

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ - UNIVERSITY OF THESSALY



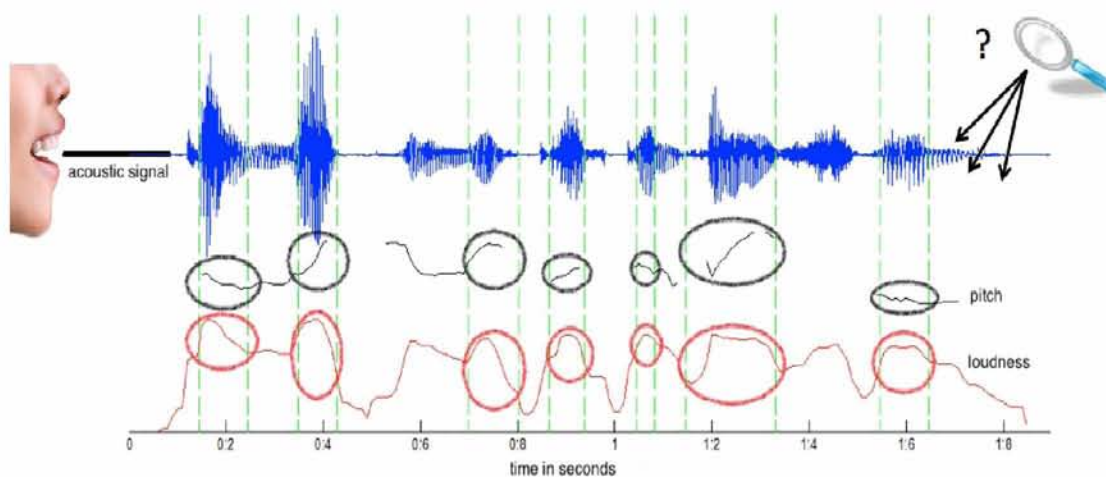
University of Thessaly

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ - DIPLOMA THESIS

## Αναγνώριση Συναισθήματος από το Σήμα του Λόγου Emotion Recognition from the Speech Signal



Αναστάσιος Ρήγας - Anastasios Rigas

Επιβλέπων: **Δρ. Γεράσιμος Ποταμιάνος**, Αναπληρωτής Καθηγητής Πανεπιστημίου Θεσσαλίας

**ΒΟΛΟΣ, ΙΟΥΝΙΟΣ 2013**

# Ευχαριστίες - Acknowledgements

Η ολοκλήρωση της παρούσας Διπλωματικής Εργασίας σηματοδοτεί και το τέλος των προπτυχιακών μου σπουδών και θα ήθελα να ευχαριστήσω κάποια άτομα, τα οποία υπήρξαν δίπλα μου όλα αυτά τα χρόνια. Σε πρώτη φάση, θα ήθελα να ευχαριστήσω τον κο Γεράσιμο Ποταμιάνο, ο οποίος ως επιβλέπων της Διπλωματικής μου Εργασίας, με βοήθησε με την καθοδήγηση και τις συμβουλές του, ως προς τη διαδικασία της συγγραφής της τρέχουσας Διπλωματικής Εργασίας, οποτεδήποτε και αν αυτό χρειάστηκε και ζητήθηκε από την πλευρά μου. Σε εκείνες τις περιπτώσεις, στήριξε την προσπάθειά μου, προτείνοντάς μου λύσεις σε τυχόν προβλήματα που προέκυπταν κατά τη διάρκεια της συγγραφής της Διπλωματικής μου Εργασίας, ακούγοντας τις ιδέες και τις προτάσεις μου, όπως τις ανέφερα κάθε φορά, και δείχνοντας ενδιαφέρον, αλλά και εμπιστοσύνη στην προσωπική δουλειά μου, σε διαστήματα μη εφικτής οπτικής επικοινωνίας. Αναμφισβήτητα όμως, η σημαντικότερη βοήθεια προήλθε από το περιβάλλον της οικογένειάς μου! Πάντα στο πλευρό μου, στις ευχάριστες, αλλά κυρίως στις δύσκολες στιγμές, δε σταμάτησαν να μου δείχνουν την αγάπη, την πίστη και την υποστήριξή τους όλα αυτά τα χρόνια των σπουδών μου και τους ευχαριστώ! Ασφαλώς, ένα μεγάλο ευχαριστώ οφείλω και στους φίλους μου, οι οποίοι υπήρξαν πάντα στο πλευρό μου, δίπλα μου και στη διάθεσή μου, όποτε τους είχα ανάγκη, πρόθυμοι να με ακούσουν, να με συμβουλεύσουν και να με ενθαρρύνουν, με διάφορους τρόπους και μέσα, ως προς την "keep walking ανηφόρα"...

Αναστάσιος Ρήγας

Βόλος, Ιούνιος 2013

# Περίληψη

Ο στόχος της παρούσας Διπλωματικής Εργασίας είναι τόσο η μελέτη όσο και η πειραματική έρευνα πάνω στην Αναγνώριση Συναισθήματος από το Σήμα του Λόγου του ομιλητή-χρήστη ενός συστήματος/μιας εφαρμογής Αλληλεπίδρασης Ανθρώπου-Υπολογιστή. Το εν λόγω θέμα αποτελεί κομμάτι της "Συναισθηματικής Υπολογιστικής" που επικεντρώνεται στη βελτίωση της φυσικότητας και της ρεαλιστικότητας των υπό ανάπτυξη εφαρμογών Αλληλεπίδρασης Ανθρώπου-Μηχανής. Αναλυτικά, όσον αφορά τη δομή της: Στο πρώτο κεφάλαιο, ο αναγνώστης της παρούσας Διπλωματικής Εργασίας εισάγεται σε βασικές έννοιες της Αναγνώρισης Συναισθήματος από το Σήμα του Λόγου. Στο δεύτερο, μελετώνται διαφορετικά είδη εξαγόμενων από το Σήμα του Λόγου γνωρισμάτων που έχει αποδειχτεί από την επιστημονική/ερευνητική κοινότητα της Επεξεργασίας Σήματος Λόγου ότι φέρουν πολύτιμη Συναισθηματική πληροφορία, και επίσης οι δύο ευρύτερα χρησιμοποιημένες μέθοδοι Ακολουθιακής Επιλογής Γνωρισμάτων: η Προς τα Εμπρός/Εμπρόσθια και η Προς τα Πίσω/Οπισθοδρομική, σε αλγοριθμικό επίπεδο, αλλά και ως προς τα πλεονεκτήματά ή/και μειονεκτήματά τους. Στο τρίτο, μελετώνται συνοπτικά οι κυριότερες πτυχές των περισσότερων, ευρέως χρησιμοποιημένων ταξινομητών, και αναλυτικότερα, των  $k$  Πλησιέστερων Γειτόνων, της γενικής μορφής της Μηχανής Διανύσματος Υποστήριξης, καθώς και των Μία-Έναντι-Μίας και Μία-Έναντι-Υπολοίπων τύπων της, που χρησιμοποιούνται για ταξινόμηση με ένα συνολικό πλήθος κλάσεων μεγαλύτερο του 2. Στο τέταρτο, ακολουθούν τα δύο πειράματα της παρούσας Διπλωματικής Εργασίας, με τα αποτελέσματά τους. Στο πέμπτο, αναλύονται ενδεχόμενα συμπεράσματά μας από τα αποτελέσματα των δύο πειραμάτων. Κλείνοντας, στο έκτο, αναφέρονται ορισμένες ενδιαφέρουσες προκλήσεις για την Αναγνώριση Συναισθήματος και σχολιάζονται ενδεχόμενες μελλοντικές επεκτάσεις μας προς έρευνα.

# Abstract

The aim of the present Diploma Thesis is the study as well as the experimental research on the Emotion Recognition from the Speech Signal of the speaker-user of a Human-Computer Interaction (HCI) system/application. This specific subject constitutes a part of the so-called research area "Affective Computing", which focuses on the improvement of the naturality and realism of the under development Human-Machine applications. Analytically, as far as its structure is concerned:

In the first chapter, the reader of the present Diploma Thesis gets introduced in basic concepts of the Emotion Recognition from the Speech Signal. In the second, different kinds of extracted from the Speech Signal features, which have been proved by the scientific/research community of Speech Signal Processing to carry important emotional information, as well as the two most widely used Sequential Feature Selection methods: Forward (SFS) and Backward (SBS), are analyzed, at algorithmic level, and according to their advantages or/and disadvantages. In the third, the most important aspects of the most widely used classifiers are briefly examined, and others, such as k Nearest Neighbors (kNN), the general form of Support Vector Machine, as well as the 1-VS-1 and 1-VS-Rest SVM types for classification of a total number of classes bigger than 2, are analyzed. In the fourth, the two experiments of the present Diploma Thesis follow, with their results. In the fifth, our potential conclusions from the results of the two experiments are extracted. Finally, in the sixth, some interesting challenges on the Emotion Recognition are referenced and our possible future research extensions are commented.

# Περιεχόμενα

<b>Ευχαριστίες-Acknowledges</b>	<b>2</b>
<b>Περίληψη</b>	<b>3</b>
<b>Abstract</b>	<b>4</b>
<b>Κεφάλαιο 1: Εισαγωγή στην Αναγνώριση Συναισθήματος από το Σήμα του Λόγου</b>	<b>10</b>
1.1 Σύγχυση Συναισθημάτων (Emotions Confusion/Ambiguity) και οι λόγοι εμφάνισής της . . .	11
1.2 Είδη Συναισθηματικού Λόγου και πλεονεκτήματά/μειονεκτήματά τους . . . . .	11
1.3 "Πρωτεύοντα"/"Αρχέτυπα" Συναισθήματα σύμφωνα με τη "Θεωρία της Παλέτας" . . . . .	13
1.4 Διαφορετικές συνεχείς 2-διάστατες αναπαραστάσεις Συναισθημάτων . . . . .	15
1.5 Είδη Συστημάτων Αυτόματης Αναγνώρισης Συναισθήματος . . . . .	18
1.6 Εφαρμογές Συστημάτων Αυτόματης Αναγνώρισης Συναισθήματος . . . . .	19
<b>Κεφάλαιο 2: Μελέτη γνωρισμάτων και μεθόδων Επιλογής Γνωρισμάτων ως προς τα πλεονεκτήματά/μειονεκτήματά τους</b>	<b>21</b>
2.1 Εξαγωγή γνωρισμάτων σε επίπεδο άρθρωσης ή πλαισίου . . . . .	21
2.2 Προσωδιακά γνωρίσματα . . . . .	21
2.3 Διαμόρφωσης Πλάτους-Διαμόρφωσης Συχνότητας (AM-FM/Formants) . . . . .	24
2.4 Mel Frequency Cepstral Coefficients (MFCCs) . . . . .	24
2.5 Linear Predictive Coefficients (LPCs) . . . . .	27
2.6 Σχετικού Φασματικού Μετασχηματισμού (Relative Spectral Perceptual Linear Predictive) .	28
2.7 Φασματικά γνωρίσματα,φασματογράμμα(Spectral features,spectrogram) . . . . .	28

2.8 Τιμές γνωρισμάτων ανάλογα με το Συναισθημα . . . . .	29
2.9 Μέθοδοι Επιλογής Γνωρισμάτων (Feature Selection Methods) . . . . .	31
2.9.1 Ακολουθιακές μέθοδοι Επιλογής Γνωρισμάτων (Sequential methods) . . . . .	32
2.9.1.1 Προς τα Εμπρός/Εμπρόσθια(Sequential Forward Feature Selection,SFS)	32
2.9.1.2 Προς τα Πίσω/Οπισθοδρομική(Sequential Backward Feature Selection,SBS)	33
<b>Κεφάλαιο 3: Μελέτη ταξινομητών ως προς τα πλεονεκτήματα/μειονεκτήματα</b>	<b>34</b>
3.1 Μηχανή Διανύσματος Υποστήριξης (Support Vector Machine, SVM) . . . . .	34
3.1.1 Μηχανή Διανύσματος Υποστήριξης Μία-Έναντι-Μίας (1-VS-1 SVM) . . . . .	37
3.1.2 Μηχανή Διανύσματος Υποστήριξης Μία-Έναντι-Υπολοίπων (1-VS-Rest SVM) . . . . .	37
3.2 k Πλησιέστεροι Γείτονες (k Nearest Neighbors, kNN) . . . . .	38
3.3 k-μέσων ομαδοποίηση (k-means clustering) . . . . .	40
3.4 Bayesian/Belief Δίκτυα (Bayesian/Belief Network, BN) . . . . .	40
3.5 Κρυμμένα Μαρκοβιανά Μοντέλα (Hidden Markov Models, HMMs) . . . . .	40
3.5 Μεικτά Γκαουσιανά Μοντέλα (Gaussian Mixture Models, GMMs) . . . . .	41
3.7 Γραμμική Διακριτή Ανάλυση (Linear Discriminant Analysis) . . . . .	41
3.8 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks, ANNs) . . . . .	41
<b>Κεφάλαιο 4: Τα πειράματά μας</b>	<b>43</b>
4.1 Βάση δεδομένων Interactive Emotion Capture (IEmoCap) . . . . .	43
4.2 Εφαρμοσμένη μεθοδολογία και για τα δύο πειράματα . . . . .	44
4.2.1 Επιλογή κατάλληλου μέρους των συνολικών αρθρώσεων της IEmoCap . . . . .	44
4.2.2 Κωδικοποίηση Συναισθημάτων σε integer labels . . . . .	45
4.2.3 Αφαίρεση μη ηχηρών τμημάτων και εξαγωγή στατιστικών σε επίπεδο άρθρωσης . . . . .	45
4.2.4 Καθορισμός training, testing sets . . . . .	46
4.3 Πρώτο πείραμα: Σύγκριση SFS, SBS, με κριτήριο σύγκλισής τους, την ελαχιστοποίηση σφάλματος ταξινόμησης του INN, στην IEmoCap . . . . .	47
4.3.1 Χρησιμοποιημένα στατιστικά στο πρώτο πείραμα . . . . .	47
4.3.2 Εφαρμογή SFS στο μεγάλο Διάνυσμα Γνωρισμάτων . . . . .	48

4.3.2.1 Πειραματισμός με επιλογή διαφορετικού πλήθους πιο σημαντικών στατιστικών κατά SFS και εύρεση του βέλτιστου . . . . .	48
4.3.2.2 Βέλτιστο πλήθος στατιστικών και συνεπαγόμενη ιεράρχηση γνωρισμάτων .	51
4.3.2.3 Αναλυτικά αποτελέσματα ταξινόμησης του συνδυασμού SFS + INN . . . . .	54
4.3.3 Εφαρμογή SBS στο μεγάλο Διάνυσμα Γνωρισμάτων . . . . .	56
4.3.3.1 Πειραματισμός με επιλογή διαφορετικού πλήθους πιο σημαντικών στατιστικών κατά SBS και εύρεση του βέλτιστου . . . . .	56
4.3.3.2 Το βέλτιστο πλήθος στατιστικών και συνεπαγόμενη ιεράρχηση γνωρισμάτων	59
4.3.3.3 Αναλυτικά αποτελέσματα ταξινόμησης του συνδυασμού SBS + INN . . . . .	62
4.4 Δεύτερο πείραμα: Σύγκριση των multi-class SVMs: 1-VS-1, 1-VS-Rest για την IEmoCap .	64
4.4.1 Χρησιμοποιημένα στατιστικά για το δεύτερο πείραμα . . . . .	64
4.4.2 Αναλυτικά αποτελέσματα ταξινόμησης αρθρώσεων με ταξινομητή 1-VS-1 SVM . . . . .	65
4.4.3 Αποτελέσματα ταξινόμησης με υλοποίηση για 1-VS-Rest SVM . . . . .	67
<b>Κεφάλαιο 5: Τα συμπεράσματά μας από τα δύο πειράματα</b>	<b>69</b>
5.1 Συμπεράσματα αποτελεσμάτων πρώτου πειράματος στην IEmoCap . . . . .	69
5.2 Συμπεράσματα αποτελεσμάτων δεύτερου πειράματος στην IEmoCap . . . . .	75
<b>Μελλοντικές ερευνητικές επεκτάσεις</b>	<b>78</b>
<b>References</b>	<b>78</b>

## Κατάλογος Σχημάτων

1	Συνεχής 2-διάστατη αναπαράσταση συναισθημάτων Valence-Arousal[1] . . . . .	16
2	Συνεχής 2-διάστατη αναπαράσταση συναισθημάτων Activation-Evaluation[2] . . . . .	17
3	Τροχός Συναισθημάτων του Plutchnik (Plutchnik's Emotions Wheel)[3] . . . . .	18

4	Κλίμακα Mel (Mel Scale)[4] . . . . .	25
5	Ταξινόμηση ενός δείγματος με ταξινομητή κλασική SVM[5] . . . . .	36
6	Σύγκριση τρόπου λειτουργίας ταξινόμησης ταξινομητών a)1-VS-Rest SVM,b)1-VS-1 SVM[17]	38
7	Ταξινόμηση ενός δείγματος με ταξινομητή 5NN[6] . . . . .	40
8	Ποσοστά χρησιμοποιημένων αρθρώσεων της IEmoCap ανά Συναισθημα . . . . .	45
9	Σύγκριση ορθά ταξινομημένων αρθρώσεων με 27, 61, 107, 159 στατιστικά κατά SFS + 1NN	50
10	Σύγκριση μέσης Αναγνώρισης Συναισθήματος με 27, 61, 107, 159 στατιστικά κατά SFS+1NN	51
11	Ιεράρχηση κατηγοριών εξαγόμενων γνωρισμάτων κατά SFS + 1NN για την IEmoCap . . .	54
12	Ορθά ταξινομημένες αρθρώσεις με το βέλτιστο πλήθος στατιστικών κατά SFS + 1NN . . .	55
13	Ποσοστά Αναγνώρισης Συναισθημάτων με το βέλτιστο πλήθος στατιστικών κατά SFS + 1NN	56
14	Σύγκριση ορθά ταξινομημένων αρθρώσεων με 27, 51, 107, 159 στατιστικά κατά SBS + 1NN	58
15	Σύγκριση μέσης Αναγνώρισης Συναισθήματος με 27, 51, 107, 159 στατιστικά κατά SBS+1NN	59
16	Ιεράρχηση κατηγοριών εξαγόμενων γνωρισμάτων κατά SBS για την IEmoCap . . . . .	62
17	Ορθά ταξινομημένες αρθρώσεις με το βέλτιστο πλήθος στατιστικών κατά SBS + 1NN . . .	63
18	Ποσοστά Αναγνώρισης Συναισθημάτων με το βέλτιστο πλήθος στατιστικών κατά SBS + 1NN	64
19	Ορθά ταξινομημένες αρθρώσεις ανά Συναισθημα με 1-VS-1 για την IEmoCap . . . . .	65
20	Ποσοστά Αναγνώρισης Συναισθημάτων με 1-VS-1 για την IEmoCap . . . . .	66
21	Ορθά ταξινομημένες αρθρώσεις ανά Συναισθημα με 1-VS-Rest SVM για την IEmoCap . .	67
22	Ποσοστά Αναγνώρισης Συναισθημάτων με 1-VS-Rest SVM για την IEmoCap . . . . .	68
23	Σύγκριση ορθά ταξινομημένων αρθρώσεων με SFS + 1NN, SBS + 1NN στην IEmoCap . .	72
24	Σύγκριση ποσοστών Αναγνώρισης Συναισθημάτων με SFS + 1NN, SBS + 1NN στην IEmoCap	73
25	Σύγκριση ορθά ταξινομημένων αρθρώσεων με επιλογή διαφορετικού πλήθους στατιστικών για τους συνδυασμούς SFS + 1NN, SBS + 1NN στην IEmoCap . . . . .	74
26	Σύγκριση ποσοστών μέσης ακρίβειας Αναγνώρισης Συναισθήματος με επιλογή διαφορετι- κού πλήθους στατιστικών για τους συνδυασμούς SFS + 1NN, SBS + 1NN στην IEmoCap .	75
27	Σύγκριση ορθά ταξινομημένων αρθρώσεων με 1-VS-1, 1-VS-Rest SVMs στην IEmoCap .	76
28	Σύγκριση ποσοστών Αναγνώρισης Συναισθημάτων με 1-VS-1, 1-VS-Rest στην IEmoCap .	77



## Κατάλογος Πινάκων

1	Κωδικοποίηση των 4 Συναισθημάτων της IEmoCap . . . . .	45
2	Χρησιμοποιημένα στατιστικά στο πρώτο πείραμα . . . . .	47
3	Ορθά ταξινομημένες αρθρώσεις με επιλογή 27, 61, 107, 159 στατιστικών κατά SFS + 1NN	49
4	Ποσοστά μέσης Αναγνώρισης Συναισθήματος με 27, 61, 107, 159 στατιστικά κατά SFS+1NN	50
5	Το βέλτιστο πλήθος στατιστικών κατά SFS + 1NN . . . . .	51
6	Ταξινόμηση ανά Συναισθημα με το βέλτιστο πλήθος στατιστικών κατά SFS + 1NN . . . . .	54
7	Ποσοστά Αναγνώρισης Συναισθημάτων με το βέλτιστο πλήθος στατιστικών κατά SFS + 1NN	55
8	Ορθά ταξινομημένες αρθρώσεις με επιλογή 27, 51, 107, 159 στατιστικών κατά SBS+1NN	57
9	Μέση Αναγνώριση Συναισθήματος με επιλογή 27, 51, 107, 159 στατιστικών κατά SBS . . .	58
10	Το βέλτιστο πλήθος στατιστικών κατά SBS + 1NN . . . . .	59
11	Ταξινόμηση αρθρώσεων με το βέλτιστο πλήθος στατιστικών κατά SBS + 1NN . . . . .	62
12	Ποσοστά Αναγνώρισης Συναισθημάτων με το βέλτιστο πλήθος στατιστικών κατά SBS + 1NN	63
13	Χρησιμοποιημένα στατιστικά για το δεύτερο πείραμα . . . . .	64
14	Αποιελέσματα ταξινόμησης αρθρώσεων με 1-VS-1 SVM για την IEmoCap . . . . .	65
15	Ποσοστά Αναγνώρισης Συναισθημάτων με 1-VS-1 για την IEmoCap . . . . .	66
16	Αποιελέσματα ταξινόμησης αρθρώσεων με 1-VS-Rest SVM για την IEmoCap . . . . .	67
17	Ποσοστά Αναγνώρισης Συναισθημάτων με ταξινομητή 1-VS-Rest SVM για την IEmoCap . .	68

# Κεφάλαιο 1: Εισαγωγή στην Αναγνώριση Συναισθήματος από το Σήμα του Λόγου

Η Αναγνώριση Συναισθήματος από το Σήμα του Λόγου έχει ως στόχο την ανάπτυξη πιο ρεαλιστικών, φυσικών και φιλικών ως προς το χρήστη (more user-friendly) εφαρμογών/διεπαφών Αλληλεπίδρασης Ανθρώπου-Υπολογιστή (Human-Computer Interaction applications/interfaces development).

Οι πρώτες μελέτες διενεργήθηκαν κατά τα μέσα της δεκαετίας του '80.

Η Αναγνώριση Συναισθήματος από το Σήμα του Λόγου υπάγεται στο φρέσκο τομέα έρευνας/επιστημονικό πεδίο της Τεχνητής Νοημοσύνης, γνωστό ως "Συναισθηματική Υπολογιστική" ("Affective Computing"). Ο Picard έχει επινοήσει αυτό τον όρο για την περιγραφή ενός νέου επιστημονικού κλάδου που ασχολείται με την αυτόματη αίσθηση, Αναγνώριση και σύνθεση του ανθρώπινου Συναισθήματος, μέσω οποιασδήποτε τροπικότητας πληροφορίας, όπως ο Λόγος ή/και η εικόνα. Στην ουσία, πρόκειται για ένα κράμα πολλών διαφορετικών επιστημονικών κλάδων.

Παρ' όλη την πρόοδο που έχει σημειωθεί στην κατανόηση των μηχανισμών των Συναισθημάτων μέσα από τον ανθρώπινο Λόγο, η πρόοδος στην ανάπτυξη και στο σχεδιασμό Συστημάτων Αυτόματης Αναγνώρισης Συναισθήματος για πρακτικές εφαρμογές βρίσκεται ακόμη στα σπάργανα και αυτό λόγω:

Της διαφωνίας μεταξύ αρκετών ερευνητών ως προς τον προσδιορισμό του ποια εξαγόμενα γνωρίσματα θεωρούνται ότι μεταφέρουν πολυτιμότερη Συναισθηματική πληροφορία από ποια άλλα

Των γλωσσικών, πολιτισμικών και μεταξύ ομιλητών διαφορών

Της έλλειψης αξιόπιστων βάσεων δεδομένων Συναισθημάτων για φυσικό/αυθόρμητο Συναισθηματικό Λό-

γο που συναντάται στην καθημερινή αλληλεπίδραση

Της διαφωνίας μεταξύ αρκετών ερευνητών ως προς το ποια συναισθήματα αρκεί να εξετάζονται. Ως προς αυτόν τον τελευταίο λόγο, η "Θεωρία της Παλέτας" ("Palet Theory") δίνει μια κατεύθυνση, προτείνοντας τον προσδιορισμό βασικών Συναισθημάτων που λειτουργούν ως "πρωτεύοντα"/"αρχέτυπα" ("primitive"). Βέβαια, από αυτά μπορεί να γίνεται συνδυασμός, για παραγωγή άλλων, όμοια με τη μείξη πρωτευόντων χρωμάτων για παραγωγή άλλων. Η "Θεωρία της Παλέτας" έχει διατυπωθεί από το Descartes.

## **1.1 Σύγχυση Συναισθημάτων (Emotions Confusion/Ambiguity) και οι λόγοι εμφάνισής της**

Μικρός λόγος Αναγνώρισης Συναισθήματος αντανακλάται μέσω του υψηλού επιπέδου Σύγχυσης του με άλλα Συναισθήματα και οφείλεται στους εξής παράγοντες:

Στην απουσία σαφών ορίων μεταξύ διαφορετικών Συναισθημάτων με αποτέλεσμα την επικάλυψή τους. Επικάλυψη δε συμβαίνει μόνο μεταξύ των γειτονικών, αλλά και αντιδιαμετρικών στο συνεχή διδιάστατο χώρο αναπαράστασης Συναισθημάτων, για παράδειγμα τα αντιδιαμετρικά Χαρά-Θυμός.

Στη μεγάλη ποικιλομορφία μεταξύ διαφορετικών ομιλητών στην έκφραση του ίδιου Συναισθήματος

Στην ισχύ του σήματος θορύβου που περιέχεται στην επιλεγμένη βάση δεδομένων Συναισθημάτων

## **1.2 Είδη Συναισθηματικού Λόγου και πλεονεκτήματά/μειονεκτήματά τους**

### **1) Φυσικός/Αυθόρμητος (Natural) Συναισθηματικός Λόγος**

#### **Παραδείγματα:**

- α) Συνεντεύξεις με ειδικούς, όπως ψυχολόγοι ή φωνολόγοι
- β) Προφορικές συνεντεύξεις εργαζομένων, όταν εξετάζονται για προαγωγή
- γ) Οι γονείς που μιλούν στα μικρής ηλικίας παιδιά τους, όταν προσπαθούν να τα κρατούν μακριά από επικίνδυνα ή βλαβερά για αυτά αντικείμενα
- δ) Συνεντεύξεις μεταξύ γιατρού και ασθενούς πριν ή μετά τη φαρμακευτική αγωγή

ε) Τηλεφωνικές κλήσεις σε τηλεφωνικά κέντρα, όπως τράπεζες ή ταξιδιωτικά πρακτορεία

### **Μειονεκτήματα του φυσικού/αυθόρμητου Συναισθηματικού Λόγου**

- α) Η χαμηλή ποιότητα των ηχογραφήσεων, επειδή ηχογραφείται μέσω τηλεφώνου
- β) Υπερβολικού βαθμού αυθόρμητα συναισθήματα εμφανίζονται σπάνια στην πραγματικότητα
- γ) Η ηχογράφηση ανθρώπων που βιώνουν αυθόρμητα υπερβολικού βαθμού συναισθήματα εγείρει σε αρκετές περιπτώσεις ζητήματα ηθικής για τη μη νομιμοποίηση ηχογράφησής τους
- δ) Δεν είναι άμεση ούτε προφανής η τμηματοποίηση του Σήματος του Λόγου σε άρθρωσεις, ώστε 1 άρθρωση να χαρακτηρίζεται από 1 Συναίσθημα.

### **2) Επαγγελματικά προσπονημένος/προσομοιωμένος (Acted) Συναισθηματικός Λόγος**

Το πιο αξιόπιστο είδος Λόγου για Αναγνώριση Συναισθήματος, επειδή χρησιμοποιούνται επαγγελματίες ηθοποιοί που μπορούν να εκφράζουν Λόγο πλούσιο σε Συναισθήματα που χαρακτηρίζεται από μεγάλη διέγερση, συνεπώς και Σήμα Λόγου μεγάλης ενέργειας ή ισχύος.

Η επιστημονική/ερευνητική κοινότητα βασίζεται συχνά στην ικανότητα επαγγελματιών ηθοποιών για αυθυποβολή, ενθουμούμενοι μια κατάσταση, όπου το επιθυμητό συναίσθημα είχε βιωθεί σε έντονο βαθμό, μια διαδικασία που καλείται μέθοδος του Stanislavski.

Οι ηχογραφήσεις πρέπει να είναι υψηλής ποιότητας, ελαχιστοποιώντας την ισχύ του σήματος του θορύβου. Χρησιμοποιούνται ανηχοικά δωμάτια (studios) και λαρυγγογράφημα.

### **Πλεονεκτήματα του προσπονημένου/προσομοιωμένου Συναισθηματικού Λόγου**

- α) Απλή απόκτηση δεδομένων
- β) Η τμηματοποίηση του Σήματος του Λόγου σε 1 άρθρωση ανά Συναίσθημα είναι τετριμμένη

### **3) Προκλημένος/επαγόμενος (Induced) Συναισθηματικός Λόγος**

Πολλοί ομιλητές ηχογραφούνται σε ένα εργαστήριο, ενώ εκτελούν μια συγκεκριμένη εργασία που είναι σχεδιασμένη να τους επάγει ένα συγκεκριμένο Συναισθημα, για παράδειγμα Θυμό ή Ενόχληση, χωρίς να γνωρίζουν ότι εξετάζεται η Συναισθηματική τους κατάσταση. Και σε αυτή την περίπτωση Συναισθηματικού Λόγου, εγείρονται ζητήματα ηθικής για τη μη νομιμοποίηση ηχογράφησης των ομιλητών.

### **1.3 “Πρωτεύοντα”/“Αρχέτυπα” Συναισθήματα σύμφωνα με τη “Θεωρία της Παλέτας”**

- 1) Χαρά
- 2) Λύπη
- 3) Φόβος
- 4) Θυμός
- 5) Έκπληξη
- 6) Αηδία
- 7) Ουδέτερη Συναισθηματική κατάσταση: Προστίθεται για να μοντελοποιεί την απουσία οποιασδήποτε εκ των παραπάνω Συναισθημάτων σε μια Συναισθηματική άρθρωση.

Παρουσιάζουν ομοιότητες μεταξύ διαφορετικών πολιτισμών και γλωσσών και παρέχουν, τουλάχιστον, ένα καλό σημείο εκκίνησης για την ταξινόμηση Συναισθημάτων.

Μη βασικά Συναισθήματα καλούνται “Συναισθήματα υψηλότερου επιπέδου” (“higher level emotions”) και σπάνια αντιπροσωπεύονται σε βάσεις δεδομένων Συναισθημάτων.

Η Ψυχολογική Παράδοση έχει αναπτυχθεί από επιφανείς ερευνητές διαφόρων επιστημονικών τομέων για τον ορισμό του Συναισθήματος:

#### **Pittam-Scherer (1992)**

Έχουν προτείνει ταξινόμηση Συναισθημάτων σύμφωνα με το βαθμό διέγερσής τους. Ο Scherer έχει προτείνει τον ακόλουθο ορισμό του Συναισθήματος, για τον οποίο υπάρχει μεγάλο consensus στην ερευνητική/επιστημονική κοινότητα:

“Τα συναισθήματα είναι επεισόδια συντονισμένων αλλαγών σε διάφορες συνιστώσες, συμπεριλαμβανομένων τουλάχιστον της νευροφυσιολογικής ενεργοποίησης και της υποκειμενικής αίσθησης, αλλά πιθανώς επίσης ροπές προς ενέργειες και νοητικές διαδικασίες, ως απόκριση εξωτερικών ή εσωτερικών ερεθισμάτων που είναι κύριας σημασίας για τον οργανισμό.”

### **Rene Descartes (Φιλοσοφία)**

Μικρό πλήθος Συναισθημάτων σχηματίζουν τη βάση για την πλειονότητά τους.

### **Charles Darwin (Βιολογία)**

Το Συναίσθημα αποτελεί ένα εξελισσόμενο φαινόμενο με σημαντικές λειτουργίες επιβίωσης που βοηθά στην επίλυση συγκεκριμένων προβλημάτων που αντιμετωπίζει το ανθρώπινο είδος.

Έχει την πεποίθηση για ύπαρξη των ίδιων Συναισθημάτων σε όλους τους ανθρώπους (καθολικότητα).

Τα Συναισθήματα αποτελούν ξεχωριστά πρότυπα ενεργειών που έχουν επιλεγεί από την εξέλιξη, χάρη στην αξία τους για την επιβίωση του ανθρώπου. Πυροδοτούνται από συγκεκριμένες καταστάσεις στη ζωή του, όπως η αντίδραση στον κίνδυνο είναι κατά κύριο λόγο η βίωση Φόβου και η αναζήτηση τρόπων διαφυγής.

### **William James (Ψυχολογία)**

Είναι αδύνατη η ύπαρξη Συναισθήματος, χωρίς σωματικές μεταβολές, που μάλιστα πάντα προηγούνται της εκδήλωσής του. Επικεντρώθηκε στη σύνδεση Συναισθήματος και σωματικής διέγερσης.

Ο άνθρωπος πρώτα αποκρίνεται σε μια κατάσταση και μετά βιώνει ένα Συναίσθημα.

### **Wilhelm Wundt (Ψυχολογία)**

Ο άνθρωπος είναι κατά κύριο λόγο Συναισθηματικός και βιώνει Συναισθήματα σε όλες τις πνευματικές και φυσικές του δραστηριότητες.

Ο Wundt πίστευε ότι είναι αδύνατο να επινοηθεί μια συνεπής λίστα βασικών Συναισθημάτων. Αυτός είναι και ο λόγος που τον οδήγησε να επινοήσει συγκεκριμένες διαστάσεις Συναισθημάτων που οδήγησαν στο μοντέλο της συνεχούς διδιάστατης αναπαράστασης Συναισθημάτων της Ενεργοποίησης-Αξιολόγησης (Activation-Evaluation continuous 2-D emotions space).

Σύμφωνα με το Wundt, το Συναίσθημα προηγείται πάντα της πνευματικής απόκρισης, γεγονός που έρχεται σε αντίθεση με τις πεποιθήσεις του William James.

## **1.4 Διαφορετικές συνεχείς 2-διάστατες αναπαραστάσεις Συναισθημάτων**

### **Σθένους-Διέγερσης (Valence-Arousal)**

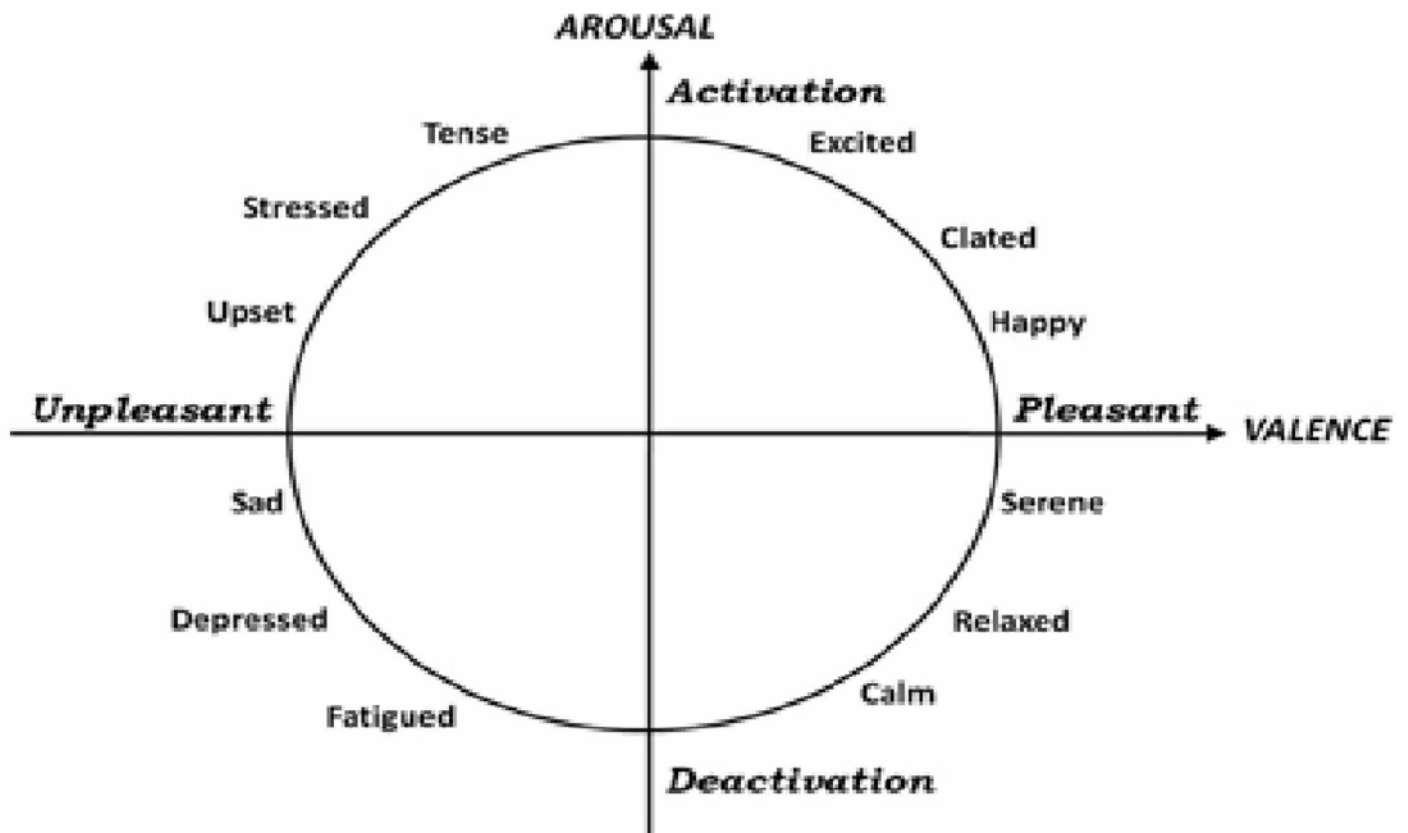
Οι διαστάσεις του επιπέδου είναι το Σθένος (Valence, V) και η Διέγερση (Arousal, A).

Το Σθένος αποτελεί το προφανέστερο συστατικό των Συναισθημάτων. Κυμαίνεται από το  $-1$  για το Δυσάρεστο ως το  $+1$  για το Ευχάριστο. Ο άξονας αυτός μπορεί να χαρακτηριστεί ως επίπεδο Ικανοποίησης.

Η Διέγερση αντικατοπτρίζει την ένταση του Συναισθήματος και κυμαίνεται από το  $-1$  για Παθητικό, Ήρεμο, Ήσυχο, Γαλήνιο) ως το  $+1$  για Ενεργό.

Κάθε Συναίσθημα, δηλαδή κάθε σημείο στο συνεχή 2-διάστατο χώρο, γίνεται αντιληπτό ως ένας γραμμικός συνδυασμός αυτών των δύο συνιστωσών.

**Παράδειγμα:** Ο Θυμός γίνεται αντιληπτός ως μια δυσάρεστη Συναισθηματική κατάσταση, συνεπώς έχει αρνητικές τιμές στον άξονα Valence, και μεγάλη ένταση, συνεπώς έχει θετικές τιμές στον άξονα Arousal.



Σχήμα 1: Συνεχής 2-διάστατη αναπαράσταση συναισθημάτων Valence-Arousal[1]

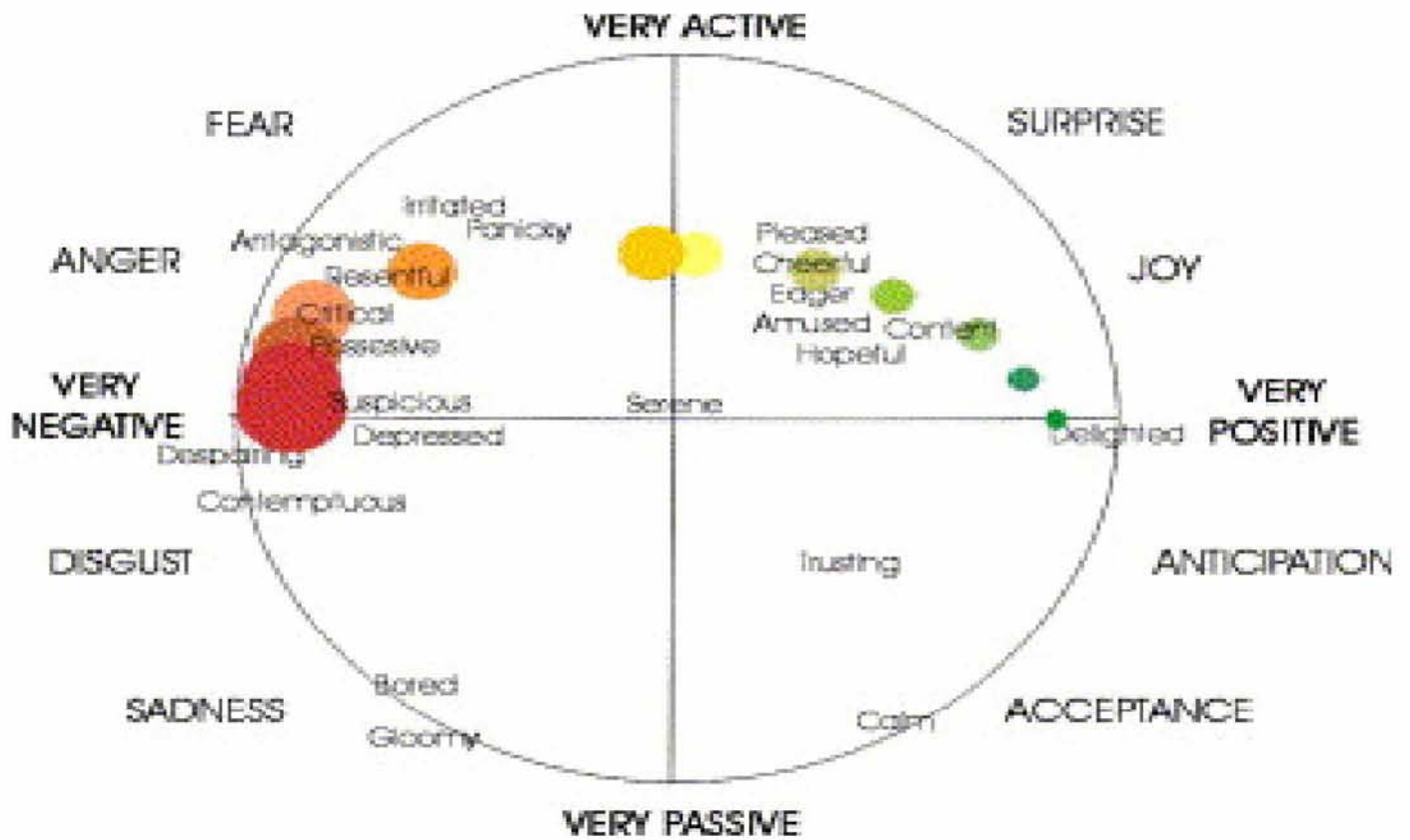
### **Ενεργοποίησης-Αξιολόγησης (Activation-Evaluation)**

Είναι ευρέως γνωστή ως Τροχός Συναισθημάτων του Whissel (1989).

Η Ενεργοποίηση αποτελεί τη δύναμη της ατομικής προδιάθεσης να κάνει κανείς μια δράση, συγκριτικά με το να μη δράσει.

Η Αξιολόγηση αντικατοπτρίζει το κατά πόσο ένα Συναίσθημα αποτιμάται ως θετικό ή αρνητικό. Χαμηλές τιμές κοντά στο 2 ανατίθενται στην Αδιαφορία και μια μεγάλη τιμή άνω του 6 στην Έκπληξη.

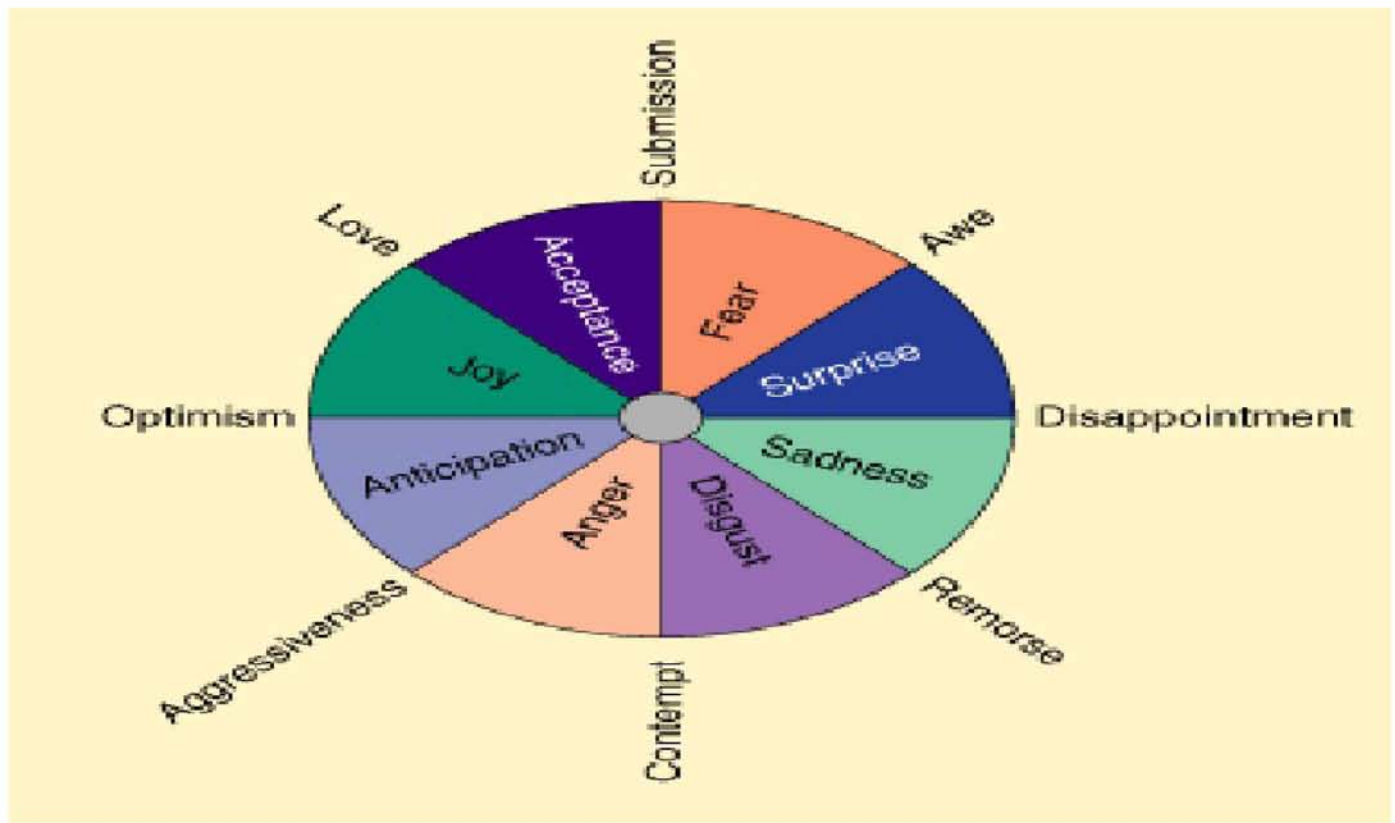




Σχήμα 2: Συνεχής 2-διάστατη αναπαράσταση συναισθημάτων Activation-Evaluation[2]

### **"Τροχός Συναισθημάτων του Plutchnik"**

Τα πιο σημαντικά Συναισθήματα τοποθετούνται σε ζεύγη αντίθετων Συναισθημάτων.



Σχήμα 3: Τροχός Συναισθημάτων του Plutchnik (Plutchnik's Emotions Wheel)[3]

## 1.5 Είδη Συστημάτων Αυτόματης Αναγνώρισης Συναισθήματος

### 1) Εξαρτημένα από τον ομιλητή

Δεδομένα από συγκεκριμένο ομιλητή χρησιμοποιούνται για το training και το testing.

Είναι λιγότερο περίπλοκα και παρέχουν μεγαλύτερη ακρίβεια Αναγνώρισης Συναισθήματος, συγκριτικά με τα ανεξάρτητα του ομιλητή.

Σε αυτά, η Αναγνώριση Συναισθήματος επηρεάζεται από τα εξής χαρακτηριστικά του ομιλητή:

Το πολιτισμικό υπόβαθρο

Τη χρησιμοποιημένη γλώσσα

Το φύλο

Την ηλικία

## **2) Ανεξάρτητα του ομιλητή**

Χαρακτηρίζονται από μικρότερη ακρίβεια Αναγνώρισης Συναισθήματος, επειδή δεν είναι άρτια συγχρονισμένα με τον κάθε ξεχωριστό ομιλητή-χρήστη τους.

Λεδομένα διαφορετικών ομιλητών χρησιμοποιούνται για training και testing.

## **3) Εξαρτημένα από τα συμφραζόμενα**

Απαιτούν κατανόηση της εκάστοτε χρησιμοποιημένης γλώσσας, ώστε να μπορούν να προσδιορίζουν τα συμφραζόμενα της τρέχουσας φράσης.

## **4) Ανεξάρτητα των συμφραζόμενων**

Προσδιορίζουν το Συναίσθημα σε μια άρθρωση, ανεξάρτητα του κειμένου που χρησιμοποιείται.

# **1.6 Εφαρμογές Συστημάτων Αυτόματης Αναγνώρισης Συναισθήματος**

- 1) Εφαρμογές διασκέδασης/ψυχαγωγίας (διαδραστικά βινεοπαιχνίδια-"interactive videogames")
- 2) Ανίχνευση προβληματικού διαλόγου για εφαρμογές Ανθρώπου-Μηχανής (Fault dialogue detection)
- 3) Ανάπτυξη πολυπρακτορικών συστημάτων (Τεχνητή Νοημοσύνη)
- 4) Φωνητικές υπηρεσίες/Τηλεφωνικά κέντρα

### **Παραδείγματα:**

- α) Που παρέχουν πληροφορίες στους πελάτες τους για τους τηλεφωνικούς λογαριασμούς τους και για τη χρέωση των τηλεφωνικών τους κλήσεων.
- β) Που καταγράφουν την Ικανοποίηση των πελατών τους για την αποτελεσματικότητα των υπηρεσιών που

τους παρέχουν.

5) Σύστημα εντός του αυτοκινήτου που δέχεται πληροφορία για τη Συναισθηματική κατάσταση του οδηγού του, ώστε να παίρνει πρωτοβουλίες για στρατηγικές ασφάλειας ή να του παρέχει κάποιου είδους βοήθεια. Υποστηρίζει την καλύτερη οδήγηση, δεδομένου ότι τα Συναισθήματα του οδηγού και η απόδοση οδήγησής του είναι συνήθως άρρηκτα συνδεδεμένα.

**Παραδείγματα :**

a) The Emotive Driver project

b) The FERMUS project

6) Ανίχνευση ψέματος από εγκληματολογικές υπηρεσίες (lie detection)

7) Για ψυχιατρική βοήθεια, ως διαγνωστικό εργαλείο, σε εικονικά συστήματα ψυχολόγων/θεραπευτών

8) Στον τομέα της εκπαίδευσης:

Κατάλληλο βοηθητικό λογισμικό (software) για την εκπαίδευση που να ανιχνεύει τα Συναισθήματα των μαθητών/φοιτητών και να επιλέγει εναλλακτικό κατάλληλο πρόγραμμα διδασκαλίας. Μπορεί να ανιχνεύει όσους διακατέχονται από Πλήξη και να επιτρέπει αλλαγή στον τρόπο διδασκαλίας και στο επίπεδο του παρεχόμενου εκπαιδευτικού υλικού ή, εναλλακτικά, να τους παρέχει Συναισθηματική εμπύχωση.

**Παράδειγμα :** ITSpoke tutoring system

9) Εκτίμηση του επείγοντος μιας κλήσης σε ένα ιατρικό κέντρο και άμεση βοήθεια με λήψη ιατρικών συμβουλών για κατ' οίκον νοσηλεία σε ασθενείς είτε από τους ίδιους είτε από συγγενείς τους

10) Ανάπτυξη πιο εύρωστων και ακριβών Συστημάτων Αυτόματης Επαλήθευσης Χρήστη, μέσω της χρήσης των ειδικών τιμών των γνωρισμάτων

11) Αυτόματη ιεράρχηση μηνυμάτων που συσσωρεύονται στον αυτόματο τηλεφωνητή, με κριτήριο το βαθμό του επείγοντος της κλήσης. Μπορεί να χρησιμοποιηθεί για την ειδοποίηση του κατόχου του αυτόματου τηλεφωνητή, ώστε να ακούει πρώτα τα σημαντικά μηνύματα.

12) Βοήθεια σε αυτιστικά άτομα, ώστε να αναπτύξουν καλύτερες δεξιότητες για την Αναγνώριση Συναισθήματος από το Λόγο του συνομιλητή τους.

# **Κεφάλαιο 2: Μελέτη γνωρισμάτων και μεθόδων Επιλογής Γνωρισμάτων ως προς τα πλεονεκτήματά/μειονεκτήματά τους**

## **2.1 Εξαγωγή γνωρισμάτων σε επίπεδο άρθρωσης ή πλαισίου**

Μακροπρόθεσμα στατιστικά καθορίζονται από κάθε άρθρωση, ενώ βραχυπρόθεσμα από ένα μικρότερο χρονικό παράθυρο που έχει συνήθως διάρκεια από 20 ως 100 ms και καλείται "χρονικό πλαίσιο Λόγου" ("time Speech frame"). Υπάρχει διαφωνία μεταξύ αρκετών ερευνητών ως προς την επιλογή του επιπέδου εξαγωγής γνωρισμάτων. Συνήθως, γίνεται σε επίπεδο άρθρωσης. Εκτιμάται ότι η εξαγωγή γνωρισμάτων σε επίπεδο άρθρωσης βοηθά περισσότερο στην Αναγνώριση Συναισθήματος από αυτή σε επίπεδο πλαισίου.

## **2.2 Προσωδιακά γνωρίσματα**

Ασχολούνται με τις "περισσότερο μουσικοειδείς" πτυχές του Λόγου.

### **Θεμελιώδης συχνότητα/Βαθμός εντάσεως (F0/Pitch)**

Ορίζεται ως το αντίστροφο του χρονικού διαστήματος μεταξύ δύο διαδοχικών ανοιγμάτων των φωνητικών χορδών, ή ισοδύναμα, ο ρυθμός ταλάντωσής τους. Εξαρτάται από την ένταση των φωνητικών χορδών και την υπογλωττιδική πίεση του αέρα.

Πολλοί από τους ήχους στο περιβάλλον έχουν μια ακουστική κυματομορφή που επαναλαμβάνεται με το χρόνο. Αυτοί οι ήχοι γίνονται συχνά αντιληπτοί, σαν να έχουν μια θεμελιώδη συχνότητα που να αντιστοιχεί στο ρυθμό επανάληψής τους. Αυτό σημαίνει ότι τα ερεθίσματα του βαθμού εντάσεως είναι συνήθως περιοδικά και η θεμελιώδης συχνότητα σχετίζεται με την περιοδικότητά τους.

Ενδεικτικά, για τη σχέση των δύο φύλων με την τιμή της θεμελιώδους συχνότητας επισημαίνεται :

Η θεμελιώδης συχνότητα των αντρών κυμαίνεται μεταξύ 80-200 Hz και των γυναικών 150-350 Hz.

### **Ρυθμός ομιλίας (Speech rate)**

$$\text{Ρυθμός ομιλίας} = \frac{\text{Πλήθος ηχηρών/έμφωνων τμημάτων}}{\text{Συνολική διάρκεια ηχηρών/έμφωνων τμημάτων}}$$

Τα κριτήρια για την ηχηρότητα (voiced) ή μη (unvoiced) ενός τμήματος του Σήματος του Λόγου βασίζονται σε κάτω κατώφλια, είτε της θεμελιώδους συχνότητας (pitch threshold) είτε της ενέργειας (energy threshold) είτε του ρυθμού Zero Crossing Rate (ZCR threshold) είτε σε διάφορους συνδυασμούς τους.

### **Ταχύτητα/πίεση του υπογλωττιδικού αέρα**

Μεγάλες τιμές της ταχύτητας/πίεσης του υπογλωττιδικού αέρα χαρακτηρίζουν ένα "μουσικοειδή" τύπο Λόγου και συναντούνται στα Συναισθήματα Χαρά και Έκπληξη. Αντίθετα, μικρές χαρακτηρίζουν ένα πιο τραχύ Λόγο και συναντούνται στα Συναισθήματα Θυμός και Αηδία.

### **Ενέργεια (Root Mean Square, RMS)/Ένταση (Intensity) και Ισχύς του Σήματος του Λόγου**

$$\text{Ενέργεια του Σήματος του Λόγου: } E[n] = T * \sum_{n=0}^{N-1} x[n]^2$$

$$\text{Ισχύς του Σήματος του Λόγου: } P[n] = \frac{1}{N} * \sum_{n=0}^{N-1} x[n]^2$$

$x[n]$ : τα δείγματα του Σήματος του Λόγου, συλλεγμένα ανά τακτά διακριτά χρονικά διαστήματα

T: η διάρκεια του Σήματος του Λόγου, στην οποία θέλουμε να υπολογίσουμε την ενέργειά του

$N$  : το συνολικό πλήθος των συλλεγμένων δειγμάτων του Σήματος του Λόγου

Η ενέργεια του Σήματος του Λόγου υπολογίζεται με χρήση της μεθόδου RMS (Root Mean Square) που υπολογίζει την ενέργεια σε τετράγωνο χρονικό πλαίσιο μικρής διάρκειας.

Σχετίζεται με το επίπεδο Διέγερσης των Συναισθημάτων. Συγκεκριμένα, ο Θυμός χαρακτηρίζεται από τη μεγαλύτερη ενέργεια του Σήματος του Λόγου που είναι μεγαλύτερη στους άντρες από τις γυναίκες.

### **Λόγος Αρμονικότητα προς Θόρυβο (Harmonic to Noise Ratio, HNR)**

$$HNR(n) = \frac{E_H(n)^2}{E_N(n)^2} = \frac{E_H(n)^2}{1 - E_H(n)^2}$$

$E_H(n)^2$  : η Αρμονική Ενέργεια του Σήματος του Λόγου

$E_N(n)^2$  : η Ενέργεια του Σήματος Θορύβου

$$0 < HNR(n) < 1$$

Ένδειξη του κατά πόσο η θεμελιώδης συχνότητα γίνεται αντιληπτή δεδομένου ενός Σήματος Θορύβου.

Η "Αρμονικότητα" χρησιμεύει εναλλακτικά και ως κριτήριο διαχωρισμού των ηχηρών/έμφωνων και των μη ηχηρών/άφωνων τμημάτων στο Σήμα του Λόγου.

### **Το μειονέκτημα της επιλογής προσωδιακών γνωρισμάτων**

Έχουν μεγάλη εξάρτηση από τα ιδιαίτερα χαρακτηριστικά του ομιλητή και συνεπώς απαιτείται κανονικοποίηση των τιμών τους σε μηδενική μέση τιμή ( $\mu = 0$ ) και μοναδιαία τυπική απόκλιση ( $\sigma = 1$ ), διαδικασία που καλείται z-κανονικοποίηση.

Συνεπώς, η επιλογή για εξαγωγή προσωδιακών γνωρισμάτων δε συνεισφέρει στη δημιουργία ενός Συστήματος Αυτόματης Αναγνώρισης Συναισθήματος με καλή ικανότητα γενίκευσης και ανεξαρτησία από τον ομιλητή, αλλά ενδείκνυται περισσότερο για δημιουργία ενός εξαρτημένου από τον ομιλητή.

## 2.3 Διαμόρφωσης Πλάτους-Διαμόρφωσης Συχνότητας (AM-FM/Formants)

Το Σήμα του Λόγου μπορεί να αναλυθεί σε  $N$  στιγμιαία σημάτα πλάτους και συχνότητας που καλούνται Formants, επειδή πρόκειται για αντηχήσεις (resonances) που συνθέτουν, σχηματίζουν (form) το συνολικό φάσμα του ακουστικού Σήματος. Εντός ενός πλαισίου του Σήματος του Λόγου, κάθε αρμονική έχει σταθερό στιγμιαίο πλάτος και σταθερή στιγμιαία συχνότητα. Κάθε Formant χαρακτηρίζεται από την κεντρική του συχνότητα (center frequency) και το εύρος ζώνης του (bandwidth).

Τα πρώτα 2 Formants  $F_1$ ,  $F_2$  σχετίζονται με την Αναγνώριση Συναισθήματος περισσότερο από τα άλλα. Υπολογίζονται με χρήση της μεθόδου Linear Predictive Analysis (LPA).

### Το πλεονέκτημα της επιλογής γνωρισμάτων Formants

Είναι περισσότερο εύρωστα από τους MFCCs έναντι στο Σήμα Αθροιστικού Θορύβου (Additive Noise).

## 2.4 Mel Frequency Cepstral Coefficients (MFCCs)

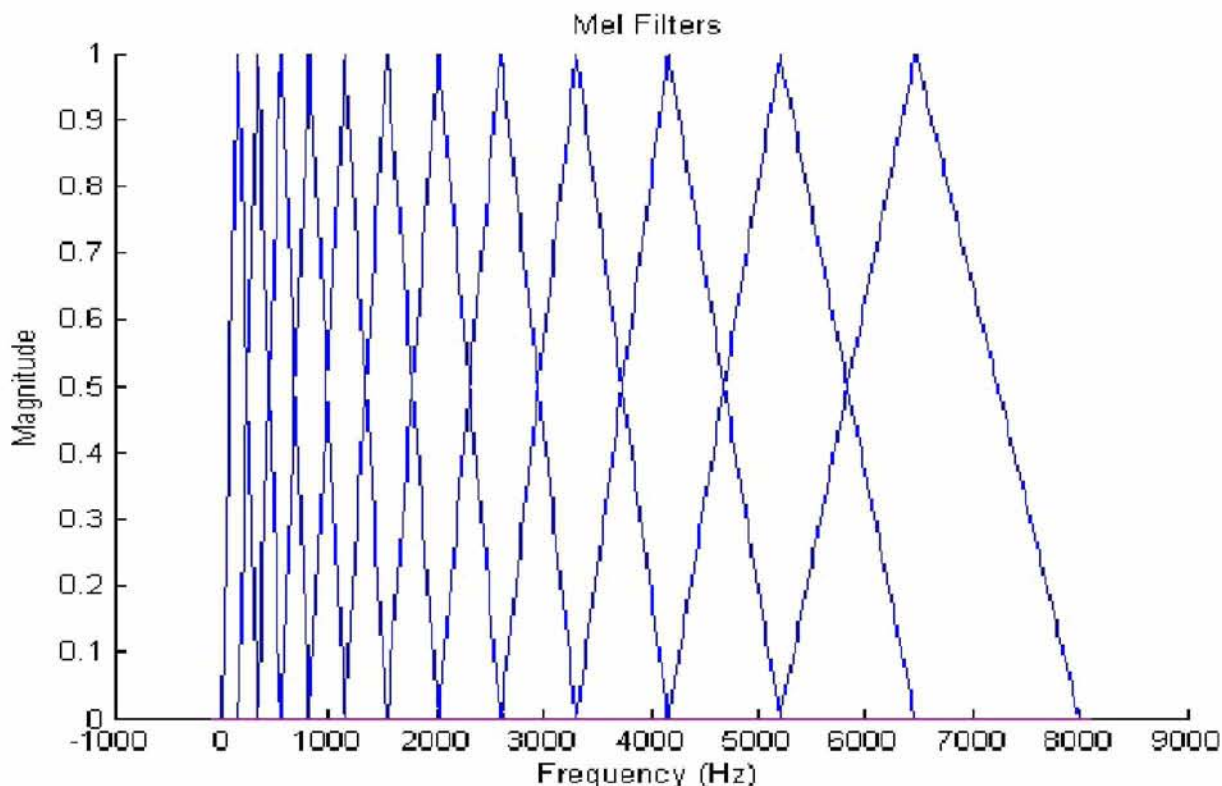
Είναι η πιο χρησιμοποιημένη φασματική αναπαράσταση του Σήματος του Λόγου.

Βασίζονται στη μεταβολή των εύρων ζώνης του ανθρώπινου συστήματος ακοής ως προς τη συχνότητα.

Φίλτρα που βρίσκονται γραμμικά σε χαμηλές συχνότητες και λογαριθμικά σε υψηλές έχουν χρησιμοποιηθεί για την ενσωμάτωση των φωνητικά σημαντικών χαρακτηριστικών του Λόγου.

Η προηγούμενη διαπίστωση εκφράζεται και στην κλίμακα Mel (Mel Scale) που είναι γραμμική, κάτω από, και λογαριθμική, πάνω από τα 1000 Hz. Οι ζώνες συχνότητας του Mel Frequency Cepstrum εκτείνονται ισόμορφα στην κλίμακα Mel, γεγονός που οδηγεί σε μια καλή προσέγγιση της ανθρώπινης ακοής.





Σχήμα 4: Κλίμακα Mel (Mel Scale)[4]

### Ανάλυση των σταδίων υπολογισμού των MFCCs

**Στάδιο 1: Μετατροπή του Σήματος του Λόγου σε πλαίσια (Speech signal frames conversion)**

**Στάδιο 2: Παραθύρωση του Σήματος του Λόγου (Speech signal Windowing)**

Χρησιμοποιείται για την ελαχιστοποίηση της παραμόρφωσης του Σήματος του Λόγου, με χρήση του παραθύρου που κάνει το Σήμα του Λόγου 0 στην αρχή και στο τέλος κάθε πλαισίου.

Τυπικά, χρησιμοποιείται το δημοφιλές παράθυρο του Hamming:

$$w(n) = 0,54 - 0,46 * \cos\left(\frac{2\pi * n}{N - 1}\right)$$

### Στάδιο 3: Εφαρμογή Discrete Time Fourier Transform, DTFT

$$X[k] = \sum_{n=0}^{N-1} x[n] * e^{\frac{j * 2\pi * k * n}{N}}, 0 \leq k < N$$

### Στάδιο 4: Mel Frequency

Το Σήμα του Λόγου αποιελείται από ήχους με διαφορετικές συχνότητες. Για κάθε ήχο με πραγματική συχνότητα  $f$ , υπολογίζεται μια υποκειμενική συχνότητα στην κλίμακα Mel. Ως σημείο αναφοράς, η συχνότητα 1 kHz, δηλαδή 40 db πάνω από το κατώφλι της ανθρώπινης ανιληπιτής ακοής, ορίζεται ως 1000 Mel. Χρησιμοποιείται η εξής φόρμουλα για υπολογισμό των Mel δοθείσης μιας συχνότητας  $f$  (Hz):

$$mel(f) = 2595 * \log\left(1 + \frac{f}{700}\right)$$

### Στάδιο 5: Cepstrum

Το φάσμα Log Mel πρέπει να μετασχηματιστεί πίσω στο πεδίο του χρόνου. Ως αποτέλεσμα προκύπτουν οι MFCCs. Η Cepstral αναπαράσταση του φάσματος του Σήματος του Λόγου παρέχει μια καλή ερμηνεία των τοπικών φασματικών ιδιοτήτων του Σήματος του Λόγου. Επειδή οι συντελεστές του φάσματος Mel είναι πραγματικοί αριθμοί και συνεπώς και οι λογάριθμοί τους, μπορούν να μετασχηματιστούν πίσω στο πεδίο του χρόνου, με χρήση του μετασχηματισμού Discrete Cosine Transform (DCT), παράγοντας τους MFCCs: Ο λογάριθμος της απόκρισης συχνότητας, ή ισοδύναμα, η ενέργεια Log Mel  $S[m]$  υπολογίζεται και εφαρμόζεται ο DCT:

$$S[m] = \ln\left[\sum_{k=0}^{N-1} X[k]^2 * H_m[k]\right], 0 < m \leq K$$

$$c[n] = \sum_{m=0}^{K-1} S[m] * \cos\left(\frac{\pi * n * (m - \frac{1}{2})}{K}\right), 0 \leq n \leq K$$

$K$  : το πλήθος των συντελεστών Mel Cepstrum. Συνήθως επιλέγεται  $K = 12$

Συνοψίζοντας, τα στάδια υπολογισμού των MFCCs αναπαριστώνται στο ακόλουθο διάγραμμα :

### **Το πλεονέκτημα της επιλογής των MFCCs**

Παρέχουν καλύτερη αναπαράσταση του Σήματος του Λόγου, συγκριτικά με τις ζώνες συχνοτήτων, καθώς εκμεταλλεύονται την ανθρώπινη ακουστική απόκριση συχνοτήτων.

### **Μειονεκτήματα της επιλογής των MFCCs**

1) Έχουν μεγάλη εξάρτηση από τα ιδιαίτερα χαρακτηριστικά του ομιλητή και συνεπώς απαιτείται  $z$ -κανονικοποίηση. Συνεπώς, η επιλογή για εξαγωγή MFCCs δε συνεισφέρει στη δημιουργία ενός Συστήματος Αυτόματης Αναγνώρισης Συναισθήματος με καλή ικανότητα γενίκευσης και ανεξαρτησία από τον ομιλητή, αλλά ενός εξαρτημένου από τον ομιλητή.

2) Είναι ευάλωτοι απέναντι στην παρουσία Σήματος θορύβου και η ικανότητα ταξινόμησής τους φθίνει με γρήγορο ρυθμό ανάλογα με την πτώση του.

$$SNR = \frac{\text{Ισχύς ωφέλιμου Σήματος Συναισθηματικής πληροφορίας}}{\text{Ισχύς Σήματος θορύβου}}$$

3) Δίνουν μεγάλο πλήθος στατιστικών στο Διάνυσμα Γνωρισμάτων (το πρόβλημα της "κατάρως διαστατικότητας" - "curse of dimensionality" problem).

## **2.5 Linear Predictive Coefficients (LPCs)**

Ένα δείγμα του Σήματος του Λόγου μπορεί να προσεγγιστεί ως γραμμικός συνδυασμός προηγούμενων δειγμάτων. Για να επιτευχθεί αυτό, είναι απαραίτητη η ελαχιστοποίηση του αθροίσματος της διαφοράς τετραγώνων μεταξύ του πραγματικού δείγματος του Σήματος του Λόγου και των συντελεστών-εκτιμήσεων. Ως αποτέλεσμα της μεθόδου Linear Predictive Analysis (LPA) που εφαρμόζεται, υπολογίζεται ένα σύνολο

μοναδικών συνιλεσιών πρόβλεψης.

### **Πλεονεκτήματα της επιλογής των LPCs**

- 1) Απαιτούν λίγο χώρο για αποθήκευση των παραμέτρων εκτίμησης-πρόβλεψης.
- 2) Υπολογίζονται αρκετά γρήγορα με χρήση του αλγορίθμου των Durbin-Levinson.

## **2.6 Σχετικού Φασματικού Μετασχηματισμού (Relative Spectral Perceptual Linear Predictive)**

Έχουν πάρει το όνομά τους από μια τεχνική ονόματι RelATive SpecTrAl (RASTA), που έχει αναπτυχθεί από το Hermansky και κάνει τα Perceptual Linear Predictive (PLP) γνωρίσματα πιο εύρωστα έναντι στο συνελικτικό θόρυβο. Ένα ζωνοπερατό φίλτρο εφαρμόζεται στην ενέργεια σε κάθε υποζώνη συχνότητας. Ως αποτέλεσμα, ο θόρυβος εξομαλύνεται (noise smoothening).

### **Το πλεονέκτημα της επιλογής των γνωρισμάτων RASTA-PLP**

Είναι πιο εύρωστα απέναντι στο θόρυβο, συγκριτικά με τα PLP.

## **2.7 Φασματικά γνωρίσματα, φασματόγραμμα (Spectral features, spectrogram)**

Εφαρμόζεται ο μετασχηματισμός Fourier σε μικρά χρονικά πλαίσια. Σε αυτά περιλαμβάνονται:

Η κλίση του φάσματος (Spectral slope)

Η απόσταση 10ου-90στού εκατοστημορίου

Η κλίση μεταξύ αδύναμων και ισχυρότερων συχνοτήτων

2 Γραμμικοί συνιλεσιές παλινδρόμησης (Linear regression coefficients)

Το κέντρο βάρους του φάσματος που υπολογίζεται από την εξής σχέση:

$$CG = \frac{\sum f_i * E_i}{\sum E_i}$$

## 2.8 Τιμές γνωρισμάτων ανάλογα με το Συναίσθημα

### Χαρά

Μεγάλη θεμελιώδης συχνότητα

Μεγάλη διασπορά (πλατύτερη καμπύλη) της θεμελιώδους συχνότητας

Μεγάλος ρυθμός ομιλίας

Λίγες συλλαβές τονίζονται.

Η τελευταία λέξη τονίζεται.

### Θυμός

Μεγάλη θεμελιώδης συχνότητα

Μεγάλη διασπορά (πλατύτερη καμπύλη) της θεμελιώδους συχνότητας

Μεγάλη ενέργεια του Σήματος του Λόγου

Μεγάλος ρυθμός ομιλίας

Μικρή διασπορά στη διάρκεια των φωνημάτων

Πολλές συλλαβές τονίζονται.

Η τελευταία λέξη δεν τονίζεται.

Μεγάλη ενέργεια υψηλής συχνότητας

### Λύπη

Μικρή θεμελιώδης συχνότητα

Μικρή διασπορά (στενότερη καμπύλη) της θεμελιώδους συχνότητας

Μικρός ρυθμός ομιλίας

Μεγάλη διασπορά στη διάρκεια των φωνημάτων

Πολύ λίγες συλλαβές τονίζονται.

Η τελευταία λέξη δεν τονίζεται.

Μικρή ενέργεια υψηλής συχνότητας

### **Ανακούφιση**

Μεγάλη θεμελιώδης συχνότητα, αλλά μικρότερη από της Χαράς

Μικρός ρυθμός ομιλίας

Μεγάλη διασπορά (πλατύτερη καμπύλη) στη διάρκεια των φωνημάτων

Πολύ λίγες συλλαβές τονίζονται.

Η τελευταία λέξη τονίζεται.

### **Φόβος**

Μεγάλη θεμελιώδης συχνότητα

Μεγάλη διασπορά (πλατύτερη καμπύλη) της θεμελιώδους συχνότητας

Μεγάλος ρυθμός ομιλίας

Μεγάλη ενέργεια υψηλής συχνότητας

### **Αηδία**

Μικρή θεμελιώδης συχνότητα

Μικρή διασπορά (στενότερη καμπύλη) της θεμελιώδους συχνότητας

## **Πλήξη**

Μικρή θεμελιώδης συχνότητα

Μικρή διασπορά (στενότερη καμπύλη) της θεμελιώδους συχνότητας

Ενδεχόμενη ομοιότητα μεταξύ ορισμένων Συναισθημάτων καταδεικνύει τη "Σύγχυση Συναισθημάτων".

## **2.9 Μέθοδοι Επιλογής Γνωρισμάτων (Feature Selection Methods)**

Το ερώτημα που τίθεται είναι ποιο υποσύνολο γνωρισμάτων ενός Διανύσματος Γνωρισμάτων ενός μεγάλου πλήθους γνωρισμάτων πρέπει να χρησιμοποιηθεί για την καλύτερη δυνατή ταξινόμηση.

Η Επιλογή Γνωρισμάτων βελτιώνει την ακρίβεια Αναγνώρισης, ενώ μειώνει και τον απαιτούμενο χρόνο.

Οι αλγόριθμοι που εφαρμόζονται χωρίζονται από ευρύτερη σκοπιά σε 2 κατηγορίες: Wrappers, Filters

### **Wrappers**

Είναι μερικές φορές επιθυμητό να βασίσουμε την απόφασή μας κατά την Επιλογή Γνωρισμάτων, όχι πάνω στις τιμές που λαμβάνει ένα κριτήριο διαχωρισιμότητας κλάσεων που έχει υιοθετηθεί, αλλά στην απόδοση του ταξινομητή αυτού καθ' εαυτού. Δηλαδή, για κάθε συνδυασμό γνωρισμάτων, πρέπει να εκτιμηθεί η πιθανότητα σφάλματος ταξινόμησης του ταξινομητή και να επιλεγεί ο συνδυασμός που δίνει την ελάχιστη πιθανότητα σφάλματος.

### **Το πλεονέκτημα της επιλογής των Wrappers**

Οδηγούν σε μεγάλη ακρίβεια Αναγνώρισης: Τα Wrappers επιτυγχάνουν καλύτερη ακρίβεια Αναγνώρισης από τα Filters, επειδή είναι προσαρμοσμένα πάνω σε συγκεκριμένο ταξινομητή και συνόλου δεδομένων.

### **Μειονεκτήματα της επιλογής των Wrappers**

- 1) Αργή εκτέλεση: Επειδή τα Wrappers κάνουν train έναν ταξινομητή για κάθε υποσύνολο γνωρισμάτων, μπορεί να γίνουν υπολογιστικά μη εφικτά. Η προσέγγιση αυτή μπορεί να αυξήσει την απαιτούμενη χρονική πολυπλοκότητα ακόμη περισσότερο, ανάλογα φυσικά και με τον τύπο του ταξινομητή.
- 2) Αδυναμία γενίκευσης: Η λύση δεν παρουσιάζει γενίκευση, καθώς δεσμεύεται από τη χρήση συγκεκριμένου ταξινομητή. Το βέλτιστο πλήθος γνωρισμάτων είναι συγκεκριμένο για το συγκεκριμένο ταξινομητή.

## 2.9.1 Ακολουθιακές μέθοδοι Επιλογής Γνωρισμάτων (Sequential methods)

### 2.9.1.1 Προς τα Εμπρός/Εμπρόσθια(Sequential Forward Feature Selection,SFS)

Αποτελεί τον απλούστερο “απλειστο” αλγόριθμο αναζήτησης. Εκκινώντας από κενό σύνολο γνωρισμάτων, προστίθεται ακολουθιακά/επαναληπτικά το γνώρισμα  $x^+$  που οδηγεί σε μεγιστοποίηση της αντικειμενικής συνάρτησης  $J(Y_k + x^+)$ , όταν συνδυαστεί με τα γνωρίσματα  $Y^k$ , που έχουν ήδη επιλεγεί. Το πλήθος των συνδυασμών, επί των οποίων η μέθοδος αναζητά λύση είναι  $l * m - l * \frac{l-1}{2}$ , όπου  $m$  το αρχικό και  $l$  το τελικό πλήθος χαρακτηριστικών του Διανύσματος Γνωρισμάτων. Πρόκειται για πλήθος γνωρισμάτων σημαντικά μικρότερο εκείνου της πλήρους/εξαντλητικής αναζήτησης. Επομένως, από τη σκοπιά της απαιτούμενης χρονικής πολυπλοκότητας, η Προς τα Πίσω (Οπισθοδρομική) αναζήτηση είναι πιο αποτελεσματική από την Προς τα Εμπρός (Εμπρόσθια), όταν η τιμή του  $l$  είναι πιο κοντά στην τιμή του  $m$  από ό,τι στο 1. Όταν η αναζήτηση βρίσκεται κοντά στο κενό σύνολο, ένα μεγάλο πλήθος καταστάσεων εκτιμάται, ενώ κοντά στο πλήρες σύνολο, η προς εξέταση περιοχή είναι μικρότερη, επειδή τα περισσότερα από τα γνωρίσματα έχουν ήδη επιλεγεί.

### Το μειονέκτημα της επιλογής της SFS

Αδυναμία απομάκρυνσης γνωρισμάτων που γίνονται ασήμαντα, ενώ προστίθενται νέα σημαντικότερα.



### **2.9.1.2 Προς τα Πίσω/Οπισθοδρομική(Sequential Backward Feature Selection/Elimination,SBS)**

Δουλεύει στην αντίθετη κατεύθυνση συγκριτικά από την SFS. Εκκινώντας από πλήρες σύνολο γνωρισμάτων, αφαιρείται ακολουθιακά/επαναληπτικά το γνώρισμα  $x^-$  που οδηγεί στη μικρότερη μείωση στην τιμή της αντικειμενικής συνάρτησης  $J(Y - x^-)$ .

Η αφαίρεση ενός γνωρίσματος μπορεί ενδεχομένως να οδηγεί σε αύξηση της τιμής της αντικειμενικής συνάρτησης  $J(Y_k - x^-) > J(Y_k)$ .

Το πλήθος των συνδυασμών, επί των οποίων η μέθοδος αναζητά λύση είναι  $1 + \frac{1}{2} * ((m + 1) * m - l * (l + 1))$ , όπου  $m$  το αρχικό και  $l$  το τελικό πλήθος χαρακτηριστικών του Διανύσματος Γνωρισμάτων. Πρόκειται για πλήθος γνωρισμάτων σημαντικά μικρότερο εκείνου της εξαντλητικής αναζήτησης.

#### **Το μειονέκτημα της SBS**

Αδυναμία για επανεκτίμηση της χρησιμότητας ενός γνωρίσματος, αν αυτό αφαιρεθεί.

Πρόκειται για άπλειστες (greedy) και υποβέλτιστες (suboptimal) τεχνικές αναζήτησης για Επιλογή Γνωρισμάτων που δεν εξετάζουν όλα τα δυνατά υποσύνολα. Άρα, δεν παρέχουν εγγύηση εύρεσης βέλτιστου υποσυνόλου, για βελτιστοποίηση ταξινόμησης και μεγιστοποίηση ακρίβειας Αναγνώρισης.

# Κεφάλαιο 3: Μελέτη ταξινομητών ως προς τα πλεονεκτήματα/μειονεκτήματα

Μετά την εξαγωγή (και) την Επιλογή Γνωρισμάτων, η Αναγνώριση Συναισθήματος από το Σήμα του Λόγου μπορεί να θεωρηθεί ως κλασικό πρόβλημα Εξόρυξης Δεδομένων (Data Mining problem).

Από ευρύτερη σκοπιά, οι ταξινομητές διακρίνονται σε 2 κατηγορίες:

- 1) Στατικοί
- 2) Δυναμικοί

## 3.1 Μηχανή Διανύσματος Υποστήριξης (Support Vector Machine, SVM)

Αναπτύχθηκε από το Vapnik (AT & T Laboratories 1995). Προσφέρει και μη γραμμική ταξινόμηση, με μετασχηματισμό του Διανύσματος Γνωρισμάτων σε ένα χώρο πολλών διαστάσεων, μέσω μιας διαδικασίας mapping που εκτελείται από την επιλεγμένη συνάρτηση πυρήνα (kernel function). Υπάρχουν οι εξής δυνατότητες επιλογής πυρήνα:

Ακτινική Συνάρτηση Βάσης (Radial Basis Function, RBF) της μορφής  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$

Γραμμική Συνάρτηση (Linear Function)

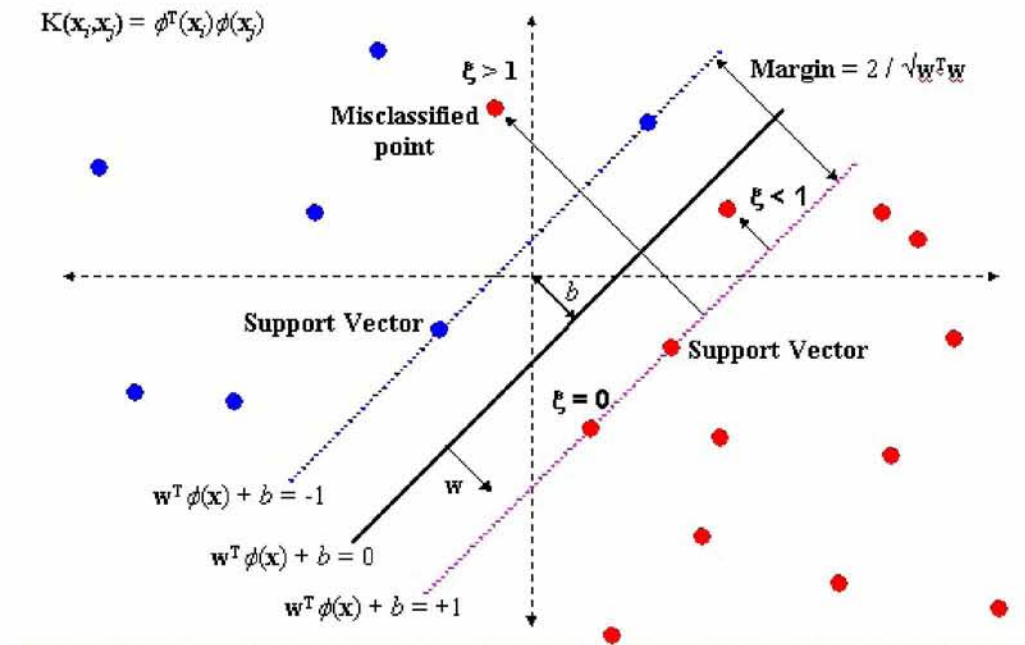
Πολυωνυμική Συνάρτηση (Polynomial Function)

Σιγμοειδής Συνάρτηση (Sigmoid Function)

Η SVM έχει σχεδιαστεί για ταξινόμηση μεταξύ 2 κλάσεων, μιας θετικών δειγμάτων +1 και μιας αρνητικών -1. Το δείγμα ταξινομείται με τη βοήθεια εύρεσης ενός διαχωριστικού υπερεπίπεδου (separating hyperplane) που μεγιστοποιεί το περιθώριο μεταξύ των 2 κλάσεων. Αυτό το μέγιστο υπερεπίπεδο έχει τη μέγιστη απόσταση από τα γειτονικά δείγματα και των 2 κλάσεων. Τα δείγματα που βρίσκονται στο περιθώριο (marginal samples) καλούνται Διανύσματα Υποστήριξης (Support Vectors).

Συνολικά, υπάρχουν 3 περιπτώσεις ταξινόμησης ενός δείγματος:

- 1) Με βεβαιότητα ως κλάση +1, αν ισχύει  $w * x + b > 1$
- 2) Με βεβαιότητα ως κλάση -1, αν ισχύει  $w * x + b < -1$
- 3) Αβεβαιότητα, αν ισχύει  $w * x + b = 0$



Σχήμα 5: Ταξινόμηση ενός δείγματος με ταξινομητή κλασική SVM[5]

### Πλεονεκτήματα της επιλογής SVM

- 1) Μεγάλη ικανότητα γενίκευσης ταξινόμησης
- 2) Μεγάλη ακρίβεια ταξινόμησης, σε ορισμένα πειράματα μάλιστα τη μέγιστη
- 3) Δυνατότητα μη γραμμικής ταξινόμησης με επιλογή μη γραμμικής συνάρτησης πυρήνα

Για ταξινόμηση περισσότερων από 2 κλάσεων, υπάρχουν 3 ευρέως διαδεδομένοι τύποι SVMs: Μία-Έναντι-Μίας (1-VS-1), Μία-Έναντι-Υπολοίπων (1-VS-Rest) και πολλαπλών επιπέδων (Multi Layered).

### 3.1.1 Μηχανή Διανύσματος Υποστήριξης Μία-Έναντι-Μίας (1-VS-1 SVM)

Συγκρίνει κάθε ζεύγος κλάσεων. Κατασκευάζει συνολικά  $\frac{n * (n - 1)}{2}$  δυαδικοί ταξινομητές, όπου  $n$  το συνολικό πλήθος των κλάσεων προς ταξινόμηση. Το αποτέλεσμα της ταξινόμησης προκύπτει μέσω μιας στρατηγικής, ενός σχήματος απόφασης "το μέγιστο κερδίζει" ("maximum wins"): Κάθε ταξινομητής βγάζει ως έξοδο μία κλάση, η έξοδος κάθε ταξινομητή υπολογίζεται για όλα τα ζεύγη κλάσεων και η κλάση που έχει προκύψει ως έξοδος κατά το μεγαλύτερο πλήθος φορών αποτελεί το αποτέλεσμα της ταξινόμησης.

#### Το μειονέκτημα της επιλογής 1-VS-1 SVM

Πρέπει να εκπαιδευτεί ένα σχετικά μεγάλο πλήθος δυαδικών ταξινομητών SVMs.

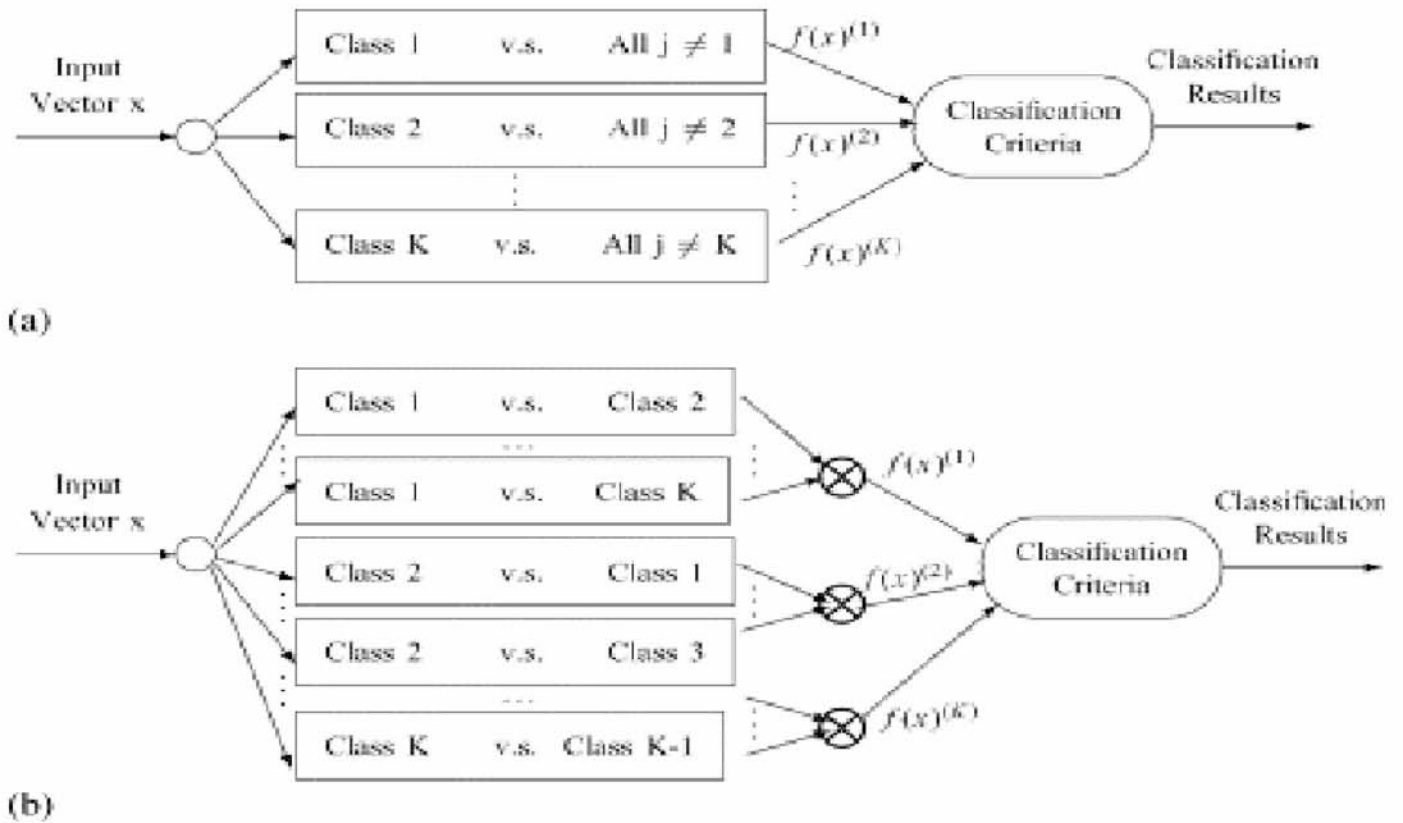
### 3.1.2 Μηχανή Διανύσματος Υποστήριξης Μία-Έναντι-Υπολοίπων (1-VS-Rest SVM)

Συγκρίνει κάθε κλάση με όλες τις υπόλοιπες. Κατασκευάζει συνολικά  $n$  ταξινομητές, συνεπώς λιγότερους από την 1-VS-1. Κάθε ταξινομητής διαχωρίζει μεταξύ μιας κλάσης και όλων των υπολοίπων. Το αποτέλεσμα της ταξινόμησης προκύπτει μέσω μιας στρατηγικής, ενός σχήματος απόφασης "ο νικητής τα παίρνει όλα" (winner takes all), κατά την οποία, η κλάση με τη μεγαλύτερη τιμή εξόδου κερδίζει.

#### Το μειονέκτημα της επιλογής 1-VS-Rest SVM

Όταν το πλήθος των κλάσεων  $n$  είναι σχετικά μεγάλο, φθίνει η ακρίβεια ταξινόμησης.

Ακολουθεί εικόνα σύγκρισης του διαφορετικού τρόπου ταξινόμησης, όπου a) 1-VS-Rest, b) 1-VS-1 SVMs:



Σχήμα 6: Σύγκριση τρόπου λειτουργίας ταξινόμησης ταξινομητών a)1-VS-Rest SVM,b)1-VS-1 SVM[17]

### 3.2 k Πλησιέστεροι Γείτονες (k Nearest Neighbors, kNN)

Πρόκειται για ένα μη παραμετρικό, σκληρό αλγόριθμο μάθησης. Δεν κάνει υποθέσεις για την υποκείμενη κατανομή των δεδομένων. Αυτό είναι αρκετά χρήσιμο, επειδή στον πραγματικό κόσμο, τα περισσότερα από τα πρακτικά δεδομένα δεν υπακούουν σε θεωρητικά μοντέλα. Σε αντίθεση, τα Μείγματα Γκαουσιανών (Gaussian Mixture) και η Γραμμική Διακριτή Ανάλυση (Linear Discriminant Analysis) αποτελούν παραμετρικούς αλγορίθμους μάθησης.

Βασίζεται στο ότι παρόμοιες παρατηρήσεις ανήκουν σε παρόμοιες κλάσεις. Το δείγμα συγκρίνεται με k δείγματα κατάρτισης, με βάση το ποιο γειτονικό δείγμα βρίσκεται πλησιέστερα σε ομοιότητα. Διάφορες μετρικές χρησιμοποιούνται για αξιολόγηση του βαθμού ομοιότητας, με πιο συχνά χρησιμοποιημένη την Ευκλείδεια απόσταση (Euclidean distance metric).

Καθοριστικός είναι ο αριθμός  $k$  που δηλώνει το πλήθος των πλησιέστερων γειτόνων που προκύπτουν με βάση την επιλεγμένη μετρική και επηρεάζουν την ταξινόμηση. Συνήθως, είναι περιττός αριθμός, αν το συνολικό πλήθος κλάσεων είναι μόνο 2. Το αποτέλεσμα ταξινόμησης διαφέρει με αλλαγή της επιλογής της τιμής του αριθμού  $k$ . Μεγαλύτερη ακρίβεια ταξινόμησης προκύπτει, όταν το πλήθος των training δεδομένων δεν είναι πολύ μεγάλο. Αν είναι, υπάρχει μεγάλη πιθανότητα σφάλματος ταξινόμησης. Η ακρίβεια ταξινόμησης ενδέχεται να αυξάνεται με αύξηση της επιλεγμένης τιμής του  $k$ , με επίπτωση όμως στην υπολογιστική πολυπλοκότητα. Η χρονική πολυπλοκότητά του είναι  $O(d * n)$ , όπου  $d$  η διάσταση του προς ταξινόμηση δείγματος και  $n$  το συνολικό πλήθος κλάσεων. Επιλογή μικρής τιμής για το  $k$  συνεπάγεται ότι το Σήμα θορύβου έχει μεγαλύτερη επίδραση στο αποτέλεσμα ταξινόμησης. Δεν υπάρχει καθολικό consensus μεταξύ των ερευνητών για το ποια τιμή πρέπει να χρησιμοποιείται για το  $k$ . Δοκιμάζονται διαφορετικές τιμές και, στο τέλος, διατηρείται αυτή που οδηγεί στη μεγιστοποίηση της ακρίβειας ταξινόμησης. Πειράματα έχουν διεξαχθεί για τιμές 1 ως 20.

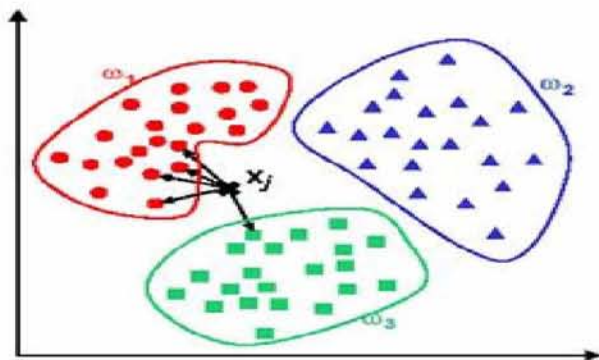
### **Το πλεονέκτημα της επιλογής $kNN$**

Αρκετά γρήγορη φάση training.

### **Μειονέκτημα της επιλογής $kNN$**

- 1) Απαιτείται σημαντική ποσότητα μνήμης για αποθήκευση όλων των δεδομένων training, επειδή χρησιμοποιούνται (σχεδόν) όλα.
- 2) Αδυναμία γενίκευσης αποτελέσματος ταξινόμησης

**Παράδειγμα:** Έστω  $k = 5$ , το συνολικό πλήθος κλάσεων είναι 3:  $\omega_1$ ,  $\omega_2$  και  $\omega_3$  και έστω ότι προκύπτουν 4 περιπτώσεις "labeled" ως  $\omega_1$ , 0 ως  $\omega_2$  και 1 ως  $\omega_3$ . Τότε, ο 5NN χρησιμοποιεί μια στρατηγική, ένα σχήμα πλειοψηφίας και ταξινομεί τελικά το δείγμα ως κλάση  $\omega_1$ .



Σχήμα 7: Ταξινόμηση ενός δείγματος με ταξινομητή 5NN[6]

### 3.3 k-μέσων ομαδοποίηση (k-means clustering)

Χρησιμοποιούν ως μετρική απόφασης την Ευκλείδεια απόσταση και ταξινομούν ένα δείγμα με κριτήριο την Ευκλείδεια απόστασή του από τη μέση τιμή κάθε κλάσης.

### 3.4 Bayesian/Belief Δίκτυα (Bayesian/Belief Network, BN)

Πρόκειται για ένα ΚΑΓ (Κατευθυνόμενο Άκυκλο Γράφο), όπου οι κορυφές του σχετίζονται με μεταβλητές κατάστασης από ένα πεπερασμένο πλήθος καταστάσεων και οι ακμές του συμβολίζουν τις δεσμευμένες (υπό συνθήκη) πιθανότητες των κορυφών και των γονεικών τους.

### 3.5 Κρυμμένα Μαρκοβιανά Μοντέλα (Hidden Markov Models, HMMs)

Πρόκειται για δυναμικό ταξινομητή και δη για το απλούστερο δυναμικό Bayesian Δίκτυο. Οι υποστηρικτές τους τονίζουν ότι τα γνωρίσματα δεν μπορεί να είναι στατικά, αλλά δυναμικά μεταβαλλόμενα.

#### Πλεονεκτήματα της επιλογής HMMs

- 1) Λαμβάνουν υπόψη τη χρονική συμπεριφορά του Σήματος του Λόγου.
- 2) Γρήγορη ταξινόμηση πραγματικού χρόνου, επειδή η προεπεξεργασία δε χρειάζεται να περιμένει για



ολόκληρη την άρθρωση.

### **3.6 Μεικτά Γκαουσιανά Μοντέλα (Gaussian Mixture Models, GMMs)**

Η ταξινόμηση ενός δείγματος γίνεται με προσέγγιση της Συνάρτησης Πυκνότητας Πιθανότητας (Probability Density Function, PDF) των γνωρισμάτων με ένα σύνολο σταθμισμένων Γκαουσιανών. Αποτελούν κατά κάποιο τρόπο τη στατική εκδοχή των δυναμικών HMMs. Συγκεκριμένα, ισχύει η εξής ισοδυναμία :

GMM = HMM μίας κατάστασης

#### **Το μειονέκτημα της επιλογής GMMs**

Περίπλοκη ταξινόμηση

### **3.7 Γραμμική Διακριτή Ανάλυση (Linear Discriminant Analysis)**

Η ταξινόμηση πραγματοποιείται με υπολογισμό γραμμικών συνδυασμών γνωρισμάτων που να διαχωρίζουν όσο το δυνατόν καλύτερα μεταξύ των υπαρκτών κλάσεων. Ένας μετασχηματισμός εφαρμόζεται για αλλαγή του συστήματος συντεταγμένων, ώστε να μεγιστοποιείται η διαφορά μεταξύ των κλάσεων.

#### **Πλεονεκτήματα της επιλογής LDA**

- 1) Απλή ταξινόμηση
- 2) Γρήγορη ταξινόμηση

### **3.8 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks, ANNs)**

Η ταξινόμηση πραγματοποιείται με χρήση μη γραμμικών συναρτήσεων.

Έχουν σχεδιαστεί ως προσομοίωση της συμπεριφοράς των ανθρώπινων νευρώνων. Αποτελούνται από ένα

πλήθος συνδεδεμένων τεχνητών νευρώνων, οργανωμένων σε 3 ή περισσότερα επίπεδα :

- 1) Εισόδου (Input layer neurons/nodes)
- 2) Κρυμμένου (Hidden layer neurons/nodes)
- 3) Εξόδου (Output layer neurons/nodes)

Έχουν χρησιμοποιηθεί πολλές υλοποιήσεις με διαφορετικό πλήθος νευρώνων/κόμβων ανά επίπεδο.

**Παραδείγματα :**

- 1) Feedforward Neural Networks (FNN), όπου υπάγεται και ο ταξινομητής Multi Layered Perceptron (MLP)
- 2) Back Propagation Neural Networks (BPNN)
- 3) Long Short Term Neural Networks (LSTN)

**Το πλεονέκτημα της επιλογής ANNs**

Μπορούν να ταξινομήσουν και δείγματα που δεν είναι γραμμικά ή μέσω ενός υπερεπίπεδου διαχωρίσιμα, χάρη στη χρήση των μη γραμμικών συναρτήσεών τους.

**Το μειονέκτημα της επιλογής ANNs**

Απαιτείται διαφορετική επιλογή αρχιτεκτονικής για κάθε συγκεκριμένο σύνολο δεδομένων, για βελτιστοποίηση της ταξινόμησης.

# Κεφάλαιο 4: Τα πειράματά μας

## 4.1 Βάση δεδομένων Interactive Emotion Capture (IEmoCap)

Εν γένει, πρέπει να τονιστεί ότι η κατάλληλη επιλογή μιας βάσης δεδομένων Συναισθημάτων προς χρήση για την Αναγνώριση Συναισθήματος θεωρείται καθοριστικής σημασίας για την ποιότητα του αντίστοιχου συστήματος που επιθυμεί κανείς να κατασκευάσει.

Και στα δύο πειράματα, χρησιμοποιήθηκε η βάση δεδομένων Συναισθημάτων IEmoCap. Δεν είναι ακόμα τόσο διαδεδομένη, επειδή αναπτύχθηκε πρόσφατα από το τμήμα Viterbi School of Engineering του University of Southern California, USC, και έχουν διεξαχθεί ελάχιστα πειράματα πάνω της.

Η IEmoCap περιέχει διτροπική πληροφορία Συναισθήματος (Λόγο και Εικόνα) συνολικά 10 ηθοποιών και των δύο φύλων, 5 αντρών και 5 γυναικών. Ωστόσο, και στα δύο πειράματα χρησιμοποιήθηκαν αποκλειστικά και μόνο οι αρθρώσεις του Σήματος του Λόγου και όχι και της Εικόνας, επειδή η εστίαση της Διπλωματικής Εργασίας είναι πάνω στην Αναγνώριση Συναισθήματος από το Σήμα του Λόγου και όχι και μέσω της σύνηξης διαφορετικών τροπικοτήτων πληροφορίας (fusion of different information modalities), όπως Λόγου και Εικόνας, που έχει αποδειχτεί ότι οδηγεί σε αύξηση της ακρίβειας Αναγνώρισης.

Κάθε άρθρωση έχει γίνει "labeled" από κριτές τόσο ως προς τη διακριτή Συναισθηματική κατηγορία της όσο και ως προς τις διαστάσεις της στο συνεχές επίπεδο αναπαράστασης Συναισθημάτων, με ανάθεση τιμών για τα πεδία Valence, Activation, Dominance. Εντούτοις, και στα δύο πειράματα, τα Συναισθήματα εξετάζονται αποκλειστικά και μόνο ως προς τη διακριτή και όχι και τη συνεχή κατηγοριοποίησή τους. Περιλαμβάνονται τα εξής 4 Συναισθήματα:

1) Θυμός

- 2) Χαρά
- 3) Λύπη
- 4) Ουδέτερη Συναισθηματική κατάσταση

Συνολικά, περιέχονται 5479 "labeled" αρθρώσεις Σήματος Λόγου, εκ των οποίων:

- 1) 1083 αρθρώσεις "labeled" ως Θυμός
- 2) 1630 αρθρώσεις "labeled" ως Χαρά
- 3) 1083 αρθρώσεις "labeled" ως Λύπη
- 4) 1683 αρθρώσεις "labeled" ως Ουδέτερη Συναισθηματική κατάσταση

## **4.2 Εφαρμοσμένη μεθοδολογία και για τα δύο πειράματα**

### **4.2.1 Επιλογή κατάλληλου μέρους των συνολικών αρθρώσεων της IEmoCap**

Επειδή το να εξεταστούν και οι 5479 αρθρώσεις θα ήταν πρακτικά χρονικά αδύνατο στο πλαίσιο της παρούσας Διπλωματικής Εργασίας, χρησιμοποιήθηκε ένα επαρκές, αρκετά ικανοποιητικό και ταυτόχρονα αντιπροσωπευτικό πλήθος τους. Συγκεκριμένα, 1096 αρθρώσεις, εκ των οποίων:

- 1) 217 αρθρώσεις "labeled" ως Θυμός
- 2) 326 αρθρώσεις "labeled" ως Χαρά
- 3) 217 αρθρώσεις "labeled" ως Λύπη
- 4) 336 αρθρώσεις "labeled" ως Ουδέτερη Συναισθηματική κατάσταση

Διατηρώντας έτσι ακριβώς το λόγο πλήθους αρθρώσεων ανά Συναίσθημα προς συνολικό πλήθος αρθρώσεων του αρχικού, πλήρους πλήθους αρθρώσεων της IEmoCap, όπως παρουσιάζεται στην επόμενη εικόνα:



Σχήμα 8: Ποσοστά χρησιμοποιημένων αρθρώσεων της IEmoCap ανά Συναισθηματικό Κατάστημα

#### 4.2.2 Κωδικοποίηση Συναισθημάτων σε integer labels

Τα 4 Συναισθηματικά Κατάστημα της IEmoCap κωδικοποιήθηκαν σε τιμές ακεραίων, όπως επιδεικνύεται παρακάτω:

Πίνακας 1: Κωδικοποίηση των 4 Συναισθημάτων της IEmoCap

Συναισθηματικό Κατάστημα	Ακέραια τιμή κωδικού Συναισθηματικού Κατάσματος
Θυμός	1
Χαρά	2
Λύπη	3
Ουδέτερη Συναισθηματική κατάσταση	4

#### 4.2.3 Αφαίρεση μη ηχηρών τμημάτων και εξαγωγή στατιστικών σε επίπεδο άρθρωσης

Σε κάθε άρθρωση, πραγματοποιήθηκε διαχωρισμός των ηχηρών/έμφωνων από τα μη ηχηρά/άφωνα τμήματα (παύσεις), αφαιρώντας τα τελευταία, κάνοντας χρήση κατάλληλου έτοιμου, ήδη υλοποιημένου αλγορίθμου-κώδικα στο Matlab, που κάνει χρήση κριτηρίου βασισμένου σε κάτω κατώφλι της τιμής της θεμελιώδους συχνότητας, την οποία και υπολογίζει, και μπορεί να βρεθεί στο: [7].

Και αυτό, επειδή τα μη ηχηρά/άφωνα τμήματα του Σήματος του Λόγου έχει αποδειχτεί ότι δε συνεισφέρουν στη Συναισθηματική πληροφορία.

Σε κάθε άρθρωση, τα απομειναντα ηχηρά/έμφωνα τμήματα συνενώθηκαν μεταξύ τους (voiced segments concatenation), μειώνοντας έτσι την αρχική χρονική διάρκεια κάθε άρθρωσης.

Από τις νέες παραχθείσες αρθρώσεις έκανα εξαγωγή στατιστικών συγκεκριμένων γνωρισμάτων. Η εξαγωγή στατιστικών γνωρισμάτων για τα δύο πειράματα της παρούσας Διπλωματικής Εργασίας έγινε σε επίπεδο άρθρωσης και όχι κάθε ξεχωριστού πλαισίου κάθε άρθρωσης. Λήφθηκαν έτσι μακροπρόθεσμα, όπως και συνηθίζεται στα περισσότερα πειράματα, και όχι βραχυπρόθεσμα στατιστικά γνωρισμάτων. Για εξαγωγή όλων των γνωρισμάτων, χρησιμοποιήθηκαν έτοιμες, ήδη υλοποιημένες ρουτίνες στο Matlab. Αναλυτικά :

1) Για εξαγωγή των LPCs, Formants και της θεμελιώδους συχνότητας, για την οποία εφαρμόστηκε η μέθοδος αυτοσυσχέισης (autocorrelation pitch extraction method) , έγινε χρήση ρουτίνων που μπορούν να βρεθούν στο: [8].

2) Για εξαγωγή PLPs + MFCCs, έγινε χρήση ρουτίνας που μπορεί να βρεθεί στο: [9].

3) Για εξαγωγή RASTA-PLPs, έγινε χρήση ρουτίνας που μπορεί να βρεθεί στο: [10].

4) Για υπολογισμό του Zero Crossing Rate (ZCR), έγινε χρήση ρουτίνας που μπορεί να βρεθεί στο: [11].

5) Για εξαγωγή της ενέργειας rms, έγινε χρήση ρουτίνας που μπορεί να βρεθεί στο: [12].

6) Τέλος, για υπολογισμό του ρυθμού HNR, έγινε χρήση ρουτίνας που μπορεί να βρεθεί στο: [13].

#### **4.2.4 Καθορισμός training, testing sets**

Το συνολικό πλήθος των χρησιμοποιημένων 1096 αρθρώσεων της IEmoCap χωρίστηκε σε 2 μέρη :

**1) Στο σύνολο εκπαίδευσης (training set), αντίστοιχο με το 80% του πλήθους των χρησιμοποιημένων αρθρώσεων ανά Συναισθημα:**

α) 174 αρθρώσεις "labeled" ως Θυμός

β) 261 αρθρώσεις "labeled" ως Χαρά

γ) 174 αρθρώσεις "labeled" ως Λύπη

δ) 289 αρθρώσεις "labeled" ως Ουδέτερη Συναισθηματική κατάσταση

**2) Στο σύνολο ελέγχου (testing set), αντίστοιχο με το 20% του πλήθους των χρησιμοποιημένων αρθρώσεων ανά Συναίσθημα :**

α) 43 αρθρώσεις "labeled" ως Θυμός

β) 65 αρθρώσεις "labeled" ως Χαρά

γ) 43 αρθρώσεις "labeled" ως Λύπη

δ) 67 αρθρώσεις "labeled" ως Ουδέτερη Συναισθηματική κατάσταση

### **4.3 Πρώτο πείραμα: Σύγκριση SFS, SBS, με κριτήριο σύγκλισής τους, την ελαχιστοποίηση σφάλματος ταξινόμησης του INN, στην IEmoCap**

#### **4.3.1 Χρησιμοποιημένα στατιστικά στο πρώτο πείραμα**

Για το πρώτο πείραμα, έκανα εξαγωγή των ακόλουθων 203 στατιστικών:

Πίνακας 2: Χρησιμοποιημένα στατιστικά στο πρώτο πείραμα

Ελάχιστο της θεμελιώδους συχνότητας
Μέγιστο της θεμελιώδους συχνότητας
Μέση τιμή της θεμελιώδους συχνότητας
Ελάχιστο των πρώτων 2 Formants F1, F2
Μέγιστο των πρώτων 2 Formants F1, F2
Μέση τιμή των πρώτων 2 Formants F1, F2
Τυπική απόκλιση των πρώτων 2 Formants F1, F2
Ελάχιστο των πρώτων 12 MFCCs
Μέγιστο των πρώτων 12 MFCCs

Πίνακας 2: Χρησιμοποιημένα στατιστικά στο πρώτο πείραμα

Μέση τιμή των πρώτων 12 MFCCs
Τυπική απόκλιση των πρώτων 12 MFCCs
Ελάχιστο των πρώτων 12 LPCs
Μέγιστο των πρώτων 12 LPCs
Μέση τιμή των πρώτων 12 LPCs
Τυπική απόκλιση των πρώτων 12 LPCs
Ελάχιστο των πρώτων 12 PLPs
Μέγιστο των πρώτων 12 PLPs
Μέση τιμή των πρώτων 12 PLPs
Τυπική απόκλιση των πρώτων 12 PLPs
Ελάχιστο των πρώτων 12 RASTA-PLPs
Μέγιστο των πρώτων 12 RASTA-PLPs
Μέση τιμή των πρώτων 12 RASTA-PLPs
Τυπική απόκλιση των πρώτων 12 RASTA-PLPs

#### 4.3.2 Εφαρμογή SFS στο μεγάλο Διάνυσμα Γνωρισμάτων

Εφαρμόστηκε SFS στο Διάνυσμα Γνωρισμάτων των 203 στατιστικών, με χρήση στο Matlab μιας έτοιμης, ήδη υλοποιημένης ρουτίνας που συγκλίνει με τερματικό κριτήριο την ελαχιστοποίηση του σφάλματος ταξινόμησης του ταξινομητή 1NN (1NN classification error minimization), δίνοντας ως παράμετρο, επιλογή 'NN' στη συγκεκριμένη υλοποίηση ρουτίνας Matlab SFS που μπορεί να βρεθεί στο: [14].

##### 4.3.2.1 Πειραματισμός με επιλογή διαφορετικού πλήθους πιο σημαντικών στατιστικών κατά SFS και εύρεση του βέλτιστου

Εν γένει, ως παράμετρος επιλογής εισόδου στη ρουτίνα featsel, μπορούν να δοθούν διαφορετικά τερματικά κριτήρια σύγκλισης, όπως για παράδειγμα, η μεγιστοποίηση της ακρίβειας ταξινόμησης με χρήση



κάποιου συγκεκριμένου ταξινομητή ή ο υπολογισμός κάποιας μετρικής απόστασης των στατιστικών μεταξύ τους. Για το τρέχον πείραμα, χρησιμοποιήθηκε η προεπιλεγμένη (default) παράμετρος επιλογής εισόδου του τερματικού κριτηρίου σύγκλισης της `featsell` που είναι η παράμετρος 'NN', για ελαχιστοποίηση του σφάλματος ταξινόμησης με ταξινομητή 1NN.

Για κάθε ένα από τα 203 γνωρίσματα αντιστοιχίστηκε μία τιμή αξιολόγησής του, στα κελιά ενός πίνακα που επιστρέφει η ρουτίνα. Όταν ήταν συμπληρωμένες και οι 203 τιμές αξιολόγησης όλων των στατιστικών, η ρουτίνα εκκινεί από κενό σύνολο στατιστικών και σε κάθε επανάληψη προσθέτει εκείνο το στατιστικό με τη μεγαλύτερη τιμή αξιολόγησης, ελέγχοντας παράλληλα αν πληροίται η τερματική συνθήκη σύγκλισης της ελαχιστοποίησης του σφάλματος ταξινόμησης του ταξινομητή 1NN. Αν πληροίται, τερματίζει και επιστρέφει το τρέχον υποσύνολο στατιστικών ως βέλτιστο, διαφορετικά, συνεχίζει με την επόμενη επανάληψη και ούτω καθεξής... Μετά λοιπόν από 61 επαναλήψεις-προσθήκες πιο σημαντικού κατά την τιμή αξιολόγησης στατιστικού, η ρουτίνα `featsell` τερμάτισε, δίνοντας το βέλτιστο πλήθος στατιστικών. Συνεπώς, η μεγαλύτερη μέση ακρίβεια Αναγνώρισης Συναισθήματος προέκυψε με επιλογή των πιο σημαντικών 61 στατιστικών. Ενδεικτικά, συγκρίναμε τη μέση ακρίβεια Αναγνώρισης Συναισθήματος με επιλογή διαφορετικού πλήθους πιο σημαντικών στατιστικών, για παράδειγμα των 27, 107 και 159 κατά SFS + 1NN. Ακολουθούν τα αποτελέσματα του πλήθους των ορθά ταξινομημένων αρθρώσεων σε κάθε περίπτωση:

Πίνακας 3: Ορθά ταξινομημένες αρθρώσεις με επιλογή 27, 61, 107, 159 στατιστικών κατά SFS + 1NN

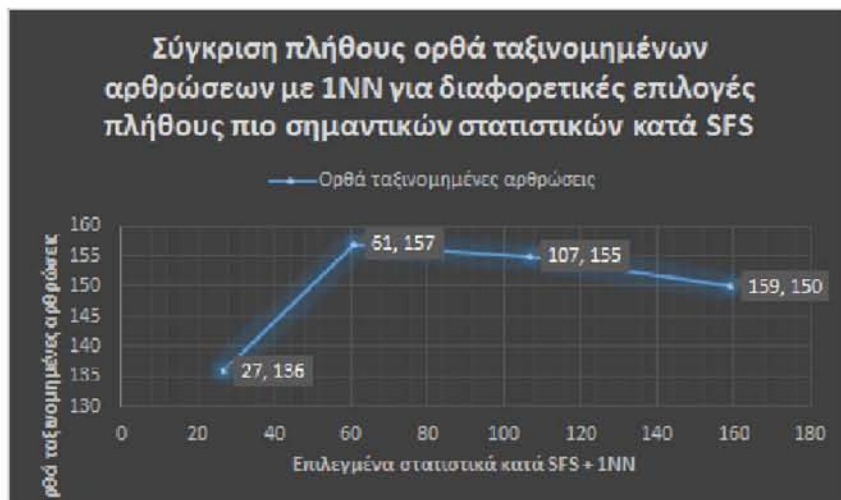
Πλήθος επιλεγμένων στατιστικών κατά SFS	Πλήθος ορθά ταξινομημένων αρθρώσεων
27	136
<b>61</b>	<b>157</b>
107	155
159	150

Ακολουθούν τα ποσοστά μέσης Αναγνώρισης Συναισθήματος, με επιλογή των πιο σημαντικών 27, 61, 107 και 159 στατιστικών κατά SFS + 1NN:

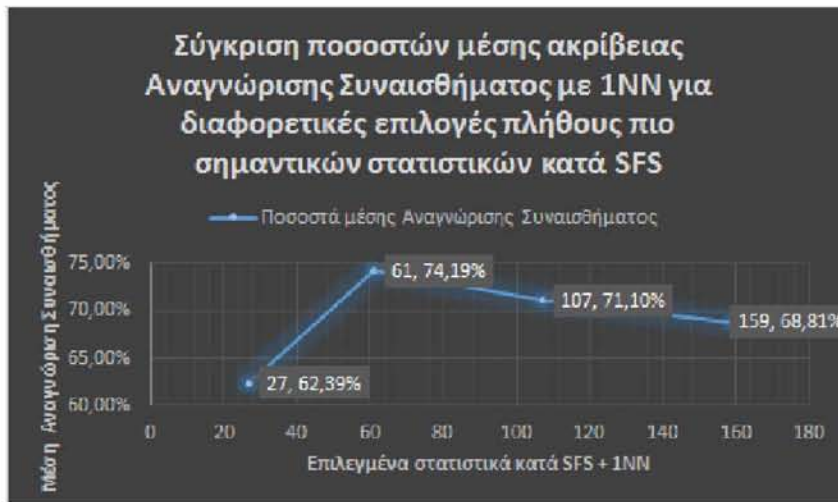
Πίνακας 4: Ποσοστά μέσης Αναγνώρισης Συναισθήματος με 27, 61, 107, 159 στατιστικά κατά SFS+1NN

Πλήθος επιλεγμένων στατιστικών κατά SFS+1NN	Πλήθος ορθά ταξινομημένων αρθρώσεων
27	62,3853%
<b>61</b>	<b>74,1870%</b>
107	71,1009%
159	68,8073%

Ακολουθούν δύο συγκριτικά ραβδογράμματα που επιδεικνύουν την επίδραση της επιλογής διαφορετικού πλήθους στατιστικών στην ορθή ταξινόμηση αρθρώσεων κατά SFS + 1NN:



Σχήμα 9: Σύγκριση ορθά ταξινομημένων αρθρώσεων με 27, 61, 107, 159 στατιστικά κατά SFS + 1NN



Σχήμα 10: Σύγκριση μέσης Αναγνώρισης Συναισθήματος με 27, 61, 107, 159 στατιστικά κατά SFS+1NN

Συνεπώς, για τη συνέχεια του πρώτου πειράματος εργαστήκαμε με τα επιλεγμένα 61 πιο σημαντικά στατιστικά κατά SFS + 1NN, τα οποία και παραθέτουμε στη συνέχεια:

#### 4.3.2.2 Βέλτιστο πλήθος στατιστικών και συνεπαγόμενη ιεράρχηση γνωρισμάτων

Πίνακας 5: Το βέλτιστο πλήθος στατιστικών κατά SFS + 1NN

1) Μέση τιμή του πρώτου Formant F1
2) Ελάχιστο του πρώτου Formant F1
3) Μέση τιμή του δεύτερου Formant F2
4) Μέση τιμή του πρώτου PLP1
5) Ελάχιστο του πρώτου PLP1
6) Μέση τιμή του δεύτερου PLP2
7) Μέση τιμή του τρίτου PLP3
8) Ελάχιστο του τρίτου PLP3
9) Τυπική απόκλιση του τρίτου PLP3
10) Τυπική απόκλιση του τέταρτου PLP4

Πίνακας 5: Το βέλτιστο πλήθος στατιστικών κατά SFS + 1NN

11) Μέση τιμή του πέμπτου PLP5
12) Μέγιστο του πέμπτου PLP5
13) Τυπική απόκλιση του πέμπτου PLP5
14) Μέση τιμή του έκτου PLP6
15) Μέγιστο του έκτου PLP6
16) Τυπική απόκλιση του έκτου PLP6
17) Μέση τιμή του έβδομου PLP7
18) Τυπική απόκλιση του έβδομου PLP7
19) Ελάχιστο του όγδοου PLP8
20) Τυπική απόκλιση του όγδοου PLP8
21) Μέση τιμή του ένατου PLP9
22) Ελάχιστο του ένατου PLP9
23) Μέγιστο του ένατου PLP9
24) Τυπική απόκλιση του ένατου PLP9
25) Μέση τιμή του δέκατου PLP10
26) Ελάχιστο του δέκατου PLP10
27) Μέγιστο του δέκατου PLP10
28) Τυπική απόκλιση του δέκατου PLP10
29) Μέση τιμή του ενδέκατου PLP11
30) Τυπική απόκλιση του ενδέκατου PLP11
31) Μέση τιμή του δωδέκατου PLP12
32) Μέγιστο του δωδέκατου PLP12
33) Τυπική απόκλιση του δωδέκατου PLP12
34) Μέση τιμή του πρώτου RASTA-PLP1
35) Ελάχιστο του πρώτου RASTA-PLP1

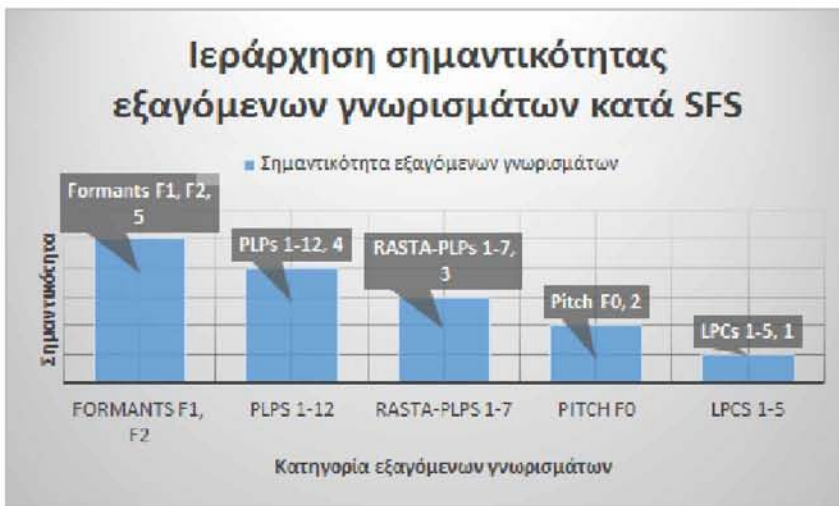
Πίνακας 5: Το βέλτιστο πλήθος στατιστικών κατά SFS + INN

36) Μέση τιμή του δεύτερου RASTA-PLP2
37) Μέση τιμή του τρίτου RASTA-PLP3
38) Ελάχιστο του τρίτου RASTA-PLP3
39) Τυπική απόκλιση του τρίτου RASTA-PLP3
40) Τυπική απόκλιση του τέταρτου RASTA-PLP4
41) Μέση τιμή του πέμπτου RASTA-PLP5
42) Μέγιστο του πέμπτου RASTA-PLP5
43) Τυπική απόκλιση του πέμπτου RASTA-PLP5
44) Μέση τιμή του έκτου RASTA-PLP6
45) Μέγιστο του έκτου RASTA-PLP6
46) Τυπική απόκλιση του έκτου RASTA-PLP6
47) Μέση τιμή του έβδομου RASTA-PLP7
48) Τυπική απόκλιση του έβδομου RASTA-PLP7
49) Ελάχιστο του όγδοου RASTA-PLP8
50) Ελάχιστο της θεμελιώδους συχνότητας
51) Μέγιστο της θεμελιώδους συχνότητας
52) Μέση τιμή του πρώτου LPC1
53) Ελάχιστο του πρώτου LPC1
54) Μέση τιμή του δεύτερου LPC2
55) Μέση τιμή του τρίτου LPC3
56) Ελάχιστο του τρίτου LPC3
57) Τυπική απόκλιση του τρίτου LPC3
58) Τυπική απόκλιση του τέταρτου LPC4
59) Μέση τιμή του πέμπτου LPC5
60) Μέγιστο του πέμπτου LPC5

Πίνακας 5: Το βέλτιστο πλήθος στατιστικών κατά SFS + INN

61) Τυπική απόκλιση του πέμπτου LPC5
--------------------------------------

Ακολουθεί μια εικόνα που επιδεικνύει την κατάρταξη ιεράρχησης των εξαγόμενων στατιστικών από την IEmoCap, ιεραρχημένα από την περισσότερο στη λιγότερο σημαντική κατηγορία :



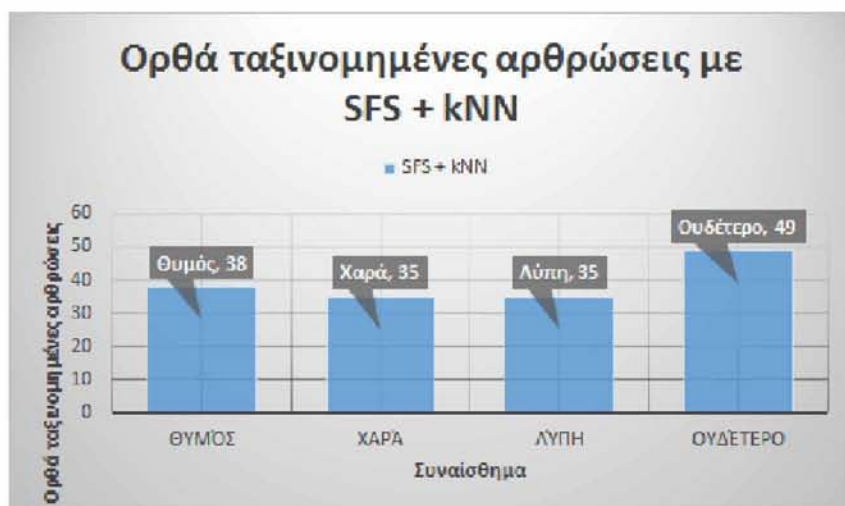
Σχήμα 11: Ιεράρχηση κατηγοριών εξαγόμενων γνωρισμάτων κατά SFS + INN για την IEmoCap

#### 4.3.2.3 Αναλυτικά αποτελέσματα ταξινόμησης του συνδυασμού SFS + INN

Πίνακας 6: Ταξινόμηση ανά Συναισθημα με το βέλτιστο πλήθος στατιστικών κατά SFS + INN

Συναίσθημα	Θυμός	Χαρά	Λύπη	Ουδέτερο	Πλήθος αρθρώσεων
<b>Θυμός</b>	<b>38</b>	2	1	2	43
<b>Χαρά</b>	15	<b>35</b>	4	11	65
<b>Λύπη</b>	2	1	<b>35</b>	5	43
<b>Ουδέτερο</b>	2	5	11	<b>49</b>	67

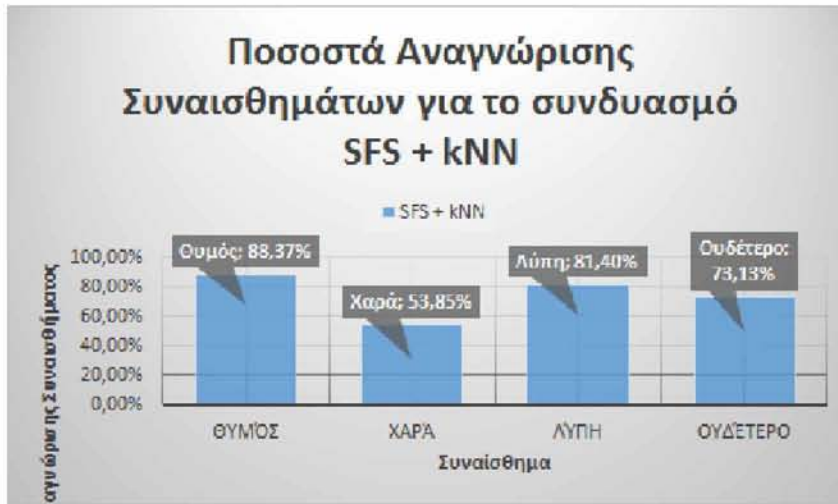
Ακολουθεί μια εικόνα που επιδεικνύει το πλήθος των ορθά ταξινομημένων αρθρώσεων:



Σχήμα 12: Ορθά ταξινομημένες αρθρώσεις με το βέλτιστο πλήθος στατιστικών κατά SFS + 1NN

Πίνακας 7: Ποσοστά Αναγνώρισης Συναισθημάτων με το βέλτιστο πλήθος στατιστικών κατά SFS + 1NN

Συναίσθημα	Θυμός	Χαρά	Λύπη	Ουδέτερο
Θυμός	<b>88,3721%</b>	4,6512%	2,3256%	4,6512%
Χαρά	23,0769%	<b>53,8462%</b>	6,1538%	16,9231%
Λύπη	4,6512%	2,3256%	<b>81,3953%</b>	11,6279%
Ουδέτερο	2,9851%	7,4627%	16,4179%	<b>73,1343%</b>



Σχήμα 13: Ποσοστά Αναγνώρισης Συναισθημάτων με το βέλτιστο πλήθος στατιστικών κατά SFS + INN

Η μέση ακρίβεια Αναγνώρισης Συναισθήματος ήταν **74,1870%**.

#### 4.3.3 Εφαρμογή SBS στο μεγάλο Διάνυσμα Γνωρισμάτων

Εναλλακτικά, εφαρμόστηκε SBS πάνω στο αρχικό Διάνυσμα Γνωρισμάτων των 203 στατιστικών, κάνοντας χρήση μιας έτοιμης, ήδη υλοποιημένης ρουτίνας Matlab που συγκλίνει με τερματικό κριτήριο την ελαχιστοποίηση του σφάλματος ταξινόμησης του ταξινομητή INN (INN classification error minimization), δίνοντας ως παράμετρο, επιλογή 'NN' στη συγκεκριμένη υλοποίηση ρουτίνας Matlab SBS που μπορεί να βρεθεί στο: [15].

##### 4.3.3.1 Πειραματισμός με επιλογή διαφορετικού πλήθους πιο σημαντικών στατιστικών κατά SBS και εύρεση του βέλτιστου

Εν γένει, ως παράμετρος επιλογής εισόδου στη ρουτίνα featselb, μπορούν να δοθούν διαφορετικά τερματικά κριτήρια σύγκλισης, όπως για παράδειγμα, η μεγιστοποίηση της ακρίβειας ταξινόμησης με χρήση κάποιου συγκεκριμένου ταξινομητή ή τον υπολογισμό κάποιας μετρικής απόστασης των γνωρισμάτων μεταξύ τους. Για το τρέχον πείραμα, χρησιμοποιήθηκε η προεπιλεγμένη (default) παράμετρος επιλογής εισόδου του τερματικού κριτηρίου σύγκλισης της featselb που είναι η παράμετρος 'NN', για ελαχιστοποίηση του σφάλματος ταξινόμησης με ταξινομητή INN.



Για κάθε ένα από τα 203 στατιστικά αντιστοιχίστηκε μία τιμή αξιολόγησής του, στα κελιά ενός πίνακα που επιστρέφει η ρουτίνα. Όταν ήταν συμπληρωμένες και οι 203 τιμές αξιολόγησης όλων των γνωρισμάτων, η ρουτίνα εκκινεί από πλήρες σύνολο στατιστικών και σε κάθε επανάληψη αφαιρεί εκείνο το στατιστικό με τη μικρότερη τιμή αξιολόγησης, ελέγχοντας παράλληλα αν πληροίται η τερματική συνθήκη σύγκλισης της ελαχιστοποίησης του σφάλματος ταξινόμησης του ταξινομητή INN. Αν πληροίται, τερματίζει και επιστρέφει το τρέχον υποσύνολο στατιστικών ως βέλτιστο, διαφορετικά, συνεχίζει με την επόμενη επανάληψη και ούτω καθεξής... Μετά λοιπόν από 152 επαναλήψεις-απαλοιφές λιγότερο σημαντικού κατά την τιμή αξιολόγησης στατιστικού, η ρουτίνα `featselb` τερμάτισε, δίνοντας το βέλτιστο πλήθος στατιστικών. Συνεπώς, η μεγαλύτερη μέση ακρίβεια Αναγνώρισης Συναισθήματος προέκυψε με επιλογή των πιο σημαντικών  $203 - 152 = 51$  στατιστικών. Ενδεικτικά, συγκρίναμε τη μέση ακρίβεια Αναγνώρισης Συναισθήματος με επιλογή διαφορετικού πλήθους πιο σημαντικών στατιστικών, για παράδειγμα των 27, 107 και 159 πιο σημαντικών κατά SBS + INN.

Ακολουθούν τα αποτελέσματα του πλήθους των ορθά ταξινομημένων αρθρώσεων σε κάθε περίπτωση:

Πίνακας 8: Ορθά ταξινομημένες αρθρώσεις με επιλογή 27, 51, 107, 159 στατιστικών κατά SBS+INN

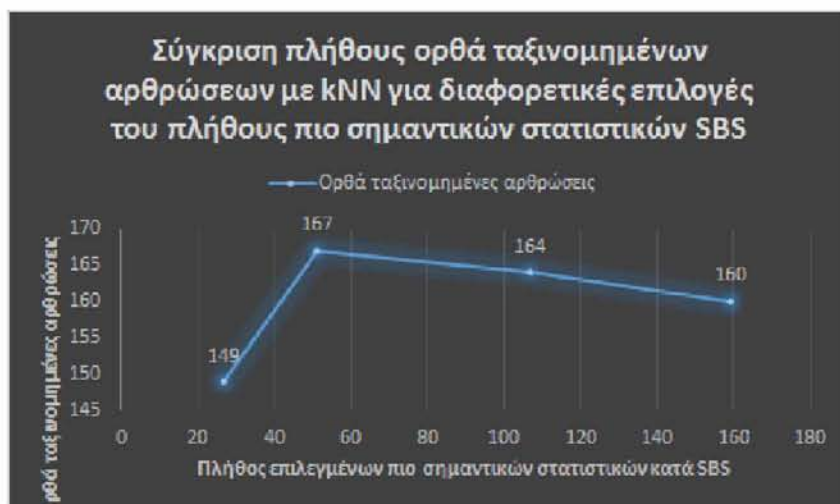
Πλήθος επιλεγμένων στατιστικών κατά SFS	Πλήθος ορθά ταξινομημένων αρθρώσεων
27	149
<b>51</b>	<b>167</b>
107	164
159	160

Ακολουθούν τα ποσοστά μέσης ακρίβειας Αναγνώρισης Συναισθήματος που προέκυψαν με επιλογή των πιο σημαντικών 27, 51, 107 και 159 στατιστικών κατά SBS + INN:

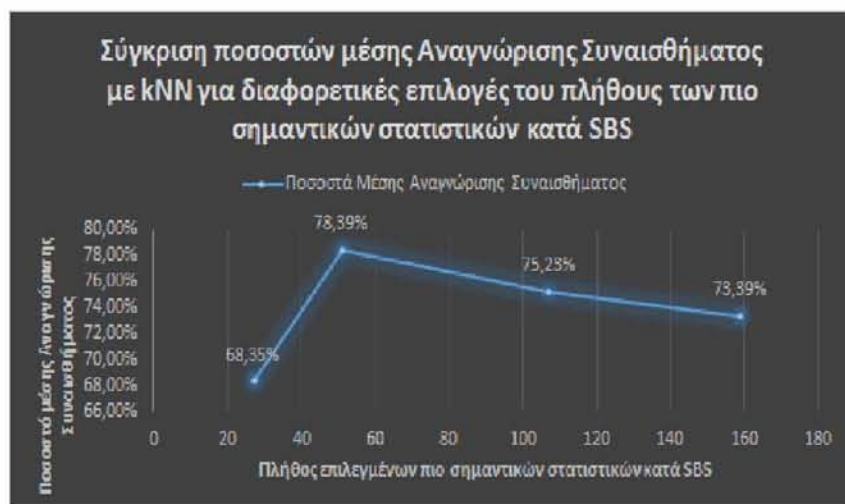
Πίνακας 9: Μέση Αναγνώριση Συναισθήματος με επιλογή 27, 51, 107, 159 στατιστικών κατά SBS

Πλήθος επιλεγμένων στατιστικών κατά SBS	Πλήθος ορθά ταξινομημένων αρθρώσεων
27	68,3486%
<b>51</b>	<b>78,3923%</b>
107	75,2294%
159	73,3945%

Ακολουθούν δύο συγκριτικές εικόνες για την επίδραση της επιλογής διαφορετικού πλήθους πιο σημαντικών στατιστικών στην ορθή ταξινόμηση αρθρώσεων για το συνδυασμό SBS + 1NN:



Σχήμα 14: Σύγκριση ορθά ταξινομημένων αρθρώσεων με 27, 51, 107, 159 στατιστικά κατά SBS + 1NN



Σχήμα 15: Σύγκριση μέσης Αναγνώρισης Συναισθήματος με 27, 51, 107, 159 στατιστικά κατά SBS+INN

Συνεπώς, για τη συνέχεια του πρώτου πειράματος εργαζόμαστε με τα επιλεγμένα 51 πιο σημαντικά στατιστικά κατά SBS + INN, τα οποία και παραθέτουμε στη συνέχεια :

#### 4.3.3.2 Το βέλτιστο πλήθος στατιστικών και συνεπαγόμενη ιεράρχηση γνωρισμάτων

Πίνακας 10: Το βέλτιστο πλήθος στατιστικών κατά SBS + INN

1) Μέση τιμή του πρώτου Formant F1
2) Ελάχιστο του πρώτου Formant F1
3) Μέση τιμή του δεύτερου Formant F2
4) Μέση τιμή του πρώτου PLP1
5) Ελάχιστο της θεμελιώδους συχνότητας
6) Μέγιστο της θεμελιώδους συχνότητας
7) Ελάχιστο του πρώτου PLP1
8) Μέση τιμή του δεύτερου PLP2

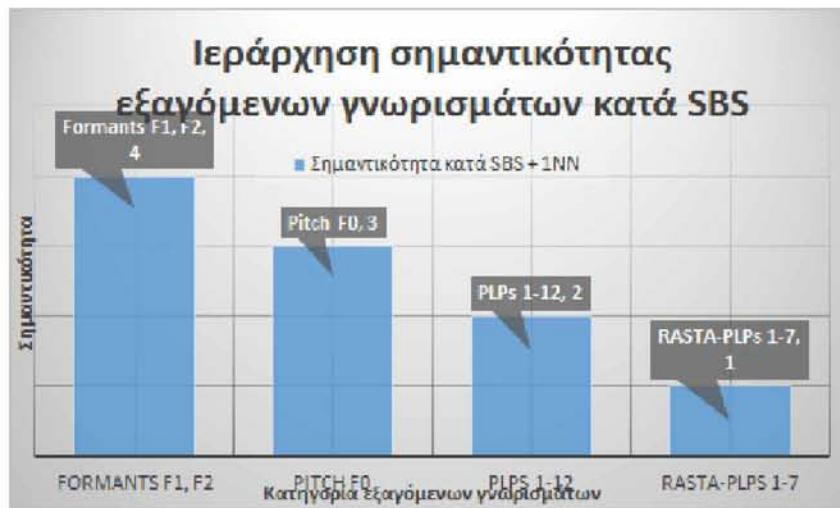
Πίνακας 10: Το βέλτιστο πλήθος στατιστικών κατά SBS + INN

9) Μέση τιμή του τρίτου PLP3
10) Ελάχιστο του τρίτου PLP3
11) Τυπική απόκλιση του τρίτου PLP3
12) Τυπική απόκλιση του τέταρτου PLP4
13) Μέση τιμή του πέμπτου PLP5
14) Μέγιστο του πέμπτου PLP5
15) Τυπική απόκλιση του πέμπτου PLP5
16) Μέση τιμή του έκτου PLP6
17) Μέγιστο του έκτου PLP6
18) Τυπική απόκλιση του έκτου PLP6
19) Μέση τιμή του έβδομου PLP7
20) Τυπική απόκλιση του έβδομου PLP7
21) Ελάχιστο του όγδοου PLP8
22) Τυπική απόκλιση του όγδοου PLP8
23) Μέση τιμή του ένατου PLP9
24) Ελάχιστο του ένατου PLP9
25) Μέγιστο του ένατου PLP9
26) Τυπική απόκλιση του ένατου PLP9
27) Μέση τιμή του δέκατου PLP10
28) Ελάχιστο του δέκατου PLP10
29) Μέγιστο του δέκατου PLP10
30) Τυπική απόκλιση του δέκατου PLP10
31) Μέση τιμή του ενδέκατου PLP11
32) Τυπική απόκλιση του ενδέκατου PLP11
33) Μέση τιμή του δωδέκατου PLP12

Πίνακας 10: Το βέλτιστο πλήθος στατιστικών κατά SBS + INN

34) Μέγιστο του δωδέκατου PLP12
35) Τυπική απόκλιση του δωδέκατου PLP12
36) Μέση τιμή του πρώτου RASTA-PLP1
37) Ελάχιστο του δεύτερου RASTA-PLP2
38) Μέση τιμή του δεύτερου RASTA-PLP2
39) Μέση τιμή του τρίτου RASTA-PLP3
40) Ελάχιστο του τρίτου RASTA-PLP3
41) Τυπική απόκλιση του τρίτου RASTA-PLP3
42) Τυπική απόκλιση του τέταρτου RASTA-PLP4
43) Μέση τιμή του πέμπτου RASTA-PLP5
44) Μέγιστο του πέμπτου RASTA-PLP5
45) Τυπική απόκλιση του πέμπτου RASTA-PLP5
46) Μέση τιμή του έκτου RASTA-PLP6
47) Μέγιστο του έκτου RASTA-PLP6
48) Τυπική του έκτου RASTA-PLP6
49) Μέση τιμή του έβδομου RASTA-PLP7
50) Τυπική απόκλιση του έβδομου RASTA-PLP7
51) Ελάχιστο του όγδοου RASTA-PLP8

Ακολουθεί μια εικόνα που επιδεικνύει την ιεράρχηση των διαφορετικών εξαγόμενων στατιστικών, από την περισσότερο στη λιγότερο σημαντική κατηγορία γνωρισμάτων για την IEmoCap.



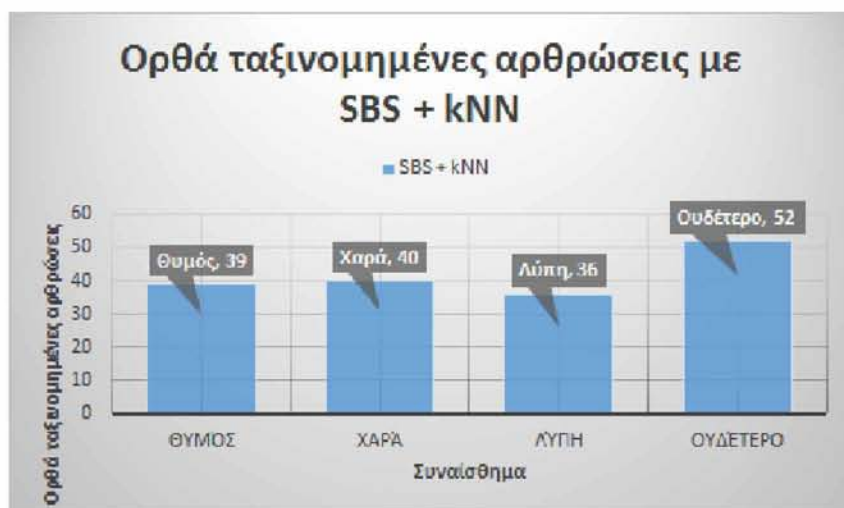
Σχήμα 16: Ιεράρχηση κατηγοριών εξαγόμενων γνωρισμάτων κατά SBS για την IEmoCap

#### 4.3.3.3 Αναλυτικά αποτελέσματα ταξινόμησης του συνδυασμού SBS + 1NN

Πίνακας 11: Ταξινόμηση αρθρώσεων με το βέλτιστο πλήθος στατιστικών κατά SBS + 1NN

Συναίσθημα	Θυμός	Χαρά	Λύπη	Ουδέτερο	Πλήθος αρθρώσεων
<b>Θυμός</b>	<b>39</b>	2	1	1	43
<b>Χαρά</b>	13	<b>40</b>	3	9	65
<b>Λύπη</b>	2	1	<b>36</b>	4	43
<b>Ουδέτερο</b>	2	6	7	<b>52</b>	67

Ακολουθεί μια εικόνα που επιδεικνύει το πλήθος των ορθά ταξινομημένων αρθρώσεων, ανά Συναίσθημα :



Σχήμα 17: Ορθά ταξινομημένες αρθρώσεις με το βέλτιστο πλήθος στατιστικών κατά SBS + INN

Πίνακας 12: Ποσοστά Αναγνώρισης Συναισθημάτων με το βέλτιστο πλήθος στατιστικών κατά SBS + INN

Συναίσθημα	Θυμός	Χαρά	Λύπη	Ουδέτερο
Θυμός	<b>90,6977%</b>	4,6512%	2,3256%	2,3256%
Χαρά	20,0000%	<b>61,5385%</b>	4,6154%	13,8462%
Λύπη	4,6512%	2,3256%	<b>83,7209%</b>	9,3023%
Ουδέτερο	2,9851%	8,9552%	10,4478%	<b>77,6119%</b>

Ακολουθεί μια εικόνα που επιδεικνύει τα ποσοστά Αναγνώρισης Συναισθημάτων:



Σχήμα 18: Ποσοστά Αναγνώρισης Συναισθημάτων με το βέλτιστο πλήθος στατιστικών κατά SBS + 1NN

Η μέση ακρίβεια Αναγνώρισης Συναισθήματος ήταν **78,3923%**.

#### 4.4 Δεύτερο πείραμα: Σύγκριση των multi-class SVMs: 1-VS-1, 1-VS-Rest για την IEmoCap

##### 4.4.1 Χρησιμοποιημένα στατιστικά για το δεύτερο πείραμα

Για το δεύτερο πείραμα, χρησιμοποίησα ένα κατά πολύ μικρότερο Διάνυσμα Γνωρισμάτων από αυτό του πρώτου πειράματος, και συγκεκριμένα, τα ακόλουθα 19 στατιστικά:

Πίνακας 13: Χρησιμοποιημένα στατιστικά για το δεύτερο πείραμα

Ρυθμός ομιλίας
Λόγος Harmonicity to Noise Ratio (HNR)
Λόγος Zero Crossing Rate (ZCR)
Μέση τιμή της θεμελιώδους συχνότητας
Μέση τιμή της ενέργειας RMS
Μέση τιμή των πρώτων 12 MFCCs
Μέση τιμή των πρώτων 2 Formants F1, F2



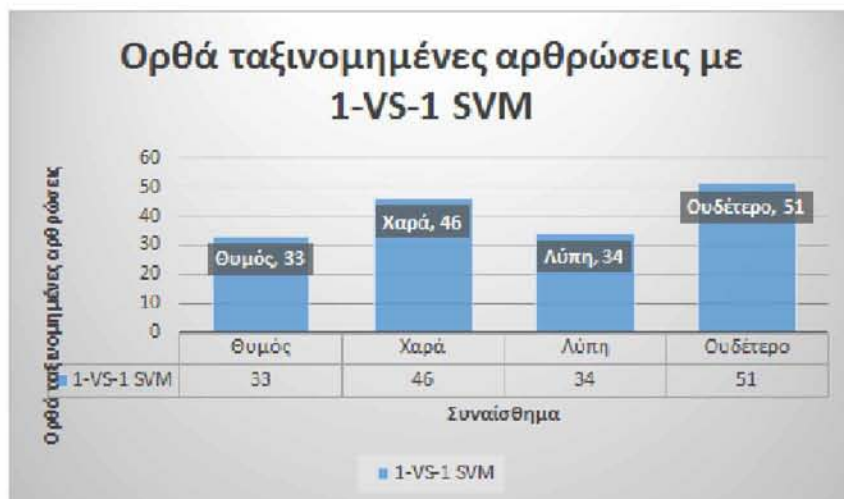
Η ταξινόμηση πραγματοποιήθηκε με συγκεκριμένες έτοιμες, ήδη υλοποιημένες ρουτίνες Matlab που υλοποιούν ταξινόμηση τύπου 1-VS-1, 1-VS-Rest SVM και που βρέθηκαν στο Matlab toolbox: [16]. Αρχικά λοιπόν, το Διάνυσμα Γνωρισμάτων των 19 στατιστικών δόθηκε ως είσοδος στη ρουτίνα svmmulti-classoneagainstone του Matlab toolbox που υλοποιεί ταξινόμηση 1-VS-1 SVM και πάρθηκαν τα αποτελέσματα ταξινόμησης αρθρώσεων ανά Συναισθημα που ακολουθούν:

#### 4.4.2 Αναλυτικά αποτελέσματα ταξινόμησης αρθρώσεων με ταξινομητή 1-VS-1 SVM

Πίνακας 14: Αποτελέσματα ταξινόμησης αρθρώσεων με 1-VS-1 SVM για την IEMoCap

Συναίσθημα	Θυμός	Χαρά	Λύπη	Ουδέτερο	Πλήθος αρθρώσεων
Θυμός	33	5	0	5	43
Χαρά	7	46	2	10	65
Λύπη	0	1	34	8	43
Ουδέτερο	0	5	11	51	67

Ακολουθεί μια εικόνα που επιδεικνύει το πλήθος των ορθά ταξινομημένων αρθρώσεων ανά Συναισθημα:



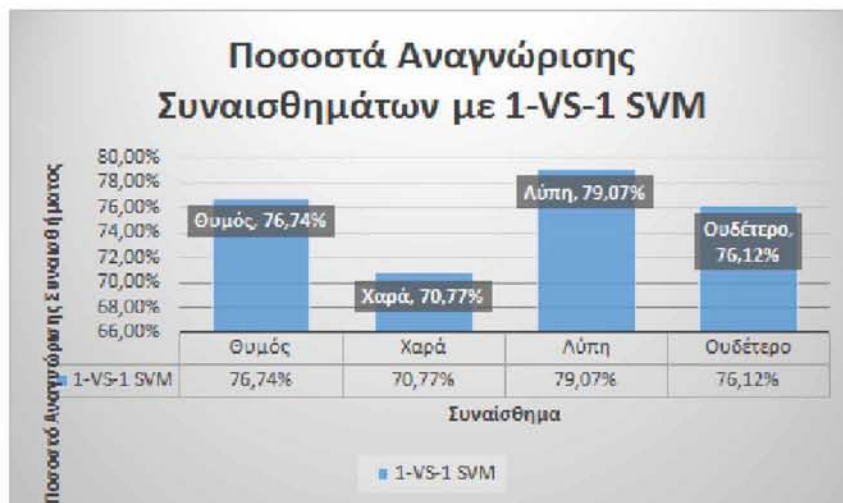
Σχήμα 19: Ορθά ταξινομημένες αρθρώσεις ανά Συναισθημα με 1-VS-1 για την IEMoCap

Ακολουθούν τα προκύπτοντα ποσοστά Αναγνώρισης Συναισθημάτων:

Πίνακας 15: Ποσοστά Αναγνώρισης Συναισθημάτων με 1-VS-1 για την IEmoCap

Συναίσθημα	Θυμός	Χαρά	Λύπη	Ουδέτερο
Θυμός	<b>76,7442%</b>	11,6279%	0%	11,6279%
Χαρά	10,7692%	<b>70,7692%</b>	3,0769%	15,3846%
Λύπη	0%	2,3256%	<b>79,0698%</b>	18,6047%
Ουδέτερο	0%	7,4627%	16,4179%	<b>76,1194%</b>

Ακολουθεί μια εικόνα που επιδεικνύει τα ποσοστά Αναγνώρισης Συναισθημάτων, ανά Συναισθημα :



Σχήμα 20: Ποσοστά Αναγνώρισης Συναισθημάτων με 1-VS-1 για την IEmoCap

Πήραμε μέση ακρίβεια Αναγνώρισης Συναισθήματος **75,6757%**.

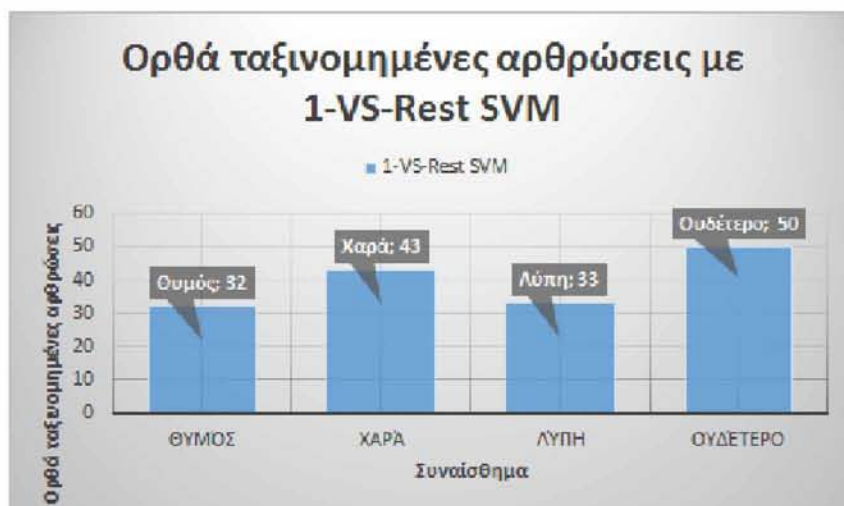
Εναλλακτικά, στη συνέχεια, το Διάνυσμα Γνωρισμάτων των 19 στατιστικών δόθηκε ως είσοδος στη ρουτίνα `svmmulticlassoneagainstall` του Matlab toolbox [16] που υλοποιεί ταξινόμηση 1-VS-Rest SVM και πάρθηκαν τα αποτελέσματα ταξινόμησης αρθρώσεων ανά Συναισθημα που ακολουθούν:

#### 4.4.3 Αποτελέσματα ταξινόμησης με υλοποίηση για 1-VS-Rest SVM

Πίνακας 16: Αποτελέσματα ταξινόμησης αρθρώσεων με 1-VS-Rest SVM για την IEmoCap

Συναίσθημα	Θυμός	Χαρά	Λύπη	Ουδέτερο	Πλήθος αρθρώσεων
Θυμός	<b>32</b>	4	0	7	43
Χαρά	5	<b>43</b>	1	16	65
Λύπη	0	0	<b>33</b>	10	43
Ουδέτερο	0	13	4	<b>50</b>	67

Ακολουθεί μια εικόνα που επιδεικνύει το πλήθος των ορθά ταξινομημένων αρθρώσεων ανά Συναισθημα :



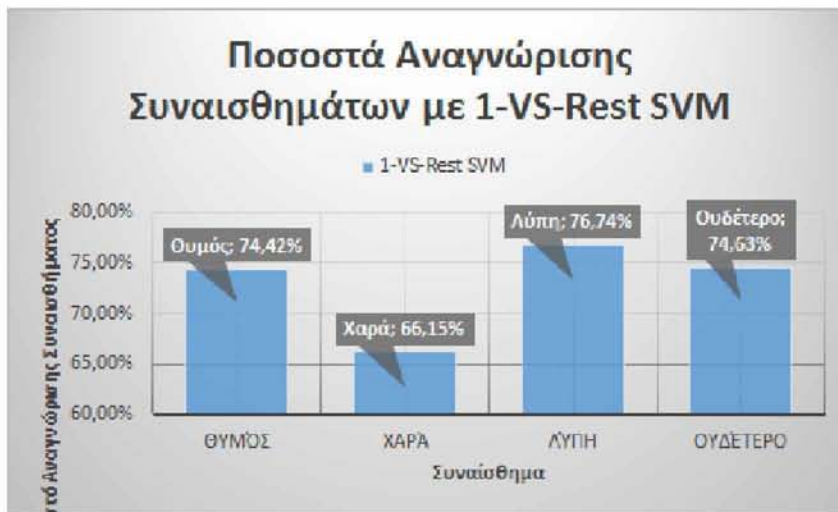
Σχήμα 21: Ορθά ταξινομημένες αρθρώσεις ανά Συναισθημα με 1-VS-Rest SVM για την IEmoCap

Ακολουθούν τα προκύπτοντα ποσοστά Αναγνώρισης Συναισθημάτων:

Πίνακας 17: Ποσοστά Αναγνώρισης Συναισθημάτων με ταξινομητή 1-VS-Rest SVM για την IEmoCap

Συναίσθημα	Θυμός	Χαρά	Λύπη	Ουδέτερο
Θυμός	<b>74,4186%</b>	9,3023%	0%	16,2791%
Χαρά	7,6923%	<b>66,1538%</b>	1,5385%	24,6154%
Λύπη	0%	0%	<b>76,7442%</b>	23,2558%
Ουδέτερο	0%	19,4030%	5,9701%	<b>74,6269%</b>

Ακολουθεί μια εικόνα που επιδεικνύει τα ποσοστά Αναγνώρισης Συναισθημάτων, ανά Συναίσθημα :



Σχήμα 22: Ποσοστά Αναγνώρισης Συναισθημάτων με 1-VS-Rest SVM για την IEmoCap

Πήραμε μέση ακρίβεια Αναγνώρισης Συναισθήματος **72,9859%**.

# Κεφάλαιο 5: Τα συμπεράσματά μας από τα δύο πειράματα

## 5.1 Συμπεράσματα αποτελεσμάτων πρώτου πειράματος στην IEmoCap

**(1) Πήραμε μεγαλύτερη μέση ακρίβεια Αναγνώρισης Συναισθήματος με επιλογή του συνδυασμού SBS + 1NN, από αυτή του SFS + 1NN, της τάξης του 4,2053%.**

Συνεπώς, μπορεί κανείς να υποστηρίξει ότι η SBS φαίνεται να προβάλλει ως καλύτερη επιλογή για μέθοδο Ακολουθιακής Επιλογής Γνωρισμάτων από την SFS, για παράμετρο επιλογής τερματικού κριτηρίου σύγκλισης και των δύο, την προεπιλεγμένη (default) παράμετρο επιλογής ελαχιστοποίησης του σφάλματος ταξινόμησης του ταξινομητή 1NN. Ως εκ τούτου, στην περίπτωση που ο δημιουργός μιας εφαρμογής Αλληλεπίδρασης Ανθρώπου-Μηχανής έχει αποφασίσει ότι θέλει να χρησιμοποιήσει ταξινομητή 1NN για Αναγνώριση Συναισθήματος και πρόκειται να εφαρμόσει μία εκ των δύο ευρύτερα απλών, εύκολων και προσβάσιμων Ακολουθιακών μεθόδων Επιλογής Γνωρισμάτων, συνίσταται να χρησιμοποιήσει SBS, αντί της SFS, είτε ενδιαφέρεται για καλύτερη μέση είτε για οποιουδήποτε από τα 4 εξεταζόμενα Συναισθήματα της IEmoCap (Θυμός, Χαρά, Λύπη, Ουδέτερη Συναισθηματική κατάσταση) την καλύτερη ανίχνευση.

**(2) Ο συνδυασμός SBS + 1NN έδωσε καλύτερη ανίχνευση Συναισθήματος για όλα τα 4 εξεταζόμενα Συναισθήματα της IEmoCap (Θυμός, Χαρά, Λύπη, Ουδέτερη Συναισθηματική κατάσταση).**

**(3) Η SFS συνέκλινε σε 61 επαναλήψεις στο τελικό βέλτιστο υποσύνολο στατιστικών, ενώ η SBS σε 152.**

Συνεπώς, η SFS χρειάστηκε το 30,0493%, ενώ η SBS το 74,8768% μιας πλήρους/εξαντλητικής αναζήτησης (exhaustive search) 203 επαναλήψεων. Καταλαβαίνει κανείς ότι, αν ο χρόνος μιας πλήρους/εξαντλητικής αναζήτησης είναι μεγάλος, ένα ποσοστό της τάξης του 44,8275% που αντιστοιχεί στο ποσοστό της διαφοράς τους, οδηγεί σε μη αμελητέα διαφορά απαιτούμενης χρονικής εκτέλεσης ως τη σύγκλιση/τον τερματισμό τους. Συγκεκριμένα, όσο αυξάνεται το πλήθος των στατιστικών του αρχικού Διανύσματος Γνωρισμάτων, στο οποίο θέλει κανείς να εφαρμόσει τις δύο Ακολουθιακές μεθόδους Επιλογής Γνωρισμάτων, τόσο αυξάνεται και η διαφορά απαιτούμενης χρονικής εκτέλεσης ως τη σύγκλιση/τον τερματισμό τους. Ως εκ τούτου, η SFS φαίνεται να προβάλλει ως πιο γρήγορη από την SBS. Ως εκ τούτου, μπορεί κανείς να υποστηρίξει ότι φαίνεται να αντισταθμίζει την απώλεια σε Αναγνώριση Συναισθήματος με τη γρηγορότερη εκτέλεσή της, σε σχέση με την SBS. Συνεπώς, προκύπτει ένα ζήτημα συμβιβαστικής απόφασης (trade-off) μεταξύ απαιτούμενου χρόνου εκτέλεσης και ακρίβειας Αναγνώρισης που ο δημιουργός μιας αλληλεπίδρασης Ανθρώπου-Μηχανής για Αναγνώριση Συναισθήματος καλείται να πάρει.

**(4) Στον αντίποδα, μια διαφορετική ανάγνωση της διαπίστωσης (2) είναι ότι η SBS οδηγεί, όταν συγκλίνει και τερματίζει, σε μικρότερο απαιτούμενο αριθμό στατιστικών προς αποθήκευση στο Διάνυσμα Γνωρισμάτων.**

Συνεπώς, η SBS μπορεί να υστερεί ως προς την πρακτική, τελική χρονική της πολυπλοκότητα ως προς την SFS, εντούτοις φαίνεται να υπερέχει ως προς τη μειωμένη χωρική πολυπλοκότητα που οδηγεί το Διάνυσμα Γνωρισμάτων μετά τη σύγκλιση/ τον τερματισμό της.

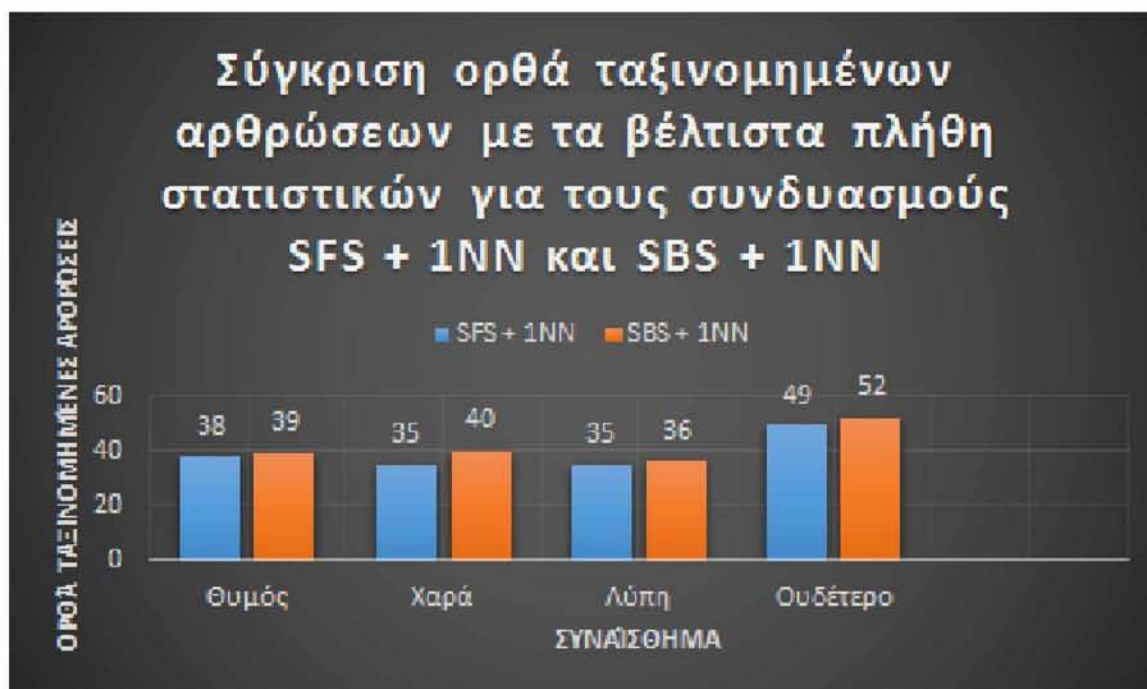
**(5) Ως πιο σημαντικά, από τις αρκετές διαφορετικές κατηγορίες γνωρισμάτων, επιλέχθηκαν κάποια στατιστικά σχετικά με τα πρώτα 2 Formants F1, F2 και με τις δύο Ακολουθιακές μεθόδους Επιλογής Γνωρισμάτων.**

Σε πρώτη εντύπωση, αυτό μας προκάλεσε κάποιο ξαφνιασμό, επειδή δεν το περιμέναμε. Από την άλλη όμως, μπορεί να κατανοηθεί, επειδή πρόκειται για τα 2 πρώτα Formants, που έχει ήδη αποδειχτεί ότι, κυρίως αυτά, και όχι τόσο τα υπόλοιπα, σχετίζονται με τη Συναισθηματική πληροφορία.

**(6) Το καλύτερα/ευκολότερα ανιχνευμένο Συναίσθημα ήταν ο Θυμός, τόσο με χρήση του συνδυασμού SFS + 1NN όσο και του SBS + 1NN.**

**(7) Το χειρότερα/δυσκολότερα ανιχνευμένο Συναίσθημα ήταν η Χαρά, τόσο με χρήση του συνδυασμού SFS + 1NN όσο και του SBS + 1NN.**

Ακολουθούν δύο συγκριτικά ραβδογράμματα που επιδεικνύουν το πλήθος των ορθά ταξινομημένων αρθρώσεων και των ποσοστών ορθής Αναγνώρισης ανά Συναίσθημα, τόσο σε περίπτωση επιλογής συνδυασμού SFS + 1NN όσο και SBS + 1NN:



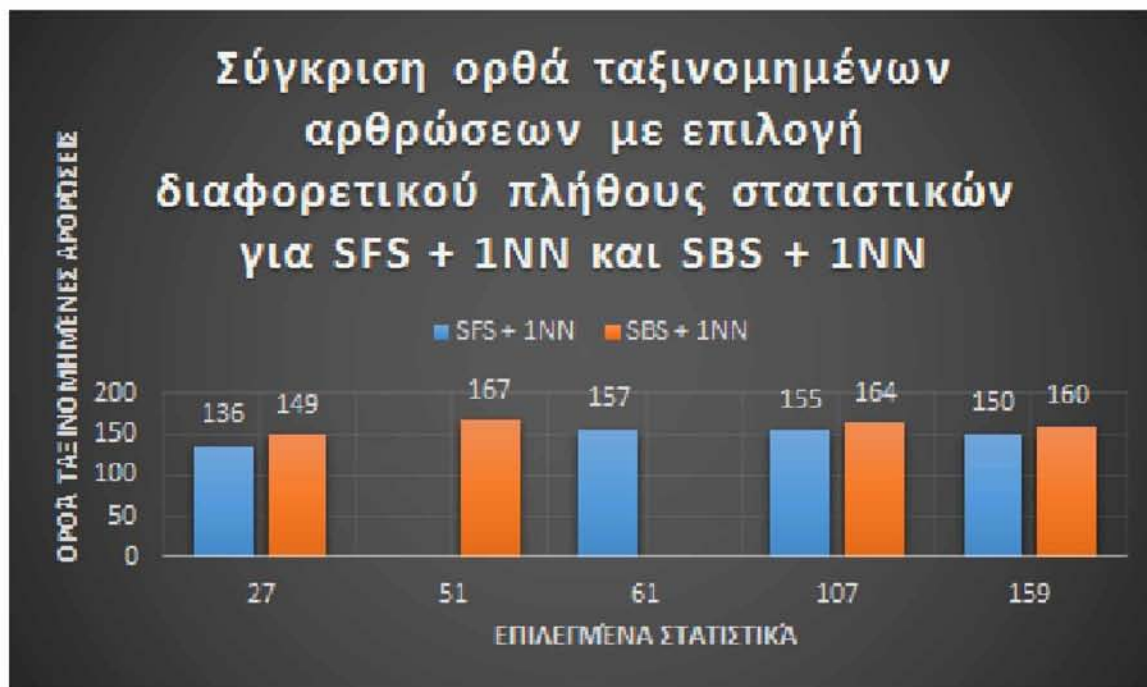
Σχήμα 23: Σύγκριση ορθά ταξινομημένων αρθρώσεων με SFS + 1NN, SBS + 1NN στην IEmoCap





Σχήμα 24: Σύγκριση ποσοστών Αναγνώρισης Συναισθημάτων με SFS + 1NN, SBS + 1NN στην IEmoCap

Τέλος, ακολουθούν δύο συγκριτικά ραβδογράμματα που επιδεικνύουν την επίδραση της επιλογής διαφορετικού πλήθους πιο σημαντικών στατιστικών στην ταξινόμηση αρθρώσεων και στα ποσοστά μέσης ακρίβειας Αναγνώρισης, για τους συνδυασμούς SFS + 1NN, SBS + 1NN:



Σχήμα 25: Σύγκριση ορθά ταξινομημένων αρθρώσεων με επιλογή διαφορετικού πλήθους στατιστικών για τους συνδυασμούς SFS + 1NN, SBS + 1NN στην IEmoCap



Σχήμα 26: Σύγκριση ποσοστών μέσης ακρίβειας Αναγνώρισης Συναισθήματος με επιλογή διαφορετικού πλήθους στατιστικών για τους συνδυασμούς SFS + 1NN, SBS + 1NN στην IEmoCap

## 5.2 Συμπεράσματα αποτελεσμάτων δεύτερου πειράματος στην IEmoCap

**(1) Πήραμε μεγαλύτερη μέση ακρίβεια Αναγνώρισης Συναισθήματος με χρήση της 1-VS-1 SVM, από αυτή της 1-VS-Rest SVM, κατά ένα ποσοστό της τάξης του 2,6898%.**

**(2) Η 1-VS-1 SVM οδήγησε σε καλύτερη ανίχνευση για όλα τα 4 Συναισθήματα της IEmoCap (Θυμός, Χαρά, Λύπη, Ουδέτερη Συναισθηματική κατάσταση), από την 1-VS-Rest SVM.**

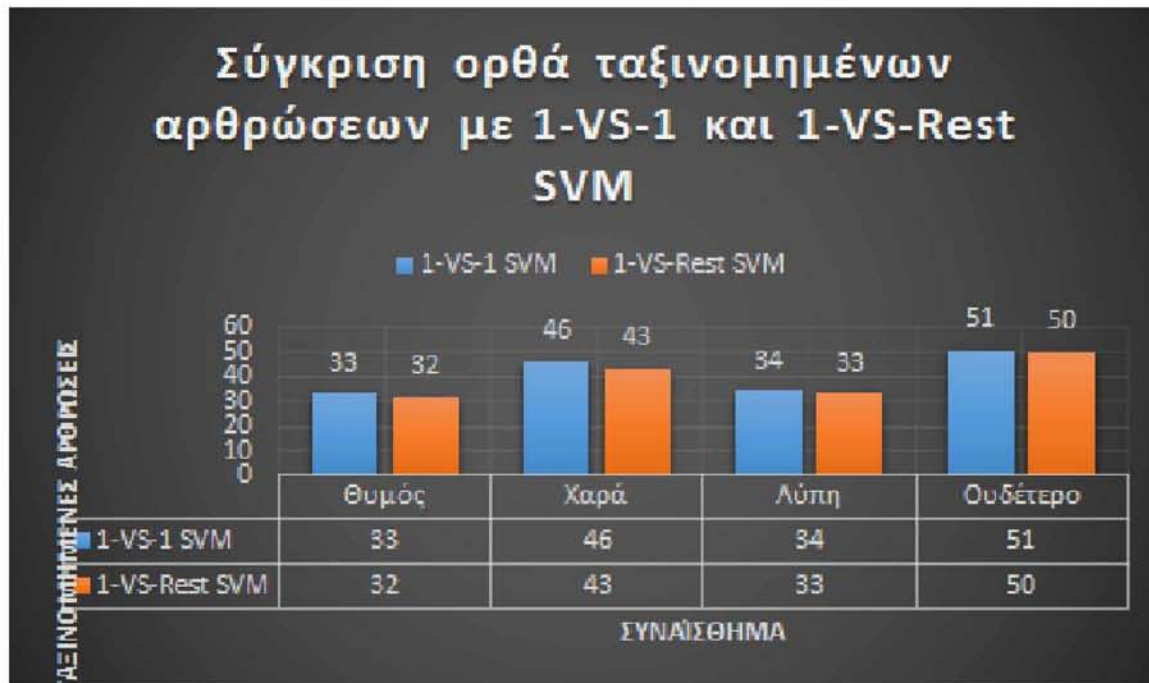
Συνεπώς, λόγω των διαπιστώσεων (1) και (2), μπορεί κανείς να υποστηρίξει ότι η 1-VS-1 SVM προβάλλει ως συγκριτικά καλύτερη επιλογή από την 1-VS-Rest SVM. Συνεπώς, σε περίπτωση που ο δημιουργός μιας εφαρμογής Αλληλεπίδρασης Ανθρώπου-Μηχανής για Αναγνώριση Συναισθήματος από το Σήμα του Λόγου έχει αποφασίσει να χρησιμοποιήσει μια εκ των δύο ταξινομήσεων τύπου multiclass SVMs, του συνιστάται να χρησιμοποιήσει την 1-VS-1 SVM, αντί της 1-VS-Rest SVM, είτε ενδιαφέρεται για την καλύ-

τερη μέση είτε για οποιοδήποτε εκ των 4 συγκεκριμένων Συναισθημάτων (Θυμός, Χαρά, Λύπη, Ουδέτερη Συναισθηματική κατάσταση) την καλύτερη ανίχνευση.

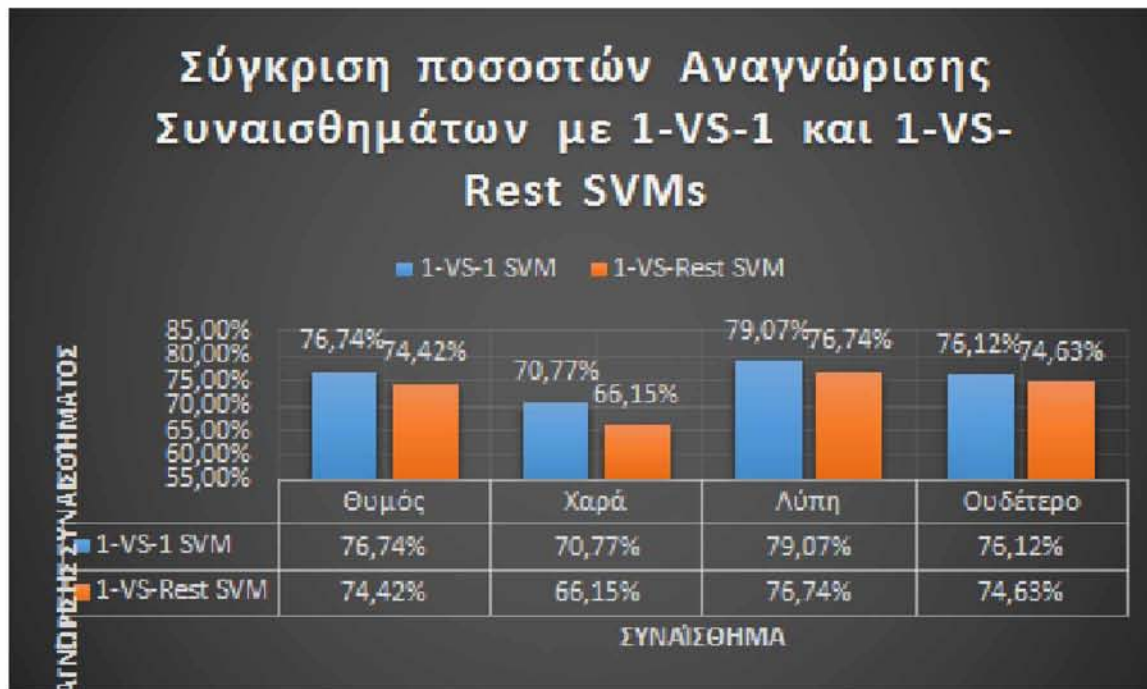
**(3) Το καλύτερα/ευκολότερα ανιχνευμένο συναίσθημα ήταν η Λύπη, τόσο με χρήση ταξινόμησης 1-VS-1 όσο και 1-VS-Rest SVM.**

**(4) Το χειρότερα/δυσκολότερα ανιχνευμένο συναίσθημα ήταν η Χαρά, τόσο με χρήση ταξινόμησης 1-VS-1 όσο και 1-VS-Rest SVM.**

Τέλος, ακολουθούν δύο εικόνες που επιδεικνύουν τη σύγκριση του πλήθους των ορθά ταξινομημένων αρθρώσεων και των ποσοστών ορθής Αναγνώρισης ανά Συναίσθημα, για τους 1-VS-1, 1-VS-Rest SVMs:



Σχήμα 27: Σύγκριση ορθά ταξινομημένων αρθρώσεων με 1-VS-1, 1-VS-Rest SVMs στην IEmoCap



Σχήμα 28: Σύγκριση ποσοστών Αναγνώρισης Συναισθημάτων με 1-VS-1, 1-VS-Rest στην IEmoCap

Συνοψίζοντας και για τα δύο πειράματά μας μαζί, παρατηρούμε ότι, παρά το γεγονός ότι οι ταξινομητές τύπου SVM, όπως αυτές στο δεύτερο πείραμα, αναμένονταν να οδηγήσουν σε μεγαλύτερη μέση ακρίβεια Αναγνώρισης Συναισθήματος, σύμφωνα με ό,τι γνωρίζουμε από τη θεωρία της Αναγνώρισης Προτύπων, εντούτοις, τα χρησιμοποιημένα μεγαλύτερα Διανύσματα Γνωρισμάτων του πρώτου πειράματος, που προκύπτουν με τις δύο Επιλογές Γνωρισμάτων, υπερνικούν το μικρότερο Διάνυσμα Γνωρισμάτων του δεύτερου πειράματος, παρά το γεγονός ότι ο ταξινομητής kNN οδηγεί τις περισσότερες φορές σε μικρότερη μέση ακρίβεια Αναγνώρισης Συναισθήματος από ταξινομητές τύπου SVM. Παρατηρούμε δηλαδή ότι με προσεκτική καλύτερη Επιλογή Γνωρισμάτων μπορούμε να υπερκαλύψουμε πλήρως την αναμενόμενη μειωμένη ικανότητα ταξινόμησης ενός ταξινομητή, συγκριτικά με έναν άλλο.

Κλείνοντας, πρέπει να σημειωθεί ότι όλα τα προηγούμενα συμπεράσματα και των δύο πειραμάτων εζηγήθηκαν πάνω στη συγκεκριμένη χρησιμοποιημένη βάση δεδομένων Συναισθημάτων IEmoCap και με την εξαγωγή συγκεκριμένων στατιστικών από συγκεκριμένα γνωρισμάτα από κάθε άρθρωση. Συνεπώς, απαιτούνται μελλοντικά νέα σχετικά πειράματα πάνω σε διαφορετικές βάσεις δεδομένων Συναισθημάτων.

## Μελλοντικές ερευνητικές επεκτάσεις

Καταρχάς, καθίσταται απαραίτητη η διενέργεια αντίστοιχων πειραμάτων σε όσο το δυνατόν περισσότερες διαφορετικού είδους βάσεις δεδομένων Συναισθημάτων για μεγαλύτερη αξιοπιστία των αποτελεσμάτων. Τόσο το πλήθος τους, στο οποίο έχει δοκιμαστεί ένα Σύστημα Αυτόματης Αναγνώρισης Συναισθήματος όσο και το πλήθος των χρησιμοποιημένων αρθρώσεων αποτελούν καθοριστικούς παράγοντες της δυναμικής των αποτελεσμάτων. Επιπλέον, καθίσταται πιο προκλητική και ενδιαφέρουσα η δοκιμή πειραμάτων σε βάσεις δεδομένων που χαρακτηρίζονται από φυσικό, αυθόρμητο και όχι επαγγελματικά προσοιημένο Συναισθηματικό Λόγο, όπως στις αρθρώσεις της IEmoCar, που χρησιμοποιήθηκαν στα δικά μας πειράματα. Κλείνοντας, ενδεχόμενες μελλοντικές ερευνητικές επεκτάσεις περιλαμβάνουν την εξέταση επιπρόσθετων τροπικοτήτων πληροφορίας, όπως τα συμφραζόμενα από το Λόγο και την ανίχνευση συγκεκριμένων λέξεων "Συναισθηματικού" περιεχομένου στις αρθρώσεις. Και αυτό, επειδή υπάρχει δυνατότητα βελτίωσης της μέσης ακρίβειας Αναγνώρισης, μέσω της πολυτροπικής σύντηξης (multimodal fusion), όπως έχει αποδειχτεί. Θεμιτό θα ήταν να συμπεριλάβουμε και τη λεκτική πληροφορία ή την πληροφορία για το αντικείμενο του εξελισσόμενου διαλόγου μεταξύ των ομιλητών κατά τη διάρκεια της αλληλεπίδρασης.

# References

- [1] <http://www.computer.org/csdl/trans/ta/2012/02/tta2012020237-abs.html>.
- [2] <http://link.springer.com/article/10.1007%2Fs10579-007-9057-1/fulltext.html>.
- [3] <http://iupuic180online.wikispaces.com/Emotion>.
- [4] [https://instruct1.cit.cornell.edu/courses/ece576/FinalProjects/f2008/pae26\\_jsc59/pae26\\_jsc59/](https://instruct1.cit.cornell.edu/courses/ece576/FinalProjects/f2008/pae26_jsc59/pae26_jsc59/).
- [5] <http://research.microsoft.com/en-us/um/people/manik/projects/trade-off/svm.html>.
- [6] [http://www.byclb.com/TR/Tutorials/neural\\_networks/ch11\\_1.htm](http://www.byclb.com/TR/Tutorials/neural_networks/ch11_1.htm).
- [7] [www.clear.rice.edu/elec532/PROJECTS96/lpc/code/vunv.m](http://www.clear.rice.edu/elec532/PROJECTS96/lpc/code/vunv.m).
- [8] [note.sonots.com/SciSoftware/Pitch.html](http://note.sonots.com/SciSoftware/Pitch.html).
- [9] [labrosa.ee.columbia.edu/matlab/rastamat/melfcc.m](http://labrosa.ee.columbia.edu/matlab/rastamat/melfcc.m).
- [10] [labrosa.ee.columbia.edu/matlab/rastamat/rastaplp.m](http://labrosa.ee.columbia.edu/matlab/rastamat/rastaplp.m).
- [11] [www.lintech.org/speechbook/Chapter6/zcr.m](http://www.lintech.org/speechbook/Chapter6/zcr.m).
- [12] [replaygain.hydrogenaudio.org/proposal/mfiles/replaygain.m](http://replaygain.hydrogenaudio.org/proposal/mfiles/replaygain.m).

- [13] [code.soundsoftware.ac.uk/embedded/emotion-detection-top-level/Code/Descriptors/Matlab/MPEG7/FromWeb/VoiceSauce/func\\_GetHNR.html](http://code.soundsoftware.ac.uk/embedded/emotion-detection-top-level/Code/Descriptors/Matlab/MPEG7/FromWeb/VoiceSauce/func_GetHNR.html).
- [14] [sonic-gesture.googlecode.com/svn/trunk/evaluate/part1/prtools/featselb.m](http://sonic-gesture.googlecode.com/svn/trunk/evaluate/part1/prtools/featselb.m).
- [15] [sonic-gesture.googlecode.com/svn/trunk/evaluate/part1/prtools/featselb.m](http://sonic-gesture.googlecode.com/svn/trunk/evaluate/part1/prtools/featselb.m).
- [16] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy, "Svm and kernel methods matlab toolbox." Perception Systemes et Information, INSA de Rouen, Rouen, France, 2005.
- [17] <http://jpl.inha.ac.kr/health%20monitoring.htm>.