής ήχου (τριβή, κρούση, ταλάντωση ή ό,τι άλλο)
- ένδειξη της φυσικής κατάστασης των σωμάτων που μετέχουν στην παραγωγή ήχου (π.χ. στερεά, υγρά)

Η επήρεια της χωρικής τοποθέτησης της πηγής, έχει μελετηθεί πριν την ανά χειράς διατριβή. Δημοσιευμένες έρευνες έδειξαν ότι, όταν μία πηγή είναι πίσω από τον ακροατή εκείνος αισθάνεται περισσότερο φόβο. Όμως, όλες οι υφιστάμενες δημοσιευμένες εργασίες δεν παρουσιάζουν αφενός την εν γένει επήρεια της χωρικής τοποθέτησης της πηγής στο συναίσθημα του ακροατή, δηλαδή τι γίνεται εάν η πηγή είναι σε κάποια θέση ανάμεσα στην ακριβώς μπροστά και ακριβώς πίσω από τον ακροατή, και, αφετέρου, εστιάζουν μόνο σε ένα συναίσθημα. Αυτό του φόβου. Η παρούσα διατριβή εστίασε στην επήρεια της χωρικής τοποθέτησης της πηγής στο συναίσθημα του ακροατή χρησιμοποιώντας ένα διαστατικό μοντέλο συναισθημάτων. Διενεργήθηκε μία σειρά πειραμάτων ακρόασης με υποκειμενική αξιολόγηση στα οποία χρησιμοποιήθηκαν ήχοι με χωρική τοποθέτηση με χρήση της αμφιωτικής τεχνολογίας. Τα

αποτελέσματα των πειραμάτων έδειξαν ότι η χωρική τοποθέτηση της πηγής μπορεί να μεταβάλλει το αρχικό συναίσθημα που εγείρει στον ακροατή ένα ηχητικό γεγονός, αναλόγως την θέση της πηγής. Για την ακρίβεια, υπάρχουν τρεις κλάσεις ηχητικών γεγονότων. Μία που περιέχει τα ηχητικά γεγονότα που, καθώς αυτά μετακινούνται περιφερειακά του ακροατή από την ακριβώς μπροστινή θέση προς την πίσω μεριά του, αυξάνουν την Διέγερσή του και μειώνουν την Ευχαρίστηση και μία άλλη που περιέχει τα ηχητικά γεγονότα όπου, για την ίδια κίνηση, δηλαδή από την ακριβώς μπροστά θέση προς την ακριβώς πίσω, αυξάνουν την Ευχαρίστηση και μειώνουν την Διέγερση. Στην τρίτη κλάση, ανήκουν όσα δεν ανήκουν σε κάποια από τις προαναφερθείσες δύο. Επίσης, δείχθηκε ότι υπάρχει και μία ειδική περίπτωση ηχητικών γεγονότων, αυτά με σημασιολογικό περιεχόμενο μουσικής, όπου το αποτέλεσμα της μετακίνησής τους προς τις πλευρικές θέσεις του ακροατή είναι να επηρεάζουν μόνο την Ευχαρίστησή του.

Συμπερασματικά, πέρα από τα τεχνικά χαρακτηριστικά του ηχητικού ερεθίσματος που μπορούν να μετρηθούν από την καθαυτού κυματομορφή, επήρεια στο συναίσθημα του ακροατή έχει και η θέση της ηχητικής πηγής. Αυτό οδηγεί στο συμπέρασμα ότι σε ένα πλαίσιο μελέτης του ηχητικού περιβάλλοντος του ανθρώπου, θα πρέπει να μελετηθούν όχι μόνο τα τεχνικά χαρακτηριστικά του ήχου που ενυπάρχουν στην κυματομορφή αλλά και τα χαρακτηριστικά της πηγής τα οποία μεταφέρονται μέσω του ηχητικού κύματος και μπορούν να γίνουν αντιληπτά από τον ακροατή. Προς αυτό το σκοπό, μπορεί να γίνει χρήση του μοντέλου της Συναισθηματικής Ακουστικής Οικολογίας όπου το βασικό του στοιχείο, δηλαδή το ηχητικό γεγονός, έχει, εξ ορισμού, τα προαναφερθέντα χαρακτηριστικά της πηγής. Για την περαιτέρω μελέτη της επήρειας της χωρικής τοποθέτησης της πηγής στο συναίσθημα του ακροατή, μέσω της έρευνας στα πλαίσια της παρούσας διατριβής, αναπτύχθηκε και διατίθεται ελεύθερα ένα σετ συναισθηματικά επισημασμένων ηχητικών γεγονότων με χωρική τοποθέτηση, το BEADS.

6.3 Πρωτοτυπία της διατριβής

Συνυπολογίζοντας όλα τα παραπάνω, φαίνεται καθαρά ότι η πρωτοτυπία της ανά χείρας διατριβής εστιάζει στο πεδίο της αναγνώρισης συναισθήματος από ηχητικά γεγονότα. Συγκεκριμένα, μέσω της παρούσας διατριβής αποδείχθηκε η καθολική επήρεια του ρυθμού, ως

ένα τεχνικό χαρακτηριστικό του ήχου που μπορεί να επηρεάσει συναισθηματικά τον ακροατή ανεξαρτήτως της μορφής ήχου (ομιλία, μουσική, ηχητικά γεγονότα). Επίσης, δείχθηκε ότι η ενέργεια έχει ισχυρή επίρεια στο συναίσθημα του ακροατή. Έτσι, ενώ για την αναγνώριση συναίσθηματος από μουσική ή ομιλία, η ενέργεια έχειδειχθεί ότι έχει σημαντική επίρεια στην Διέγερση του ακροατή, στα ηχητικά γεγονότα αποδείχθηκε ότι δεν ισχύει αυτό. Τα γεγονότα αυτά δεν είχαν αποδειχθεί πριν την παρούσα διατριβή και μπορούν να συμβάλουν ως επιστημονικό υπόβαθρο για την μετέπειτα εξέλιξη της έρευνας πάνω στην αναγνώριση συναίσθηματος από ηχητικά δεδομένα.

Ακόμα, μέσα από την έρευνα που διενεργήθηκε στα πλαίσια της παρούσας διατριβής, αναπτύχθηκε και διατίθεται ελεύθερα το πρώτο παγκοσμίως σετ συναισθηματικά επισημασμένων ηχητικών γεγονότων με χωρική τοποθέτηση, το BEADS. Η διάθεσή του γίνεται μέσω του ιστοχώρου του Εργαστηρίου Επεξεργασίας Οπτικοακουστικών Σημάτων του Τμ. Τεχνών Ήχου και Εικόνας, του Ιονίου Πανεπιστημίου¹³ και το σετ αυτό αποτελεί ένα από τα τρία διαθέσιμα σετ με συναισθηματικά επισημασμένα ηχητικά γεγονότα.

Τέλος, με την έρευνα στα πλαίσια της παρούσας διατριβής αναπτύχθηκε ένα αναλυτικό μοντέλο επίρειας της χωρικής τοποθέτησης της πηγής στο συναίσθημα του ακροατή. Ένα αποτέλεσμα που δεν έχει παρουσιαστεί ξανά και αποτελεί την βάση για την έρευνα αφενός πάνω σε μία ολιστική μελέτη της επίρειας του ηχητικού περιβάλλοντος του ανθρώπου στον άνθρωπο και, αφετέρου, πάνω στην επίρεια στο συναίσθημα του ακροατή των χαρακτηριστικών της πηγής που μπορούν να μεταφερθούν με το ηχητικό ερέθισμα (π.χ. χωρική τοποθέτηση, μηχανισμός παραγωγής ήχου κ.α.).

6.4 Μελλοντικές προεκτάσεις και εφαρμογές

Η εξόρυξη συναισθηματικής πληροφορίας από ηχητικά γεγονότα θα μπορούσε να αξιοποιηθεί με δύο τρόπους: είτε για την διαχείριση και μελέτη της συναισθηματικής επίρειας ενός δεδομένου ηχητικού περιβάλλοντος, είτε για την διαχείριση της καθαυτού συναισθηματικής πληροφορίας.

Προς αυτό το σκοπό όμως θα πρέπει να ερευνηθεί ένα σύνολο ερωτημάτων, με κυρίαρχο την

¹³<http://epoasi.eu/gr/research/publications/research-material/>

σχέση σημασιολογικού περιεχομένου, τεχνικών χαρακτηριστικών και εγχειρόμενου συναισθήματος. Βάσει των αποτελεσμάτων και συμπερασμάτων της παρούσας διατριβής, υπάρχει το ενδεχόμενο αλληλεξάρτησης αυτών των τριών. Το τελευταίο σημαίνει ότι είναι πολύ πιθανό όταν ένα ηχητικό γεγονός, με ένα συγκεκριμένο σημασιολογικό περιεχόμενο, δύναται να επηρεάσει με έναν συναισθηματικό τρόπο τον ακροατή τότε ήχοι με παρόμοιας συναισθηματικής επίδρασης σημασιολογικό περιεχόμενο ενδέχεται να μοιράζονται, έως κάποιο βαθμό, κοινές τιμές στα κάποια τεχνικά χαρακτηριστικά τους. Από την άλλη, αυτό σημαίνει ότι θα ήταν επίσης πιθανό εάν πρέπει να επικοινωνηθεί κάποιο συγκεκριμένο σημασιολογικό περιεχόμενο με ένα ηχητικό ερέθισμα το οποίο θα έχει μία αναμενόμενη συναισθηματική επίρεια, τότε το ερέθισμα αυτό ενδέχεται να έχει τεχνικά χαρακτηριστικά που έχουν άλλα ηχητικά ερεθίσματα με παρόμοια συναισθηματική επίρεια. Ακόμα, θα πρέπει να αυξηθεί το ποσοστό αναγνώρισης μέσω της χρήσης περισσότερων τεχνικών χαρακτηριστικών. Μένουν πτυχές του ηχητικού γεγονότος, όπως η τονικότητα και η χροιά, που θα μπορούσαν να συνεισφέρουν στην εξόρυξη συναισθηματικής πληροφορίας. Αυτά τα χαρακτηριστικά θα μπορούσαν επίσης να χρησιμοποιηθούν και για τον δεύτερο τρόπο.

Συγκεκριμένα, με την κατάρτιση ενός πλήρους σετ τεχνικών χαρακτηριστικών που ευθύνονται για την συναισθηματική επίδραση των ηχητικών γεγονότων πάνω στον ακροατή, θα γίνει εφικτό να υπάρξουν πειράματα για την δημιουργία ηχητικών δεδομένων κατάλληλων να εγείρουν συγκεκριμένα συναισθήματα στον ακροατή. Τέτοια ηχητικά γεγονότα θα μπορούσα να χρησιμοποιηθούν σε όλες τις εφαρμογές που έχουν ηχητική διεπαφή, δίνοντας την δυνατότητα για στοχευμένη συναισθηματική αλληλεπίδραση με τον χρήστη. Αυτό το γεγονός θα σήμαινε μία πιο άμεση και ελεύθερη σημασιολογικού περιεχομένου επικοινωνία με την εκάστοτε συσκευή ή εφαρμογή, κάνοντας έτσι την επικοινωνία αυτή ανεξάρτητη από γλώσσα και πολιτιστικό υπόβαθρο. Επίσης, με βάσει αυτή την γνώση θα μπορούσε να υπάρξει αναγνώριση και αντιστοίχιση της συναισθηματικής διάστασης από το ηχητικό κανάλι επικοινωνίας σε κάποιο άλλο, π.χ. οπτικό. Κάτι τέτοιο θα έκανε δυνατή την πλήρη συναισθηματική επικοινωνία ανθρώπων με απώλεια ακοής με το ηχητικό περιβάλλον τους. Τέλος, η αντιστοίχιση αυτή θα μπορούσε να επεκταθεί και στην μεταφορά συναισθηματικής πληροφορίας από το οπτικό κανάλι στο ακουστικό, βοηθώντας έτσι ανθρώπους που δεν μπορούν να δουν, να αντιλαμβάνονται την συναισθηματική πληροφορία που ενδέχεται να υπάρχει στο περιβάλλον τους, και μεταδίδεται μέσω του οπτικού καναλιού, μέσω ηχητικών ερεθισμάτων.

ΒΙΒΛΙΟΓΡΑΦΊΑ

- [1] J. Blauert, “Analysis and synthesis of auditory scenes,” in *Communication Acoustics*, J. Blauert, Ed. Springer Berlin Heidelberg, 2005, pp. 1–25.
- [2] K. Drossos, A. Floros, K. Agavanakis, N.-A. Tatlas, and N.-G. Kanellopoulos, “Emergency voice/stress-level combined recognition for intelligent house applications,” in *Audio Engineering Society Convention 132*, Apr 2012.
- [3] K. Drossos, A. Floros, and N.-G. Kanellopoulos, “Affective acoustic ecology: Towards emotionally enhanced sound events,” in *Proceedings of the 7th Audio Mostly Conference: A Conference on Interaction with Sound*, ser. AM ’12. New York, NY, USA: ACM, 2012, pp. 109–116.
- [4] M. Marcell, M. Malatanos, C. Leahy, and C. Comeaux, “Identifying, rating, and remembering environmental sound events,” *Behavior Research Methods*, vol. 39, no. 3, pp. 561–569, 2007.
- [5] W. W. Gaver, “What in the world do we hear? an ecological approach to auditory event perception,” *Ecological Psychology*, vol. 5, pp. 1–29, 1993.
- [6] M. M. Bradley and P. J. Lang, “Affective reactions to acoustic stimuli.” *Psychophysiology*, vol. 37, no. 2, pp. 204–215, Mar 2000.

- [7] A. Wiczorkowska, "Towards extracting emotions from music," in *Intelligent Media Technology for Communicative Intelligence*, ser. Lecture Notes in Computer Science, L. Bolc, Z. Michalewicz, and T. Nishida, Eds. Springer Berlin Heidelberg, 2005, vol. 3490, pp. 228–238.
- [8] K. Scherer, R. Banse, H. Wallbott, and T. Goldbeck, "Vocal cues in emotion encoding and decoding," *Motivation and Emotion*, vol. 15, no. 2, pp. 123–148, 1991.
- [9] A. P. Oliveira and A. Cardoso, "Modeling affective content of music: A knowledge base approach," in *5th Sound and Music Computing Conference*, M. Supper and S. Weinzierl, Eds., Jul-Aug 2008, p. 6.
- [10] K. Mulligan and K. R. Scherer, "Toward a working definition of emotion," *Emotion Review*, vol. 4, no. 4, pp. 345–357, 2012.
- [11] K. R. Scherer, "Toward a dynamic theory of emotion: The component process model of affective states," *Geneva Studies in Emotion and Communication*, vol. 1, pp. 1–98, 1987.
- [12] —, "The neuropsychology of emotion," pp. 137–162, 2000.
- [13] F. Weninger, F. Eyben, B. W. Schuller, M. Mortillaro, and K. R. Scherer, "On the acoustics of emotion in audio: What speech, music and sound have in common," *Frontiers in Psychology*, vol. 4, no. 292, 2013.
- [14] P. N. Juslin and L. Petri, "Communication of emotions in vocal expression and music performance: different channels, same code?" *Psychological Bulletin*, vol. 129, pp. 770–814, Sept. 2003.
- [15] K. R. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Communication*, vol. 40, no. 1-2, pp. 227–256, Apr. 2003.
- [16] —, "Which emotions can be induced by music? what are the underlying mechanisms? and how can we measure them?" *Journal of New Music Research*, vol. 33, no. 3, pp. 239–251, 2004.
- [17] P. N. Juslin and K. R. Scherer, "Vocal expression of affect," in *The New Handbook of*

- Methods in Nonverbal Behavior Research*, J. A. Harrigan, R. Rosenthal, and K. R. Scherer, Eds. Oxford University Press, 2005, ch. 3, pp. 65–135.
- [18] E.-S. Jee, C. H. Kim, S.-Y. Park, and K.-W. Lee, “Composition of musical sound expressing an emotion of robot based on musical factors,” in *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, Aug 2007, pp. 637–641.
- [19] Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H. Chen, “A regression approach to music emotion recognition,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, no. 2, pp. 448–457, Feb 2008.
- [20] K. Markov and T. Matsui, “Music genre and emotion recognition using gaussian processes,” *Access, IEEE*, vol. 2, pp. 688–697, 2014.
- [21] K. Drossos, A. Floros, S. Potirakis, N.-A. Tatlas, and G. Tuna, “A socially-intelligent multi-robot service team for in-home monitoring,” in *Information, Intelligence, Systems and Applications, IISA 2014, The 5th International Conference on*, July 2014, pp. 159–164.
- [22] M. Ruxanda, B. Y. Chua, A. Nanopoulos, and C. S. Jensen, “Emotion-based music retrieval on a well-reduced audio feature space,” in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, April 2009, pp. 181–184.
- [23] K. Drossos and A. Floros, “Investigation of the possibility of mapping emotional information from audio to visual channel,” in *7th National Conference "Acoustics 2014"*, Oct 2014.
- [24] Y.-H. Yang and H. H. Chen, “Machine recognition of music emotion: A review,” *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 3, pp. 40:1–40:30, May 2012.
- [25] T. Li and M. Ogihara, “Toward intelligent music information retrieval,” *Multimedia, IEEE Transactions on*, vol. 8, no. 3, pp. 564–574, June 2006.
- [26] W. Zheng, M. Xin, X. Wang, and B. Wang, “A novel speech emotion recognition

- method via incomplete sparse least square regression,” *Signal Processing Letters, IEEE*, vol. 21, no. 5, pp. 569–572, May 2014.
- [27] J. sik Park, J.-H. Kim, and Y.-H. Oh, “Feature vector classification based speech emotion recognition for service robots,” *Consumer Electronics, IEEE Transactions on*, vol. 55, no. 3, pp. 1590–1596, August 2009.
- [28] L. Lu, D. Liu, and H.-J. Zhang, “Automatic mood detection and tracking of music audio signals,” *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 1, pp. 5–18, Jan 2006.
- [29] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull, “Music emotion recognition: A state of the art review,” in *11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, 2010, pp. 255–266.
- [30] K. Drossos, R. Kotsakis, G. Kalliris, and A. Floros, “Sound events and emotions: Investigating the relation of rhythmic characteristics and arousal,” in *Information, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on*, July 2013, pp. 1–6.
- [31] S. R. Livingstone and A. R. Brown, “Dynamic response: Real-time adaptation for music emotion,” in *Proceedings of the Second Australasian Conference on Interactive Entertainment*, ser. IE '05. Sydney, Australia, Australia: Creativity & Cognition Studios Press, 2005, pp. 105–111.
- [32] D. Casacuberta, “Dj el niño: expressing synthetic emotions with music,” *AI & SOCIETY*, vol. 18, no. 3, pp. 257–263, 2004.
- [33] S. Kim and E. André, “E.: Composing affective music with a generate and sense approach,” in *In: Proceedings of Flairs 2004 - Special Track on AI and Music*, AAAI Press, 2004.
- [34] I. Luengo, E. Navas, I. Hernáez, and J. Sánchez, “Automatic emotion recognition using prosodic parameters,” in *in Proceedings of INTERSPEECH*, 2005, pp. 493–496.

- [35] S. Koolagudi and K. Rao, “Emotion recognition from speech: a review,” *International Journal of Speech Technology*, vol. 15, no. 2, pp. 99–117, 2012.
- [36] V. A. Petrusin, “Emotion in speech: Recognition and application to call centers,” in *Conference of Artificial Neural Networks in Engineering (ANNIE '99)*, Nov 1999, pp. 7–10.
- [37] I. Ekman and R. Kajastila, “Localization cues affect emotional judgments - results from a user study on scary sound,” in *Audio Engineering Society Conference: 35th International Conference: Audio for Games*, Feb 2009.
- [38] P. N. Juslin and D. Västfjäll, “Emotional responses to music: The need to consider underlying mechanisms,” *Behavioral and Brain Sciences*, vol. 31, pp. 559–575, Oct. 2008.
- [39] S. R. Michalski, G. J. Carbonell, and M. T. Mitchell, Eds., *Machine Learning an Artificial Intelligence Approach Volume II*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1986.
- [40] C. Bishop, *Pattern Recognition and Machine Learning*, ser. Information Science and Statistics, M. Jordan, R. Nowak, and B. Schölkopf, Eds. Springer-Verlag New York, 2006, no. 1.
- [41] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*, ser. Adaptive Computation and Machine Learning series. The MIT Press, 2012.
- [42] Z. Ghahramani, “Unsupervised learning,” in *Advanced Lectures on Machine Learning*, ser. Lecture Notes in Artificial Intelligence, O. Bousquet, U. von Luxburg, and G. Rätsch, Eds. New York, NY, USA: Springer-Verlag New York, Inc., 2003, ch. 5.
- [43] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed., ser. Springer Series in Statistics. Springer-Verlag New York, 2009.
- [44] W. Fan, H. Wang, P. Yu, and S. Ma, “Is random model better? on its accuracy and efficiency,” in *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*, Nov 2003, pp. 51–58.

- [45] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The weka data mining software: An update,” *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009.
- [46] Z. Shiwei, Z. Li, W. Suyu, L. Xiaoguang, and S. Lansun, “Pornographic image recognition in compressed domain based on multi-cost sensitive decision tree,” in *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*, vol. 4, July 2010, pp. 225–229.
- [47] J.-T. Chien, C.-H. Huang, and S.-J. Chen, “Compact decision trees with cluster validity for speech recognition,” in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 1, May 2002, pp. I-873–I-876.
- [48] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain,” *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958.
- [49] C. Neocleous and C. Schizas, “Artificial neural network learning: A comparative review,” in *Methods and Applications of Artificial Intelligence*, ser. Lecture Notes in Computer Science, I. Vlahavas and C. Spyropoulos, Eds. Springer Berlin Heidelberg, 2002, vol. 2308, pp. 300–313. [Online]. Available:
- [50] D. O. Hebb, *The Organization of Behavior*. Wiley & Sons, 1949.
- [51] S. Russell and P. Norvi, *Artificial Intelligence: A Modern Approach*, 3rd ed. Pearson, December 2009.
- [52] D. Ma, J. Liu, and Z. Wang, “The pattern classification based on fuzzy min-max neural network with new algorithm,” in *Advances in Neural Networks – ISNN 2012*, ser. Lecture Notes in Computer Science, J. Wang, G. Yen, and M. Polycarpou, Eds. Springer Berlin Heidelberg, 2012, vol. 7368, pp. 1–9. [Online]. Available:
- [53] P. Simpson, “Fuzzy min-max neural networks. i. classification,” *Neural Networks, IEEE Transactions on*, vol. 3, no. 5, pp. 776–786, Sep 1992.
- [54] K. Rennolls, A. Soper, P. Robbins, and R. Guthrie, “Stochastic neural networks,” in *ICANN '93*, S. Gielen and B. Kappen, Eds. Springer London, 1993, pp. 481–481.

- [55] G. Winter, *Genetic Algorithms in Engineering and Computer Science*. Wiley, 1995.
- [56] R. Linsker, “Self-organization in a perceptual network,” *Computer*, vol. 21, no. 3, pp. 105–117, March 1988.
- [57] T.-W. Lee, *Independent Component Analysis: Theory and Applications*, 1st ed. Springer US, 1998.
- [58] S. Marinai, M. Gori, and G. Soda, “Artificial neural networks for document analysis and recognition,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 1, pp. 23–35, Jan 2005.
- [59] L. Goodenday, K. Cios, and I. Shin, “Identifying coronary stenosis using an image-recognition neural network,” *Engineering in Medicine and Biology Magazine, IEEE*, vol. 16, no. 5, pp. 139–144, Sept 1997.
- [60] S.-Y. Kung and J.-N. Hwang, “Neural networks for intelligent multimedia processing,” *Proceedings of the IEEE*, vol. 86, no. 6, pp. 1244–1272, Jun 1998.
- [61] J. Benitez, J. Castro, and I. Requena, “Are artificial neural networks black boxes?” *Neural Networks, IEEE Transactions on*, vol. 8, no. 5, pp. 1156–1164, Sep 1997.
- [62] D. A. Freedman, *Statistical Models: Theory and Practice*, 2nd ed. Cambridge University Press, Apr. 2009.
- [63] G. A. F. Seber and A. J. Lee, *Linear Regression Analysis*, 1st ed. Wiley & Sons, 2003.
- [64] D. A. S. Fraser, A. Wong, and J. Wu, “Regression analysis, nonlinear or nonnormal: Simple and accurate p values from likelihood analysis,” *Journal of the American Statistical Association*, vol. 94, no. 448, pp. 1286–1295, Dec. 1999.
- [65] V. Bewick, L. Cheek, and J. Ball, “Statistics review 14: Logistic regression,” *Crit Care*, vol. 9, no. 1, pp. 112–118, 2005.
- [66] O. Rudovic, I. Patras, and M. Pantic, “Regression-based multi-view facial expression recognition,” in *Pattern Recognition (ICPR), 2010 20th International Conference*

- on, Aug 2010, pp. 4121–4124.
- [67] W. Lee, C. Sekhar, K. Takeda, and F. Itakura, “Multiple regression using support vector machines for recognition of speech in a moving car environment,” in *Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference on*, vol. 2, Nov 2002, pp. 904–908 vol.2.
- [68] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
- [69] G. Mountrakis, J. Im, and C. Ogole, “Support vector machines in remote sensing: A review,” *{ISPRS} Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247 – 259, 2011.
- [70] P. Mantero, G. Moser, and S. Serpico, “Partially supervised classification of remote sensing images through svm-based probability density estimation,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 3, pp. 559–570, March 2005.
- [71] L. Guo, J. Wang, Q. Fang, X. Gu, and J. Fu, “Motion recognition for unsupervised hand rehabilitation using support vector machine,” in *Biomedical Circuits and Systems Conference (BioCAS), 2012 IEEE*, Nov 2012, pp. 104–107.
- [72] C. Wang, L. Lan, Y. Zhang, and M. Gu, “Face recognition based on principle component analysis and support vector machine,” in *Intelligent Systems and Applications (ISA), 2011 3rd International Workshop on*, May 2011, pp. 1–4.
- [73] B. Sonkamble and D. Doye, “An overview of speech recognition system based on the support vector machines,” in *Computer and Communication Engineering, 2008. ICCCE 2008. International Conference on*, May 2008, pp. 768–771.
- [74] R. Vilalta and Y. Drissi, “A perspective view and survey of meta-learning,” *Artificial Intelligence Review*, vol. 18, no. 2, pp. 77–95, 2002.
- [75] L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [76] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*, 1st ed. Chapman &

- Hall/CRC, 2012.
- [77] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119 – 139, 1997.
- [78] M. Ichino, N. Komatsu, W. Jian-Gang, and Y. W. Yun, “Speaker gender recognition using score level fusion by adaboost,” in *Control Automation Robotics Vision (ICARCV), 2010 11th International Conference on*, Dec 2010, pp. 648–653.
- [79] C. Ruan, Q. Ruan, and X. Li, “Real adaboost feature selection for face recognition,” in *Signal Processing (ICSP), 2010 IEEE 10th International Conference on*, Oct 2010, pp. 1402–1405.
- [80] Y. Li, B. Zhang, S. Shan, X. Chen, and W. Gao, “Bagging based efficient kernel fisher discriminant analysis for face recognition,” in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 3, 2006, pp. 523–526.
- [81] Z. Bin, “Handwritten chinese similar characters recognition based on adaboost,” in *Control Conference, 2007. CCC 2007. Chinese*, July 2007, pp. 576–579.
- [82] H. Nishimura, M. Kobayashi, M. Maruyama, and Y. Nakano, “Off-line character recognition using hmm by multiple directional feature extraction and voting with bagging algorithm,” in *Document Analysis and Recognition, 1999. ICDAR '99. Proceedings of the Fifth International Conference on*, Sep 1999, pp. 49–52.
- [83] X.-H. Han, Y. wei Chen, and X. Ruan, “Image recognition by learned linear subspace of combined bag-of-features and low-level features,” in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, Sept 2010, pp. 1049–1052.
- [84] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*, 1st ed. Prentice Hall, 1982.
- [85] D. Sander, D. Grandjean, and K. R. Scherer, “2005 special issue: A systems approach to appraisal mechanisms in emotion,” *Neural Networks*, vol. 18, no. 4, pp. 317–352, May 2005.

- [86] K. Drossos, A. Floros, and A. Giannakouloupoulos, “Beads: A dataset of binaural emotionally annotated digital sounds,” in *Information, Intelligence, Systems and Applications, IISA 2014, The 5th International Conference on*, July 2014, pp. 158–163.
- [87] T. D. Wager, L. F. Barrett, E. Bliss-Moreau, K. Lindquist, , S. Duncan, H. Kober, J. Joseph, M. Davidson, and J. Mize, “The neuroimaging of emotion,” pp. 249–271, 2000.
- [88] P. Ekman, “An argument for basic emotions,” *Cognition & Emotion*, vol. 6, no. 3-4, pp. 169–200, Jan 1992.
- [89] R. Adolphs, “Neural systems for recognizing emotion,” *Current Opinion in Neurobiology*, vol. 12, no. 2, pp. 169 – 177, 2002.
- [90] D. Hagemann, S. R. Waldstein, and J. F. Thayer, “Central and autonomic nervous system integration in emotion,” *Brain and Cognition*, vol. 52, no. 1, pp. 79 – 87, 2003, affective Neuroscience.
- [91] S. Koelsch, “Investigating emotion with music neuroscientific approaches,” *Annals of the New York Academy of Science*, vol. 1060, pp. 1–7, 2005.
- [92] A. O. A. Ortony and T. J. Turner, “What’s basic about basic emotions?” *Psychological Review*, vol. 97, no. 3, pp. 315–331, Jul 1990.
- [93] J. Posner, J. A. Russell, and B. S. Peterson, “The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology.” *Development and Psychopathology*, vol. 17, no. 3, pp. 715–734, Sept. 2005.
- [94] K. Hevner, “Experimental studies of the elements of expression in music,” *The American Journal of Psychology*, vol. 48, no. 2, pp. 246–268, Apr. 1936.
- [95] A. Wiczorkowska, P. Synak, R. Lewis, and Z. W. Raś, “Extracting emotions from music data,” in *Foundations of Intelligent Systems*, ser. Lecture Notes in Computer Science, M.-S. Hacid, N. V. Murray, Z. W. Raś, and S. Tsumoto, Eds. Springer

- Berlin Heidelberg, 2005, vol. 3488, pp. 456–465.
- [96] T. Li and M. Ogihara, “Detecting emotion in music,” in *Proceedings of the International Symposium on Music Information Retrieval*, 2003, pp. 239–240.
- [97] J. A. Russel, “A circumplex model of affect,” *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.
- [98] J. Sanghoon, S. Rho, B.-j. Han, and E. Hwang, “A fuzzy inference-based music emotion recognition system,” in *Visual Information Engineering, 2008. VIE 2008. 5th International Conference on*, July 2008, pp. 673–677.
- [99] K. Drossos, A. Floros, A. Giannakoulopoulos, and N. Kanellopoulos, “Investigating the impact of sound angular position on the listener affective state,” *Transactions on Affective Computing*, vol. 6, no. 1, pp. 27–42, Jan 2015.
- [100] C. A. Smith and P. C. Ellsworth, “Patterns of cognitive appraisal in emotion,” *Journal of Personality & Social Psychology*, vol. 48, no. 4, pp. 813–838, Apr. 1985.
- [101] C. Stickel, M. Ebner, S. Steinbach-Nordmann, G. Searle, and A. Holzinger, “Emotion detection: Application of the valence arousal space for rapid biological usability testing to enhance universal access,” in *Universal Access in Human-Computer Interaction. Addressing Diversity*, ser. Lecture Notes in Computer Science, C. Stephanidis, Ed. Springer Berlin Heidelberg, 2009, vol. 5614, pp. 615–624.
- [102] M. M. Bradley and P. J. Lang, “Measuring emotion: The self-assessment manikin and the semantic differential,” *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49 – 59, 1994.
- [103] —, “International affective digitized sounds (iads): Stimuli, instruction manual and affective ratings,” The Center for Research in Psychophysiology, University of Florida, Gainesville, Fl, Tech. Rep. B-2, 1999.
- [104] S. Brilis, E. Gkatzou, A. Koursoumis, K. Talvis, K. Kermanidis, and I. Karydis, “Mood classification using lyrics and audio: A case-study in greek music,” in *Artificial Intelligence Applications and Innovations*, ser. IFIP Advances in Informa-

- tion and Communication Technology, L. Iliadis, I. Maglogiannis, H. Papadopoulos, K. Karatzas, and S. Sioutas, Eds. Springer Berlin Heidelberg, 2012, vol. 382, pp. 421–430.
- [105] C. Laurier, J. Grivolla, and P. Herrera, “Multimodal music mood classification using audio and lyrics,” in *International Conference on Machine Learning and Applications*, San Diego, California (USA), 11/12/2008 2008.
- [106] O. Lartillot, P. Toivainen, and T. Eerola, “A matlab toolbox for music information retrieval,” in *Data Analysis, Machine Learning and Applications*, ser. Studies in Classification, Data Analysis, and Knowledge Organization, C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker, Eds. Springer Berlin Heidelberg, 2008, pp. 261–268.
- [107] M. Barthelet, G. Fazekas, and M. Sandler, “Music emotion recognition: From content- to context-based models,” in *From Sounds to Music and Emotions*, ser. Lecture Notes in Computer Science, M. Aramaki, M. Barthelet, R. Kronland-Martinet, and S. Ystad, Eds. Springer Berlin Heidelberg, 2013, vol. 7900, pp. 228–252.
- [108] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” *Speech and Audio Processing, IEEE Transactions on*, vol. 10, no. 5, pp. 293–302, Jul 2002.
- [109] P. Mermelstein, “Distance measures for speech recognition-psychological and instrumental,” in *Pattern recognition and artificial intelligence*, C. H. Chen, Ed. Academic Press, 1976, pp. 374–388.
- [110] M. Hossan, S. Memon, and M. Gregory, “A novel approach for mfcc feature extraction,” in *Signal Processing and Communication Systems (ICSPCS), 2010 4th International Conference on*, Dec 2010, pp. 1–5.
- [111] K. Krishna Kishore and P. Krishna Satish, “Emotion recognition in speech using mfcc and wavelet features,” in *Advance Computing Conference (IACC), 2013 IEEE 3rd International*, Feb 2013, pp. 842–847.
- [112] T. Pohle, E. Pampalk, and G. Widmer, “Evaluation of frequently used audio features for classification of music into perceptual categories.” in *In Proceedings of the Fourth*

- International Workshop on Content-Based Multimedia Indexing (CBMI'05)*, Jun. 2005.
- [113] L. Cai, M. Xiu-Li, W. Wang-gen, and Z. Yan-fei, “A research of objective evaluation method for audio brightness,” in *Audio Language and Image Processing (ICALIP), 2010 International Conference on*, Nov 2010, pp. 69–73.
- [114] P. N. Juslin, “Cue utilization in communication of emotion in music performance: relating performance to perception,” *Journal of Experimental Psychology*, vol. 26, no. 6, pp. 1797–1813, 2000.
- [115] P. Laukka, P. Juslin, and R. Bresin, “A dimensional approach to vocal expression of emotion,” *Cognition & Emotion*, vol. 19, no. 5, pp. 633–653, 2005.
- [116] A. Kameoka and M. Kuriyagawa, “Consonance theory part i: Consonance of dyads,” *The Journal of the Acoustical Society of America*, vol. 45, no. 6, pp. 1451–1459, 1969.
- [117] J. Foote and S. Uchihashi, “The beat spectrum: a new approach to rhythm analysis,” in *Multimedia and Expo, 2001. ICME 2001. IEEE International Conference on*, Aug 2001, pp. 881–884.
- [118] J. Foote, M. Cooper, and U. Nam, “Audio retrieval by rhythmic similarity,” in *Proceedings of the International Conference on Music Information Retrieval*, 2002, pp. 265–266.
- [119] E. Pampalk, A. Rauber, and D. Merkl, “Content-based organization and visualization of music archives,” in *Proceedings of the Tenth ACM International Conference on Multimedia*, ser. MULTIMEDIA '02. New York, NY, USA: ACM, 2002, pp. 570–579.
- [120] H. Fastl, “Fluctuation strength and temporal masking patterns of amplitude-modulated broadband noise,” *Hearing Research*, vol. 8, no. 1, pp. 59 – 69, 1982.
- [121] L. Austin, “Sound diffusion in composition and performance: An interview with denis smalley,” *Comput. Music J.*, vol. 24, no. 2, pp. 10–21, Jul. 2000.
- [122] G. Kendall and M. Ardila, “The artistic play of spatial organization: Spatial at-

- tributes, scene analysis and auditory spatial schemata,” in *Computer Music Modeling and Retrieval. Sense of Sounds*, ser. Lecture Notes in Computer Science, R. Kronland-Martinet, S. Ystad, and K. Jensen, Eds. Springer Berlin Heidelberg, 2008, vol. 4969, pp. 125–138.
- [123] F. Wiering, “Can humans benefit from music information retrieval?” in *Adaptive Multimedia Retrieval: User, Context, and Feedback*, ser. Lecture Notes in Computer Science, S. Marchand-Maillet, E. Bruno, A. Nürnberger, and M. Detyniecki, Eds. Springer Berlin Heidelberg, 2007, vol. 4398, pp. 82–94.
- [124] D. Kelly, “Methods for evaluating interactive information retrieval systems with users,” *Found. Trends Inf. Retr.*, vol. 3, no. 1-2, pp. 1–224, Jan. 2009.
- [125] J. Downie, “The scientific evaluation of music information retrieval systems: Foundations and future,” *Computer Music Journal*, vol. 28, no. 2, pp. 12–23, June 2004.
- [126] M. Lesaffre, M. Leman, and J.-P. Martens, “A user-oriented approach to music information retrieval,” in *Content-Based Retrieval*, ser. Dagstuhl Seminar Proceedings, T. Crawford and R. C. Veltkamp, Eds., no. 06171. Dagstuhl, Germany: Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 2006.
- [127] S. Zhang, Q. Tian, S. Jiang, Q. Huang, and W. Gao, “Affective mtv analysis based on arousal and valence features,” in *Multimedia and Expo, 2008 IEEE International Conference on*, June 2008, pp. 1369–1372.
- [128] P. Ekman, R. W. Levenson, and W. V. Friesen, “Autonomic nervous system activity distinguishes among emotions,” *Science*, vol. 221, no. 4616, pp. 1208–1210, Sep 1983.
- [129] J. T. Cacioppo, G. G. Berntson, K. M. Poehlmann, and T. A. Ito, “The psychophysiology of emotion,” in *Handbook of Emotions*, 3rd ed., M. Lewis, J. M. Haviland-Jones, and L. F. Barrett, Eds. Guilford Press, 2000, ch. 11, pp. 180–195.
- [130] K. R. Scherer, “Appraisal theory,” in *Handbook of Cognition and Emotion*. John Wiley & Sons, 1999, ch. 30, pp. 637–663.

- [131] —, “Expression of emotion in voice and music,” *Journal of Voice*, vol. 9, no. 3, pp. 235 – 248, 1995.
- [132] T. Li and M. Ogihara, “Content-based music similarity search and emotion detection,” in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, vol. 5, May 2004, pp. V-705–8 vol.5.
- [133] C.-H. Yeh, H.-H. Lin, and H.-T. Chang, “An efficient emotion detection scheme for popular music,” in *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*, May 2009, pp. 1799–1802.
- [134] D. Liu, L. Lu, and H.-J. Zhang, “Automatic mood detection from acoustic music data,” in *ISMIR (International Conference on Music Information Retrieval) 2003*. The Johns Hopkins University, Oct 2003, pp. 81–87.
- [135] M. Bartoszewski, H. Kwasnicka, U. Markowska-Kaczmar, and P. Myszkowski, “Extraction of emotional content from music data,” in *Computer Information Systems and Industrial Management Applications, 2008. CISIM '08. 7th*, June 2008, pp. 293–299.
- [136] Y.-H. Yang, C.-C. Liu, and H. H. Chen, “Music emotion classification: A fuzzy approach,” in *Proceedings of the 14th Annual ACM International Conference on Multimedia*, ser. MULTIMEDIA '06. New York, NY, USA: ACM, 2006, pp. 81–84.
- [137] Y. H. Yang, Y. C. Lin, H. T. Cheng, and H. H. Chen, “Mr. emo: Music retrieval in the emotion plane,” in *Proceedings of the 16th ACM International Conference on Multimedia*, ser. MM '08. New York, NY, USA: ACM, 2008, pp. 1003–1004.
- [138] Y. Feng, Y. Zhuang, and Y. Pan, “Popular music retrieval by detecting mood,” in *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, ser. SIGIR '03. New York, NY, USA: ACM, 2003, pp. 375–376.
- [139] K. F. Macdorman, S. Ough, and C. chang Ho, “Automatic emotion prediction of song excerpts: Index construction, algorithm design, and empirical comparison,” *Journal of New Music Research*, vol. 36, no. 4, pp. 281–299, 2007.

- [140] C. Laurier¹, O. Lartillot, T. Eerola, and P. Toivaine, “Exploring relationships between audio features and emotion in music,” in *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009)*, Aug. 2009, pp. 260–264.
- [141] W. Muyuan, Z. Naiyao, and Z. Hancheng, “User-adaptive music emotion recognition,” in *Signal Processing, 2004. Proceedings. ICSP '04. 2004 7th International Conference on*, vol. 2, Aug 2004, pp. 1352–1355 vol.2.
- [142] W. L. Cheung and G. Lu, “Music emotion annotation by machine learning,” in *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*, Oct 2008, pp. 580–585.
- [143] S. Krothapalli and S. Koolagudi, “Speech emotion recognition: A review,” in *Emotion Recognition using Speech Features*, ser. SpringerBriefs in Electrical and Computer Engineering. Springer New York, 2013, pp. 15–34.
- [144] S. Ramakrishnan, “Recognition of emotion from speech: A review,” in *Speech Enhancement, Modeling and Recognition- Algorithms and Applications*, S. Ramakrishnan, Ed. InTech, 2012, ch. 7, pp. 121–138.
- [145] R. B. Lanjewar and D. S. Chaudhari, “Speech emotion recognition: A review,” *Internation Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 2, no. 4, pp. 68–71, Mar. 2013.
- [146] T. Vogt, E. André, and J. Wagner, “Automatic recognition of emotions from speech: A review of the literature and recommendations for practical realisation,” in *Affect and Emotion in Human-Computer Interaction*, ser. Lecture Notes in Computer Science, C. Peter and R. Beale, Eds. Springer Berlin Heidelberg, 2008, vol. 4868, pp. 75–91.
- [147] R. Banse and K. R. Scherer, “Acoustic profiles in vocal emotion expression,” *Journal of Personality and Social Psychology*, vol. 70, no. 3, pp. 614–636, 1996.
- [148] A. Sarrafzadeh, S. Alexander, F. Dadgostar, C. Fan, and A. Bigdeli, ““how do you know that i don’t understand?” a look at the future of intelligent tutoring systems,”

- Computers in Human Behavior*, vol. 24, no. 4, pp. 1342 – 1363, 2008, including the Special Issue: Integration of Human Factors in Networked Computing.
- [149] P. Vlamos, A. Floros, M. N. Giannakos, and K. Drossos, “Towards an interactive e-learning system based on emotions and affective cognition,” in *Readings in Technology and Education: Proceedings of ICICTE 2010*, Jul 2010.
- [150] B. Schuller, S. Hantke, F. Weninger, W. Han, Z. Zhang, and S. Narayanan, “Automatic recognition of emotion evoked by general sound events,” in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, March 2012, pp. 341–344.
- [151] M. Grimm, K. Kroschel, E. Mower, and S. Narayanan, “Primitives-based evaluation and estimation of emotions in speech,” *Speech Commun.*, vol. 49, no. 10-11, pp. 787–800, Oct. 2007.
- [152] F. Hagman, “Emotional response to sound - influence of spatial determinants,” Master’s thesis, Chalmers University of Technology, Göteborg, 2010.
- [153] T. Bänziger, M. Mortillaro, and K. R. Scherer, “Introducing the geneva multimodal expression corpus for experimental research on emotion perception (gemep),” *Emotion*, vol. 12, no. 5, pp. 1161–1179, 2012.
- [154] C. Stickel, M. Ebner, S. Steinbach-Nordmann, G. Searle, and A. Holzinger, “Emotion detection: Application of the valence arousal space for rapid biological usability testing to enhance universal access,” in *Universal Access in Human-Computer Interaction. Addressing Diversity*, ser. Lecture Notes in Computer Science, C. Stephanidis, Ed. Springer Berlin Heidelberg, 2009, vol. 5614, pp. 615–624.
- [155] S. Giripunje and N. Bawane, “Anfis based emotions recognition in speech,” in *Knowledge-Based Intelligent Information and Engineering Systems*, ser. Lecture Notes in Computer Science, B. Apolloni, R. Howlett, and L. Jain, Eds. Springer Berlin Heidelberg, 2007, vol. 4692, pp. 77–84.
- [156] B. Y. Newman, “And now, acoustic ecology,” *Optometry - Journal of the American Optometric Association*, vol. 76, no. 11, pp. 629–631, November 2005.

- [157] R. M. Schafer, *The Tuning of the World*. Knopf, 1977.
- [158] K. Drossos and A. Floros, “An approach for modeling affective acoustic ecology in city environments,” in *EchoPolis - Days of Sound : “Sounds, Noise and Music in re-thinking sustainable city and the eco-neighborhoods”*, Sept. 29-30 2013.
- [159] J. Blauert, *Spatial Hearing - Revised Edition: The Psychophysics of Human Sound Localization*, revised ed., J. S. Allen, Ed. Massachusetts, U.S.A.: M.I.T. Press, 1983.
- [160] K. R. Scherer, “Vocal affect signalling: A comparative approach,” in *Advances in the study of behavior*, J. Rosenblatt, C. Beer, M.-C. Busnel, and P. J. B. Slater, Eds. Academic Press, 1985, vol. 15, pp. 189–224.
- [161] M. Borchert and A. Dusterhoft, “Emotions in speech - experiments with prosody and quality features in speech for use in categorical and dimensional emotion recognition environments,” in *Natural Language Processing and Knowledge Engineering, 2005. IEEE NLP-KE '05. Proceedings of 2005 IEEE International Conference on*, Oct 2005, pp. 147–151.
- [162] T. L. Nwe, S. W. Foo, and L. C. D. Silva, “Speech emotion recognition using hidden markov models,” *Speech Communication*, vol. 41, no. 4, pp. 603 – 623, 2003.
- [163] R. Kotsakis, G. Kalliris, and C. Dimoulas, “Investigation of broadcast-audio semantic analysis scenarios employing radio-programme-adaptive pattern classification,” *Speech Communication*, vol. 54, no. 6, pp. 743 – 762, 2012.
- [164] C. Dimoulas, G. Papanikolaou, and V. Petridis, “Pattern classification and audiovisual content management techniques using hybrid expert systems: A video-assisted bioacoustics application in abdominal sounds pattern analysis,” *Expert Systems with Applications*, vol. 38, no. 10, pp. 13 082 – 13 093, 2011.
- [165] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, 2nd ed., W. A. Shewhart and S. S. Wilks, Eds. John Wiley & Sons, Inc., 2005.
- [166] K. Drossos, A. Giannakouloupoulos, and A. Floros, “Web based platform for emotion

- evaluation from audio data,” in *6th National Conference ACOUSTICS 2012*, 2012, pp. 250–257.
- [167] M. Bryden, R. G. Ley, and J. Sugarman, “A left-ear advantage for identifying the emotional quality of tonal sequences,” *Neuropsychologia*, vol. 20, no. 1, pp. 83 – 87, 1982.
- [168] R. Schatz, S. Egger, and K. Masuch, “The impact of test duration on user fatigue and reliability of subjective quality ratings,” *J. Audio Eng. Soc*, vol. 60, no. 1/2, pp. 63–73, 2012.
- [169] W. Iba and P. Langley, “Induction of one-level decision trees,” in *Proceedings of the Ninth International Workshop on Machine Learning*, ser. ML '92. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1992, pp. 233–240.
- [170] W. Mohamed, M. Salleh, and A. Omar, “A comparative study of reduced error pruning method in decision tree algorithms,” in *Control System, Computing and Engineering (ICCSCE), 2012 IEEE International Conference on*, Nov 2012, pp. 392–397.
- [171] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.
- [172] Y.-H. Yang and H. H. Chen, *Music Emotion Recognition*. Boca Raton, FL, USA: CRC Press, Taylor and Francis Group, 2011.