



www.uom.gr

ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ
ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ



ΤΜΗΜΑ ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΥΣΤΗΜΑΤΑ ΥΠΟΛΟΓΙΣΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Με τίτλο

**Αναγνώριση Συναισθημάτων για τη Βελτιστοποίηση της
Επικοινωνίας Ανθρώπου – Υπολογιστή**



Όνοματεπώνυμο φοιτητή: Ηλίου Θεόδωρος

Επιβλέπων καθηγητής: Μαυρίδης Ιωάννης

Θεσσαλονίκη, Σεπτέμβριος 2011

(KENH)



ΤΜΗΜΑ ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

με τίτλο

**Αναγνώριση Συναισθημάτων για τη Βελτιστοποίηση της
Επικοινωνίας Ανθρώπου – Υπολογιστή
του
ΗΛΙΟΥ ΘΕΟΔΩΡΟΥ**

Επιβλέπων: Μαυρίδης Ιωάννης
Επίκουρος Καθηγητής Πανεπιστημίου Μακεδονίας

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 11 Σεπτεμβρίου 2011

.....

Μαυρίδης Ιωάννης

Επίκουρος

.....

Γεωργιάδης Χρήστος

Επίκουρος

.....

Στειακάκης Εμμανουήλ

Λέκτορας

Θεσσαλονίκη, Σεπτέμβριος 2011

ΠΕΡΙΛΗΨΗ

Η επικοινωνία είναι μια πολύ σημαντική ικανότητα, που βασίζεται όχι μόνο στο γλωσσολογικό μέρος αλλά και στο συναισθηματικό. Στο πεδίο της Αλληλεπίδρασης Ανθρώπου Υπολογιστή η αναγνώριση συναισθημάτων είναι μια πρόκληση, και γι αυτό πολλές έρευνες γίνονται πάνω στο πεδίο αυτό.

Ο πιο εκφραστικός τρόπος με τον οποίο οι άνθρωποι εκδηλώνουν τα συναισθήματά τους είναι τα χαρακτηριστικά της φωνής και οι εκφράσεις του προσώπου. Πρόσφατα, η πληροφορία που παρέχεται από κάμερες και μικρόφωνα καθώς και οι προχωρημένες τεχνικές επεξεργασίας εικόνας και ήχου, κάνουν τον υπολογιστή να «βλέπει» και να «ακούει».

Η εργασία αυτή επικεντρώνεται στις μεθόδους με τις οποίες προσπαθούμε να κάνουμε ένα υπολογιστικό σύστημα να αναγνωρίζει το συναίσθημα του χρήστη με τον οποίο αλληλεπιδρά.

Η πρώτη μέθοδος είναι αυτή της αναγνώρισης των εκφράσεων του προσώπου, η δεύτερη είναι μέσω της φωνής, η τρίτη είναι με την επεξεργασία κειμένου - διαλόγων ενώ η τέταρτη αποτελεί συνδυαστική μέθοδο των παραπάνω. Στο κυρίως μέρος της εργασίας παρουσιάζεται κάθε τεχνική αναλυτικά καθώς και διάφορα μοντέλα αναγνώρισης συναισθημάτων και γίνεται αναφορά στη σχετική βιβλιογραφία. Οι περισσότερες μέχρι τώρα έρευνες γίνονται σε βάσεις δεδομένων ηχογραφημένες από ηθοποιούς γιατί όπως καταλαβαίνουμε είναι δύσκολο να συλλεχθούν δεδομένα από τυχαίους ομιλητές.

Σύμφωνα με τη βιβλιογραφία, τα αποτελέσματα υποδεικνύουν ότι ένα σύστημα αναγνωρίζει τη συναισθηματική κατάσταση του χρήστη σε ένα αρκετά μεγάλο ποσοστό, παρόλα αυτά όμως πολλά προβλήματα όπως η διαφορετική έκφραση συναισθήματος από άτομα με διαφορετική κουλτούρα καθώς και η έλλειψη ερευνών σε πραγματικά δεδομένα είναι μερικά από τα ζητήματα που πρέπει να εξεταστούν.

Λέξεις κλειδιά: *emotion recognition, neural networks, speech processing.*

ΕΥΧΑΡΙΣΤΙΕΣ

Ο δρόμος για την «απόκτηση γνώσης» είναι δύσκολος και αποτελείται από πολλά στάδια. Καθένα από αυτά τα βήματα απαιτεί δύναμη, εμπιστοσύνη, θέληση μα πάνω από όλα πάθος προκειμένου να αποτελέσει βάση για τη μετάβαση στο επόμενο στάδιο. Τίποτα ωστόσο δεν θα είχε το ίδιο αποτέλεσμα εάν δεν είχα τη βοήθεια και την υποστήριξη των ανθρώπων που ήταν δίπλα μου κατά τη διάρκεια αυτού του δύσκολου αλλά άκρως εποικοδομητικού ταξιδιού.

Πρώτα από όλα, θα ήθελα να ευχαριστήσω τον επιβλέπων Καθηγητή μου, Μαυρίδη Ιωάννη, όχι μόνο για την υπομονή και επιμονή του, αλλά και για τις γνώσεις και την ηθική στήριξη που απλόχερα μου προσέφερε. Η βοήθειά του υπήρξε πολύτιμη για μένα.

Δεύτερον, δεν θα μπορούσα να παραλείψω το Διδακτικό Προσωπικό του Μεταπτυχιακού. Ο καθένας τους, με τη δική του εμπειρία και προσωπικότητα, συνέβαλε στην πραγματοποίηση αυτής της εργασίας.

Επιπλέον, θα ήθελα να ευχαριστήσω τους συναδέλφους μου, οι οποίοι μοιράστηκαν μαζί μου αυτό το ταξίδι. Οι κοινοί στόχοι και οι κοινές ανησυχίες μας οδήγησαν στην δημιουργία δυνατών σχέσεων, μέσω των οποίων ανταλλάξαμε δύναμη και θάρρος πολλές φορές κατά τη διάρκεια του περασμένου έτους.

Τελευταία αλλά όχι αμελητέα, θα θέλαμε να ευχαριστήσω την οικογένειά μου που με στήριξε στο δύσκολο αυτό έργο μου και σε όλη τη διάρκεια των μέχρι τώρα σπουδών μου.

Περιεχόμενα

ΠΕΡΙΛΗΨΗ	iv
ΕΥΧΑΡΙΣΤΙΕΣ	v
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ	vii
1. ΕΙΣΑΓΩΓΗ	8
1. ΕΠΙΚΟΙΝΩΝΙΑ ΑΝΘΡΩΠΟΥ-ΥΠΟΛΟΓΙΣΤΗ	9
1.1 ΕΙΣΑΓΩΓΗ.....	9
1.2 ΟΡΟΛΟΓΙΑ.....	9
1.3 ΣΤΟΧΟΙ.....	9
1.4 ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ.....	10
1.5 ΔΙΕΠΙΦΑΝΕΙΕΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ.....	11
1.5.1 Οι διεπιφάνειες περιέχουν κάποιο μήνυμα για τον χρήστη.....	11
1.5.2 Με τι πρέπει να επικοινωνούν οι υπολογιστές.....	12
1.5.3 Επικοινωνία Ανθρώπου Υπολογιστή, Κίνητρα για την Αναγνώριση Συναισθημάτων.....	14
2 ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ	14
2.1 Συναισθήματα στην διάδραση ανθρώπου-υπολογιστή.....	14
2.1.2 Τι είναι τα συναισθήματα.....	16
2.1.3 Πολυλειτουργικότητα δεικτών και πολυεπίπεδος χαρακτηρισμός συναισθημάτων.....	18
2.1.4 Ανάλυση Συναισθημάτων.....	19
2.1.5 Συναισθήματα, διαθέσεις και γνωστικές λειτουργίες.....	20
2.2 ΤΙ ΜΕΛΕΤΑΜΕ ΣΤΗΝ ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ.....	21
2.2.1 Συναισθηματικά δείγματα-οι ηθοποιοί.....	21
2.2.2 Πραγματική ζωή-ανθρώπινα όντα.....	21
2.3 ΠΡΟΒΛΗΜΑΤΑ ΣΤΗΝ ΑΝΑΛΥΣΗ ΤΗΣ ΓΛΩΣΣΑΣ ΤΩΝ ΣΥΝΑΙΣΘΗΜΑΤΩΝ.....	22
2.3.1 Μεθοδολογικά προβλήματα στην ανάλυση της συναισθηματικής γλώσσας.....	22
2.3.2 Ψέμα και συναισθήματα.....	22
2.3.3 Χειρισμός λανθασμένων συναισθηματικών εκφράσεων από τον χρήστη.....	24
2.4 ΠΑΓΚΟΣΜΙΟΤΗΤΑ ΚΑΙ ΠΟΛΙΤΙΣΜΙΚΗ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΗΝ ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ.....	26
2.4.1 Διαφορές στην έκφραση μεταξύ των συναισθημάτων.....	27
3. ΠΡΟΣΕΓΓΙΣΕΙΣ ΑΝΑΓΝΩΡΙΣΗΣ ΣΥΝΑΙΣΘΗΜΑΤΩΝ	28
3.1 ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΑΠΟ ΚΕΙΜΕΝΟ.....	28
3.1.2 Επιλογή Χαρακτηριστικών και Μοντέλα Αναπαράστασης Εγγράφων.....	29
3.1.3 TF-IDF score.....	29
3.1.4 Bag of words (BoW).....	30
3.1.5 Προσέγγιση Επιλογής Χαρακτηριστικών – POS Vs WordNet Database.....	31
3.2 Εναλλακτικές τεχνικές υπολογισμού γλωσσικών μοντέλων.....	33
3.2.1 Μοντέλα Παράλειψης (Skipping Models).....	33
3.2.2 Μοντέλα Ομαδοποίησης (Clustering Models).....	33
3.2.3 Μοντέλα Μνήμης (Caching Models).....	33
3.2.4 Λέξεις πρόκλησης (Word triggers).....	34

3.2.5 Μοντέλα μίξης προτάσεων (Sentence mixture models)	34
3.2.6 Προσαρμογή γλωσσικού μοντέλου	34
3.3 Τεχνικές στατιστικής μοντελοποίησης της γλώσσας με την χρήση γραμματικής πληροφορίας	35
3.3.1 Μοντέλα βασισμένα σε μέρη του λόγου (POS).....	35
3.3.2 Υπολογισμός πιθανοτήτων απλών N-grams μοντέλων	36
3.3.3 N-Grams.....	37
3.3.4 Συντακτική δομή.....	38
3.4 ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΜΕΣΩ ΤΟΥ ΠΡΟΣΩΠΟΥ	38
3.4.1 Περιγραφικές διαδικασίες για συναισθηματικές εκφράσεις μέσω του προσώπου..	39
3.4.2 Στόχοι και εκφράσεις που συσχετίζονται με την έκφραση συναισθημάτων	39
3.4.3 Πρόσωπα και φυσιολογία συναισθημάτων.....	40
3.4.4 Προσεγγίσεις συναισθηματικής αναγνώρισης μέσω προσώπου	41
3.5 ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΜΕΣΩ ΤΗΣ ΦΩΝΗΣ.....	42
3.6 ΣΥΝΔΥΑΣΤΙΚΗ ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ	44
4. ΤΑΞΙΝΟΜΗΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΚΑΙ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ	45
4.1 Η ΤΑΞΙΝΟΜΗΣΗ ΤΩΝ ΣΥΝΑΙΣΘΗΜΑΤΩΝ	45
4.2 ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΟΜΙΛΙΑΣ.....	48
4.2.1 Βάσεις δεδομένων συναισθηματικού λόγου, στα Αγγλικά	48
4.2.2 Βάσεις δεδομένων στα Γερμανικά.....	50
4.2.3 Βάσεις δεδομένων στα Ιαπωνικά.....	51
4.2.4 Βάσεις δεδομένων στα Ολλανδικά	52
4.2.5 Συναισθηματικές βάσεις στα Ισπανικά.....	53
4.2.6 Συναισθηματικές βάσεις στα Δανέζικα	53
4.2.7 Βάσεις συναισθηματικού λόγου στα Εβραϊκά.....	53
4.2.8 Βάσεις δεδομένων στα Σουηδικά	53
4.2.9 Βάσεις δεδομένων στα Κινέζικα.....	53
4.2.10 Βάσεις δεδομένων στα Ρωσικά.....	54
4.2.11 Πολυγλωσσική βάση δεδομένων συναισθηματικού λόγου.....	54
5. ΧΡΟΝΙΚΗ ΑΝΑΔΡΟΜΗ ΣΤΑ ΣΥΣΤΗΜΑΤΑ ΑΝΑΓΝΩΡΙΣΗΣ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΚΑΙ ΣΧΕΤΙΚΕΣ ΜΕΛΕΤΕΣ	54
5.1 Σχετικές Έρευνες	54
5.1.1 Με Προσωδικά χαρακτηριστικά.....	54
5.1.2 Γλωσσολογική- Λεκτική Πληροφορία	70
ΒΙΒΛΙΟΓΡΑΦΙΑ	75

Κατάλογος Εικόνων

Εικόνα 1: Εύρος συχνότητας νοηματικού σήματος	44
Εικόνα 2: Τα συναισθήματα σύμφωνα με το σθένος και τη διέγερση	46
Εικόνα 3: Τυπική διαδικασία ER με χρήση λεκτικών χαρακτηριστικών	71

«Αυτό που ενεργοποιεί ανθρώπους δεν είναι γεγονότα και πληροφορίες, αλλά συναισθήματα, ιστορίες και πάνω απ' όλα άλλοι άνθρωποι...»

Manfred Spitzer 2002

1. ΕΙΣΑΓΩΓΗ

Η ανάπτυξη της τεχνολογίας τα τελευταία χρόνια, έχει κάνει μέρος της ζωής μας την χρήση των υπολογιστών. Η αναζήτηση της φυσικής επικοινωνίας μεταξύ ανθρώπου-υπολογιστή είναι μια ανάγκη που απορρέει από αυτή την καθημερινή σχέση. Ο χώρος των συναισθημάτων είναι εν μέρει ανεξερεύνητος από τους επιστήμονες, οι οποίοι προσπαθούν να ερμηνεύσουν, να διαχωρίσουν και να καθορίσουν τις συναισθηματικές καταστάσεις του ανθρώπου. Τα τελευταία χρόνια γίνεται μια προσπάθεια για την σύζευξη της τεχνολογίας με την συναισθηματική κατάσταση του ατόμου. Η αναγνώριση συναισθημάτων μέσω υπολογιστή, αναφέρεται στην προσπάθεια ενός συστήματος να αναγνωρίσει την συναισθηματική κατάσταση ενός χρήστη που έρχεται σε επαφή με αυτόν. Αυτό, επιτυγχάνεται με διαφορετικές μεθόδους. Η πρώτη είναι αυτή της αναγνώρισης των εκφράσεων του προσώπου, η δεύτερη είναι μέσω της φωνής, η τρίτη είναι με την επεξεργασία κειμένου-διαλόγων ενώ η τέταρτη αποτελεί συνδυαστική μέθοδο των παραπάνω.

Η φωνή είναι ένας από τους βασικούς μηχανισμούς επικοινωνίας που χρησιμοποιεί ο άνθρωπος για να επικοινωνήσει και να εκφράσει συναισθήματα. Το πεδίο της αναγνώρισης συναισθημάτων μέσω της φωνής είναι ένα σχετικά νέο πεδίο για την επιστημονική έρευνα. Η μέτρηση παραμέτρων όπως η ενέργεια και η θεμελιώδης συχνότητα είναι μια μέθοδος που ακολουθείται στην καταγραφή των χαρακτηριστικών του σήματος ομιλίας έτσι ώστε να ταξινομηθεί και να ακολουθήσει η μοντελοποίηση.

Για την μελέτη των σημάτων ομιλίας είναι αναγκαία η καταγραφή τους. Επειδή αυτό στην καθημερινή ζωή είναι δύσκολο να πραγματοποιηθεί, έχουν δημιουργηθεί κάποιες βάσεις δεδομένων προσποιούμενου λόγου (acted speech) από ηθοποιούς ώστε να διεξαχθεί έρευνα στην αναγνώριση συναισθημάτων από την ομιλία.

1. ΕΠΙΚΟΙΝΩΝΙΑ ΑΝΘΡΩΠΟΥ-ΥΠΟΛΟΓΙΣΤΗ

1.1 ΕΙΣΑΓΩΓΗ

Είναι γεγονός ότι οι υπολογιστές αποτελούν αναπόσπαστο κομμάτι της καθημερινής μας ζωής. Έτσι, έχει δημιουργηθεί η ανάγκη για την αποτύπωση ενός πιο φυσικού περιβάλλοντος επικοινωνίας μεταξύ ανθρώπων και υπολογιστών. Για την επίτευξη αυτού του στόχου, ένας υπολογιστής θα πρέπει να είναι σε θέση να αντιλαμβάνεται την εκάστοτε κατάσταση και να ανταποκρίνεται με διαφορετικούς τρόπους ανάλογα με αυτό που αντιλαμβάνεται κάθε φορά [1]. Για να γίνει η επικοινωνία ανθρώπου υπολογιστή φυσική και φιλική, θα ήταν χρήσιμο να δώσουμε στους υπολογιστές την ικανότητα να αναγνωρίζουν καταστάσεις με τον ίδιο τρόπο που τις αναγνωρίζουν και οι άνθρωποι. Στα σημερινά υπολογιστικά συστήματα που αφορούν την διάδραση ανθρώπου-υπολογιστή, οι μηχανές μπορούν να αναγνωρίσουν τον ομιλητή καθώς και το περιεχόμενο της ομιλίας, με την χρήση τεχνικών που αφορούν την αναγνώριση της ταυτότητας του ομιλητή και του λόγου. Εάν οι μηχανές εφοδιαστούν με τεχνικές αναγνώρισης συναισθημάτων, θα είναι σε θέση να γνωρίζουν τον τρόπο έκφρασης του λόγου – το «πώς λέγεται»- και να αντιδρούν κατάλληλα, κάνοντας την επικοινωνία ανθρώπου-μηχανής να μοιάζει πιο ομαλή και φυσική.

1.2 ΟΡΟΛΟΓΙΑ

Η επικοινωνία ανθρώπου-υπολογιστή, είναι μια μελέτη μεταξύ των ανθρώπων-χρηστών και των υπολογιστών. Επίσης θεωρείται ως η σύζευξη της επιστήμης των υπολογιστών με την επιστήμη της Ψυχολογίας που μελετά τις συμπεριφορές. Η αλληλεπίδραση χρηστών-υπολογιστών διαδραματίζεται στην επιφάνεια εργασίας του χρήστη, η οποία αποτυπώνεται από την συνεργασία του υλικού εξοπλισμού και του λογισμικού [2]. Πιο περιεκτικά, θα λέγαμε ότι η επικοινωνία ανθρώπου-υπολογιστή είναι ένας τομέας που αφορά τον σχεδιασμό, την αξιολόγηση και την απόδοση των διαδραστικών υπολογιστικών συστημάτων για την χρήση από ανθρώπους, και την συνεχή εξέλιξη ως προς την ευχρηστία αυτών των συστημάτων.

1.3 ΣΤΟΧΟΙ

Ένας βασικός στόχος της Επικοινωνίας Ανθρώπου Υπολογιστή είναι η βελτίωση της διάδρασης μεταξύ των χρηστών και των υπολογιστών, κάνοντας τους υπολογιστές πιο εύχρηστους και δεκτικούς στις ανάγκες των χρηστών. Συγκεκριμένα , η επικοινωνία ανθρώπου υπολογιστή ασχολείται με [4][5]:

- Μεθοδολογίες και μηχανισμούς για τον σχεδιασμό διεπιφανειών (Τον σχεδιασμό της καλύτερης δυνατής διεπιφάνειας με δεδομένο το κοινό που απευθύνεται και την θεματολογία.)
- Μεθόδους για υλοποίηση διεπιφανειών (εργαλειοθήκες λογισμικού και βιβλιοθήκες, αποδοτικοί αλγόριθμοι)
- Τεχνικές αξιολόγησης και σύγκρισης διεπιφανειών
- Δημιουργία νέων διεπιφανειών και τεχνικών αλληλεπίδρασης
- Δημιουργία περιγραφικών και προγνωστικών μοντέλων και θεωριών αλληλεπίδρασης

Ένας μακροπρόθεσμος στόχος της αλληλεπίδρασης ανθρώπου υπολογιστή είναι ο σχεδιασμός συστημάτων που ελαχιστοποιούν την απόσταση μεταξύ του ανθρώπινου γνωστικού μοντέλου ως προς το «τι θέλουν να πραγματοποιήσουν», και στην ικανότητα του υπολογιστή να κατανοήσει την ανάγκη του χρήστη.

Επαγγελματίες που ασχολούνται με τον κλάδο είναι συνήθως σχεδιαστές που σχετίζονται με την πρακτική εφαρμογή των σχεδιαστικών μεθοδολογιών στα πραγματικά προβλήματα. Η δουλειά τους συχνά περιστρέφεται γύρω από τον σχεδιασμό γραφικών διεπαφών χρήστη. Οι ερευνητές, ενδιαφέρονται στην ανάπτυξη νέων μεθοδολογιών σχεδίασης, και πειραματίζονται συνεχώς με εξελισσόμενο υλικό εξοπλισμό, κατασκευάζοντας νέα συστήματα λογισμικού, ερευνώντας νέα πρότυπα για την διάδραση και αναπτύσσοντας μοντέλα και θεωρίες που αφορούν την διάδραση.

1.4 ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ

Ένας αριθμός διαφόρων μεθοδολογιών που σκιαγραφούν τις τεχνικές για επικοινωνία ανθρώπου-υπολογιστή έχουν προκύψει από το 1980, με την ανοδική πορεία αυτού του κλάδου. Οι περισσότερες μεθοδολογίες σχεδιασμού, πηγάζουν από ένα μοντέλο που αφορά το πώς οι χρήστες, οι σχεδιαστές και τα τεχνικά συστήματα μπορούν να αλληλεπιδράσουν.

Το πρώτο βιβλίο που εκφράζει τα συναισθήματα στους ανθρώπους και τα ζώα είχε γραφτεί από τον Charles Darwin τον δέκατο ένατο αιώνα [3]. Έπειτα από αυτή την δουλειά ορόσημο, οι ψυχολόγοι σταδιακά συγκέντρωσαν το ενδιαφέρον τους και την γνώση τους πάνω στο κομμάτι αυτό. Ένα νέο κύμα ενδιαφέροντος έχει αναπτυχθεί τελευταία από ψυχολόγους και ειδικούς που ασχολούνται με την τεχνητή νοημοσύνη. Οι λόγοι ανάπτυξης αυτού του νέου κύματος είναι προφανείς [4]. Αρχικά είναι η τεχνολογική πρόοδος, σε ότι αφορά την καταγραφή, την αποθήκευση και την επεξεργασία οπτικοακουστικού υλικού. Έπειτα, η ανάπτυξη μη εμφανών αισθητήρων, η έλευση φορητών υπολογιστών και η ανάγκη για τον

εμπλουτισμό της επιφάνειας διάδρασης ανθρώπου-υπολογιστή από το point-and-click (δείξε και πάτα) στο sense-and-feel (εντόπισε και ψηλάφισε). Τέλος, η εισαγωγή άψυχων αντικειμένων που μοιάζουν με «ζωντανά», στη ζωή μας, όπως ρομποτάκια που μοιάζουν με ζώα –Furbies Tiger & Sony’s Aibo- τα οποία υποτίθεται ότι πρέπει να είναι ικανά να εκφράσουν, να έχουν και να καταλαβαίνουν συναισθήματα. Ένα νέο κομμάτι έρευνας στην Τεχνητή Νοημοσύνη γνωστό ως Συναισθηματική Υπολογιστική (affective computing) έχει πρόσφατα δημιουργηθεί [5].

Σε ότι αφορά την έρευνα αποκωδικοποίησης και απεικόνισης συναισθημάτων μέσα από την ομιλία, ψυχολόγοι έχουν κάνει πολλά πειράματα και έχουν προτείνει πολλές θεωρίες. Από την άλλη πλευρά, οι ερευνητές της τεχνητής νοημοσύνης κάνουν βήματα στους τομείς της συναισθηματικής σύνθεσης ομιλίας, στην αναγνώριση συναισθημάτων, και στην χρήση «πρακτόρων» για την αποκωδικοποίηση και την έκφραση συναισθημάτων [6][7][8][9].

1.5 ΔΙΕΠΙΦΑΝΕΙΕΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ

1.5.1 Οι διεπιφάνειες περιέχουν κάποιο μήνυμα για τον χρήστη

Για πολύ καιρό έχουμε συνειδητοποιήσει την χρήση του υπολογιστή σαν εργαλείο. Η λέξη «εργαλείο» ως μεταφορά, υπαγορεύει ότι ο υπολογιστής θα χρησιμοποιηθεί σαν γραφομηχανή, υπολογιστής, ή κουτί για ηλεκτρονικές κάρτες εύρεσης πληροφοριών. Το 1988 στο συνέδριο «Ανάπτυξη λογισμικού και κατασκευή πραγματικότητας» (Software Development and Reality Construction) συζητήσαν κι άλλες μεταφορές για τον υπολογιστή και καθόρισαν άλλο ρόλο, αυτόν της βοήθειας προσανατολισμού [10]. Πέντε χρόνια αργότερα η Brenda Laurel περιέγραψε τους υπολογιστές σαν το θέατρο και εισήγαγε νέες μεταφορές στην συζήτηση [11]. Η ιδέα αυτή είναι πέραν της μεταφοράς της λέξης «εργαλείο» και φαίνεται να είναι κατάλληλη για όλες τις νέες εμπειρίες επικοινωνίας και συνεργασίας (όπως είναι ομάδες πληροφοριών, δωμάτια συνομιλιών, εικονικές κοινότητες κ.α.). Ο υπολογιστής και ιδιαίτερα η διεπιφάνεια, σαν την οριογραμμή ανάμεσα στον χρήστη και στην λειτουργικότητα οποιασδήποτε μορφής, έχει γίνει ένα μέσον και ένα περιβάλλον μέσα στο οποίο η επικοινωνία και η συνεργασία είναι εφικτή [12]. Επιπλέον, ο νέος ρόλος της διεπιφάνειας, είναι να κάνει την επικοινωνία μεταξύ ανθρώπων μέσω του διαδικτύου όσο το δυνατόν πιο ευχάριστη [13]. Αυτή η διεπιφάνεια όχι μόνο μας βοηθάει να φτάσουμε στους επικοινωνιακούς μας στόχους αλλά και συνδράμει στην επικοινωνία.

Οι Byron Reeves & Clifford Nass απέδειξαν στην μελέτη τους The Media Equation ότι οι άνθρωποι διαδρούν με τους υπολογιστές ακριβώς με τον ίδιο τρόπο που διαδρούν με άλλους

ανθρώπους. Αυτό σημαίνει ότι μεταφέρουμε τα θεμελιώδη πρότυπα της κοινωνικής συμπεριφοράς στους υπολογιστές, ακόμη κι αν γνωρίζουμε ότι εκείνοι είναι «απλά μηχανές». Μπορούμε να εξαπατηθούμε από ένα μηχάνημα πολύ εύκολα όταν θέλουμε να το εκλάβουμε σαν πρότυπο επικοινωνίας [14][15]. Οι Reeves & Nass συνοψίζουν στην μελέτη τους «Έχουμε βρει ότι η διάδραση των ατόμων με τους υπολογιστές, την τηλεόραση και τα νέα μέσα είναι θεμελιωδώς κοινωνική και φυσική, όπως η διάδρασή τους στην καθημερινή ζωή» [14]. Το δεύτερό τους συμπέρασμα αναφέρει ότι «Οι άνθρωποι αποκρίνονται κοινωνικά και φυσικά στα μέσα ακόμη κι αν πιστεύουν ότι δεν είναι λογικό να το κάνουν, ακόμη κι αν δεν νομίζουν ότι αυτές οι αποκρίσεις τους χαρακτηρίζουν» [14].

Για να καταλάβουμε καλύτερα τις θέσεις των Reeves και Nass, πρέπει να τις δούμε από την θέση της πραγματικής θεωρίας της επικοινωνίας [16][17]. Αυτή η θεωρία προτείνει στα αξιώματα της επικοινωνίας ότι για να καταλάβει κανείς κάθε συμπεριφορά σαν μια μορφή επικοινωνίας, και να καταλάβει κάθε επικοινωνία κάτω από την οπτική του περιεχομένου και την οπτική των σχέσεων, οι οποίες κυριαρχούν στην οπτική του περιεχομένου, είναι μια μορφή «μετα-επικοινωνίας» [17]. Η επικοινωνία είναι πάντα περισσότερο μια ανταλλαγή πληροφοριών, αλλά είναι μια κοινωνική διαδικασία που δείχνει και αντισταθμίζει τις σχέσεις μεταξύ των ανθρώπων.

Εάν μεταφέρουμε την κοινωνική και επικοινωνιακή συμπεριφορά μας στην χρήση των υπολογιστών, αντιμετωπίζουμε φαινόμενα όπως η εμπειρία της χρήσης της μηχανής όχι μόνο σε επίπεδο περιεχομένου αλλά και κάτω από την οπτική των σχέσεων. Είναι σημαντικό να σημειώσουμε την λέξη «εμπειρία» γιατί οι υπολογιστές δεν επικοινωνούν μαζί μας –ούτε σε επίπεδο περιεχομένου ούτε σε επίπεδο σχέσεων. Μας δίνουν σημάδια που εμείς μπορούμε να τα μεταφράσουμε με σκοπό να οδηγηθούμε σε συμπεράσματα. Ωστόσο φαίνεται να αντιμετωπίζουμε μια μορφή πιο προσιτής επικοινωνίας κατά την διάδραση με τον υπολογιστή, κάτι που είναι αποτέλεσμα του σχεδιασμού της επιφάνειας διάδρασης του υπολογιστή. Λόγω αυτού, ο σχεδιασμός της διεπιφάνειας είναι πολύ σημαντικός σε σχέση με την αναπαράσταση της λειτουργίας [14][15][16][17].

1.5.2 Με τι πρέπει να επικοινωνούν οι υπολογιστές

Εάν οι συναισθηματικές συνιστώσες παίζουν κυρίαρχο ρόλο για τα ανθρώπινα όντα, και αν οι άνθρωποι αντιλαμβάνονται την χρήση των υπολογιστών σαν μια μορφή διάδρασης, τότε πρέπει να αναρωτηθούμε τι σημαίνει αυτό για τον σχεδιασμό της επιφάνειας διάδρασης. Για την απάντηση αυτής της ερώτησης πρέπει να καθορίσουμε την αντίληψή μας για τον χρήστη κάτω από ανθρωπολογικές οπτικές. Βασιζόμενοι στην έρευνα του Humberto

Maturana, και στην πρόταση του Frank Thissen η κατανόηση των ανθρωπίνων όντων γίνεται κάτω από τις τρεις οπτικές: Την συναισθηματική οπτική, την γνωστική οπτική και την οπτική της δομικής σύζευξης [12][18]. Προϋποθέσεις αυτών των οπτικών είναι οι ανθρώπινοι παράγοντες, οι βιολογικές δομές, τα κληρονομικά και επίκτητα χαρακτηριστικά, οι προτιμήσεις, τα κίνητρα και οι γνώσεις.

Η συναισθηματική οπτική βοηθά –όπως προαναφέραμε– τον να επιβιώνει [19][20] [21][22][23][24]. Είναι κυρίαρχο. Ο στόχος του σχεδιαστή της διεπιφάνειας είναι να δημιουργήσει μια κατάσταση που δίνει έμφαση σε αισθήματα που παρέχουν σιγουριά, και γενικότερα θετικά συναισθήματα. Περιμένουμε να δούμε ευχάριστες καταστάσεις και να αποφύγουμε τις δυσάρεστες. Μια διεπιφάνεια που υιοθετεί αυτή την οπτική πρέπει να «στέλνει» μηνύματα που εκφράζουν ασφάλεια («Είσαι ασφαλής»), που δημιουργούν εμπιστοσύνη («Είμαι αξιόπιστο και αποδοτικό»), που προσελκύουν την προσοχή («Αυτό είναι ενδιαφέρον για σένα») και επίσης να περιέχει στοιχεία παιχνιδιού («Αυτό είναι διασκεδαστικό»). Σε αυτό το περιεχόμενο βασίζονται οι ιδέες των παιχνιδιών εκμάθησης, και φαίνεται πολύ σημαντικό για τον μελλοντικό σχεδιασμό διεπιφανειών [25].

Η δεύτερη οπτική βασίζεται στην ιδέα ότι οι άνθρωποι σαν συστήματα, σταθερά οργανώνουν τους εαυτούς τους και κάνοντας αυτό χτίζουν την γνώση τους πάνω στον κόσμο και την γνώση τους για τους εαυτούς τους [18]. Εδώ βρίσκουμε μια σταθερή ερμηνεία και ταξινόμηση της αντίληψης, πράγμα που σημαίνει ότι δεν υπάρχει «αντικειμενικότητα» και ουδέτερη κατανόηση για το τι είναι ο κόσμος, ή ακόμη για το τι είναι η γνώση. Η δομή της γνώσης είναι πάντα μεμονωμένη και η επιφάνεια διάδρασης πρέπει να συμφωνεί με αυτή την μοναδικότητα. Πρέπει να είναι προσαρμοστική και ο χρήστης πρέπει να έχει την ικανότητα να την καθορίσει σύμφωνα με τις δικές του προσδοκίες. Η διεπιφάνεια πρέπει να περιλαμβάνει διαφορετικούς τύπους αντίληψης (οπτικούς, ηχητικούς κ.α.). Πρέπει να περιέχει δραματουργικά στοιχεία, με ρυθμικά στοιχεία αφήγησης τα οποία ταιριάζουν στην αντίληψη του ατόμου[23] [26] [27] [28].

Η τρίτη οπτική, σημαίνει ότι το ανθρώπινο όν πρέπει να διαδρά με άλλα ανθρώπινα όντα σε κοινωνικές καταστάσεις [29] [16] [17] [13]. Αυτά τα μηνύματα είναι τα «Ανήκεις σε αυτή την ομάδα», «Είσαι σημαντικός», «Είσαι αποδεκτός». Όπως μπορούμε να δούμε σε εικονικές κοινότητες η κοινωνική διάδραση μπορεί να έχει ισχυρή δυναμική. Μια διεπιφάνεια που ενεργοποιεί την κοινωνική επικοινωνία και την ομαδική δυναμική βοηθά τους χρήστες της να οδηγηθούν πέρα από τις κλασσικές διεπιφάνειες.

1.5.3 Επικοινωνία Ανθρώπου Υπολογιστή, Κίνητρα για την Αναγνώριση Συναισθημάτων

Υπάρχουν πολλά κίνητρα για την αναγνώριση της συναισθηματικής κατάστασης από την ανθρώπινη φωνή. Στην αλληλοεπίδραση που μπορεί να υπάρξει ανάμεσα σε άνθρωπο και μηχάνημα, το μηχάνημα μπορεί να καταφέρει να παράγει πιο κατάλληλες απαντήσεις αν η συναισθηματική κατάσταση του ανθρώπου ταυτοποιηθεί με ακρίβεια [30]. Τα τελειότερα συστήματα αναγνώρισης αυτόματης ομιλίας καταφεύγουν στην κατανόηση της (φυσικής) μητρικής γλώσσας, για να μπορέσουν να βελτιώσουν την ακρίβεια της αναγνώρισης των λέξεων που έχουν ειπωθεί ή καταγραφεί. Βελτίωση μπορεί να παρουσιαστεί αν η συναισθηματική κατάσταση αποσπαστεί και έτσι ενισχυθεί η ακρίβεια του συστήματος. Γενικώς η μετάφραση αυτή απαιτείται για να διεκπεραιωθεί η επικοινωνία χρησιμοποιώντας διαφορετικές μορφές γλώσσας. Οι συναισθηματικές καταστάσεις με βάση τις στατιστικές έχουν συγκεκριμένο χρονικό διάγραμμα [31].

Πιο χαρακτηριστικά, οι άνθρωποι με συναισθηματικές διαταραχές όπως η μανιοκατάθλιψη ή παθολογικό άγχος μπορεί να είναι σε αυτήν την κατάσταση για ένα μεγάλο χρονικό διάστημα (μήνες ή και χρόνια) . Επίσης κάποιος μπορεί να είναι σε μια κακή διάθεση για κάποιες εβδομάδες ή μήνες και συναισθήματα όπως ο θυμός και η χαρά, μπορεί να είναι μεταβατικά και να μη διαρκούν περισσότερο από λίγα λεπτά, σε αυτές τις περιπτώσεις. Επομένως, το συναίσθημα εδώ έχει μια ευρύτερη αλλά και μια πιο στενή έννοια και επίδραση. Η ευρύτερη έννοια αφορά το μακροπρόθεσμο συναίσθημα ενώ η στενή έννοια αναφέρεται σε βραχυπρόθεσμο ερεθισμό του εγκεφάλου που ενεργοποιείται στους ανθρώπους και τους ωθούν να δράσουν. Στην αυτόματη αναγνώριση των συναισθημάτων ένας μηχανισμός δεν θα μπορούσε να διακρίνει αν η συναισθηματική κατάσταση οφείλεται σε μακροπρόθεσμη ή βραχυπρόθεσμη επίδραση εφόσον αντικατοπτρίζεται στην ομιλία ή στην έκφραση του προσώπου του ομιλητή. Ωστόσο, οι προσπάθειες για αναγνώριση συναισθημάτων είναι σημαντικές για την βελτίωση των συστημάτων διάδρασης ανθρώπου-υπολογιστή.

2 ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ

2.1 Συναισθήματα στην διάδραση ανθρώπου-υπολογιστή

Σύμφωνα με τις έρευνες που έκαναν οι Reeves & Nass το 1996 προτείνουν ότι οι άνθρωποι διαδρούν με τους υπολογιστές σύμφωνα με τους ίδιους κανόνες κοινωνικής διάδρασης όπως κάνουν οι άνθρωποι μεταξύ τους [14]. Για παράδειγμα, σε ένα συγκεκριμένο πείραμα τους έδειξαν ότι οι άνθρωποι έχουν μια ευγενική διάθεση προς τους υπολογιστές. Η χρήση αυτή των κοινωνικών κανόνων συμπεριφοράς-ευγένειας, φαίνεται να είναι υποσυνείδητη αφού τα ενδιαφερόμενα πρόσωπα αρνήθηκαν ότι χρησιμοποίησαν κανόνες

κοινωνικής συμπεριφοράς σε μια συνέντευξη μετά από το παραπάνω πείραμα [32]. Απ' ότι φαίνεται, οι άνθρωποι αυτόματα χρησιμοποιούν τους ίδιους κανόνες διάδρασης στις σχέσεις ανθρώπου με άνθρωπο αλλά και ανθρώπου με υπολογιστή. Για να κάνουμε την διάδραση μεταξύ ανθρώπου και υπολογιστή αβίαστη, ο υπολογιστής θα πρέπει να ανταποκρίνεται σε αυτήν την φυσική τάση που υπάρχει [33]. Η σύνθεση ομιλίας έχει αποδειχθεί ότι αποτελεί ένα σημαντικό μέρος της διάδρασης ανθρώπου-υπολογιστή στο μέλλον. Παρόλα αυτά φαίνεται ότι πριν η χρήση της σύνθεσης ομιλίας γίνει ευρέως γνωστή, η ποιότητα της ομιλίας θα πρέπει να βελτιωθεί. Ένα χαρακτηριστικό που κάνει την σύνθεση ομιλίας να ξεχωρίζει από την ανθρώπινη ομιλία, είναι η ακαταλληλότητα της προσωδίας και η εκφραστικότητα [34].

Σε ότι αφορά την επικοινωνία συναισθημάτων, η προσωδία της ομιλίας παίζει κεντρικό ρόλο σύμφωνα με έρευνες που έχουν γίνει το 1995 από τους Bachorowski & Owren & Scherer [35] [36]. Ο καθορισμός των φωνητικών δεικτών των συναισθημάτων έχει αποδειχθεί μια πολύ περίπλοκη εργασία [37]. Όμως, κάποιες παράμετροι φωνητικών ενδείξεων των συναισθημάτων έχουν ανακαλυφθεί. Μια βασική παράμετρος φωνητικού σήματος είναι η θεμελιώδης συχνότητα F_0 που είναι ο βαθμός στον οποίο οι φωνητικές διακυμάνσεις δονούνται (αυξομείωση της έντασης). Ο θυμός και η χαρά, για παράδειγμα, φαίνεται να παράγουν μια αύξηση στην μέση θεμελιώδη συχνότητα ενώ από την άλλη πλευρά η λύπη φαίνεται να συσχετίζεται με την μείωση στην μέση θεμελιώδη συχνότητα [38]. Άλλα χαρακτηριστικά του λόγου είναι για παράδειγμα η μελωδία, η παύση, η ένταση, και ο ρυθμός ομιλίας, τα οποία παίζουν έναν ρόλο στην φωνητική επικοινωνία των συναισθημάτων [37]. Έχει προταθεί ότι στο μέλλον η σύνθεση ομιλίας με τις κατάλληλες και αποτελεσματικές προσωδίες, μπορούν να χρησιμοποιηθούν σε συστήματα υπολογιστών που είναι εύκολα στην χρήση [39][36][38]. Παρόλα αυτά συνδυάζοντας συναισθήματα με την σύνθεση ομιλίας είναι μια περίπλοκη διαδικασία, αφού ένα σημαντικό εμπόδιο είναι η ποικιλομορφία της έκφρασης του ίδιου συναισθήματος [39]. Επομένως, ένα μηχάνημα σύνθεσης ομιλίας ικανό να έχει φυσικούς ήχους και συναισθηματική ομιλία απαιτεί ακόμη περισσότερη εργασία. Όμως, τα μηχανήματα σύνθεσης ομιλίας μπορούν ήδη να χρησιμοποιηθούν στην ανάλυση συναισθηματικών μηνυμάτων χρησιμοποιώντας σημασιολογικά στοιχεία, όπως για παράδειγμα στο περιεχόμενο της ομιλίας.

Στην μοντελοποίηση της σχέσης ανάμεσα στα φωνητικά στοιχεία και στο γλωσσολογικό περιεχόμενο στην φωνητική έκφραση δύο μοντέλα έχουν προταθεί, το μοντέλο συνδιακύμανσης (covariation model) και το μοντέλο διαμόρφωσης (configuration model) [40][41]. Το μοντέλο συνδιακύμανσης θεωρεί ότι η συναισθηματική σημασία αλλά και συγκεκριμένα φωνητικά στοιχεία μεταβάλλονται κατευθείαν και ανεξάρτητα από το

γλωσσολογικό περιεχόμενό του. Σύμφωνα με το μοντέλο αυτό, οι ακουστικοί παράμετροι ποικίλουν με την δύναμη μιας συγκεκριμένης κατάστασης. Το μοντέλο διαμόρφωσης από την άλλη μεριά, συστήνει ότι το συναισθηματικό μήνυμα αναλύεται σαν σχηματισμός ανάμεσα στα φωνητικά στοιχεία και στην γλωσσολογική δομή. Έχει προταθεί, ότι το πρώτο μοντέλο είναι καλύτερο στην επεξήγηση συναισθημάτων που έχουν σχέση με την διέγερση (θυμός, χαρά) ενώ το δεύτερο μοντέλο είναι πιο επαρκές όταν η κατάσταση του ομιλητή έχει σχέση με διαδικασία σκέψης (αμφισβήτηση).

2.1.2 Τι είναι τα συναισθήματα

Το συναίσθημα είναι μια περίπλοκη διαδικασία και δεν υπάρχει σαφής ορολογία γι' αυτό. Πολλοί ερευνητές συμφωνούν ότι τα συναισθήματα περιλαμβάνουν τα ακόλουθα και βασικά στοιχεία. Νευροφυσιολογική δραστηριότητα, Νευρομυική δραστηριότητα, Υποκειμενική εμπειρία και το στοιχείο της συμπεριφοράς [42][43][44][45][46]. Η έρευνα στον ρόλο του κεντρικού νευρικού συστήματος για τα συναισθήματα έχει αυξηθεί καθώς πολλές τεχνικές απεικόνισης έχουν αναπτυχθεί. Είναι επίσης γνωστό ότι η αμυγδαλή¹ παίζει έναν κεντρικό ρόλο τουλάχιστον σε συναισθηματικές διαδικασίες που έχουν σχέση με τον φόβο [47][48][49][50]. Το συγκεκριμένο σημείο, έχει ονομαστεί 'ο συναισθηματικός υπολογιστής του εγκεφάλου' και είναι εξαιρετικής σημασίας όταν το ερέθισμα συσχετίζεται με μια συναισθηματική σημασία [49]. Η νευρομυική δραστηριότητα παράγει εξωτερικές εκφράσεις συναισθημάτων, για παράδειγμα του προσώπου και φωνητικές εκφράσεις [51][40]. Η υποκειμενική εμπειρία ενός συναισθήματος (όπως για παράδειγμα η λύπη ο φόβος, η χαρά) μπορεί να θεωρηθεί ότι περιλαμβάνει τρία στοιχεία: Τη γνώση της δομής της συναισθηματικής κατάστασης, την αναγνώριση των ερεθισμάτων του σώματος (αυτόνομη διέγερση) και αναγνώριση ετοιμότητας κίνησης (μια τάση της προσέγγισης ή της απομάκρυνσης). Όλα αυτά τα στοιχεία συνθέτουν μια πλήρη συναισθηματική εμπειρία [42].

Σύμφωνα με έρευνες από τους Lang και λοιπούς, τα συναισθήματα μπορεί να αντιμετωπιστούν σαν προδιάθεση κινήσεων και όχι οι ίδιες οι κινήσεις που εντοπίζονται σε χώρο δύο διαστάσεων [52][53]. Αυτός ο χώρος απαρτίζεται από συντεταγμένες σθένους και διέγερσης. Υπάρχουν δύο ξεχωριστές εγκεφαλικές λειτουργίες που είναι υπεύθυνες για τις αλλαγές ανάμεσα στην προσέγγιση (σεξουαλικότητα, ανατροφή) και τάσεις απομάκρυνσης (άμυνα και προστασία) από έναν οργανισμό παράγοντας έτσι κίνητρο για την συμπεριφορά του. Η ενεργοποίηση αυτών των εγκεφαλικών λειτουργιών συσχετίζεται με δραστηριότητες

¹ Ονομάστηκε έτσι επειδή το σχήμα της μοιάζει με αμύγδαλο. Είναι μέρος του λιμπικού συστήματος του εγκεφάλου, και ευθύνεται για την ρύθμιση των συναισθημάτων. Η αμυγδαλή συνδέεται κυρίως με συναισθήματα όπως αυτό του φόβου και του άγχους αλλά και με έναν αρνητικό τρόπο, με το συναίσθημα της ευχαρίστησης.

του φλοιού του εγκεφάλου, της αυτονομίας και της συμπεριφοράς. Το στοιχείο του σθένους μιας συναισθηματικής εμπειρίας οφείλεται στην ενεργοποίηση αυτών των εγκεφαλικών λειτουργιών. Η διέγερση φαίνεται σαν να αντιπροσωπεύει το επίπεδο δραστηριότητας των λειτουργιών του εγκεφάλου που έχουν σχέση με την προσέγγιση ή την απομάκρυνση ή την συνενεργοποίηση των δύο παραπάνω συστημάτων.

Τα συναισθήματα μπορεί να περιγραφούν σαν ξεχωριστές κατηγορίες (όπως ευχαρίστηση/ευτυχία, θυμό, φόβο, λύπη και αηδία) και όλα έχουν τις ξεχωριστές τους εκφράσεις συναισθηματικές εμπειρίες και επιπτώσεις συμπεριφοράς [54][55]. Η άποψη μεγέθους από την άλλη πλευρά καθορίζει τα συναισθήματα με βάση ευρύτερες διαστάσεις. Υπάρχει ακόμη αντιπαράθεση που αφορά τον αριθμό των απαιτούμενων διαστάσεων αλλά το σθένος (από αρνητικό σε θετικό) και η διέγερση (από ήρεμη κατάσταση σε έναν ερεθισμό συναισθημάτων) είναι δύο διαστάσεις που χρησιμοποιούνται συχνότατα. Η θέση ισχύος (ο βαθμός ελέγχου που κάποιος συγκεκριμένα αισθάνεται ότι έχει σε μια συναισθηματική κατάσταση) είναι η τρίτη διάσταση που έχει χρησιμοποιηθεί για την περιγραφή ενός συναισθήματος [56][57]. Έχει προταθεί ότι το σθένος και η διέγερση είναι πρωταρχικές διαστάσεις στην περιγραφή συναισθηματικής εμπειρίας [58][56]. Πιο πρόσφατα, οι προσεγγίσεις σε κατηγορία και σε διάσταση συνήθως φαίνονται συμπληρωματικές η μια της άλλης [42][46][55]. Οι διαστάσεις του σθένους και της διέγερσης θεωρούνται σαν στοιχεία που τονίζουν όλα τα ξεχωριστά συναισθήματα. Για παράδειγμα, υπάρχουν συχνές καταστάσεις όπου η πληροφόρηση είναι τόσο λίγη για την λήψη αποφάσεων για το ποιο ξεχωριστό συναίσθημα εξετάζουμε. Σε αυτές τις περιπτώσεις οι άνθρωποι πιθανόν χρησιμοποιούν πληροφόρηση διαστάσεων για την διαδικασία επεξήγησης των συναισθημάτων [46]. Έχει επίσης επισημανθεί ότι ερευνητές οι οποίοι σθεναρά υποστηρίζουν την άποψη του διακριτικού συναισθήματος, χρησιμοποιούν κάποιες φορές μεθοδολογία η οποία πηγάζει από προσέγγιση διαστάσεων [55]. Αυτό συμβαίνει διότι κατά μέσο όρο είναι ευκολότερο να αποκτήσει κάποιος αξιόπιστες μετρήσεις σε ένα ευρύ επίπεδο διαστάσεων απ' ότι σε ένα επίπεδο διακριτικών συναισθημάτων.

Αν και έχει αποδειχθεί ότι είναι δύσκολο να καθοριστεί το συναίσθημα από κάποια θεωρία που να τη στηρίζουν οι ερευνητές ομόφωνα, υπάρχουν κάποιες σχετικές αντιλήψεις οι οποίες πρέπει να διαφοροποιηθούν από τα συναισθήματα [46][54][55]. Τα συναισθήματα εξωτερικεύονται γρήγορα και τυπικά διαρκούν από λίγα δευτερόλεπτα μέχρι λίγα λεπτά το πολύ, ενώ οι «διαθέσεις» μας διαρκούν για ημέρες, εβδομάδες ή μήνες. Επιπλέον οι διαθέσεις δηλώνονται σαν απλές καταστάσεις αισθημάτων αλλά δεν έχουν καμία σχέση με τις παγκόσμιες εκφράσεις που χαρακτηρίζουν το πρόσωπο. Όμως, η έκφραση του προσώπου, ή

τουλάχιστον η μυϊκή δραστηριότητα του προσώπου αποτελεί ένα βασικό μέρος της συναισθηματικής αντίδρασης [45][46][59]. Άλλο σημείο διάκρισης, έχει προταθεί να είναι το αντικείμενο μιας συναισθηματικής κατάστασης. Όπως έχει αναφερθεί, τα συναισθήματα πάντα έχουν να κάνουν με κάτι (παράδειγμα, όταν κάποιος είναι θυμωμένος για κάτι) αλλά οι διαθέσεις δεν έχουν κανένα συγκεκριμένο αντικείμενο [42]. Οι διαθέσεις αυτές μπορούν να θεωρηθούν προδιαθέσεις για να ανταποκριθούν με ένα συγκεκριμένο τρόπο σε κάποιο συναισθηματικό ερέθισμα. Για παράδειγμα, κάποιος άνθρωπος που είναι σε μια ευχάριστη διάθεση, είναι πιο πιθανό να ανταποκριθεί σε ένα ερέθισμα ευτυχίας και χαράς απ' ό,τι ένας άνθρωπος που είναι σε κακή διάθεση [42][46]. Η διάκριση «συναισθηματικών επεισοδίων» από διαθέσεις και συναισθήματα, έχει επίσης καταγραφεί [42]. Σύμφωνα με αυτή την καταγραφή, τα συναισθηματικά επεισόδια χαρακτηρίζονται από αλληλοεπιδράσεις μεταξύ του υποκειμένου (του προσώπου) και του αντικειμένου που απαρτίζονται από μια σειρά διαφορετικών συναισθημάτων. Αυτές είναι καταστάσεις συναισθηματικής εμπλοκής οι οποίες διαρκούν για ώρες ή και μέρες. Επειδή τα συναισθηματικά επεισόδια διαθέτουν μια εσωτερική συνοχή (μια υποκειμενική γνώση για να αντιμετωπιστεί ένα συγκεκριμένο θέμα και μια διαρκής κατάσταση συναισθηματικής εμπλοκής) προτείνεται ότι πρέπει να θεωρηθούν σαν συναισθηματικές ομάδες οι οποίες είναι ξεχωριστές από τις διαθέσεις και τα συναισθήματα [42].

2.1.3 Πολυλειτουργικότητα δεικτών και πολυεπίπεδος χαρακτηρισμός συναισθημάτων

Υπάρχει μια πληθώρα μελετών που δείχνουν ότι η προσωδία είναι πολυλειτουργική [60]. Το ίδιο χαρακτηριστικό μπορεί να χρησιμοποιηθεί για την υπόδειξη του τονισμού, της έκφρασης, ή του ύφους της πρότασης ή μπορεί να χρησιμοποιηθεί για τον καθορισμό των χαρακτηριστικών της γλώσσας ή του χρήστη. Συνομιλώντας, μια λειτουργία μπορεί όχι μόνο να σηματοδοτηθεί από την προσωδία αλλά και από άλλα γλωσσολογικά χαρακτηριστικά. Για παράδειγμα, ο τρόπος έκφρασης μιας πρότασης στα Γερμανικά μπορεί να υποδεικνύεται όχι μόνο από την προσωδία αλλά και από το λεκτικό περιεχόμενο, την σειρά των λέξεων, και την μορφολογία. Δεν υπάρχει κάποιος λόγος γιατί θα πρέπει να είναι διαφορετικά με συναισθήματα.

Τα αποτελέσματα της έρευνας δείχνουν ότι οι ομιλητές πράγματι χρησιμοποιούν έναν αριθμό από διαφορετικά γλωσσολογικά χαρακτηριστικά για να εκδηλώσουν την συναισθηματικότητά τους [60]. Πάλι, εάν η πρόταση που μελετάται παραμένει σταθερή, η γλωσσολογική διάταξη παραμένει σταθερή και εκείνη, και τα γλωσσολογικά επίπεδα παρά την προσωδία, δε μπορούν να δράσουν σαν δείκτες του συναισθήματος. Σχεδόν πάντα, μόνο τα ακουστικά δεδομένα έχουν μελετηθεί σε βασικές έρευνες που αφορούν το συναίσθημα, και όχι

το λεκτικό περιεχόμενο, η γραμματική δομή, ή η δομή του διαλόγου κατά την διάδραση. Παρόλα αυτά, εάν θέλουμε να ταυτοποιήσουμε ορθώς εκφράσεις σε διαλόγους, η πολυλειτουργικότητα των δεικτών και ο πολυεπίπεδος χαρακτηρισμός συναισθημάτων πρέπει να συμπεριληφθούν.

2.1.4 Ανάλυση Συναισθημάτων

Η δημιουργία συναισθημάτων για την έκφραση των συναισθηματικών καταστάσεων τουλάχιστον σε καταστάσεις συναλλαγής στις δυτικές κοινωνίες, αλλά ακόμη περισσότερο σε κάθε κοινωνία και κουλτούρα- είναι επηρεασμένη σε μεγάλο βαθμό από τους νόμους και τους κανόνες. Αυτό σημαίνει ότι έχουμε να κάνουμε με μια παραλλαγή-συγκάλυψη συναισθημάτων [61]. Επομένως μιλάμε για κανόνες έκφρασης. Θα ήταν ωστόσο, πολύ απλό να πιστεύουμε ότι ένα μοναδικό συναίσθημα θα μπορούσε να συγκρατηθεί, αυτό σημαίνει, ότι προϋπάρχει ένα γνήσιο συναίσθημα που απλά δεν θα έχει εκφραστεί σε όλη του την ένταση. Η μελέτη «lost luggage» των Scherer και Ceschi για παράδειγμα, παρουσιάζει πειστικά ότι σε μια τέτοια κατάσταση της πραγματικής ζωής – που οι επιβάτες πρέπει να περιμένουν μάταια για τις βαλίτσες τους στο αεροδρόμιο και να παραπονιούνται στο γραφείο που χειρίζεται τα θέματα των αποσκευών- ένα μοναδικό και ατόφιο συναίσθημα δεν παρατηρήθηκε ούτε από το άτομο που συνομιλούσαν ούτε αναφέρθηκε από τους ίδιους τους ανθρώπους [62][63]. Κατά συνέπεια, η μη ύπαρξη ποικιλομορφίας συναισθηματικών καταστάσεων που είχε παρατηρηθεί από προηγούμενες μελέτες είναι περισσότερο ένα αποτέλεσμα του πειραματικού σχεδιασμού. Εάν ένας ηθοποιός διαβάζει προτάσεις με διαφορετικά συναισθήματα, πρέπει να διαφοροποιήσει την προφορά των συλλαβών με κάποιο σταθερό τρόπο ώστε να εκφράζει δυνατά συναισθήματα. Στο παράδειγμα των αποσκευών που αναφέρθηκε παραπάνω οι ομιλητές μπορούσαν να επιλέξουν ανάμεσα σε διαφορετικές μεθόδους.

Από την πλευρά των εφαρμογών, τα πράγματα είναι ακόμη πιο περίπλοκα. Ας υποθέσουμε ότι στοχεύουμε να ταυτοποιήσουμε τα συναισθήματα που προφέρονται περισσότερο, αυθεντικά ή ανάμικτα, σε μια εφαρμογή που αναφέρεται στην πραγματική ζωή, όπως για παράδειγμα σε ένα διάλογο τηλεφωνικού κέντρου. Εάν οι ομιλητές παρουσιάζονται και εκφράζουν τον απροκάλυπτο θυμό τους με περίπλοκο τρόπο, θα είναι αναμφίβολα πολύ αργά για το σύστημα να μπορέσει να ανταποκριθεί με τέτοιο τρόπο ώστε να σώσει τον διάλογο. Επομένως, αυτό που βλέπουμε είναι όχι ο ‘πλήρης θυμός’ αλλά όλες οι μορφές που υποδηλώνουν ελάχιστο ή μέτριο εκνευρισμό που να αποτελεί ένδειξη ενός κριτικού πνεύματος στον διάλογο που μπορεί να οδηγήσει σε θυμό εάν δεν ληφθούν μέτρα.

2.1.5 Συναισθήματα, διαθέσεις και γνωστικές λειτουργίες

Υπάρχουν στοιχεία ότι τα συναισθήματα μπορούν να έχουν σημαντικές επιρροές σε μερικές πλευρές των γνωστικών λειτουργιών. Έχει βρεθεί ότι ρυθμίζουν για παράδειγμα την μνήμη, την λήψη μιας απόφασης, και τις διαδικασίες των αισθήσεων [58][64][65]. Έχουν προταθεί δύο ερμηνείες για τον λόγο της επιρροής των συναισθημάτων στις γνωστικές λειτουργίες [66].

Η υπόθεση πόλωσης μέσω συμπερασμάτων (inferred bias) προτείνει ότι η συναισθηματική προδιάθεση στις γνωστικές λειτουργίες καλλιεργείται μέσω της μάθησης. Σύμφωνα με αυτή την άποψη, οι συσχετισμοί μεταξύ των συναισθημάτων και των λειτουργικών εκβάσεων μαθαίνονται. Ένα αρνητικό συναίσθημα μπορεί καταρχήν να έχει δημιουργηθεί από μια μη επιτυχημένη δραστηριότητα, αλλά υποσυνείδητα μπορεί να προκαθορίσει την γνωστική διαδικασία. Αυτό μπορεί να οδηγήσει, για παράδειγμα, σε αρνητικές συναισθηματικές καταστάσεις που ενεργοποιούν αναπαραστάσεις μη επιτυχημένων γεγονότων. Αυτές οι αναπαραστάσεις, με την σειρά τους, στιγματίζουν την υποσυνείδητη γνωστική διαδικασία και μπορεί να οδηγήσουν σε φτωχότερη γνωστική λειτουργία.

Η άποψη της ολοκληρωτικής πόλωσης (integral bias) από την άλλη πλευρά, βλέπει τις επιπτώσεις των συναισθημάτων σε γνωστικά θέματα ως αποτέλεσμα κληρονομικών χαρακτηριστικών των συναισθημάτων. Οι Martin και Jones εξέτασαν τις δύο προτεινόμενες υποθέσεις με την χρήση μιας τροποποιημένης έκδοσης ενός Stroop² ως μέτρο της γνωστικής διαδικασίας. Με την προαναφερθείσα μέθοδο, το υποκείμενο που υποβαλλόταν στη μελέτη, έπρεπε να ονομάσει το χρώμα του μελανιού από το οποίο απαρτιζόταν το ερέθισμα. Έχει βρεθεί ότι η ταχύτητα ονομασίας του χρώματος μειώνεται εάν το ερέθισμα περιλαμβάνει αντιθετικά ονόματα χρωμάτων, σε σύγκριση με το ερέθισμα που απαρτίζεται από μη λεκτικούς χαρακτήρες. Στην δική τους εκδοχή για την μέθοδο Stroop, ζητήθηκε από αραχνοφοβικούς και ελεγκτές, να ονομάσουν τα χρώματα που σχετίζονταν με την φοβία και τα χρώματα που σχετίζονταν με τον έλεγχο από αντίστοιχες φωτογραφίες (σπιτιών και αραχνών). Τα αποτελέσματά τους έδειξαν ότι η μείωση στο επίπεδο της εκτέλεσης (η πιο αργή ονομασία των χρωμάτων της εκάστοτε φοβίας) δεν επηρεαζόταν από την ηλικία του ατόμου, ή την διάρκεια της φοβίας ή από την διαφορετικές αιτίες της φοβίας (άμεση αντίδραση, μοντελοποίηση, οδηγίες/πληροφορίες). Όλες αυτές οι μεταβλητές σχετίζονται με την αύξηση των πιθανοτήτων για να μάθηση των συσχετισμών μεταξύ συναισθημάτων και των γνωστικών συνεπειών.

² Η μέθοδος αυτή βασίζεται στην παρατήρηση ότι οι άνθρωποι μπορούν πιο εύκολα να διαβάσουν λέξεις παρά να ξεχωρίζουν και να λένε τα χρώματα με το ονομά τους.

Επειδή απέτυχαν να προκαλέσουν διαφορές στην γνωστική διαδικασία, οι συντάκτες κατέληξαν ότι τα αποτελέσματά τους ταιριάζουν καλύτερα με την υπόθεση integral bias.

2.2 ΤΙ ΜΕΛΕΤΑΜΕ ΣΤΗΝ ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ

2.2.1 Συναισθηματικά δείγματα-οι ηθοποιοί

Μια κατανοητή ανασκόπηση από ερευνητικές προσεγγίσεις σε ότι αφορά την φωνητική έκφραση συναισθημάτων, δίδεται από τον Tischer ο οποίος αναφέρεται σε 53 μελέτες [67]. Έντεκα από αυτές βασίζονται σε δεδομένα από την πραγματική ζωή, συνήθως περιορισμένα σε συγκεκριμένα είδη συναισθημάτων, όπως ο φόβος και ο θυμός. Οχτώ μελέτες διεξήχθησαν σε πειραματικό περιβάλλον, στο οποίο ζητήθηκε από τα αντικείμενα της μελέτης να φανταστούν συγκεκριμένες καταστάσεις τις οποίες θα εξωτερίκευαν με την χρήση φωνητικών εκφράσεων. Οι περισσότερες από τις μελέτες χρησιμοποιούν φωνητικές εκφράσεις από προσποιούμενα συναισθήματα από ηθοποιούς, και σε δεκαέξι μελέτες τα ηχητικά χαρακτηριστικά χρησιμοποιήθηκαν με σκοπό να προκαλέσουν διαφορετικές αντιδράσεις στα αντικείμενα μελέτης (ανασύνθεση φωνητικών εκφράσεων). Ο κλασικός πειραματικός σχεδιασμός για μελέτες εργαστηρίου που αφορούν τα συναισθήματα –καθώς και οι τελευταίες μελέτες δείχνουν ότι τα πράγματα δεν έχουν αλλάξει πολύ [68][69][70][71]- φαίνεται να είναι ο ακόλουθος: Οι έμπειροι ομιλητές ηθοποιοί αντιδρούν σαν να ήταν σε μια κατάσταση έξαψης, σαν να ήταν χαρούμενοι, θυμωμένοι, λυπημένοι κλπ. Με σκοπό να κρατηθεί μια αίσθηση σταθερότητας στη μελέτη, γίνεται χρήση του ίδιου φορέα της πρότασης και των αντικειμένων προς μελέτη [60]. Αυτή η πειραματική ρύθμιση είναι παρόμοια με αυτές που χρησιμοποιήθηκαν και στα φωνητικά/φωνολογικά πειράματα [72].

2.2.2 Πραγματική ζωή-ανθρώπινα όντα

Ο στόχος όλων αυτών των προσπαθειών είναι φυσικά η μοντελοποίηση του λόγου των «φυσιολογικών» ανθρώπινων όντων στην πραγματική ζωή της διάδρασης ανθρώπου-υπολογιστή. Η εξαγωγή δεδομένων από πραγματικές καταστάσεις μπορεί να αντιμετωπίσει δύο βασικά προβλήματα [60]. Πρώτον, είναι δύσκολη η παρακολούθηση και καταγραφή σκηνικών της πραγματικής ζωής λόγω ηθικών περιορισμών και επειδή πραγματικά συστήματα αυτόματων διαλόγων είναι ακόμη σπάνια. Δεύτερον, οι στόχοι είναι κινούμενοι. Εάν αλλάξουμε την εφαρμογή στην οποία τα δεδομένα ηχογραφούνται έστω και ελάχιστα, μπορεί να επηρεαστεί η γλωσσολογική ή η συναισθηματική συμπεριφορά του χρήστη σε μεγάλο εύρος [73]. Ας σημειωθεί ότι ακόμη και η ανομοιομορφία μεταξύ της γλώσσας μπορεί να προκαλέσει πρόβλημα σε ότι αφορά την αναγνώριση των λέξεων. Η διαφορά είναι όμως ότι τέτοιες ανομοιομορφίες είναι γνωστές αν και δεν μπορούν να μοντελοποιηθούν με κατάλληλο

τρόπο. Για την συναισθηματική συμπεριφορά απλών (ανεπιτήδευτων) ανθρώπων σε καταστάσεις της πραγματικής ζωής, δεν γνωρίζουμε το εύρος της ανομοιομορφίας. Μια αντιπροσωπευτική κλίμακα από δεδομένα της πραγματικής ζωής είναι πολύ δύσκολο να επιτευχθεί.

2.3 ΠΡΟΒΛΗΜΑΤΑ ΣΤΗΝ ΑΝΑΛΥΣΗ ΤΗΣ ΓΛΩΣΣΑΣ ΤΩΝ ΣΥΝΑΙΣΘΗΜΑΤΩΝ

2.3.1 Μεθοδολογικά προβλήματα στην ανάλυση της συναισθηματικής γλώσσας

Οι Batliner, Fischer, Huber, Silker και Noth εστιάζουν στο ότι οι περισσότερες από τις υπάρχουσες μελέτες για τα συναισθήματα, παραλείπουν ή κρατούν αμετάβλητους τους παρεμβατικούς παράγοντες μέσω του πειραματικού σχεδιασμού [60]. Υπάρχει όμως ένας αριθμός προβλημάτων που σχετίζεται με αυτή την διαδικασία, κατά την οποία γίνεται μίμηση των συναισθημάτων. Κατά συνέπεια, η χρήση δεδομένων από τον τρόπο ομιλίας των ηθοποιών είναι προβληματική σε ορισμένες περιπτώσεις. Το πιο σημαντικό είναι ότι δεν είναι ξεκάθαρο ότι αυτό που θα κάνουν οι ηθοποιοί όταν θα τους ζητηθεί να εκφράσουν ένα συγκεκριμένο συναίσθημα, θα ανταποκρίνεται στο τι θα κάνει ένας άνθρωπος στην πραγματική ζωή. Καθώς το καθήκον του ηθοποιού έγκειται στην προσποίηση μιας θετικής συναισθηματικής κατάστασης, δεν είναι αυταπόδεικτο ότι οι ομιλητές στην πραγματική ζωή εκφράζουν τα συναισθήματά τους [74][75]. Επιπροσθέτως, δεν είναι ξεκάθαρο εάν θα χρησιμοποιήσουν τα ίδια γλωσσολογικά μέσα που θα χρησιμοποιήσουν οι ηθοποιοί. Αυτό βέβαια αφορά τα αποτελέσματα αυτών των μελετών που δεν υιοθετούν τον λόγο των ηθοποιών αλλά επιτρέπουν μόνο προσωδιακή έκφραση του συναισθήματος που αναφέρονται, για παράδειγμα, την ανάγνωση προτάσεων. Στην πραγματική ζωή οι ομιλητές δεν περιορίζονται στη χρήση μόνο προσωδίας αλλά μπορούν να επιλέξουν ανάμεσα σε ένα αριθμό από διαφορετικές διαθέσιμες στρατηγικές. Τελικώς οι ηθοποιοί προσποιούνται μόνο ότι έχουν ένα συναίσθημα και δεν γνωρίζουμε με ποιους τρόπους αυτό επιδρά στους τρόπους που αυτοί εκφράζουν την συναισθηματικότητά τους. Εντούτοις, γενικώς για τέτοια πειράματα μια πολύ καλή απόδοση αναφέρεται: Οι πειραματιζόμενοι μπορούν να διαπιστώσουν τα επιδιωκόμενα συναισθήματα με μεγάλη αξιοπιστία και οι αυτόματοι ταξινομητές να αποδίδουν μεγάλο βαθμού αναγνωριστικότητα [60].

2.3.2 Ψέμα και συναισθήματα

Στην επικοινωνία μεταξύ ανθρώπων, τα συναισθήματα είναι το νούμερο ένα θέμα για το οποίο οι άνθρωποι ψεύδονται και μελέτες έχουν δείξει ότι μέχρι το 30% της κοινωνικής διάδρασης που διαρκεί πάνω από δέκα λεπτά περιέχει ψεύδος [76]. Κατασκευάζοντας τα

τεχνολογικά συστήματα ως ενσωματωμένους ομιλητικούς συντελεστές με την ικανότητα να εντοπίζουν, να αναπαράγουν, να γενικεύουν ή και να δείχνουν συναισθήματα, είναι συνεπώς αναπόφευκτο να διερευνηθούν τα κρίσιμα ερωτήματα για τον χειρισμό των λανθασμένων συναισθηματικών εκφράσεων από τον χρήστη [78].

Όλοι οι χρήστες τείνουν να χειρίζονται τους υπολογιστές σαν ανθρώπινα όντα, συχνά συμπεριφέρονται με ένα τρόπο που θεωρούν κοινωνικά επιθυμητό και αποφασίζουν να μην πουν στο σύστημα τις αληθινές τους σκέψεις [14]. Για παράδειγμα, έχει παρατηρηθεί ότι κάποιοι χρήστες προσπάθησαν να είναι ευγενικοί όταν το παιδαγωγικό σύστημα τους ρώτησε εάν τους αρέσει το σύστημα, παρότι ήταν αρκετά φανερό ότι ψεύδονταν [79]. Είναι φανερό ότι ένα παιδαγωγικό σύστημα για παράδειγμα χρειάζεται να είναι ευαίσθητο στις πραγματικές συναισθηματικές καταστάσεις των μαθητών με σκοπό να επιτύχει τους παιδαγωγικούς του στόχους. Από την άλλη πλευρά, δεν έχει ιδιαίτερο νόημα να προσομοιώνουμε τα «πραγματικά» συναισθήματα ενός παιδαγωγού μέσα από μια εκπαιδευτική εφαρμογή. Βέβαια, το σύστημα πρέπει να εκφράζει έντονα τα συναισθήματα με στόχο να παρακινεί τους μαθητές και με αυτό τον τρόπο να βοηθάει στην μάθηση.

Οι περισσότερες μελετημένες ενδείξεις για το ψέμα ή την εξαπάτηση μέχρι στιγμής είναι ορατές. Στην φωνή, όμως, η παρατήρηση της ομιλίας εξαπάτησης μπορεί να βρεθεί ως ψευδής αιτία συναισθηματικής διέγερσης στον ψεύτη. Αλλά στον λόγο, έχουμε να διαχωρίσουμε μεταξύ ικανών και μη ψευτών [80]. Οι ικανοί ψεύτες μπορούν να κρύψουν τα συναισθήματά τους με μεγάλη ικανότητα, για παράδειγμα όχι μόνο να πείσουν το άτομο στο οποίο ψεύδονται αλλά ακόμη και τον ίδιο τους τον εαυτό για την αλήθεια του ψέματος που λένε. Οι κακοί ψεύτες μπορούν επίσης να επιτύχουν, αλλά η συμπεριφορά τους διαφέρει ανάλογα με το αν οι συνομιλητές τους είναι ενδοτικοί ή υποψιασμένοι. Όταν ψεύδεται σε κάποιον που είναι ενδοτικός, η λεκτική συμπεριφορά είναι συνήθως υψηλά ελεγχόμενη με ακριβή προφορά, βαθύτερη φωνή, λιγότερες επαναλήψεις και διορθώσεις και πολλές παύσεις. Όταν ψεύδεται σε κάποιον που είναι υποψιασμένος, από την άλλη πλευρά, τοποθετεί το ψέμα μέσα σε μια κατάσταση υψηλής συναισθηματικής διέγερσης που χαρακτηρίζεται από υψηλή και ποικιλόμορφη διακύμανση, με πολλές επαναλήψεις και διορθώσεις.

Στον τομέα της επιφάνειας διάδρασης ανθρώπου-υπολογιστή, ιδιαίτερα στον τομέα των ενσωματωμένων ομιλούμενων συντελεστών, τα ψέματα και η εξαπάτηση είναι θέματα που σχεδόν δεν υπάρχουν. Ο Kenzie περιγράφει τους εξαπατούμενους ομιλητές ως εκπαιδευόμενους συνεργάτες στον στρατιωτικό τομέα (check point) [81]. Αλλά δεν παρουσιάζουν έναν εκτελούμενο σύστημα, απλά κάποιες ιδέες για τα θετικά ενός τέτοιου εκπαιδευτικού σεναρίου. Οι Castelfranchi και Poggi εξέλιξαν μια θεωρία εξαπάτησης στην

επικοινωνία η οποία βοηθά στην μοντελοποίηση της εξαπάτησης και της επιλογής της στρατηγικής εξαπάτησης η οποία είναι κατάλληλη για ένα δεδομένο πλαίσιο [82]. Αυτή η θεωρία στηρίχτηκε στην αρχικοποίηση ενός εργαλείου μοντελοποίησης της εξαπάτησης, στο οποίο αυτός που εξαπατά και αυτός που λαμβάνει το μήνυμα έχουν μοντελοποιηθεί [83]. Ο Carofiglio μοντελοποίησε κάποιες στρατηγικές εξαπάτησης με την πεποίθηση ότι τα δίκτυα μπορούν να ενσωματωθούν στους ομιλούμενους συντελεστές. Περισσότερη ανάλυση μπορεί να βρεθεί στον τομέα των ομιλούμενων συστημάτων και των πολυσυντελεστικών συστημάτων όπου διαφορετικές στρατηγικές εξαπάτησης και οι επιπτώσεις τους εξετάζονται [84]. Οι Ward και Hexmoor, όπως και ο Castelfranchi παρουσιάζουν πολυσυντελεστικές προσομοιώσεις ως εξεταστήρια για την διερεύνηση των επιπτώσεων, των θετικών και αρνητικών, από διαφορετικές μορφές εξαπάτησης στην διάδραση των συντελεστών [85][86].

Άλλοι ερευνητές έχουν εστιάσει σε αντίθετο πρόβλημα, που το ονομάζουν 'πώς να πιστοποιηθούν τα αληθινά συναισθήματα του ανθρώπου' (όπως αντιτίθενται σε αυτά που εκφράζονται σκόπιμα). Οι περισσότερες από αυτές τις προσεγγίσεις εστιάζουν στην ανάλυση των βιο-σημάτων τα οποία είναι –σε σύγκριση με τις εξωτερικές ερμηνείες των εκφράσεων– σχεδόν συνειδητά ελεγχόμενα από τον χρήστη και έτσι πολύ επιτηδευμένα. Οι μετρήσεις που έγιναν ξεχωρίζουν μεταξύ των συναισθημάτων και περιλαμβάνουν τον παλμό της καρδιάς, την αγωγιμότητα του δέρματος, την θερμοκρασία του δέρματος, την δραστηριότητα των μυών και τον ρυθμό της αναπνοής. Παραδείγματα γνωστών εφαρμογών είναι οι ανιχνευτές αλήθειας (lie-detectors).

Οι Bosma & Andre περιγράφουν μια προσπάθεια να αποκαλύψουν την πρόθεση του χρήστη από μη σαφείς ενέργειες διαλόγου, ερμηνεύοντας την συναισθηματική κατάσταση του χρήστη [79]. Ιδιαίτερα, οι μικρές προτάσεις τείνουν να είναι αρκετά ασαφείς όταν λαμβάνεται υπ' όψη μόνο οι γλωσσολογικές πληροφορίες. Μία έκφραση όπως η λέξη «tight» μπορεί να ερμηνευθεί ως επιβεβαίωση και την ίδια στιγμή σαν αποδοκιμασία, εάν προφερθεί κυνικά ή με την έλλειψη προφοράς.

2.3.3 Χειρισμός λανθασμένων συναισθηματικών εκφράσεων από τον χρήστη

Η αναγνώριση του ψέματος ή άλλων μορφών εξαπάτησης του χρήστη, απαιτεί την επεξεργασία των σημάτων έκφρασης και την ερμηνεία τους σύμφωνα με το κλίμα που επικρατεί την ώρα που γίνεται η επικοινωνία [78]. Τα γνωστικά μοντέλα του χρήστη και η αναπαράσταση ιδιοτήτων των συμφραζομένων βοηθά στην μείωση του υψηλού επιπέδου ανακρίβειας σε αυτή την διαδικασία αναγνώρισης, παρέχοντας μια απάντηση στις ερωτήσεις του τύπου : 'Είναι η έκφραση του χρήστη αληθοφανής, αξιόπιστη, λογική και σύμφωνη με την

προηγούμενή του/της συμπεριφορά;’ Η ‘Έχει ο χρήστης κανένα όφελος να μην πει την αλήθεια, στην συγκεκριμένη κατάσταση;’

Η ακουστική ανάλυση του λόγου του χρήστη μπορεί να υποστηρίξει αμέσως την αυτόματη αναγνώριση της εξαπάτησης. Επίσης, αυτό δε μπορεί να τύχει μεγάλης επιστημονικής προσοχής ακόμη (εκτός από τις προσεγγίσεις της συναισθηματικής αναγνώρισης που χρησιμοποιούν ως υλικό τον λόγο ηθοποιών που προσποιούνται τα συναισθήματα πράγμα το οποίο εμπεριέχει ένα είδος εξαπάτησης). Η αυτόματη ανίχνευση του ψεύδους μπορεί να βασιστεί στην αναγνώριση του συναισθήματος που τονίζεται, που μπορεί να είναι για παράδειγμα ο φόβος να εντοπιστεί ο ψεύτης ή μια κατάσταση υψηλού στρες. Η αυτόματη εξαγωγή των χαρακτηριστικών της προσωδίας του ψεύδους χρειάζεται όμως ακόμη χρόνο για να επιτευχθεί [87].

Επειδή τα κοινωνικά ψέματα χρησιμοποιούνται σε συγκεκριμένες καταστάσεις με σκοπό την προστασία άλλων ατόμων ή άλλων σχέσεων, η αναγνώριση της εκάστοτε κατάστασης βοηθά στον εντοπισμό λανθασμένων συναισθηματικών εκφράσεων από τον χρήστη. Ανάλογα με το γενικό πλαίσιο της διάδρασης, το σύστημα πρέπει να αντιδρά στην λανθασμένη συναισθηματική αντίδραση του χρήστη. Στις περιπτώσεις που είναι πιο επιτόλαιες, όπου ο χρήστης είναι απλός προς το σύστημα και χρησιμοποιεί ένα λευκό ψέμα, το σύστημα μπορεί να επιλέξει να αγνοήσει αυτό το ψέμα, ή οι πληροφορίες για το ψέμα του χρήστη μπορούν να χρησιμοποιηθούν έτσι ώστε να αναβαθμιστεί το μοντέλο του χρήστη προωθώντας θετικά συναισθήματα στην πλευρά του συστήματος επειδή τον εξέλαβε σοβαρά ως έναν συνομιλητή. Σε ορθολογιστικές καταστάσεις από την άλλη πλευρά όπως για παράδειγμα στην ιατρική συμβουλή, μπορεί να είναι επικίνδυνο να αγνοήσει το σύστημα τις λανθασμένες δηλώσεις που είχε καταλάβει σωστά. Μια πιο λεπτομερής και πιθανόν απλή εξήγηση μπορεί να είναι η σωστή αντίδραση γι’ αυτή την κατάσταση.

Οι Oliviero και τα μέλη του WP8 στην εργασία τους Report on Basic Cues and Open Research Topics in Communication and Emotions δίνουν έμφαση στην ισχυρή εξάρτηση της αντίδρασης στα ψέματα του χρήστη [78]. Έτσι, μια από τις μεγαλύτερες προκλήσεις απαρτίζεται από την αναβάθμιση του γενικού πλαισίου των αναπαραστάσεων και των γνωστικών μοντέλων του χρήστη έτσι ώστε να εκτιμούν την κατάσταση και να την χειρίζονται με μια κατάλληλη αντίδραση όπως είναι η άγνοια, η αναβάθμιση του μοντέλου του χρήστη ή η επεξεργασία.

2.4 ΠΑΓΚΟΣΜΙΟΤΗΤΑ ΚΑΙ ΠΟΛΙΤΙΣΜΙΚΗ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΗΝ ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ

Ψυχολόγοι, έχουν θέσει το ερώτημα, αν τα συναισθήματα είναι παγκόσμια ή αν ποικίλουν ανά πολιτισμό. Αν και υπάρχουν πολλοί θεωρητικοί που έχουν λάβει ακραίες θέσεις και έχουν προκαλέσει ζωνρότατες αντιπαραθέσεις υπάρχουν πρόσφατα θεωρητικά μοντέλα που έχουν αποπειραθεί να αναλύσουν και την παγκοσμιότητα αλλά και την πολιτιστική διαφοροποίηση που υπάρχει με το να συγκεκριμενοποιούν ειδικές όψεις των συναισθημάτων και να δείξουν ομοιότητες και διαφορές σε όλα τα επίπεδα πολιτισμών [88][89][90][91][92][93][94]. Οι Mesquita & Frijda ισχυρίζονται ότι αν και τα συναισθήματά μας είναι βιολογικά προγραμματισμένα, η διαδικασία της εκμάθησης ελέγχου τόσο της έκφρασης αλλά και της αντίληψης τους, εξαρτάται σε υψηλό βαθμό από πολιτισμικούς παράγοντες [91][92]. Πρέπει να αναφερθεί ότι η συναισθηματική έκφραση ανήκει στην κατηγορία της συναισθηματικής συμπεριφοράς. Οι εξωτερικές εκφράσεις των ανθρώπων συνοδεύονται τις πιο πολλές φορές από συναισθηματικές εμπειρίες. Επειδή αυτές οι συναισθηματικές συμπεριφορές είναι δυνατόν να εντοπιστούν, είναι στην διάθεσή μας για περαιτέρω έρευνα ιδιαίτερα δε από επιστήμονες οι οποίοι έχουν εξετάσει και τις ίδιες τις εκφράσεις αλλά και την ακριβή αναγνώρισή τους. Σαν συνέπεια, υπάρχει πλούσιο υλικό στην διαπολιτισμική συναισθηματική αναγνώριση που μπορεί εμπειρικά να ερευνηθεί με σκοπό να γίνει καλύτερα κατανοητή η παγκοσμιότητα και η πολιτισμική συγκεκριμενοποίηση της συναισθηματικής αναγνώρισης.

Πολλές επεξηγήσεις σχετικά με την πολιτισμική παγκοσμιότητα και διαφοροποίηση έχουν προωθηθεί με απόλυτη ακρίβεια και υπάρχουν πολλοί παράγοντες που κάνουν προσπάθεια να εξετάσουν το συγκεκριμένο θέμα. Ο πρώτος παράγοντας αναφέρεται στην παρουσίαση και αποκωδικοποίηση κανόνων που έχουν σχέση με την συναισθηματική αναγνώριση και προτείνει ότι η πολιτισμική διαφοροποίηση, η έκθεση και αποκωδικοποίηση κανόνων χρησιμοποιούνται σαν εφαρμογές για την καταλληλότητα της αποτύπωσης και κρίσης συναισθημάτων [88][95]. Δεύτερος παράγοντας είναι η γλώσσα. Η πολιτισμική ποικιλομορφία για την ακρίβεια της συναισθηματικής αναγνώρισης έχει επίσης αποδοθεί και στις διαφορές που εμφανίζονται στην γλώσσα. Οι λέξεις που χρησιμοποιούνται για να μεταφραστούν συναισθηματικές ιδέες και ταμπέλες μπορεί να υπονοούν διαφορετικές έννοιες σε κάποιες κουλτούρες οι οποίες θα είναι διαφορετικές μεταξύ τους [92]. Η ποικιλομορφία στην διαπολιτισμική συναισθηματική αναγνώριση αποδίδεται στην ομοιομορφία που παρουσιάζει μια ομάδα. Είναι ευνόητο ότι η ακρίβεια της αναγνώρισης συναισθημάτων είναι

υψηλότερη όταν τα συναισθήματα εκφράζονται και γίνονται αντιληπτά από τα μέλη της ίδιας κοινωνικής και πολιτισμικής ομάδας.

2.4.1 Διαφορές στην έκφραση μεταξύ των συναισθημάτων

Έντονες συζητήσεις στην γενικότερη συναισθηματική φιλολογία έχουν γίνει για το ποια συγκεκριμένα συναισθήματα μπορούν παγκοσμίως να αναγνωριστούν. Για παράδειγμα, έχει υπάρξει κάποια αντιπαλότητα για το αν η περιφρόνηση θα πρέπει να συμπεριληφθεί με τα έξι βασικά συναισθήματα –θυμό, απέχθεια, φόβος, ευτυχία, λύπη και έκπληξη- όπως έχουν αναγνωριστεί από τον Ekman [88]. Επιπλέον, ο Rusell ισχυρίζεται ότι ευρύτερες διαστάσεις όπως είναι το σθένος και η διέγερση, είναι παγκόσμια συναισθήματα ενώ ειδικές κατηγορίες όπως θυμός ή έκπληξη μπορούν να διαφέρουν από πολιτισμό σε πολιτισμό [96]. Αυτή η άποψη δεν έχει αμφισβητηθεί, και γι' αυτό το λόγο μπορεί να είναι χρήσιμο να εξετάσουμε την πολιτισμική ποικιλομορφία που θα συσχετίζεται με την αναγνώριση και των ευρύτερων διαστάσεων των θετικών και αρνητικών συναισθημάτων καθώς και συγκεκριμένων διακριτικών συναισθημάτων [97] [98].

Υπάρχουν λοιπόν ενδιαφέρων συντελεστές στην διαπολιτισμική συναισθηματική αναγνώριση, που χρησιμοποιούνται στην γενικότερη συναισθηματική φιλολογία [91] [92] [93] [94]. Αυτοί συμπεριλαμβάνουν:

- Την έκθεση ανάμεσα σε διάφορους πολιτισμούς
- Την θέση των μειονοτήτων και των πλειονοτήτων σε έναν πολιτισμό
- Συμπεριφορά των ερευνών
- Άλλοι δημογραφικοί παράγοντες

Σε ότι αφορά την έκθεση ανάμεσα σε διάφορους πολιτισμούς, είναι πολύ πιθανό ότι η ακρίβεια συναισθηματικής αναγνώρισης μπορεί να μετριαστεί από το πόσο έχει εκτεθεί κάποιος στους κανόνες και τους στόχους κάποιας κουλτούρας. Υπάρχουν σημαντικές εικασίες ότι το είδος συναισθηματικής έκφρασης μπορεί να διαφέρει στις διαπολιτισμικές ομάδες και ότι τα άτομα μεμονωμένα είναι δυνατόν να καταλαβαίνουν καλύτερα τα συναισθήματα που εκφράζονται με τον τρόπο που οι ίδιοι έχουν συνηθίσει. Μεγαλύτερη και καλύτερη επαφή και οικειότητα μπορούν να συσχετιστούν με μεγαλύτερη ακρίβεια στην αναγνώριση συναισθημάτων ανάμεσα στις διάφορες κουλτούρες.

Υπάρχει μακροχρόνια διαμάχη για το αν τα μέλη μιας πλειονότητας μπορούν με μεγαλύτερη ακρίβεια να κρίνουν τα συναισθήματα αλλά και την συμπεριφορά των μελών κάποιας μειοψηφικής ομάδας αλλά και το αντίστροφο. Ομάδες μειονοτήτων και πλειονοτήτων μπορούν να διαφέρουν εξαιτίας του βαθμού έκθεσης που έχουν η μια στην άλλη, βασιζόμενοι

στον αριθμό των μελών αλλά και στις ατομικές εμπειρίες από την μεγαλύτερη έκθεση των μελών της πλειοψηφικής ομάδας.

Υπάρχουν πολλές απόψεις για την μεθοδολογία μελέτης συναισθηματικής αναγνώρισης [97] [98] [95]. Το αντίκτυπο διαφόρων παραγόντων στο σχεδιασμό έρευνας των ερεθισμάτων σε επίπεδο διαπολιτισμικής ακρίβειας αλλά και το πλεονέκτημα που υπάρχει να ανήκει κανείς σε μια ομάδα, αποτελούν παράγοντες έρευνας για την αναγνώριση συναισθημάτων.

Τέλος, σε ότι αφορά τους άλλους δημογραφικούς παράγοντες υπάρχει η ηλικία των συμμετεχόντων, το φύλο για την διαπολιτισμική ακρίβεια και το πλεονέκτημα της συμμετοχής σε κάποια ομάδα, για την συναισθηματική αναγνώριση.

3. ΠΡΟΣΕΓΓΙΣΕΙΣ ΑΝΑΓΝΩΡΙΣΗΣ ΣΥΝΑΙΣΘΗΜΑΤΩΝ

Η αναγνώριση των συναισθημάτων είναι ένας τομέας προς διερεύνηση τα τελευταία χρόνια. Οι μελέτες που έχουν γίνει πάνω σε αυτόν το ερευνητικό τομέα ποικίλουν ως προς την βάση που χρησιμοποιούν για την διερεύνησή του. Ο μεγαλύτερος κλάδος που έχει αναλυθεί είναι αυτός της αναγνώρισης μέσω της φωνής. Λιγότερη προσοχή έχει δοθεί σε ότι αφορά την αναγνώριση συναισθημάτων μέσω των εκφράσεων του προσώπου, ενώ τα τελευταία χρόνια γίνονται προσπάθειες για συνδυαστική λειτουργία των δύο μεθόδων. Παρακάτω, γίνεται ανάλυση των μεθόδων αυτών και των προσεγγίσεων που έχουν δοθεί μέχρι στιγμής.

3.1 ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΑΠΟ ΚΕΙΜΕΝΟ

Η αναγνώριση συναισθημάτων (ER) από έγγραφα κειμένων είναι ουσιαστικά ένα πρόβλημα ταξινόμησης βασισμένο στο περιεχόμενο που περιλαμβάνει έννοιες από περιοχές της επεξεργασίας φυσικής γλώσσας (Natural Language Processing NLP) καθώς επίσης και της εκμάθησης μηχανών (Machine Learning ML). **Διάφορες τεχνικές από την εκμάθηση μηχανών** έχουν υιοθετηθεί για την επίτευξη της αναγνώρισης συναισθημάτων στα έγγραφα κειμένων, όπως τα δέντρα απόφασης, οι Naïve Bayesian ταξινομητές, οι διανυσματικές μηχανές υποστήριξης SVM και τα νευρωνικά δίκτυα. Εντούτοις, ένα σημαντικό μέρος του αλγορίθμου ταξινόμησης είναι η **κατάλληλη αναπαράσταση των εγγράφων κειμένων** για την **εισαγωγή σε αυτούς τους ταξινομητές**.

Υπάρχουν δύο σημαντικά βήματα για την επίτευξη της αναγνώρισης συναισθημάτων στα έγγραφα κειμένων μέσω των τεχνικών της NLP και ML.

1. Το **πρώτο βήμα**, δηλαδή η επεξεργασία του εγγράφου κειμένου για τη μετατροπή του σε μια κατάλληλη μορφή ώστε να μπορεί να εισαχθεί σε ένα αλγόριθμο ταξινόμησης αντιμετωπίζεται από την NLP.

2. **Το δεύτερο βήμα**, που είναι η ανάπτυξη και η εκπαίδευση των μοντέλων ταξινόμησης αντιμετωπίζεται από το ML.

Κάθε ένας από τους ταξινομητές που έχουν δοκιμαστεί μέχρι τώρα απαιτεί το έγγραφο κειμένου να **αναπαρίσταται υπό μορφή διανύσματος χαρακτηριστικών** (feature vector). Συνήθως, το μοντέλο **Σάκκος των λέξεων (Bag of Words BoW)** και οι **τιμές TF-IDF (term frequency - inverse document frequency)** χρησιμοποιούνται για αυτόν το λόγο. Εντούτοις, η επιλογή των χαρακτηριστικών αποτελεί ένα σημαντικό τομέα έρευνας που προσπαθεί να απομονώσει μόνο τη χρήσιμη πληροφορία από τα γλωσσικά συστατικά του κειμένου και να απορρίψει τα περιττά στοιχεία [99][100][101][102][103][111][112].

3.1.2 Επιλογή Χαρακτηριστικών και Μοντέλα Αναπαράστασης Εγγράφων

Το ζήτημα των χαρακτηριστικών γνωρισμάτων περιλαμβάνει δύο σημαντικά ζητήματα:

Το πρώτο είναι η εξαγωγή μόνο εκείνων των τμημάτων του εγγράφου που είναι σχετικά με την ταξινόμηση, ενώ το δεύτερο εξετάζει την ανάθεση των κατάλληλων βαρών σε αυτά τα χαρακτηριστικά ώστε να απεικονιστεί ο βαθμός σχετικότητας για κάθε ένα από αυτά. Αυτά τα βάρη πρέπει να δώσουν τα ακόλουθα μέτρα:

(α) ένα μέτρο για το πόσο σχετικό είναι το χαρακτηριστικό μέσα σε ένα δεδομένο έγγραφο με σεβασμό προς τα άλλα χαρακτηριστικά του ίδιου εγγράφου και

(β) ένα μέτρο πόσου σχετικού είναι το χαρακτηριστικό γνώρισμα για το δεδομένο έγγραφο όσον αφορά άλλα έγγραφα στην ίδια βάση δεδομένων. Η TF-IDF τιμή φαίνεται να συλλαμβάνει αυτές τις δύο πληροφορίες ικανοποιητικά από την άποψη της συχνότητας του χαρακτηριστικού γνωρίσματος.

3.1.3 TF-IDF score

- **TF or term frequency:** Αναφέρεται στη συχνότητα εμφάνισης ενός όρου μέσα στο ίδιο έγγραφο. Υπολογίζεται ως εξής:

$$tf_{i,j} = n_{i,j} / \sum_k n_{k,j} \quad (1)$$

Το $tf_{i,j}$ αναπαριστά τη συχνότητα εμφάνισης του όρου i στο έγγραφο j , το $n_{i,j}$ αναφέρεται στο πόσες φορές ο όρος i εμφανίζεται στο έγγραφο j και ο παρονομαστής μας δίνει το σύνολο των όρων στο έγγραφο j .

- **IDF or inverse document frequency:** Αυτό δίνει ένα μέτρο για το πόσα έγγραφα περιέχουν ένα συγκεκριμένο όρο σε σχέση με το σύνολο των εγγράφων της βάσης. Υπολογίζεται ως εξής:

$$idf_{i,j} = \log(|D| / |\{d : t_i \in d\}|) \quad (2)$$

Το idf_i αναπαριστά την αντίστροφη συχνότητα εμφάνισης του όρου i στο έγγραφο j , το $|D|$ είναι το σύνολο των εγγράφων στη βάση και ο παρονομαστής το σύνολο των εγγράφων της βάσης που περιέχουν τον όρο i .

Η συνολική **TF-IDF** τιμή τότε υπολογίζεται ως προϊόν των παραπάνω δύο τιμών:

$$tf-idf_{i,j} = tf_{i,j} \times idf_i \quad (3)$$

Το TF μέρος δίνει πληροφορία για τη σχετικότητα ενός όρου μέσα σε ένα έγγραφο και το IDF μέρος ανεβάζει ή κατεβάζει την τιμή των βαρών σε συνάρτηση με τη σχετικότητα σε ολόκληρη τη βάση. Αυτή η τιμή επίσης εκφράζει πόσο συγκεκριμένος είναι ένας όρος για ένα έγγραφο σε σχέση με ολόκληρη τη βάση [102][104][105][106][107][108][109][110].

3.1.4 Bag of words (BoW)

Ο σάκκος των λέξεων είναι η πιο βασική μορφή όλων των προτύπων αντιπροσώπευσης εγγράφων. Θεωρεί κάθε έγγραφο ως άτακτο σύνολο λέξεων που εμφανίζονται στο κείμενο. Αυτή η προσέγγιση δεν λαμβάνει υπόψη οποιαδήποτε σημασιολογία του κειμένου και εξετάζει τις λέξεις ως ανεξάρτητες οντότητες. Η πληροφορία που είναι σχετική με την ακολουθία ή/και τη γραμματική των λέξεων χάνεται εντελώς. Εντούτοις, εάν η επιλογή χαρακτηριστικών γνωρισμάτων γίνει κατάλληλα, βρίσκεται να αποδίδει αρκετά καλά με μερικούς ταξινομητές όπως τον Naïve Bayesian.

Το έγγραφο αναπαριστάται ως ένα διάνυσμα χαρακτηριστικών (Feature Vector), όπως βλέπουμε παρακάτω. Θεωρούμαι ότι το λεξικό L είναι μια συλλογή από n όρους-λέξεις:

$$L = \{t_1, t_2, \dots, t_n\}$$

Ένα έγγραφο κειμένου i της βάσης δίνεται σαν διάνυσμα με κάθε διάστασή του να αποτελεί ένα μοναδικό χαρακτηριστικό όπως βλέπουμε παρακάτω:

$$d_i = \langle w_{1i}, w_{2i}, \dots, w_{ni} \rangle$$

όπου w_{ki} είναι το βάρος της λέξης k που σχετίζεται με το έγγραφο i το οποίο συνήθως είναι το $tf-idf$ score. Αν η λέξη k εμφανίζεται στο έγγραφο i , τότε το w_{ki} είναι μη μηδενική τιμή, αλλιώς είναι μηδέν.

Ο **πίνακας λέξεων-εγγράφων** (**term – document matrix**) ορίζεται για όλα τα έγγραφα της βάσης, όπου κάθε γραμμή του είναι το διάνυσμα χαρακτηριστικών ενός συγκεκριμένου εγγράφου (d_i), και κάθε στήλη του είναι ένας μοναδικός όρος-λέξη της βάσης. Ο πίνακας λέξεων-εγγράφων είναι γενικά πολύ μεγάλος και έχει πολλά μηδενικά. Τυπικά το μέγεθος του μπορεί να είναι μερικές χιλιάδες και για τις δύο διαστάσεις. Η αφαίρεση λέξεων με την τεχνική stemming κατά τη διάρκεια της μετατροπής ενός εγγράφου σε BoW οδηγεί σε

αφαίρεση άχρηστης πληροφορίας και στη βελτίωση των αποτελεσμάτων. Η τεχνική **stemming** λαμβάνει υπόψιν ότι κάθε λέξη μπορεί να χρησιμοποιείται σε διαφορετικούς τύπους (ουσιαστικό, επίθετο, επίρρημα, ενικός, πληθυντικός, κτλ) και γι αυτό μετατρέπει όλες τις λέξεις στη βασική μορφή. Επίσης κόβει λέξεις όπως (a, the, and, and many more) οι οποίες δεν παρέχουν χρήσιμη πληροφορία για το έγγραφο, και έτσι ελαττώνει σημαντικά το μέγεθος του πίνακα λέξεων-εγγράφων και βελτιώνει την επάρκεια στους υπολογισμούς [113][114][115][116][117].

3.1.5 Προσέγγιση Επιλογής Χαρακτηριστικών – POS Vs WordNet Database

Εξετάζοντας ιδιαίτερα την περίπτωση της ταξινόμησης εγγράφων σε κατηγορίες συναισθημάτων, βλέπουμε ότι υπάρχουν πολλές λέξεις που δεν συμβάλλουν σε οποιαδήποτε μορφή συναισθήματος όπως ορισμένα ουσιαστικά (αντικείμενα και υποκείμενα), αντωνυμίες, συνδετικά, κ.λπ. Το κόστος υπολογισμού μπορεί να μειωθεί αρκετά εάν αυτές οι άχρηστες λέξεις αφαιρεθούν στο βήμα προεπεξεργασίας εγγράφων. Αυτό μπορεί να γίνει με την εξαγωγή μόνο των σχετικών μερών του λόγου (POS) από τις προτάσεις του εγγράφου. Τώρα, ποιο POS μπορεί καλύτερα να αντιπροσωπεύσει το συναίσθημα σε ένα έγγραφο κειμένων: τα επίθετα και τα επιρρήματα είναι πιθανότερο να καθορίσουν το συναίσθημα από άλλα μέρη του λόγου όπως τα ουσιαστικά ή το ρήματα, όπως στην πρόταση “κουβέντιασαν ευτυχισμένοι για πολύ”. Εντούτοις, σε μερικές περιπτώσεις τα επίθετα ή/και τα επιρρήματα μπορούν να φέρουν άχρηστες πληροφορίες και αφήνουν έξω τις σημαντικές λέξεις. Ο συνυπολογισμός των ουσιαστικών και των ρημάτων για την ταξινόμηση περιλαμβάνει πολλές άχρηστες πληροφορίες και κατά συνέπεια, η προσέγγιση POS για την εξαγωγή χαρακτηριστικών δεν φαίνεται να αποδίδει αρκετά. Μια άλλη προσέγγιση που χρησιμοποιεί τη βάση δεδομένων WordNet μπορεί να εφαρμοστεί για την εξαγωγή των χρήσιμων πληροφοριών από το έγγραφο και την απόρριψη αυτών που είναι άσχετες για την ταξινόμηση στις κατηγορίες συναισθημάτων.

Η βάση δεδομένων WordNet είναι μια ελεύθερα και δημόσια - διαθέσιμη βάση δεδομένων αγγλικών λέξεων, η οποία ομαδοποιεί τις λέξεις βασισμένες στη λεξικολογική και σημασιολογική σχέση. Οι λέξεις ομαδοποιούνται σε ομάδες (synsets), κάθε μια από τις οποίες καθορίζει ένα συναίσθημα. Από το WordNet λεξικό λαμβάνεται ένας κατάλογος λέξεων (με μορφή ρίζας) που συσχετίζεται με μια δεδομένη κατηγορία συγκίνησης. Ένα πλήρες σύνολο καταλόγων από αυτόν τον κατάλογο δημιουργείται ενσωματώνοντας τα συνώνυμα κάθε μιας από αυτές τις λέξεις-ρίζες (μέσω των synsets). Κατά τη διάρκεια του βήματος προεπεξεργασίας εγγράφων, κάθε λέξη στη βάση συγκρίνεται με αυτούς τους καταλόγους και

μόνο εκείνες οι λέξεις που είναι παρούσες σε αυτούς τους καταλόγους συμπεριλαμβάνονται στην αντιπροσώπευση BoW του δεδομένου εγγράφου. Όλες οι άλλες λέξεις του εγγράφου απορρίπτονται. Αυτό οδηγεί στη δημιουργία των προτύπων εγγράφων που είναι τα πλέον κατάλληλα για την ταξινόμηση εγγράφων σε κατηγορίες συναισθημάτων [118][119][120].

3.1.6 Μελλοντική Έρευνα -Ζητήματα σε emotion – based text ταξινόμηση

Εκτός από το πληροφοριακό ή/και το επεξηγηματικό περιεχόμενο, πολλά από τα έγγραφα κειμένων περιέχουν κάποιες συναισθηματικές πληροφορίες. Η ταξινόμηση των εγγράφων βασισμένων σε αυτήν την πληροφορία μπορεί να εξυπηρετήσει σημαντικές λειτουργίες σε πολλές περιοχές εφαρμογών. Παρόλα αυτά, η ταξινόμηση συναισθημάτων είναι πολύ διαφορετική από την απλή ταξινόμηση, γιατί δεν είναι μια απλή σύγκριση του περιεχομένου του εγγράφου με ένα σύνολο λέξεων κλειδιών ή φράσεων.

Παρακάτω είναι μερικά από τα ζητήματα που κάνουν την συναισθηματική ταξινόμηση των εγγράφων ένα πολύ δύσκολο πρόβλημα και κατά συνέπεια δίνουν κίνητρα για μελλοντική έρευνα:

- Δεν υπάρχει κανένα συγκεκριμένο σύνολο λέξεων που μπορεί εντελώς να καθορίσει ένα συγκεκριμένο συναίσθημα.
- Δεδομένου ενός συνόλου λέξεων, περισσότερα από ένα συναισθήματα μπορούν να εξαχθούν από τις λέξεις `happy και `excited, όπως το συναίσθημα της ευτυχίας και της έκπληξης.
- Πολλές φορές, ένα ορισμένο σύνολο λέξεων μπορεί να μην περιέχει συναισθηματική πληροφορία, αλλά όταν τοποθετείται σε πρόταση και στο περιεχόμενο του εγγράφου, δηλώνει κάποιο συναίσθημα. Παράδειγμα: “δεν καταλαβαίνω γιατί οι άνθρωποι λένε ψέματα!” «Δεν υπάρχει καμία συναισθηματική λέξη σε αυτήν την πρόταση εντούτοις, ένα συναίσθημα θλίψης είναι σαφές από την πρόταση. Κατά συνέπεια η σημασιολογία μιας δεδομένης πρότασης συνολικά είναι πιο σημαντική από τις μεμονωμένες λέξεις [121][122][123][124].
- Παρουσία ασάφειας - μια δεδομένη πρόταση μπορεί να δηλώνει περισσότερα από ένα συναισθήματα. Για παράδειγμα η παραπάνω πρόταση μπορεί να δηλώνει και ένα συναίσθημα θυμού. Τα συμφραζόμενα της πρότασης διαδραματίζουν έναν σημαντικό ρόλο εδώ, παρά τη σημασιολογία μιας πρότασης.
- Ένα έγγραφο μπορεί να περιέχει περισσότερα από ένα συναισθήματα.

- Μπορεί να μας μπερδέψει η άρνηση κάποιου συναισθήματος όπως η φράση “Δεν μ αρέσει αυτό” ή “δεν είμαι λυπημένος”.
- Ιδιαίτερα σε μεγάλα έγγραφα κειμένων, διαφορετικά μέρη του εγγράφου μπορεί να περιέχουν διαφορετικό συναίσθημα.
- Η ταξινόμηση σε μια προκαθορισμένη κατηγορία συναισθήματος μπορεί να μην είναι εφικτή σε ορισμένες περιπτώσεις.

3.2 Εναλλακτικές τεχνικές υπολογισμού γλωσσικών μοντέλων

3.2.1 Μοντέλα Παράλειψης (Skipping Models)

Ουσιαστικά είναι απίθανο μια αλληλουχία πέντε λέξεων που εμφανίζεται στα δεδομένα εκπαίδευσης να συμβεί και στα πειραματικά δεδομένα. Παρόλα αυτά είναι λιγότερο απίθανο παρόμοιες αλληλουχίες ίδιων λέξεων να εμφανιστούν. Ειδικότερα, με την χρήση των μοντέλων παράλειψης γίνεται προσπάθεια να υπολογισθούν οι πιθανότητες του τύπου

$$P(w_i | w_{i-4}w_{i-3}w_{i-1}) \text{ και } P(w_i | w_{i-4}w_{i-2}w_{i-1}). \quad (4)$$

Όπως γίνεται κατανοητό σε μια σειρά πέντε λέξεων οι δυνατοί συνδυασμοί που μπορούν να επιτευχθούν στο πλαίσιο της τεχνικής παράλειψης, είναι πολλοί.

3.2.2 Μοντέλα Ομαδοποίησης (Clustering Models)

Ας θεωρήσουμε την πιθανότητα ενός τριγράμματος $P(\text{Thursday} | \text{party on})$. Πιθανόν τα δεδομένα εκπαίδευσης δεν περιλαμβάνουν την φράση “party on Thursday”, αν και άλλες φράσεις του τύπου “party on Friday” μπορεί να περιλαμβάνονται. Συνεπώς αν βάλουμε ομάδες λέξεων και τοποθετήσουμε την λέξη Tuesday στην ομάδα λέξεων Weekday μπορούμε να υπολογίσουμε την πιθανότητα αποσυνθέτοντας την ως εξής:

$$\begin{aligned} p(\text{Thursday} | \text{party on}) \\ = p(\text{Weekday} | \text{party on}) \times p(\text{Thursday} | \text{party on Weekday}) \end{aligned} \quad (5)$$

3.2.3 Μοντέλα Μνήμης (Caching Models)

Εάν θεωρηθεί ότι ένας ομιλητής χρησιμοποιεί μια λέξη είναι πιθανόν ότι αυτή τη λέξη θα την χρησιμοποιήσει και στο μέλλον. Σύμφωνα με τους Kuhn, De Mori, Kuriec και Jelinek αυτή η παρατήρηση είναι η αρχή των μοντέλων μνήμης. Ειδικότερα, στην περίπτωση ενός πίνακα μνήμης μονογραμμάτων, σχηματίζεται ένα μοντέλο μονογραμμάτων από τις πιο πρόσφατες λέξεις (αυτές που συνοδεύουν ένα συγκεκριμένο άρθρο ή αν δεν υπάρχει κάτι τέτοιο ένα συγκεκριμένο αριθμό από προγενέστερες λέξεις). Αυτό το μοντέλο μνήμης μονογραμμάτων μπορεί να παρεμβληθεί γραμμικά με ένα συμβατικό N-gram. Επίσης μπορεί

να χρησιμοποιηθούν και άλλα είδη από μνήμες, παράδειγμα ένα τρίγραμμα εξομάλυνσης μπορεί να σχηματιστεί από προηγούμενες λέξεις και να παρεμβληθεί [124][125][126][127].

3.2.4 Λέξεις πρόκλησης (Word triggers)

Μια γενίκευση της ιδέας των μοντέλων μνήμης για την συσχέτιση 2 διαφορετικών λέξεων οδηγεί σε μια νέα περιοχή έρευνας που ονομάζεται λέξεις πρόκλησης (Rosenfeld 1996, Beeferman et al. 1997). Επί της αρχής, οι σχέσεις μεταξύ οποιουδήποτε ζευγαριού λέξεων ή φράσεων μπορεί να μοντελοποιηθεί. Δυστυχώς, οι υπολογιστικές απαιτήσεις για την εκπαίδευση τέτοιου μοντέλου αυξάνει υπέρ-γραμμικά σε σχέση με τον αριθμό των ανεξάρτητων μοντελοποιημένων ζευγαριών λέξεων πρόκλησης, κάνοντας ακόμη και αποτρεπτική την χρήση ενός μικρού αριθμού ζευγαριών [128][129].

3.2.5 Μοντέλα μίξης προτάσεων (Sentence mixture models)

Οι Iyer και Ostendorf παρατήρησαν ότι μέσα σε ένα σώμα κειμένων υπάρχουν διαφορετικοί τύποι προτάσεων που μπορεί να ομαδοποιηθούν βάσει του θέματος τους, του ύφους τους και άλλων κριτηρίων [130]. Ας υποθέσουμε ότι στα δεδομένα της Wall Street Journal υπάρχουν τρία διαφορετικά είδη προτάσεων, π.χ προτάσεις οικονομικού ενδιαφέροντος (με λέξεις αριθμούς και ονόματα μετοχών), προτάσεις επιχειρηματικού ενδιαφέροντος (με λέξεις συνενώσεις εταιρειών, τρόποι προβολής) και προτάσεις με διάφορες ιστορίες. Έτσι μπορεί να υπολογισθεί η πιθανότητα μιας πρότασης βάσει κάθε ενός τύπου πρότασης και μετά να ληφθεί ένα άθροισμα με διαφορετικά βάρη για κάθε μια από αυτές της πιθανότητες.

3.2.6 Προσαρμογή γλωσσικού μοντέλου

Σε γενικές γραμμές η τεχνική των N-grams δίνει καλά αποτελέσματα όταν παρέχεται ικανοποιητικός όγκος δεδομένων εκπαίδευσης σε ένα συγκεκριμένο αντικείμενο. Όταν τα δεδομένα εκπαίδευσης είναι διαφορετικά από τα δεδομένα δοκιμής τότε τα αποτελέσματα δεν είναι ικανοποιητικά και η απόδοση της τεχνικής υστερεί. Για να αντιμετωπιστεί το πρόβλημα αυτό έχουν γίνει πολλές προσπάθειες να χρησιμοποιηθούν μείγματα γλωσσικών μοντέλων ώστε να βελτιωθεί η απόδοση τους. Μια τυπική μέθοδος μίξης των γλωσσικών μοντέλων είναι αυτή που απορρέει από τον συνδυασμό γλωσσικών μοντέλων που έχουν εκπαιδευτεί σε ένα μικρό κομμάτι δεδομένων εκπαίδευσης, συγκεκριμένου αντικειμένου με γλωσσικά μοντέλα που βασίζονται σε μεγαλύτερα και λιγότερο εξειδικευμένα δεδομένα εκπαίδευσης. Για παράδειγμα ένα γλωσσικό μοντέλο που βασίζεται σε μεγάλο όγκο δεδομένων από εφημερίδες μπορεί να παρεμβληθεί με γλωσσικά μοντέλα που έχουν δομηθεί πάνω σε λιγοστά δεδομένα

από κείμενα ραδιοφωνικής και τηλεοπτικής ειδησεογραφίας. Τα βάρη που θα χρησιμοποιηθούν στην παρεμβολή μπορούν να υπολογισθούν με την χρήση κάποιων δεδομένων, ή αλλιώς μπορούν να προσαρμοστούν δυναμικά σαν συνάρτηση της τρέχουσας ιστορίας. Η εφαρμογή της τρέχουσας ιστορίας εμφανίζεται και σε τεχνικές βελτίωσης των απλών γλωσσικών μοντέλων όπως αυτό το μοντέλο μνήμης που περιγράφεται παραπάνω [131][132].

3.3 Τεχνικές στατιστικής μοντελοποίησης της γλώσσας με την χρήση γραμματικής πληροφορίας

3.3.1 Μοντέλα βασισμένα σε μέρη του λόγου (POS)

Σε ένα μοντέλο N-gram, το λεξικό είναι μια λίστα από λέξεις που ανήκουν σε πάνω από μια κατηγορίες. Το σίγουρο είναι ότι σε μια γλώσσα οι λέξεις σχηματίζουν σύνθετες και πολλές φορές μη κατανοητές λεκτικές σχέσεις. Όπως καταλαβαίνει κανείς η λέξη “THUESDAY” είναι πιο κοντά στην λέξη “WEDNESDAY” παρά στην λέξη “CHAIR”.

Η πιο απλή και αποτελεσματική μέθοδος για να αναδειχθεί η λεξική σχέση των λέξεων είναι η χρήση της πληροφορίας του παρέχεται με την ιδιότητα του Μέρους του Λόγου. Άρα η πληροφορία που φέρει ένας POS tagger μπορεί να συμπεριληφθεί στον υπολογισμό των N-grams όπως αναφέρει ο Jelinek (1989). Παράδειγμα για ένα τρίγραμμα μπορούμε να γράψουμε το εξής:

$$P(w_n | w_{n-2} w_{n-1}) = P(w_n | POS_n) \cdot P(POS_n | POS_{n-2}, POS_{n-1}) \quad (6)$$

Όπου n POS είναι η POS κατηγορία του w_n .

Το κύριο κίνητρο της ενσωμάτωσης ενός POS σε N-grams είναι να μειωθεί ο αριθμός των παραμέτρων και κατά συνέπεια η διασπορά της εκτίμησης. Ένα από τα πιο πρακτικά προβλήματα της γλώσσας είναι η πολυσημασία των λέξεων με αποτέλεσμα η κατάταξη μιας λέξης σε μια κατηγορία βάσει του POS να είναι πολύ δύσκολη. Ένας πολύ σύγχρονος POS αναλυτής παρουσιάζει ποσοστό ακρίβειας γύρω στο 95-97% κάτω από ιδανικές συνθήκες. Όμως, διάφορα πειράματα αποδεικνύουν ότι τα μοντέλα αυτά συνήθως δεν αποδίδουν ικανοποιητικά, όπως αυτό φαίνεται από την βελτίωση του βαθμού αβεβαιότητας σε σχέση με τα κανονικά N-grams που βασίζονται μόνο σε λέξεις. Το συμπέρασμα που εξάγεται είναι ότι αυτό που είναι χρήσιμο στην γλωσσική πληροφορία δεν είναι και κατ' επέκταση αποτελεσματικό στις τεχνικές πρόβλεψης λέξεων [120][121][123][128][131].

3.3.2 Υπολογισμός πιθανοτήτων απλών N-grams μοντέλων

Ο σκοπός ενός γλωσσικού μοντέλου είναι να ορισθεί η πιθανότητα $P(w_1^n)$ μιας ακολουθίας λέξεων $w_1^n = w_1, w_2, w_3, \dots, w_n$. Μπορούμε να χρησιμοποιήσουμε τον κανόνα της αλυσίδας των πιθανοτήτων για να υπολογίσουμε την πιθανότητα:

$$P(w_1^n) = P(w_1)P(w_2 | w_1)P(w_3 | w_1^2)P(w_4 | w_1^3) \dots P(w_n | w_1^{n-1}) = \prod_{k=1}^n P(w_k | w_1^{k-1}) \quad (7)$$

Το ερώτημα που γεννάται είναι το πώς μπορούμε να υπολογίσουμε την πιθανότητα $P(w_n | w_1^{n-1})$. Από την στιγμή που είναι δύσκολο να υπολογισθεί η πιθανότητα του τύπου $P(w_n | w_1^{n-1})$ για μεγάλα n μπορούμε να υπολογίσουμε την πιθανότητα μιας λέξης που εξαρτάται μόνο από 2 προγενέστερες λέξεις, και μιλάμε για την θεωρία των τριγραμμάτων

όπου ισχύει ότι $P(w_n | w_1^{n-1}) \approx P(w_n | w_{n-2}, w_{n-1})$, και στην πράξη δουλεύει ικανοποιητικά.

Η υπόθεση αυτή που θέλει την πιθανότητα της λέξης να εξαρτάται μόνο από τις προηγούμενες 2 λέξεις καλείται υπόθεση Markov. Η υπόθεση αυτή δέχεται ότι η πιθανότητα ενός μελλοντικού γεγονότος μπορεί να προβλεφθεί κοιτώντας το άμεσο παρελθόν του και όχι το πολύ μακρινό παρελθόν του.

Το N-grams γλωσσικό μοντέλο, χρησιμοποιεί τις προηγούμενες $N-1$ λέξεις (τυπικά μια ή δύο) σαν μια προσέγγιση της ιστορίας. Συνεπώς ένα δίγραμμα καλείται σαν πρώτης τάξης Markov μοντέλο (κοιτάζει μια λέξη στο παρελθόν), ενώ το τρίγραμμα σαν δεύτερης τάξης Markov μοντέλο και σε γενικές γραμμές κάθε N-gram καλείται σαν $N-1$ τάξης Markov μοντέλο.

Άρα η γενική εξίσωση για κάθε N-gram υπολογίζοντας την υπό-συνθήκη πιθανότητα μιας επόμενης λέξης σε μια ακολουθία είναι:

$$P(w_n | w_1^{n-1}) \approx P(w_n | w_{n-N+1}^{n-1}) \quad (8)$$

Η γενική εξίσωση αυτή δείχνει ότι η πιθανότητα μιας λέξης w_n βάσει όλων των προγενέστερων λέξεων μπορεί να προσεγγιστεί με την πιθανότητα βάσει των N τελευταίων λέξεων. Βάσει ενός διγραμμικού μοντέλου η συνολική πιθανότητα του string μέσω της προηγούμενης εξίσωσης, είναι ίση με:

$$P(w_1^n) \approx \prod_{k=1}^n P(w_k | w_{k-1}) \quad (9)$$

Τα N-grams μοντέλα μπορούν να εκπαιδευτούν μετρώντας και κανονικοποιώντας τις

εμφανίσεις τους σε ένα σώμα κειμένων εκπαίδευσης. Παίρνουμε ένα σώμα κειμένου και υπολογίζουμε τον αριθμό ενός συγκεκριμένου διγράμματος και τον διαιρούμε με το άθροισμα όλων των διγραμμάτων που μοιράζονται την ίδια πρώτη λέξη:

$$P(w_n | w_{n-1}) = \frac{c(w_{n-1}w_n)}{\sum_w c(w_{n-1}w)} \quad (10)$$

Το άθροισμα όλων των διγραμμάτων που ξεκινούν με αυτήν την λέξη ισούται με τον αριθμό εμφάνισης της λέξης. Άρα μπορούμε να γράψουμε ότι ισχύει:

$$P(w_n | w_{n-1}) = \frac{c(w_{n-1}w_n)}{c(w_{n-1})} \quad (11)$$

Και στην γενικευμένη μορφή της, στην περίπτωση των N-grams,

$$P(w_n | w_{n-N+1}^{n-1}) = \frac{c(w_{n-N+1}^{n-1}w_n)}{c(w_{n-N+1}^{n-1})} \quad (12)$$

Για τα τριγράμματα η πιθανότητα αυτή γράφεται ως εξής:

$$P(w_3 | w_1, w_2) = \frac{C(w_1, w_2, w_3)}{C(w_1, w_2)} \quad (13)$$

Η τελευταία εξίσωση υπολογίζει την πιθανότητα του N-gram διαιρώντας τον αριθμό εμφάνισης μιας ακολουθίας λέξεων με τον αριθμό εμφάνισης των προηγούμενων λέξεων. Ο λόγος αυτός καλείται σχετική συχνότητα. Οι εκτιμήσεις της πιθανότητας N-gram μπορούν να υπολογισθούν χρησιμοποιώντας τις σχετικές πιθανότητες, που αποκαλούνται ως εκτιμήσεις μέγιστης πιθανότητας, δηλ. οι κανονικοποιημένοι αριθμοί εμφάνισης των N-grams σε ένα συγκεκριμένο σώμα κειμένου εκπαίδευσης [124][126][127][128][131][132].

3.3.3 N-Grams

Μια συνηθισμένη προσέγγιση στα γλωσσολογικά μοντέλα είναι η χρήση των n- grams. Στην αναγνώριση συναισθήματος βασισμένη στην ερμηνεία της γλώσσας μόνο unigrams έχουν εφαρμοσθεί έως τώρα. Μας προσφέρουν τη δυνατότητα να αντιλαμβανόμαστε το συναίσθημα κάθε μόνης παρατηρούμενης λέξης χωρίς να μοντελοποιούμε την εξάρτηση από τις γειτονικές. Για να χειριστούμε μόνο λέξεις κλειδιά που τους αποδίδονται συναισθήματα οι Lee και Pieruccini πρότειναν την ταξινόμηση των αφηρημένων όρων που δεν τους αποδίδονται συναισθήματα αρχικά. Έπειτα η συχνότητα εμφάνισης της κάθε παρατηρούμενης λέξης για κάθε υπάρχον συναίσθημα και την παραγόμενη συναισθηματική αιχμή υπολογίζονται στη φάση της εκπαίδευσης. Αν όμως, δεν επιθυμούμε να παράγουμε απλά τη συχνότητα κάθε παρατηρούμενης λέξης αλλά να μελετήσουμε και τη συσχέτισή της με τις

γειτονικές της θα πρέπει να χρησιμοποιήσουμε tri- grams. Με τα tri- grams έχουμε τη δυνατότητα να μελετήσουμε την κάθε παρατηρούμενη λέξη σε σχέση με αυτές που βρίσκονται πίσω και μπροστά της. Αυτό φαίνεται ιδιαίτερος χρήσιμο αν σκεφτούμε πως το νόημα μιας λέξης μπορεί να διαφοροποιηθεί ριζικά λόγω των γειτονικών λέξεων. Για παράδειγμα, αν μιας λέξης προηγείται το «δεν» τότε εκφράζεται το αντίθετο από αυτό που θα εκφραζόταν αν ήταν μόνη της [113][114][115][116][117][118].

3.3.4 Συντακτική δομή

Πολλές προσπάθειες έχουν γίνει στο να ενσωματωθούν πληροφορίες σχετικά με την σύνταξη στο γλωσσικό μοντέλο. Τα N-grams μοντέλα εκμεταλλεύονται τη στατιστική εμφάνισης των λέξεων στα σώματα κειμένων, αλλά δεν έχουν αντίληψη της γραμματικής σε αποστάσεις μεγαλύτερης από N. Ένα εναλλακτικό γλωσσικό μοντέλο, είναι η Ανεξάρτητη από τα Συμφραζόμενα Πιθανοτική Γραμματική (Probabilistic Context Free Grammar) (PCFG), η οποία αποτελείται από μια Ανεξάρτητη από τα Συμφραζόμενα Γραμματική (CFG) μέσα στην οποία κάθε κανόνας αναδιατύπωσης έχει μια συσχετισμένη πιθανότητα. Το άθροισμα όλων των πιθανοτήτων για όλους τους κανόνες με την ίδια αριστερή πλευρά ισούται με 1. Στο μοντέλο PCFG, η πιθανότητα μιας συμβολοσειράς ($P(w_n)$), είναι απλώς το άθροισμα των πιθανοτήτων των συντακτικών δένδρων της. Η πιθανότητα ενός δεδομένου δένδρου είναι το γινόμενο των πιθανοτήτων όλων των κανόνων που απαρτίζουν τους κόμβους του δένδρου [110][111][112][113][114][115][116].

3.4 ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΜΕΣΩ ΤΟΥ ΠΡΟΣΩΠΟΥ

Σύμφωνα με τον Δαρβίνο, οι εκφράσεις προσώπου είναι παγκοσμίως αποδεκτές και έχουν μια εξελικτική βάση [99]. Σύμφωνα με αυτόν, οι εκφράσεις προσώπου που συσχετίζονται με τα συναισθήματα έχουν εξελιχθεί να υπηρετήσουν έναν φυσικό σκοπό πολύ περισσότερο από μια απλή επικοινωνία. Δηλαδή, όταν τα νήπια ουρλιάζουν από τον πόνο η κυκλοφορία του αίματος επηρεάζεται και τα μάτια τους τείνουν να γεμίσουν με αίμα. Κατά συνέπεια οι μύες που βρίσκονται γύρω από τα μάτια συσπώνται έντονα για προστασία. Στους ενήλικες αν και το ουρλιαχτό στις πιο πολλές περιπτώσεις καταπνίγεται, οι μύες γύρω από τα μάτια τείνουν να συσπώνονται ακόμα και κατά τη διάρκεια μιας απλής κατάστασης ανησυχίας [100]. Ο Ekman δηλώνει ότι οι εκφράσεις προσώπου των συναισθημάτων έχουν μια εθελούσια πλευρά που δίνει στους ανθρώπους τη δυνατότητα να επικοινωνούν τα συναισθήματά τους σκόπιμα [97]. Επίσης, οι εκφράσεις προσώπου δείχνουν συναισθήματα ακούσια (Κατά την επεξεργασία του βίντεο της απολογίας του για τη Monica Lewinsky ο Μπιλ Κλίντον «προδόθηκε» από ένα στιγμιαίο ειρωνικό χαμόγελο που είχε καταγραφεί κατά τη διάρκειας

απολογίας του). Ο Ekman δηλώνει ότι οι εκφράσεις προσώπου είναι βασικές στον καθορισμό και στον κανονισμό των διαπροσωπικών σχέσεων. Για παράδειγμα, άνθρωποι με εκ γενετής παράλυση προσώπου έχουν τεράστιες δυσκολίες στην ανάπτυξη και διατήρηση απλών σχέσεων [54].

3.4.1 Περιγραφικές διαδικασίες για συναισθηματικές εκφράσεις μέσω του προσώπου

Τα προφανή στοιχεία στην αναγνώριση συναισθημάτων από πρόσωπα είναι παράλληλα, κατά κάποιον τρόπο, με θέματα που συσχετίζονται με τον ήχο της φωνής. Πρέπει όμως να επισημανθεί ότι διαφέρουν σε κάποια άλλα σημεία. Όπως συμβαίνει στον λόγο, πολλοί ερευνητές κάνουν προσπάθεια να καθοριστεί η έκφραση των συναισθημάτων σε ότι έχει σχέση με τους λεγόμενους ποιοτικούς στόχους, δηλαδή σταθερές θέσεις οι οποίες μπορούν να εμφανιστούν σε κάποια φωτογραφία [101]. Το θέμα που απεικονίζεται στην φωτογραφία συνήθως «συλλαμβάνει» κάποια έκφραση δηλαδή τη στιγμή που το συγκεκριμένο συναίσθημα καταγράφεται. Προσφάτως η έρευνα έχει στραφεί προς τις περιγραφές που δίνουν έμφαση σε χειρονομίες και συγκεκριμένες κινήσεις του προσώπου [95][96][97][98]. Στην πράξη, όμως, η τεχνική έρευνα πάνω στις εκφράσεις του προσώπου για τα συναισθήματα έχει λάβει πολύ σοβαρά υπόψιν τον όρο συναίσθημα στην ευρύτερη έννοιά του.

3.4.2 Στόχοι και εκφράσεις που συσχετίζονται με την έκφραση συναισθημάτων

Η ανάλυση της έκφρασης των συναισθημάτων ενός ανθρώπινου προσώπου απαιτεί έναν αριθμό διαδικαστικών βημάτων που θα επιχειρήσουν να εντοπίσουν συγκεκριμένα στοιχεία πάνω στο πρόσωπο. Όπως παραδείγματος χάρη τα μάτια το στόμα και την μύτη και παρακολουθούν την κίνηση και τις συσπάσεις του προσώπου λαμβάνοντας τα σαν χαρακτηριστικά σημεία.

Τα χαρακτηριστικά του προσώπου μπορούν να παρουσιαστούν, ως στατικά όπως π.χ. χρώμα του δέρματος- τα οποία μπορεί να μετατρέπονται σταδιακά όπως μόνιμες ρυτίδες ή σε πιο γρήγορους ρυθμούς, όπως όταν ένας άνθρωπος ανεβάζει τα φρύδια σαν ένδειξη έκπληξης [102]. Ο εντοπισμός της θέσης και του σχήματος του στόματος, των ματιών, των βλεφάρων, των ρυτίδων γίνεται μέσω σημείων-στόχων, που χρησιμοποιούνται για να εντοπιστούν συγκεκριμένες διαθέσεις από ανθρώπους. Γενικά οι τεχνικές που χρησιμοποιούνται για να εντοπίσουν εκφράσεις του προσώπου ή γενικότερα συναισθήματα, αντιμετωπίζουν το πρόβλημα να εντοπίζουν και να αποσπών τις περιοχές ή τα χαρακτηριστικά του προσώπου που έχουν σχέση με την γενικότερη κίνηση του κεφαλιού. Συστήνουν λοιπόν τη χρήση

συγκεκριμένων μυών που μπορούν να περιγράψουν χαρακτηριστικές χειρονομίες ή εκφράσεις στο πρόσωπο του ανθρώπου.

3.4.3 Πρόσωπα και φυσιολογία συναισθημάτων

Σε αντίθεση με τον λόγο, η επιστήμη της νευροβιολογίας μας προσφέρει σημαντικό υλικό για την ανακάλυψη συναισθημάτων που προέρχονται από το πρόσωπο. Είναι γνωστό ότι το δεξί κροταφικό σημείο του ανθρώπου περιλαμβάνει έναν τομέα που έχει πολύ μεγάλη ευαισθησία σε ότι αφορά το πρόσωπο. Φθορά της συγκεκριμένης περιοχής μπορεί να οδηγήσει στο φαινόμενο της προσωπαγωγσίας, δηλαδή στην ανικανότητα να αναγνωρίζονται πρόσωπα [101]. Πιο συγκεκριμένος εντοπισμός των περιοχών που βρίσκονται στον ανθρώπινο εγκέφαλο και έχουν σχέση με το πρόσωπο, πραγματοποιείται με την χρήση απεικόνισης εγκεφάλου. Αυτό το συγκεκριμένο στοιχείο υποδεικνύει ότι υπάρχουν συγκεκριμένοι χώροι στις ινιακές και κροταφικές περιοχές οι οποίες υπερδραστηριοποιούνται κατά την διάρκεια της κίνησης του προσώπου [103].

Οι εκφράσεις προσώπου μπορούν να μετρηθούν με την ανάλυση εμφανών αλλαγών στην μυϊκή δραστηριότητα του προσώπου χρησιμοποιώντας ειδικά συστήματα [104]. Όμως η μυϊκή δραστηριότητα του προσώπου μπορεί να είναι τόσο μικρή σε μέγεθος που θα είναι δύσκολο να εντοπιστεί. Με την μέτρηση της ηλεκτρικής δραστηριότητας των μυών του προσώπου ακόμη και η μη παρατηρήσιμες μυϊκές δραστηριότητες (μικρές δραστηριότητες ή κινήσεις που περιορίζονται γρήγορα) μπορούν να μετρηθούν [105][106]. Το ηλεκτρομυογράφημα είναι μια μέθοδος που μετρά την ηλεκτρική δραστηριότητα ενός μύος είτε κατευθείαν από τον συγκεκριμένο μυ με ηλεκτρόδια σε μορφή βελόνας ή από το δέρμα που βρίσκεται πάνω από τον μυ με ηλεκτρόδια επιφάνειας, το οποίο χρησιμοποιείται συχνά στην ψυχοφυσιολογική έρευνα [107].

Στις μελέτες που γίνονται με το ηλεκτρομυογράφημα, έχουν γίνει τις τελευταίες δεκαετίες συσχετισμοί ανάμεσα σε πολλαπλούς και διαφορετικούς μύες του προσώπου και διαφορετικά συναισθήματα. Τα πιο αξιόπιστα ευρήματα δείχνουν μια αύξηση στην δραστηριότητα του μέγα ζυγωματικού (zygomaticus major)³ στην μυϊκή περιοχή κατά την διάρκεια θετικού συναισθηματικού ερεθισμού αλλά και αποτελούν αύξηση στην δραστηριότητα του corrugators supercili (ο μύς που «κεντά» και χαμηλώνει το φρύδι παράγοντας το σχήμα της λύπης) σε μυϊκή περιοχή κατά την διάρκεια αρνητικού ερεθισμού [108][109][110]. Ο συσχετισμός κίνησης των δύο παραπάνω μυϊκών περιοχών αλλά και η εμπειρία θετικών και αρνητικών συναισθημάτων αντίστοιχα έχει επίσης ερμηνευθεί σε πολλές

³ ο μύς που τραβά τις γωνίες του χείλους παράγοντας έτσι το χαμόγελο

μελέτες [53][105]. Συνοψίζοντας οι έρευνες δείχνουν ότι η δραστηριότητα που αναπτύσσεται σε αυτές τις δύο μυϊκές περιοχές συσχετίζεται με την συναισθηματική αξία του ερεθίσματος αλλά και στην υποκειμενική εμπειρία συναισθημάτων.

Οι μελέτες στα αποτελέσματα ηλεκτρομυογραφημάτων προσώπου, με ερέθισμα συναισθηματικού ήχου ήταν σπάνιες. Στην εργασία τους, οι Jancke και λοιποί μελέτησαν τρία ερεθίσματα περιβαλλοντικού συναισθηματικού ήχου πάνω σε πολλούς μύες του προσώπου [111]. Αν και τα ερεθίσματα διέφεραν στην υποκειμενική εκτίμηση σθένους και έξαψης δεν παρήγαγαν ξεκάθαρες συναισθηματικά συσχετιζόμενες αλλαγές σε ηλεκτρομυογράφημα προσώπου. Για παράδειγμα, η δραστηριότητα της μυϊκής περιοχής *zygomaticus major* παρουσίασε αξιοσημείωτη αύξηση κατά την διάρκεια όλων των συναισθηματικών ερεθισμών σε σύγκριση με την αρχική του θέση, ανεξάρτητα από τις μετρήσεις σθένους. Ο Jancke και οι συνάδελφοί του μελέτησαν επίσης, τις επιπτώσεις υψηλών και χαμηλών τόνων έντασης στο ηλεκτρομυογράφημα του προσώπου, και βρήκαν ότι η υψηλοί τόνοι έντασης και οι ήχοι (που χαρακτηρίστηκαν δυσάρεστοι) παρήγαγαν τις πιο δυνατές απαντήσεις στους επάνω μύες του προσώπου (στην μυϊκή περιοχή του *corrugators supercili*). Οι επιστήμονες αυτοί μελέτησαν τις επιπτώσεις μιας απλής λέξης που αναφέρθηκε από ηθοποιούς που προσομοίαζαν διαφορετικά συναισθήματα. Τα αποτελέσματά τους έδειξαν ότι τα στοιχεία προσωδίας ήταν ένας αποτελεσματικός τρόπος στο να μεταφέρουν συναισθηματικές πληροφορίες και επαρκούσαν για να παράγουν όμοιες αλλαγές προσώπου στο ηλεκτρομυογράφημα στο αντικείμενό μας.

Σε γενικές γραμμές το πιο σημαντικό συμπέρασμα είναι ότι ο σχηματισμός που εμφανίζεται για τις εκφράσεις συναισθημάτων στο πρόσωπο πραγματοποιείται σε διαφορετικούς χώρους και βασίζεται σε διαφορετικά στοιχεία.

3.4.4 Προσεγγίσεις συναισθηματικής αναγνώρισης μέσω προσώπου

Για όλες τις προσεγγίσεις που έχουν να κάνουν με την ανάλυση των συναισθημάτων που εκφράζονται μέσω του προσώπου, η πρώτη ενέργεια είναι ο εντοπισμός του προσώπου (tracking). Πολλές μέθοδοι έχουν αναπτυχθεί για τον συγκεκριμένο σκοπό χωρίς όμως να έχουν επιτευχθεί ικανοποιητικά αποτελέσματα [112][113][114]. Άλλες μέθοδοι χρησιμοποιούν χρωματικούς τόνους που παρουσιάζονται στην περιοχή του προσώπου (χρώμα επιδερμίδας) αλλά η ακρίβεια των αποτελεσμάτων δεν επιτυγχάνεται πάντα λόγω ανομοιογένειας στον φωτισμό.

Μεγάλες ψυχολογικές έρευνες πάνω στις εκφράσεις προσώπου έχουν πραγματοποιηθεί χρησιμοποιώντας φωτογραφίες ανθρώπων σε κοντινό πλάνο που αποτυπώνουν την έκφραση τους εκείνη τη στιγμή [115]. Αυτές οι φωτογραφίες είναι ενδεικτικές διότι επιτρέπουν σε

κάποιον να εντοπίσει συγκεκριμένα στατικά στοιχεία όπως είναι οι ρυτίδες καθώς και σχήματα και θέσεις των άλλων στοιχείων του προσώπου μας. Όμως πρέπει να επισημανθεί σε αυτή την περίπτωση ότι η απόσπαση αυτών των συγκεκριμένων στοιχείων μπορεί να είναι πολύ δύσκολη ενώ υπάρχουν ελάχιστες τεχνικές κατηγοριοποίησης εκφράσεων προσώπου που μπορούν να θεωρηθούν επιτυχείς σε πραγματικές συνθήκες φωτισμού.

3.5 ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΜΕΣΩ ΤΗΣ ΦΩΝΗΣ

Η ανθρώπινη φωνή είναι ένας από τους βασικούς παράγοντες μεταφοράς της κοινωνικής και συναισθηματικής επικοινωνίας [116]. Από τα αρχικά στάδια της εξέλιξης, τα νήπια επικοινωνούν με συναισθηματικά-επιφορτισμένα ηχητικά επιφωνήματα με της μητέρες τους [117]. Η συναισθηματική ομιλία παραμένει ένα βασικό κανάλι της έκφρασης των συναισθημάτων κατά την εξέλιξη και την πορεία της ζωής μας, ίσως περισσότερο τώρα σε σχέση με το παρελθόν, δεδομένης της συνεχώς αυξανόμενης επικοινωνίας μέσω του τηλεφώνου [118]. Παρά το γεγονός ότι οι ηχητικές συναισθηματικές εκφράσεις παρουσιάζονται τόσο συχνά όσο και οι συναισθηματικές εκφράσεις μέσω του προσώπου στην καθημερινή ζωή, και αναγνωρίζονται ανάμεσα σε πολιτισμούς σε ίδιο βαθμό με αυτές του προσώπου, οι πρώτες δεν έχουν λάβει της απαιτούμενης προσοχής από τους ψυχολόγους και τους επιστήμονες που ασχολούνται με την νευρολογία όπως συμβαίνει με τις δεύτερες [94]. Σαν συνέπεια, σήμερα είναι πολύ λίγα τα πράγματα που γνωρίζουμε για τα νεύρα που αναφέρονται στην αντίληψη των συναισθημάτων στην φωνή.

Ο Hughlings παρατήρησε ότι ασθενείς με ποικίλες γλωσσολογικές δυσκολίες που οφείλονταν στην βλάβη του αριστερού ημισφαιρίου του εγκεφάλου, είχαν ακόμη την ικανότητα να επικοινωνούν συναισθηματικά μέσω της φωνής, και έτσι προτάθηκε ότι το αριστερό ημισφαίριο πρέπει να υπηρετεί τέτοιες λειτουργίες [119]. Αυτό υποστηρίχθηκε από τον Tucker και Watson σε διάφορες μελέτες που παρατηρούσαν την αντίληψη της συναισθηματικής προσωδίας σε άτομα που είχαν βλάβη στο δεξί ημισφαίριο του εγκεφάλου, σε σύγκριση με άτομα που είχαν βλάβη στο αριστερό μέρος του εγκεφάλου [120].

Υπάρχουν δύο ευρύτατες πηγές πληροφόρησης στην ομιλία μας. Κατά πρώτον, υπάρχει το σημασιολογικό κομμάτι της ομιλίας που μεταφέρει γλωσσολογικές πληροφορίες, σε τέτοιο βαθμό που οι κουβέντες διαμορφώνονται σύμφωνα με τους κανόνες της προφοράς που ισχύει στην κάθε γλώσσα [30]. Από την άλλη υπάρχει η παραγλωσσολογική πληροφόρηση που αναφέρεται σε αναλυτικά μηνύματα όπως η συναισθηματική κατάσταση του ομιλητή. Αυτά τα στοιχεία και ιδιαίτερα το δεύτερο, αποτελούν το πρώτο σημαντικό βήμα για την παρουσίαση της συναισθηματικής κατάστασης του ομιλητή.

Από ιατρικής απόψεως όμως, υπάρχουν ενδείξεις που παρουσιάζουν το συμπαθητικό νευρικό σύστημα να ερεθίζεται με αισθήματα θυμού φόβου και χαράς [121]. Σε αυτές τις περιπτώσεις, οι παλμοί της καρδιάς και η πίεση αυξάνονται, το στόμα στεγνώνει και οι μύες περιστασιακά τρέμουν. Η ομιλία είναι δυνατή, γρήγορη και διατυπώνεται με υψηλή συχνότητα. Από την άλλη μεριά, με τον ερεθισμό του παρασυμπαθητικού νευρικού συστήματος, σε συναισθήματα όπως η λύπη, οι παλμοί της καρδιάς, η πίεση μειώνονται το σάλιο αυξάνει και ο λόγος-ομιλία που παράγεται είναι σε υψηλή ένταση αλλά αργός.

Η προσωδία είναι ο κύριος μεταφορέας της κατάστασης του ομιλητή και ως εκ τούτου είναι σημαντική στην ανάκτηση των πληροφοριών που είναι θεμελιώδεις για την επικοινωνία. Ο όρος «προσωδία» ορίζεται ως το σύνολο των χαρακτηριστικών της φωνής που προσδίδουν φυσικότητα στο σήμα και χαρακτηρίζονται από χρονική διάρκεια μεγαλύτερη από αυτή των στοιχειωδών μονάδων της φωνής (φωνημάτων) [7][121][122]. Τα χαρακτηριστικά αυτά είναι:

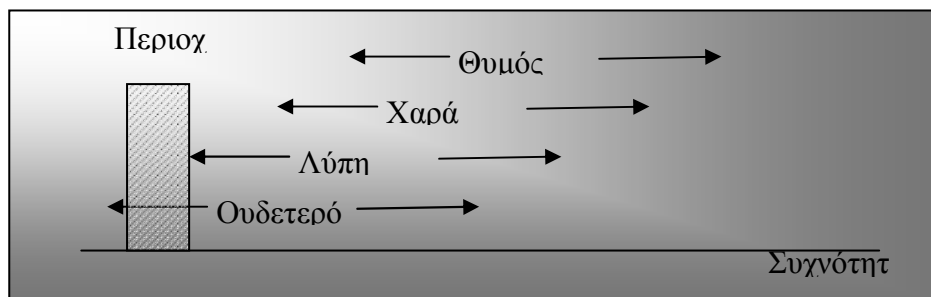
1. Θεμελιώδης συχνότητα – Pitch (Η θεμελιώδης συχνότητα F0 ορίζεται ως η χαμηλότερη συχνότητα στην οποία το φωνητικό σήμα επαναλαμβάνεται).
2. Ενέργεια σήματος – Energy (Η ενέργεια είναι το ακουστικό ανάλογο της έντασης της φωνής, αν και η σχέση τους δεν είναι γραμμική επειδή εξαρτάται έντονα από την ευαισθησία του ανθρώπινου ακουστικού συστήματος στις διαφορετικές συχνότητες).
3. Συντονισμοί φωνητικού καναλιού – Formants (Ως formants ορίζουμε τις συχνότητες συντονισμού του φωνητικού καναλιού, δηλαδή τις συχνότητες εκείνες που ενισχύει το φωνητικό κανάλι κατά τις διάρκεια της ομιλίας. Οι συχνότητες των formants εξαρτώνται απ' το σχήμα και τις διαστάσεις της φωνητικής οδού. Γύρω απ' τη συχνότητα ενός formant, στην κυματομορφή του σήματος ομιλίας, παρουσιάζεται μια συγκέντρωση της ακουστικής ενέργειας. Κατά συνέπεια τα formants γίνονται φανερά ως κορυφές στο φάσμα του σήματος ομιλίας και αντίστοιχα ως σκουρόχρωμες περιοχές σε ένα φασματογράφημα υποδεικνύοντας την συγκέντρωση ενέργειας στις αντίστοιχες περιοχές).
4. Συντελεστές MFCC (Στην επεξεργασία ήχου, το Φάσμα συχνοτήτων Mel (Mel frequency cepstrum - MFC) είναι μια αναπαράσταση του βραχυπρόθεσμου φάσματος έντασης ενός ήχου, βασισμένου στην κλίμακα συχνοτήτων Mel, μια μη γραμμική κλίμακα συχνοτήτων, που προσεγγίζει τη συμπεριφορά του συστήματος ακοής. Οι συντελεστές του φάσματος

συχνότητας Mel (MFCCs – Mel Frequency Cepstrum Coefficients) είναι οι συντελεστές εκείνοι που αποτελούν στο σύνολο τους το φάσμα Φάσμα συχνότητων Mel).

5. Αρμονικότητα – Harmonicity (Η αρμονικότητα είναι ένα μέτρο της δύναμης της αντίληψης του pitch για έναν ήχο. Με άλλα λόγια η αρμονικότητα αντιπροσωπεύει την ακουστική περιοδικότητα ενός σήματος).

6. Ρυθμός ομιλίας (Το ποσοστό του αριθμού των λέξεων ή των φωνημάτων στη μονάδα του χρόνου ή το αντίστροφο της μέσης διάρκειας της ηχηρής ομιλίας (voiced speech)).

Γενικά, η ομιλία του ανθρώπου έχει ένα μεγάλο εύρος συχνότητας. Ωστόσο, το ενδιαφέρον εύρος συχνότητας ή αλλιώς το νοηματικό εύρος συχνότητας είναι από 100Hz μέχρι 5000Hz [122]. Αυτό το εύρος, έχει παρατηρηθεί ότι ποικίλει μεταξύ των ανθρώπων και μεταξύ των διαφορετικών συναισθημάτων όπως φαίνεται και στην Εικόνα 1. Με την χρήση αυτής της άποψης μπορούμε να έχουμε κάποιου είδους αναγνώριση συναισθημάτων από την φωνή.



Εικόνα 1 Εύρος συχνότητας νοηματικού σήματος

3.6 ΣΥΝΔΥΑΣΤΙΚΗ ΑΝΑΓΝΩΡΙΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ

Σχετικά, έχουν πραγματοποιηθεί κάποιες προσπάθειες στην εφαρμογή ακουστικής και οπτικής πληροφορίας στα συστήματα αναγνώρισης συναισθημάτων. Ο De Silva πρότεινε ένα σύστημα αναγνώρισης συναισθημάτων βασισμένο σε έναν οπτικοακουστικό κανόνα, στον οποίο τα δεδομένα εξόδου του ταξινομητή συνδέονται στο επίπεδο της απόφασης [123]. Από τα ηχητικά δεδομένα χρησιμοποιήθηκαν τα χαρακτηριστικά της προσωδίας, ενώ από τα οπτικά δεδομένα χρησιμοποιήθηκαν οι μέγιστες αποστάσεις και οι ταχύτητες μεταξύ έξι συγκεκριμένων σημείων του προσώπου. Μία παρόμοια προσέγγιση παρουσιάζεται και από τον Chen, στην οποία η κυρίαρχη μοντελοποίηση, σύμφωνα με τα υποκειμενικά πειράματα που έγιναν από τον De Silva & Miyasato, χρησιμοποιήθηκε για να καθοριστούν οι διαφορές

μεταξύ των αποτελεσμάτων των συστημάτων που ασχολούνται μόνο με την μια μέθοδο [124] [125]. Και στις δύο μελέτες συμπεριέλαβαν ότι η εκτέλεση του συστήματος βελτιωνόταν κατά την ταυτόχρονη χρήση και των δύο μεθόδων.

Ο Yoshitomi και λοιποί πρότεινε ένα σύστημα που περιλάμβανε και τις δύο μεθόδους. Δεν χρησιμοποιούσε όμως μόνο την φωνητική και την οπτική πληροφορία αλλά και την διανομή της θερμότητας που προέκυπτε από μια κάμερα υπερύθρων [126]. Υπήρχαν αμφισβητήσεις, ότι οι εικόνες υπερύθρων δεν είναι τόσο ευαίσθητες όταν υπάρχουν συνθήκες έντονου φωτισμού, πράγμα που είναι ένα από τα βασικά προβλήματα όταν οι εκφράσεις του προσώπου καταγράφονται με την χρήση συμβατικών καμερών. Χρησιμοποίησαν μια βάση δεδομένων που κατεγράφη από μια γυναίκα ομιλήτη που διάβαζε μια μόνο λέξη, ενώ υποκρινόταν ότι βρισκόταν σε πέντε διαφορετικές συναισθηματικές καταστάσεις. Συνέδεσαν αυτούς τους τρεις τρόπους στο επίπεδο των αποφάσεων με την χρήση του καθορισμού εμπειρικής βαρύτητας. Η απόδοση του συστήματος ήταν καλύτερη όταν οι τρεις μέθοδοι χρησιμοποιούνταν συνδυαστικά.

Από τους Huang και Chen ένα σύστημα συνδυαστικής αναγνώρισης συναισθημάτων προτάθηκε για την αναγνώριση έξι συναισθημάτων, στα οποία το οπτικοακουστικό υλικό ενοποιούταν στο επίπεδο της εξαγωγής των χαρακτηριστικών [127] [128]. Χρησιμοποίησαν τα χαρακτηριστικά της προσωδίας από τα ηχητικά δεδομένα, και την θέση και την κίνηση των οργάνων του προσώπου από τα δεδομένα του βίντεο. Τα καλύτερα χαρακτηριστικά από τον συνδυασμό των παραπάνω, χρησιμοποιήθηκαν για την είσοδό τους στον ταξινομητή. Οι αποδόσεις για το συνδυαστικό σύστημα έφταναν το 97,2%. Ωστόσο χρησιμοποίησαν μια μικρή βάση δεδομένων με μόνο έξι καταγραφές για κάθε συναίσθημα, πράγμα που περιορίζει την γενίκευση των αποτελεσμάτων όταν πρόκειται για μεγαλύτερο πακέτο δεδομένων.

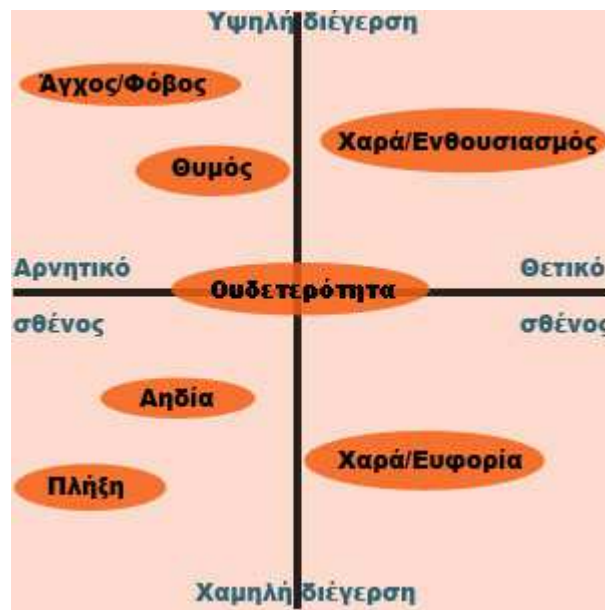
Όλες οι παραπάνω μελέτες έχουν δείξει ότι η δημιουργία συστημάτων αναγνώρισης συναισθημάτων μπορεί να βελτιωθεί με την συνδυαστική χρήση των προαναφερθέντων μεθόδων. Παρόλα αυτά δεν είναι ξεκάθαρο το ποια τεχνική είναι κατάλληλη για τον συνδυασμό των τεχνικών αυτών. Η μελέτη του Busso και λοιπών προσπαθεί να δώσει μιαν απάντηση σε αυτόν τον προβληματισμό συγκρίνοντας τις αποφάσεις και τα χαρακτηριστικά σε επίπεδο τεχνικών αναβάθμισης στα πλαίσια της εκτέλεσης ενός συστήματος [129] .

4. ΤΑΞΙΝΟΜΗΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΚΑΙ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ

4.1 Η ΤΑΞΙΝΟΜΗΣΗ ΤΩΝ ΣΥΝΑΙΣΘΗΜΑΤΩΝ

Σήμερα, η ανάγκη για την μοντελοποίηση των συναισθημάτων απορρέει από την ανάγκη μιας «ομιλούσας» και πολυμορφικής εφαρμογής επικοινωνίας. Για παράδειγμα,

ποικίλοι παράγοντες λογισμικού χρειάζονται τα συναισθηματικά μοντέλα για να επικοινωνήσουν με τους ανθρώπους με πιο φυσικό τρόπο. Οι παράγοντες λογισμικού, χρειάζονται τα συναισθηματικά μοντέλα για αναγνώριση συναισθημάτων με την χρήση του προσώπου, της κίνησης του σώματος, του λόγου και ακόμη και για την σύνθεση συναισθημάτων στην εμπύχωση του προσώπου ενός παράγοντα του λογισμικού, την ζωντάνια στις κινήσεις του σώματος και στην σύνθεση του λόγου [130].



Εικόνα 2. Τα συναισθήματα σύμφωνα με το σθένος και τη διέγερση.

Η πιο ευρέως αποδεκτή ιδέα στην έρευνα που αφορά τα συναισθήματα, είναι ότι συγκεκριμένα συναισθήματα είναι πρωταρχικά κι άλλα δευτερεύοντα. Τα δευτερεύοντα συναισθήματα προκύπτουν από τον συνδυασμό των πρωταρχικών. Αυτή η ιδέα προέρχεται από τον Descartes (1970). Δεν υπάρχει κάποια συγκεκριμένη λίστα με τα βασικά συναισθήματα. Όμως, υπάρχει μια γενική αποδοχή μιας βασικής λίστας συναισθημάτων: Θυμός, αηδία, φόβος, χαρά, λύπη, έκπληξη [131]. Οι Pittam & Scherer το 1992 πρότειναν την ταξινόμηση των συναισθημάτων από την διέγερση [38]. Χαρακτηρίζουν την χαρά και τον θυμό ως συναισθήματα με μεγάλη διέγερση, ενώ την λύπη και τον ουδέτερο λόγο ως συναισθήματα με χαμηλή διέγερση όπως αποτυπώνεται στην Εικόνα 2.

Υπάρχουν τέσσερις διαφορετικές γενικές θεωρητικές οπτικές στην ψυχολογία για την εξήγηση και οριοθέτηση των συναισθημάτων. Αυτές οι οπτικές είναι η Δαρβίνεια, η οπτική κατά James, η γνωστική και τέλος η οπτική της κοινωνικής δομής [132]. Η γενική ιδέα της Δαρβίνειας οπτικής είναι ότι τα συναισθήματα αποτελούν ένα βαθμιαία εξελισσόμενο φαινόμενο με ενδιαφέρουσες λειτουργίες επιβίωσης που βοηθούν στην λύση συγκεκριμένων προβλημάτων που αντιμετωπίζουμε σαν είδη. Σύμφωνα με τη θεωρία αυτή θα πρέπει να

εντοπίζουν τα ίδια συναισθήματα σε όλους τους ανθρώπους. Υπάρχουν πολλοί ερευνητές που έχουν ακολουθήσει την Δαρβίνεια οπτική, συμπεριλαμβανομένου του Paul Ekman, Alan Fridlund και Sylvan Tomkins [88][133][134]. Ο υποκινητής για την οπτική κατά James ήταν ο William James. Επέμενε ότι θα ήταν αδύνατον να υπάρχουν συναισθήματα χωρίς αλλαγές στο σώμα και ότι αυτές πάντα έρχονται πρώτες [135]. Στην γνωστική οπτική, η κεντρική ιδέα είναι ότι οι σκέψεις και τα συναισθήματα είναι αχώριστα. Όλα τα συναισθήματα πέρασαν σε αυτήν την οπτική σύμφωνα με αυτό που ο Arnold αποκαλούσε αξιολόγηση, η διαδικασία με την οποία τα γεγονότα στο περιβάλλον κρίνονται καλά ή κακά για εμάς [136]. Υποστηρικτές αυτής της οπτικής ήταν εκτός από τον Arnold και οι Frijda, Zajonc και ο LeDoux [42][49] [137]. Αυτοί που υιοθετούσαν την τέταρτη οπτική, πίστευαν ότι τα συναισθήματα είναι πολιτιστικά προϊόντα. Η ερμηνεία των συναισθημάτων εμφανίζεται λόγω των κοινωνικών κανόνων που διδάσκονται. Οι Catharine Lutz και Claire Armon-Jones έχουν κάνει ενδιαφέρουσες συνεργασίες πάνω σε αυτές τις ιδέες περί κοινωνικής κατασκευής των συναισθημάτων [138][133].

Ο Kramer έκανε μια ανασκόπηση σε έναν αριθμό μελετών που περιγράφουν ότι ποικίλες συναισθηματικές και φυσικές καταστάσεις των ομιλητών, συμπεριλαμβανομένου του φύλλου, της ηλικίας της εμφάνισης, της εξυπνάδας και της προσωπικότητας μπορούν να προσδιοριστούν από την φωνή [139]. Πολλοί ερευνητές από τότε έχουν προσπαθήσει να δημιουργήσουν αυτόματα συστήματα ανάλυσης συναισθημάτων με την χρήση του λόγου, την αναγνώριση συναισθημάτων από τον λόγο και πρόσφατα την αυτόματη σύνθεση του συναισθηματικού λόγου [133][139].

Ο στόχος των συστημάτων σύνθεσης λόγου είναι η αναπαραγωγή ομιλίας που να είναι φυσική στο άκουσμά της. Αυτό είναι το ίδιο που επιδιώκουν και οι ηθοποιοί [140]. Η καταγραφή μιας συλλογής από φυσικό λόγο που περιέχει συναισθήματα που εκφράζονται με φυσικό τρόπο, είναι ένα πολύπλοκο πρόβλημα. Κάποιοι ερευνητές έχουν κάνει ανάλυση του φυσικά εκφραζόμενου ή αυθόρμητου συναισθήματος [121].

Η αναγνώριση συναισθημάτων εξελίσσεται σε διαφορετικές κατευθύνσεις. Αυτοί που ασχολούνται με την δημιουργία προγραμμάτων που αναγνωρίζουν συναισθήματα, χρησιμοποιούν διαφορετικούς τύπους βάσεων δεδομένων, διαφορετικές ιδιότητες, και μεθόδους αναγνώρισης. Οι βάσεις δεδομένων συνήθως προορίζονται από μία μόνο γλώσσα και έχουν ένα μικρό αριθμό καταγραφών ή αριθμό συναισθημάτων. Αυτό περιπλέκει το πρόβλημα σε ότι έχει να κάνει με τα ποσοστά επιτυχίας της αναγνώρισης των συναισθημάτων.

4.2 ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΟΜΙΛΙΑΣ

Σύμφωνα με έρευνα των Βεργίδη και Κωτρόπουλου έχουν καταγραφεί τριάντα δύο βάσεις δεδομένων, συναισθηματικού λόγου [141]. Κάθε βάση αποτελείται από ένα κύριο μέρος ανθρώπινου λόγου, το οποίο έχει προφερθεί κάτω από διαφορετικές συναισθηματικές καταστάσεις. Παρακάτω υπάρχει μια πλήρης περιγραφή κάθε βάσης, και των εφαρμογών τους. Η παραπάνω έρευνα των Βεργίδη και Κωτρόπουλου έδειξε ότι η αυτόματη αναγνώριση συναισθημάτων πάνω σε αυτές τις βάσεις δεδομένων δε μπορεί να επιτύχει μια σωστή ταξινόμηση που να ξεπερνά την τάξη του 80% για τα τέσσερα από τα βασικά συναισθήματα. Επίσης, καταλήγουν ότι τα φυσικά συναισθήματα δε μπορούν να ταξινομηθούν εύκολα όπως αυτά που έχουν προέλθει από προσομοίωση (επιτηδευμένη προσποίηση συναισθημάτων-ηθοποιία). Τρίτον, τα πιο γνωστά συναισθήματα που έχουν ερευνηθεί για την μείωση της συχνότητας εμφάνισης είναι ο θυμός, η λύπη, η ευτυχία, ο φόβος, η αηδία, η χαρά, η έκπληξη και η πλήξη.

4.2.1 Βάσεις δεδομένων συναισθηματικού λόγου, στα Αγγλικά

1. Η βάση δεδομένων καταγράφηκε στο Faculty of Electrical Engineering and Computer Science, University of Maribor στην Σλοβενία [142]. Περιέχει συναισθηματικό λόγο τριών κατηγοριών συναισθημάτων, όπως η αηδία, η έκπληξη, η χαρά, ο φόβος, ο θυμός και η λύπη. Επίσης, συμπεριλαμβάνονται δύο ουδέτερα συναισθήματα. Το «γρήγορο και δυνατό» και το «χαμηλό και απαλό». Οι κατηγορίες αυτές έχουν καταγραφεί σε MPEG-4 [143]. Η βάση δεδομένων περιλαμβάνει 186 διαφορετικές αρθρώσεις για κάθε συναισθηματική κατηγορία.
2. Ο R.Cowie κατασκεύασε αυτή την βάση δεδομένων στο Queen's University του Belfast [144-146]. Οι αναγνώστες ήταν 40 εθελοντές ηλικίας από 18 έως 69 ετών. Οι εθελοντές διάβασαν 5 κείμενα των 7-8 προτάσεων γραμμένες σε συναισθηματικά κατάλληλο τόνο και με περιεχόμενο ανάλογο για το κάθε συναίσθημα. Κάθε κείμενο έχει άμεση σχέση με την αντίστοιχη συναισθηματική κατάσταση.
3. Ο R.Cowie και ο M.Schroder κατασκεύασαν την βάση Belfast Natural Database στο Queens University. Υπήρξαν δύο είδη καταγραφών [147]. Η μια έγινε σε στούντιο και η άλλη κατευθείαν από τα τηλεοπτικά προγράμματα. Το χρονικό μέγεθος των ήχων ήταν λίγο μεγάλο έτσι ώστε να μπορεί να δείξει την εξέλιξη των συναισθημάτων μέσα από αυτή την διάρκεια. Οι καταγραφές του στούντιο περιείχαν δύο κομμάτια. Το πρώτο κομμάτι περιέχει συζητήσεις μεταξύ των μαθητών. Το δεύτερο μέρος περιέχει οπτικοακουστικές καταγραφές από συνεντεύξεις.

4. Ο R.Cole και οι βοηθοί του στο Πανεπιστήμιο του Colorado, δημιούργησαν μια βάση δεδομένων για παιδιά (Kids Audio Speech Corpus NSF/ITR Reading Project) [148]. Ο στόχος αυτής της έρευνας ήταν η συλλογή επαρκών οπτικοακουστικών δεδομένων από παιδιά με σκοπό την ενεργοποίηση της εξέλιξης για οπτικοακουστικά συστήματα αναγνώρισης, τα οποία θα μπορούν να διευκολύνουν την διάδραση πρόσωπο με πρόσωπο με ηλεκτρονικούς καθηγητές. Μόνο 1000 από τις 45000 προτάσεις είχαν συναισθηματικό προσανατολισμό.
5. Η βάση δεδομένων Emotional Prosody Speech and Transcripts δημιουργήθηκε από τον M. Liberman στο Παν/μιο της Pennsylvania [149]. Αυτή η βάση περιέχει 9 ώρες συναισθηματικού υλικού. Περιέχει λόγο σε 15 συναισθηματικές κατηγορίες. Αυτή η βάση χρησιμοποιείται από την LDC [149].
6. SUSAS (Speech Under Simulated and Actual Stress), είναι μια βάση δεδομένων που πραγματοποιήθηκε στο University of Colorado Boulder από τον J.Hansen [149]. Η βάση δεδομένων περιλαμβάνει φωνές από 32 ομιλητές που οι ηλικίες τους ήταν ανάμεσα στα 22 και τα 76. Επίσης, τέσσερις πιλότοι στρατιωτικών ελικοπτέρων καταγράφηκαν κατά τη διάρκεια μιας μάχης. Λέξεις από το λεξιλόγιο 35 λέξεων επικοινωνίας αεροσκαφών απαρτίζουν την βάση δεδομένων.
7. Η έβδομη βάση δεδομένων που έχει καταγραφεί είναι αυτή από το Πανεπιστήμιο Macquire και κατασκευάστηκε από την Papeira [150]. Η βάση απαρτίζεται από 40 προτάσεις που έχουν ειπωθεί από δύο ηθοποιούς σε πέντε συναισθηματικές κατηγορίες. Υπάρχουν δύο επαναλήψεις από αυτές τις 40 προτάσεις, πράγμα που μας κάνει 80 δεδομένα. Στην μελέτη, 31 άτομα, έκριναν τις προτάσεις. Αυτοί που άκουσαν τις προτάσεις, τις κατηγοριοποίησαν σε έξι σκάλες έντασης της κλίμακας Likert.
8. Ο Edgington στο BT Labs της Μ.Βρετανίας, δημιούργησε μια βάση για την εκπαίδευση ενός μηχανήματος σύνθεσης φωνής [151]. Δεκατρείς κριτές πιστοποίησαν τα συναισθήματα με 79.3% ακρίβεια. Η βάση επίσης περιέχει την ενέργεια σήματος, την διάρκεια των συλλαβών, και την θεμελιώδη συχνότητα για κάθε φώνημα.
9. Οι Polzin & A.H. Waibel στον πανεπιστήμιο Carnegie Mellon, κατασκεύασαν επίσης μια βάση δεδομένων [152]. Το κύριο μέρος απαρτιζόταν από 291 λέξεις σύμβολα για κάθε συναίσθημα και κάθε ομιλητή. Το μήκος της πρότασης ποικίλει από δύο μέχρι 12 λέξεις. Αυτές οι λέξεις περιέχονταν σε ερωτήσεις, σε δηλώσεις ή σε διαταγές. Η βάση αξιολογήθηκε από άλλα άτομα. Η αναγνώριση του κάθε συναισθήματος έφτανε το 70%.

10. Η V.Petrushin στο Center for Strategic Technology Research, Accenture δημιούργησε αυτή την βάση με σκοπό να εκπαιδεύσει νευρωνικά δίκτυα στην αναγνώριση συναισθημάτων. Είναι χωρισμένη σε δύο μελέτες [153]. Η πρώτη μελέτη περιέχει ως βασικό μέρος 700 μικρές προτάσεις οι οποίες έχουν εκφραστεί από 30 επαγγελματίες ηθοποιούς. Στην δεύτερη μελέτη, ηχογραφήθηκαν 56 τηλεφωνικά μηνύματα.
11. Ο R.Fernandez κατασκεύασε στο MIT μια βάση συναισθηματικού λόγου [154]. Τα πρόσωπα που χρησιμοποιήθηκαν για την μελέτη ήταν δύο οδηγοί που ρωτήθηκαν να προσθέσουν δύο αριθμούς κατά την διάρκεια της οδήγησης. Οι δύο ανεξάρτητες μεταβλητές σ αυτό το πείραμα ήταν η ταχύτητα οδήγησης και η συχνότητα με την οποία έλυνε κάθε οδηγός τις μαθηματικές ερωτήσεις.

4.2.2 Βάσεις δεδομένων στα Γερμανικά

12. Η βάση δεδομένων Verbmobil ηχογραφήθηκε στο Πανεπιστήμιο του Hamburg. Η βάση περιέχει φωνές από 58 γηγενείς Γερμανούς ομιλητές (29 άντρες και 29 γυναίκες) καθώς μιλούσαν σε ένα υποτιθέμενο ARS⁴ σύστημα[155][156]. Από απόσταση, ο ερευνητής ρύθμιζε τις απαντήσεις του ARS με τέτοιο τρόπο ώστε ο ομιλητής να πιστέψει ότι μιλάει στην μηχανή. Οι παραπάνω διάλογοι προέρχονται από την παράσταση του «Μάγου του OZ» (Wizard of Oz). Οι ιδιότητες της προσωπίας ενός συναισθήματος, όπως η επιμήκυνση των συλλαβών και η έμφαση στις λέξεις σημειώνονται.
13. Η βάση SmartKom Multimodal Corpus κατασκευάστηκε στο Institute for Phonetik and Oral Communication στο Munich [155][157]. Η βάση περιέχει διαλόγους τύπου «Wizard of Oz» στα αγγλικά και τα γερμανικά. Η βάση περιέχει πολλαπλά κανάλια ήχου και δύο κανάλια βίντεο (πρόσωπο και σώμα από το πλάι). Ο στόχος αυτής της μελέτης ήταν η δημιουργία ενός μοντέλου, για τις διεπιφάνειες ανθρώπου-υπολογιστή, από τις φηγούρες και την φωνή.
14. Στο Technical University of Berlin δημιουργήθηκε αυτή η βάση δεδομένων από τον W.F.Sendlmeier et al [158][159]. Κάθε ένας από τους δέκα επαγγελματίες ηθοποιούς εξέφραζαν δέκα λέξεις και πέντε προτάσεις σε όλες τις κατηγορίες συναισθημάτων. Το περιεχόμενο αξιολογήθηκε από 25 κριτές οι οποίοι κατηγοριοποίησαν κάθε συναίσθημα με ποσοστό επιτυχίας 80%
15. Η δημιουργία αυτής της βάσης δεδομένων έγινε στο Max-Planck-Institute of Cognitive Neuroscience από τον K.Alter [160]. Καταγράφηκαν επίσης και τα

⁴ Αυτόματο σύστημα απόκρισης (Automatic Response System)

ηλεκτροεγκεφαλογραφήματα. Ο στόχος της μελέτης αυτής ήταν ο συσχετισμός των συναισθημάτων τα οποία αναγνωρίζονται από τον λόγο με μια περιοχή στον ανθρώπινο εγκέφαλο. Χρησιμοποιήθηκε μια εκπαιδευμένη και ευφραδής γυναίκα. Είκοσι άτομα, έκριναν και το σημασιολογικό περιεχόμενο και τα χαρακτηριστικά της προσωδίας, και τα ταξινόμησαν σε μια κλίμακα από ένα έως πέντε.

16. Στο Πανεπιστήμιο της Γενεύης (University of Geneva) πραγματοποιήθηκε η δημιουργία αυτής της βάσης δεδομένων από τον κ.Κ.Scherer [161]. Ο σκοπός αυτής της μελέτης ήταν να φέρει στο φως τις διαφορές του συναισθηματικού λόγου στην αντίληψη μεταξύ ανθρώπων που προέρχονται από διαφορετικές χώρες. Οι προτάσεις προέρχονταν από μια τεχνητή γλώσσα η οποία κατασκευάστηκε από ένα επαγγελματία φθογγολόγο.
17. Η βάση Magdeburger Prosodie Korpus από τους B.Wendt & H.Scheich στο Institute of Neurobiology κατασκευάστηκε με στόχο την κατασκευή ενός «χάρτη εγκεφάλου» των συναισθημάτων [162]. Η βάση περιέχει επίσης την προφορά των λέξεων, το μήκος τους, την ταχύτητά τους, αφαιρετική/πραγματική τάση, κατηγοριοποιήσεις.
18. Ο M.Schoroeder ηχογράφησε μια βάση διφώνων που μπορούν να χρησιμοποιηθούν στην σύνθεση του συναισθηματικού λόγου [163]. Χρησιμοποιήθηκε ένας άντρας Γερμανός, για την παραγωγή ενός σετ Γερμανικών διφώνων, ένα για κάθε έναν από τους τρεις βαθμούς φωνητικής έκφρασης «απαλός» «μέσος» «δυνατός». Τέσσερις ειδικοί έλεγξαν την φωνητική έκφραση, την εγκυρότητα του τόνου, και το αν ήταν φωνητικά σωστό.
19. Ο M.Schoroeder κατασκεύασε μια βάση η οποία περιέχει «συναισθηματικές εκρήξεις» [164]. Η μελέτη του δείχνει ότι οι συναισθηματικές εκρήξεις, που παρουσιάζονται χωρίς περιεχόμενο, μπορούν να μεταφέρουν μια ευπροσδιόριστη συναισθηματική ερμηνεία. Οι επαγγελματίες επιλέγουν τις συναισθηματικές εκρήξεις από την Γερμανική λογοτεχνία. Η βάση απαρτίζεται από περίπου 80 διαφορετικές συναισθηματικές εκρήξεις. Η βάση περιέχει λόγο σε 10 διαφορετικές συναισθηματικές κατηγορίες.

4.2.3 Βάσεις δεδομένων στα Ιαπωνικά

20. Οι Nakatsu et al. στα ATR Laboratories κατασκεύασαν αυτή βάση [165]. Η βάση περιέχει λόγο σε 8 κατηγορίες συναισθημάτων. Για την διεξαγωγή της μελέτης χρησιμοποιήθηκαν 100 ντόπιοι ομιλητές (50 άνδρες και 50 γυναίκες) και ένας επαγγελματίας ομιλητής ραδιοφώνου. Ο επαγγελματίας ομιλητής έπρεπε να διαβάσει

100 ουδέτερες προτάσεις σε 8 συναισθηματικές παραλλαγές. Έπειτα, ζητήθηκε από τους απλούς ομιλητές να μιμηθούν τον επαγγελματία ηθοποιό και να πουν τις ίδιες λέξεις.

21. Η βάση κατασκευάστηκε στο Kyoto Institute of Technology από τον Y.Niimi et al [166]. Αποτελείται από τμήματα VCV (vowel consonant vowel- φωνήεν σύμφωνο φωνήεν) για καθεμία από τις τρεις συναισθηματικές κατηγορίες. Αυτά τα τμήματα μπορούν να παράγουν κάθε παραλλαγή προφοράς από τους Γιαπωνέζους. Τα τμήματα συγκεντρώθηκαν από μια συλλογή 400 γλωσσολογικά ανεξάρτητων προτάσεων. Οι προτάσεις αναλύθηκαν για την παραγωγή ενός οδηγού για την σχεδίαση τέτοιων βάσεων δεδομένων, και την παραγωγή μιας εξίσωσης για κάθε φώνημα, η οποία μπορεί να προβλέψει την διάρκειά της βασιζόμενη στο γλωσσολογικό και φωνητικό περιεχόμενό της. Δώδεκα άνθρωποι έκριναν την βάση και οργάνωσαν κάθε συναίσθημα με ποσοστό επιτυχίας 84.
22. Οι A.Iida & N.Campbell στα ATR Laboratories κατασκεύασαν τη βάση στην οποία τα συναισθήματα προσομοιώνονται αλλά δεν μεγαλοποιούνται [167]. Η βάση περιέχει κείμενα μονολόγου που έχουν συλλεχθεί από εφημερίδες, από το διαδίκτυο, από αυτοβιογραφίες κινητικά ανάπηρων ατόμων, από εκθέσεις και στήλες. Κάποιες εκφράσεις χαρακτηριστικές για κάθε συναίσθημα, εισάγονταν σε κατάλληλες θέσεις με σκοπό την αύξηση της έκφρασης κάθε συναισθήματος-στόχου.

4.2.4 Βάσεις δεδομένων στα Ολλανδικά

23. Βάση δεδομένων Groninger. Έχει κατασκευαστεί στο Psychology School at Groningen University in Netherlands και ταξινομήθηκε από την ERLA⁵ [156]. Περιλαμβάνει 20 ώρες Ολλανδικού λόγου. Η βάση προσανατολίζεται μερικώς στα συναισθήματα. Περιλαμβάνεται ηλεκτρογλωσσογράφημα και ορθογραφική καταγραφή. Οι ομιλητές δεν είναι ηθοποιοί και τα συναισθήματα είναι επιτηδευμένα και όχι φυσικά. Η βάση περιέχει μικρά κείμενα, μικρές προτάσεις, ψηφία, μονοσύλλαβες λέξεις, και μεγάλης διάρκειας φωνήεντα.
24. Ο Mozziconacci σύλλεξε την βάση αυτή δεδομένων με σκοπό να μελετήσει την σχέση μεταξύ της ταχύτητας ομιλίας και των συναισθημάτων [168][169]. Το υλικό που χρησιμοποιήθηκε για την μελέτη είναι 315 προτάσεις. Κάθε ένας από τους τρεις ομιλητές διαβάζει πέντε προτάσεις, οι οποίες έχουν σημασιολογικά ουδέτερο

⁵ European Language Resources Association www.era.info

περιεχόμενο. Εικοσιτέσσερις κριτές αξιολόγησαν τις προτάσεις και δύο ειδικοί που ασχολούνταν με τον τονισμό τις κατηγοριοποίησαν.

4.2.5 Συναισθηματικές βάσεις στα Ισπανικά

25. Η Spanish Emotional Speech Database δημιουργήθηκε από τον J.M.Montero και τους βοηθούς του [170]. Δεκαπέντε κριτές ταυτοποίησαν κάθε συναίσθημα με ποσοστό 85%. Ο χαρακτηρισμός της βάσης έγινε ημιαυτόματα. Το κύριο μέρος αποτελείται από τρία μικρά αποσπάσματα (4-5 προτάσεων), δεκαπέντε μικρές προτάσεις και τριάντα απομονωμένες λέξεις. Όλες είχαν ουδέτερο λεξιλογικό, συντακτικό και σημασιολογικό περιεχόμενο.
26. Η I. Iriundo από το Πανεπιστήμιο της Βαρκελώνης δημιούργησε αυτή βάση [171]. Ο λόγος κρίθηκε από 1054 μαθητές κατά την διάρκεια μιας εξέτασης αντίληψης. Από τις 336 συνομιλίες μόνο οι 34 πέρασαν την εξέταση.

4.2.6 Συναισθηματικές βάσεις στα Δανέζικα

27. Οι I.F.Engberg, T.Brondsted και A.V.Hansen στο Center for Person Kommunikation at Aalborg University ηχογράφησαν την Δανέζικη Βάση Δεδομένων Συναισθηματικού Λόγου (DES) [172]. Η κατασκευή της ήταν μέρος της έρευνας Voice Attitudes and Emotions in Speech Synthesis. Είκοσι κριτές –γηγενείς ομιλητές από 18 μέχρι 25 ετών- πιστοποίησαν τα συναισθήματα με ποσοστό 67%

4.2.7 Βάσεις συναισθηματικού λόγου στα Εβραϊκά

28. Στην πανεπιστημιακή σχολή Holon Academic Institute of Technology στο Ισραήλ, ο N.Amir et al ηχογράφησε μια συναισθηματική πολυμορφική βάση δεδομένων [173]. Η βάση αποτελείται από συναισθηματικό λόγο, ηλεκτρομυογράφημα ενός μυός του πάνω μέρους του προσώπου ο οποίος βοηθά στην έκφραση των συναισθημάτων, καταγραφή του παλμού της καρδιάς, αντοχή στους αιφνιδιασμούς και τις ενδείξεις τους.

4.2.8 Βάσεις δεδομένων στα Σουηδικά

29. Ο Abelin et al ηχογράφησε μια βάση δεδομένων, στην οποία συμμετείχαν διαφορετικής εθνικότητας ακροατές οι οποίοι ταξινόμησαν τις προτάσεις στις συναισθηματικές καταστάσεις [174]. Η ομάδα των ακροατών αποτελούταν από 35 Σουηδούς ομιλητές και 78 μετανάστες.

4.2.9 Βάσεις δεδομένων στα Κινέζικα

30. Αυτή η βάση δεδομένων περιέχει αποσπάσματα προτάσεων από Κινέζικα τηλεπαιχνίδια [175]. Δημιουργήθηκε από τον/την F.Yu et al στο Microsoft Research

China. Τέσσερα άτομα ταξινόμησαν 2000 προτάσεις. Όταν πάνω από δύο άτομα ταξινομούσαν μια πρόταση σε ένα συναίσθημα, τότε η πρόταση αυτή ανήκε στο συγκεκριμένο συναίσθημα

4.2.10 Βάσεις δεδομένων στα Ρωσικά

31. Η βάση RUSSian LANguage Affective Speech (RUSSLANA) δημιουργήθηκε από τους V.Makarova & V.Petrushin στο Πανεπιστήμιο Meikai στην Ιαπωνία [176]. Οι συνολικές προτάσεις ήταν 3660 από 61 (12 άνδρες) ντόπιους Ρώσους ομιλητές ηλικίας από 16 μέχρι 28. Συμπεριλαμβάνονται στην βάση τα χαρακτηριστικά του λόγου όπως η ενέργεια, οι κυματομορφές και καμπύλες.

4.2.11 Πολυγλωσσική βάση δεδομένων συναισθηματικού λόγου.

32. Η βάση «lost luggage study» που δημιουργήθηκε από τον K.Scherer, ηχογραφήθηκε στον διεθνή αερολιμένα της Γενεύης [177]. Οι ομιλητές είναι 109 επιβάτες που περιμένουν μάταια να έρθουν οι βαλίτσες τους.

5. ΧΡΟΝΙΚΗ ΑΝΑΔΡΟΜΗ ΣΤΑ ΣΥΣΤΗΜΑΤΑ ΑΝΑΓΝΩΡΙΣΗΣ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΚΑΙ ΣΧΕΤΙΚΕΣ ΜΕΛΕΤΕΣ

5.1 Σχετικές Έρευνες

5.1.1 Με Προσωδιακά χαρακτηριστικά

Η επιλογή παραμέτρων για την αναγνώριση συναισθημάτων από την ομιλία βασίζεται κυρίως στα προσωδιακά χαρακτηριστικά γνωρίσματα, όπως η θεμελιώδης συχνότητα, η ενέργεια, τα πέντε πρώτα Formants, οι συντελεστές Cepstral συχνότητας (Mel Frequency Cepstral Coefficients MFCCs), ο ρυθμός ομιλίας, τα χρονικά χαρακτηριστικά (πχ παύσεις), παράμετροι ποιότητας φωνής, η άρθρωση καθώς και μετρικές και χαρακτηριστικά γνωρίσματα που πηγάζουν από όλα τα παραπάνω, όπως ο μέσος όρος, τυπική απόκλιση, το μέγιστο, ελάχιστο, ρυθμός αλλαγής, διασπορά, διάμεσος και ούτω καθεξής.

Οι Tsuyoshi Moriyama και Shinji Ozawa δημιούργησαν ένα σύστημα ικανό να αναγνωρίζει και να συνθέτει συναισθηματικό περιεχόμενο στον λόγο [178]. Αρχικά συγκέντρωσαν το περιεχόμενο που σχετίζεται με τα φυσικά χαρακτηριστικά του συναισθηματικού λόγου. Αυτό, αξιολογήθηκε με γραμμικές στατιστικές μεθόδους από άτομα που το άκουσαν, και έπειτα εφαρμόστηκε στο σύστημα. Στο σύστημα, οι καμπύλες της θεμελιώδους συχνότητας (pitch ή F0) εκφράζονται από τα μοντέλα που έχουν προταθεί από τον Fujisaki (με επτά παραμέτρους) [196]. Το σύστημα κρίθηκε ικανό να αναγνωρίζει τα συναισθήματα από τα φυσικά χαρακτηριστικά του λόγου και να προσθέτει κάποια

συναισθήματα στον ουδέτερο λόγο. Ιδιαίτερα σε ότι αφορά τον «θυμό», την «έκπληξη», την «λύπη» τόσο το σύστημα αναγνώρισης όσο και το σύστημα σύνθεσης δούλευαν σωστά, πράγμα που τους οδήγησε στο συμπέρασμα ότι αυτά τα συναισθήματα περιγράφονται καλά από τα συστατικά της προσωδίας του λόγου.

Οι A.Paeschke, M.Kienast, και W.F.Sendlmeie αναφέρουν ότι τα συναισθήματα επηρεάζουν τον τρόπο που ομιλεί κάποιος και είναι πιθανόν να αναγνωρίζουμε την συναισθηματική κατάσταση ενός ομιλητή απλώς με το άκουσμα μερικών λέξεων [69]. Ο σκοπός της έρευνας ήταν ο διαχωρισμός των βασικών συναισθημάτων από τα χαρακτηριστικά της προσωδίας, και συγκεκριμένα από χαρακτηριστικά της θεμελιώδους συχνότητας F_0 , των πέντε πρώτων Formants, των MFCCs και άλλων. Κατέληξαν ότι αυτοί οι παράμετροι μεταξύ των άλλων αποτελούν την κατεύθυνση και την τομή των αλλαγών της συχνότητας σε ένα επίπεδο συλλαβών, και τις διακυμάνσεις της θεμελιώδους συχνότητας πάνω σε ολόκληρες κουβέντες μετά από μια διαδικασία μορφοποίησης. Επιπλέον, η συλλαβική διακύμανση θεμελιώδους συχνότητας χωρίς μορφοποίηση ερευνήθηκε [131][132]. Η πλειοψηφία των αποτελεσμάτων χρησιμοποιήθηκε στον διαχωρισμό ανάμεσα σε συναισθήματα φόβου, θυμού, ευτυχίας, λύπης, βαριεστημάρας και απέχθειας και απέδειξε ότι η F_0 των έξι προαναφερθέντων συναισθημάτων που ερευνήθηκαν διαχωρίζει το ένα από το άλλο. Επίσης, το επίπεδο έντασης διαφέρει μεταξύ τους. Η κατεύθυνση και η τομή της θεμελιώδους συχνότητας σε επίπεδο συλλαβών βρέθηκε επίσης να είναι σημαντική για την εύρεση του συναισθήματος. Και τέλος, εμφανέστατα αποτελέσματα προέκυψαν για την F_0 δια μέσου ολόκληρων φράσεων και προτάσεων.

Οι Vladimir Hozjan και Zdravko Kacic, ανέλυσαν ένα πολυγλωσσικό σύστημα αναγνώρισης συναισθημάτων από χαρακτηριστικά συγκεκριμένων βάσεων δεδομένων [133]. Η αναγνώριση έγινε σε Αγγλικές, Σλοβένικες, Ισπανικές και Γαλλικές βάσεις δεδομένων [193]. Οι βάσεις αυτές περιελάμβαναν ουδέτερο λόγο και έξι συναισθήματα όπως αηδία, χαρά, έκπληξη, φόβο, θυμό και λύπη. Η εξαγωγή των χαρακτηριστικών για την αναγνώριση έγινε με δύο τρόπους. Ο πρώτος τρόπος αφορούσε τον καθορισμό των χαρακτηριστικών χαμηλού επιπέδου (low level features) και ο δεύτερος τον υπολογισμό των χαρακτηριστικών υψηλού επιπέδου (high level features) από τα χαρακτηριστικά χαμηλού επιπέδου. Η αναγνώριση των συναισθημάτων έγινε με την χρήση τεχνητών νευρωνικών δικτύων. Η ακρίβεια στην αναγνώριση των συναισθημάτων ήταν μεγαλύτερη για αναγνώριση σε μοντέλα εξαρτημένου χρήστη, μικρότερη για μοντέλα που περιείχαν μόνο μια γλώσσα και ακόμη μικρότερη για μοντέλα που ήταν πολυγλωσσικά.

Η Valery A.Petrushin περιγράφει μια πειραματική μελέτη πάνω στην ηχητική συναισθηματική έκφραση, στην αναγνώριση και την εξέλιξη ενός υπολογιστικού συστήματος για την αναγνώριση συναισθημάτων [4]. Η μελέτη στηρίζεται σε έναν κορμό δεδομένων που υπολογίζεται στις 700 μικρές προτάσεις που εκφράζουν πέντε συναισθήματα. Την χαρά, τον θυμό, την λύπη, τον φόβο, και την φυσιολογική (μη συναισθηματική, ουδέτερη) κατάσταση, οι οποίες υποδύονται από τριάντα άτομα. Οι προτάσεις αξιολογήθηκαν από είκοσι τρεις κριτές, είκοσι από τους οποίους πήραν μέρος και στην ηχογράφηση. Η ακρίβεια στην αναγνώριση των συναισθημάτων στον λόγο ήταν 61% για την χαρά, 72,2% για τον θυμό, 68,3% για την λύπη, 49,5% για τον φόβο και 66,3 για την ουδετερότητα. Η ανθρώπινη δυνατότητα για τον διαχωρισμό των συναισθημάτων ήταν περίπου στα ίδια επίπεδα όπως φάνηκε από την έρευνα αλλά η τυπική απόκλιση ήταν πολύ μεγαλύτερη. Επίσης αξιολογήθηκε η ανθρώπινη ικανότητα να ξεχωρίζουν τα δικά τους συναισθήματα. Όπως αποδείχθηκε οι άνθρωποι έχουν αυξημένη ικανότητα στην αναγνώριση του θυμού της λύπης και του φόβου, ενώ δε συμβαίνει το ίδιο και με την αναγνώριση της χαράς και της ουδετερότητας. Τα χαρακτηριστικά που χρησιμοποιήθηκαν στην έρευνα ήταν η διακύμανση και η ενέργεια της φωνής, ενώ έγινε χρήση διαφόρων τύπων ταξινομητών έτσι ώστε να συγκριθούν τα αποτελέσματα. Τα καλύτερα αποτελέσματα δόθηκαν από την χρήση τεχνητών νευρωνικών δικτύων τα οποία οδήγησαν σε ποσοστά επιτυχίας της τάξεως του 70%.

Η Valery A.Petrushin προχώρησε στην δημιουργία ενός συντελεστή αναγνώρισης συναισθημάτων που ήταν σε θέση να αναλύει τηλεφωνικά σήματα και να τα διαχωρίζει σε δύο συναισθηματικές καταστάσεις «διαταραχής» και «ηρεμίας». Το σύστημα είχε ποσοστά επιτυχίας στο 77% και χρησιμοποιήθηκε ως μέρος περιορισμού ανεπιθύμητων μηνυμάτων σε τηλεφωνικά κέντρα.

Οι R.Nakatsu, J.Nicholson και N.Tosa εξέτασαν την αναγνώριση των συναισθημάτων που εμπλέκονται με τον ανθρώπινο λόγο [179]. Πρότειναν ένα αλγόριθμο αναγνώρισης βασισμένο σε νευρωνικά δίκτυα και ακόμη μια μέθοδο συλλογής βάσεως δεδομένων που περιέχει συναισθήματα. Πειραματίστηκαν με την βοήθεια των νευρωνικών δικτύων εκπαιδεύοντας την συγκεκριμένη βάση δεδομένων. Το ποσοστό επιτυχίας κυμάνθηκε γύρω στο 50% για εφαρμογή ανεξάρτητου χρήστη σε οχτώ συναισθηματικές καταστάσεις (θυμός, λύπη, χαρά, φόβος, έκπληξη, αηδία, όρεξη για παιχνίδι, ουδετερότητα).

Οι Balomenos και λοιποί παρουσίασαν οπτικές που αναφέρονται στην εξέλιξη ενός τεχνητού συστήματος που θα μπορεί να καταχωρεί το συναίσθημα του χρήστη και να αντιδρά ανάλογα [180]. Αυτές οι οπτικές εξετάστηκαν στα πλαίσια της έρευνας ERMIS (Emotion-Rich- Man –machine Interaction Systems) [180]. Αυτή η έρευνα είναι ενδιαφέρουσα αφού

βοηθά στην διερεύνηση αυτού που θεωρούμε ως την φυσική προσέγγιση στην εξέλιξη των συναισθηματικά-ευαίσθητων συστημάτων.

Οι Katarzyna, Nikolaou, Masiak και Treasure, ασχολήθηκαν με την αναγνώριση συναισθημάτων ατόμων που πάσχουν από την ασθένεια της νευρικής ανορεξίας [181]. Η έρευνα των παραπάνω εστιάστηκε όχι μόνο στην αναγνώριση μέσω της φωνής αλλά και σε αυτή μέσω των εκφράσεων του προσώπου. Για τους σκοπούς της έρευνας, χρησιμοποιήθηκαν τριάντα άτομα με νευρική ανορεξία σε σύγκριση με τριάντα υγιείς γυναίκες και εξετάστηκαν σε ηχητικά και όχι μόνο ερεθίσματα που εξέφραζαν συγκεκριμένα συναισθήματα. Τα αποτελέσματα της έρευνας έδειξαν ότι η ομάδα των γυναικών που έπασχαν από νευρική ανορεξία ήταν φτωχή σε ότι αφορά την αναγνώριση συναισθημάτων. Αυτό σημειώθηκε περισσότερο για τα αρνητικά συναισθήματα στα πρόσωπα, ενώ παρατηρήθηκε και για τα δύο συναισθήματα στις φωνές.

Οι C.M. Lee, S.Narayanan και R.Pieraccini, ασχολήθηκαν με την αυτόματη ταξινόμηση ομιλούμενων εκφράσεων βασισμένη στην συναισθηματική κατάσταση του ομιλητή [182]. Τα δεδομένα που χρησιμοποίησαν προέρχονταν από μια βάση που ηχογραφήθηκε από διάλογους μεταξύ ανθρώπου-υπολογιστή μιας διαφημιστικής εφαρμογής της SpeechWorks. Τα συναισθήματα που εξετάστηκαν ήταν τα αρνητικά και τα μη αρνητικά. Τα χαρακτηριστικά που χρησιμοποιήθηκαν από τους ταξινομητές ήταν η θεμελιώδης συχνότητα και η ενέργεια του ομιλούμενου σήματος. Για την βελτίωση της διαδικασίας ταξινόμησης, χρησιμοποιήθηκαν δύο συγκεκριμένες μέθοδοι επιλογής χαρακτηριστικών. Επίσης έγινε ανάλυση κυρίων συνιστωσών (Principal component analysis) με σκοπό την μείωση των διαστάσεων των χαρακτηριστικών πράγμα που έδινε αυξημένη ακρίβεια ταξινόμησης. Τα αποτελέσματα ήταν θετικά αφού έφτασαν τα επίπεδα του 70%-80%.

Οι A.Batliner, K.Fischer, R.Huber, J.Spilker, E.Nothing, έκαναν χρήση τριών βάσεων δεδομένων οι οποίες προέκυψαν από διαφορετικές πηγές [60]. Η πρώτη, αφορούσε ηθοποιούς οι οποίοι προσποιούνταν κάποια συναισθήματα. Η δεύτερη αφορούσε δεδομένα που προέκυπταν από ανθρώπους στην πραγματική ζωή, ενώ η τρίτη απαρτιζόταν από δεδομένα ανθρώπων που είχαν την αίσθηση ότι συνομιλούσαν με ένα σύστημα αυτόματου διαλόγου (WOZ study) [60]. Με την χρήση ίδιας μεθόδου ταξινόμησης σύγκριναν τα αποτελέσματα που προέκυψαν από τις τρεις βάσεις, και κατέληξαν ότι όσο πιο κοντά είναι τα δεδομένα στην πραγματικότητα τόσο μειώνεται η αξιοπιστία της προσωδίας ως ένδειξη της συναισθηματικής κατάστασης του ομιλητή. Σαν συνέπεια πρότειναν την αλλαγή του στόχου έτσι ώστε να πάνε να ψάχνουν για μια ακολουθία συγκεκριμένων συναισθημάτων στον λόγο του χρήστη, αλλά αντίθετα να ψάχνουν για ενδείξεις «προβλημάτων στην συνομιλία». Έτσι πρότειναν μια

δομή Απεικόνισης της Συναισθηματικής Κατάστασης του Χρήστη (Monitoring of User State Emotion) στην οποία ένας ταξινομητής προσωδίας συνδυάζεται με άλλες πηγές γνώσης, όπως η ιδιόμορφη γλωσσολογική συμπεριφορά κατά την συζήτηση, για παράδειγμα τη χρήση επαναλήψεων.

Οι Tim Lay New, Say Wei Foo και Liyanage C.De Silva, πρότειναν μια μέθοδο ταξινόμησης συναισθημάτων ανεξάρτητη από το περιεχόμενο του κειμένου [30]. Η μέθοδος αυτή κάνει χρήση των μικρού μήκους λογαριθμικών συντελεστών συχνότητας (LFPC-Log Frequency Power Coefficients) για την αναπαράσταση των σημάτων ομιλίας, και τα κρυμμένα Μαρκοβιανά μοντέλα (HMM) ως ταξινομητή. Τα συναισθήματα χωρίστηκαν σε έξι κατηγορίες. Ο προσδιορισμός της ονομασίας των συναισθημάτων έγινε με τα αρχέτυπα συναισθήματα του Θυμού, της Αηδίας, της Χαράς, του Φόβου, της Λύπης και της Έκπληξης. Η βάση δεδομένων αποτελούταν από 60 προτάσεις, οι οποίες ειπώθηκαν από δώδεκα ομιλητές και χρησιμοποιήθηκαν στην εκπαίδευση και την εξέταση του συστήματος. Η διαδικασία LFPC συγκρίθηκε με μεθόδους που χρησιμοποιούνται ευρέως στην αναγνώριση συναισθημάτων όπως η LPCC (Linear Prediction Cepstral Coefficients) και η MFCC (Mel-Frequency Cepstral Coefficients). Τα αποτελέσματα έδειξαν ότι το προτεινόμενο σύστημα έχει απόδοση 78% και μέγιστη απόδοση 96% στην κατηγοριοποίηση των συγκεκριμένων συναισθημάτων. Από τις αποδόσεις οι ερευνητές οδηγήθηκαν στο συμπέρασμα ότι η μέθοδος LFPC που χρησιμοποιήθηκε είναι πιο αποδοτική από τις παραδοσιακές παραμέτρους χαρακτηριστικών.

Οι Oh-Wook Kwon, Kwokleung Chan, Jiucang Hao και Te-Won Lee, επέλεξαν για την αναγνώριση των συναισθημάτων την ενέργεια, την διακύμανση, τις συνιστώσες MFCC, ως τα βασικά στοιχεία αναγνώρισης και προσέθεσαν την ταχύτητα/επιτάχυνση της διακύμανσης και των MFCC για τον σχηματισμό των επιπέδων των χαρακτηριστικών [183]. Εξήγαγαν στατιστικές που χρησιμοποιήθηκαν για ταξινομητές υψηλής ακρίβειας, θεωρώντας ότι κάθε επίπεδο είναι ένα μονοδιάστατο σήμα. Τα χαρακτηριστικά που προέκυψαν αναλύθηκαν με την χρήση Quadratic Discriminant Analysis (QDA) και Support Vector Machine (SVM). Τα αποτελέσματα τις έρευνας έδειξαν ότι η ενέργεια και η θεμελιώδης συχνότητα ήταν οι πιο σημαντικοί παράγοντες. Οι ερευνητές έκανα χρήση δύο διαφορετικών βάσεων δεδομένων και σύγκριναν τα αποτελέσματά τους σε διάφορους ταξινομητές όπως ο SVM, LDA (linear discriminant analysis), QDA και HMM (Hidden Markov Models). Με την πρώτη βάση δεδομένων SUSAS (μη εξαρτώμενη από το κείμενο) και τον ταξινομητή Gaussian SVM, έφτασαν σε ποσοστά επιτυχίας της τάξεως του 96,3% για συναισθήματα όπως

η ουδετερότητα και το άγχος [24]. Με την βάση δεδομένων AIBO έφτασαν ποσοστά επιτυχίας 42.3% για πέντε κλάσεις συναισθηματικής αναγνώρισης [25].

Οι Sherif Yacoub, Steve Simke, Xiaofan Lin και John Burns μελέτησαν την αναγνώριση συναισθημάτων από σήματα ομιλίας δίνοντας ιδιαίτερη έμφαση στην εξαγωγή των χαρακτηριστικών των συναισθημάτων από μικρές προτάσεις από εφαρμογές Διαδραστικής Φωνητικής Απόκρισης (IVR, Interactive Voice Response applications) [184]. Εστίασαν στον διαχωρισμό του θυμωμένου λόγου σε αντιπαράθεση με τον ουδέτερο λόγο, καθώς πρωταγωνιστούν στις εφαρμογές που αφορούν τηλεφωνικά κέντρα. Αναφέρθηκαν στην ταξινόμηση άλλων τύπων συναισθημάτων όπως η λύπη, η πλήξη, η χαρά, και ο ελαφρύς θυμός. Συνέκριναν τα αποτελέσματα που προέκυψαν από την χρήση νευρωνικών δικτύων, SVM, K-NN (K-Nearest Neighbour) και δένδρα αποφάσεων. Η βάση δεδομένων που χρησιμοποίησαν οι ερευνητές ήταν αυτή από το Linguistic Data Consortium at University of Pennsylvania, η οποία είχε ηχογραφηθεί με την βοήθεια οκτώ ηθοποιών που εξέφραζαν δεκαπέντε συναισθήματα. Τα αποτελέσματα έδειξαν ότι ο εμφανής θυμός και η ουδετερότητα μπόρεσαν να διαχωριστούν με ποσοστό ακρίβειας της τάξεως του 90%.

Οι Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neuman, και Shrikanth Narayanan ανέλυσαν τα θετικά και αρνητικά στοιχεία των συστημάτων που βασίζονται μόνο στις οπτικές είτε στις ακουστικές πληροφορίες [185]. Περιγράφουν ακόμη δύο προσεγγίσεις που χρησιμοποιήθηκαν στην σύνδεση αυτών των εκδοχών: το επίπεδο της απόφασης και το επίπεδο της ενσωμάτωσης των χαρακτηριστικών. Χρησιμοποίησαν μια βάση δεδομένων η οποία ηχογραφήθηκε από μια ηθοποιό και ταξινόμησαν τα εξής συναισθήματα: λύπη, θυμός, χαρά και ουδετερότητα. Με την χρήση σημαδιών στο πρόσωπο της ηθοποιού κατέγραψαν και οπτικά τις μεταβολές, σε συνδυασμό με την καταγραφή των ηχητικών δεδομένων. Τα αποτελέσματα έδειξαν ότι το σύστημα που βασιζόταν στις εκφράσεις του προσώπου είχε μεγαλύτερη απόδοση σε σχέση με το σύστημα που βασιζόταν αποκλειστικά στις φωνητικές καταγραφές. Από τα αποτελέσματα αυτά φάνηκε ότι οι δύο εκδοχές αλληλοσυμπληρώνονται και όταν χρησιμοποιούνται συνδυαστικά μπορούν να δώσουν ένα πολύ δυνατό σύστημα αναγνώρισης συναισθημάτων.

Η Astrid Paeschke εξέτασε την ολική ροπή της θεμελιώδους συχνότητας [186]. Ο πρωταρχικός της στόχος ήταν η εξέταση αυτής της παραμέτρου σε ότι αφορά τον χαρακτηρισμό συγκεκριμένων συναισθημάτων. Ο δεύτερός της στόχος ήταν να φέρει νέους προβληματισμούς σε σχέση με τις αιτίες απόκλισης και τον βαθμό επιρροής στην παραγωγή του λόγου. Τα αποτελέσματα την οδήγησαν ότι η ροπή αυτή δείχνει πιο κατάλληλη στην περιγραφή του συναισθήματος της πλήξης και της λύπης, ενώ λιγότερο κατάλληλη για το

άγχος και την απέχθεια, και χωρίς νόημα για την χαρά και τον θυμό. Κατέληξε λοιπόν ότι αυτή η μέτρηση βοηθά στον καθορισμό των συναισθημάτων χαμηλής διέγερσης.

Οι Eun Ho Kim, Kyung Hak Hyun και Yoon Keun Kwak πρότειναν την εξέταση του εύρους συχνότητας του σήματος ομιλίας, για την αναγνώριση των συναισθημάτων μέσω της φωνής [187]. Σε σύγκριση με τα άλλα χαρακτηριστικά αυτό είναι ανεξάρτητο από την έκταση του σήματος και είναι αποδοτικό σε περιβάλλον που έχει θόρυβο. Μέσα από πειράματα απέδειξαν ότι μετά την επεξεργασία και το βαθυπέρατο φιλτράρισμα των σημάτων, το μόνο συναίσθημα που συνεχίζει να έχει σήμα είναι αυτό της ουδετερότητας. Υποστήριξαν πως, όταν το μέγεθος ενός σήματος αλλάζει λόγω του βάθους της φωνής του ομιλητή, ή η απόσταση μεταξύ του ομιλητή και του μικροφώνου ή τα χαρακτηριστικά του μικροφώνου, η διαδικασία αναγνώρισης γίνεται ακόμη πιο δύσκολη. Η επεξεργασία που χρησιμοποίησαν, λοιπόν, κάνει το σήμα να είναι ανεξάρτητο με το μέγεθος ή το μικρόφωνο που χρησιμοποιείται, κάνοντας την διαδικασία αναγνώρισης πιο ισχυρή. Η βάση δεδομένων που χρησιμοποίησαν είχε δημιουργηθεί στο University's Media and Communication Signal Processing Laboratory με την υποστήριξη του Korea Research Institute of Standards and Science και τα ποσοστά επιτυχίας ήταν περίπου στο 76%.

Οι Kostoulas και Fakotakis το 2007 προσπάθησαν να παρουσιάσουν μια προσέγγιση για την αναγνώριση συναισθημάτων. Εστίασαν στην αξιολόγηση της τεχνικής αναγνώρισης εξαρτημένου χρήστη [177]. Χρησιμοποίησαν 20 χαρακτηριστικά και τον αλγόριθμο C4.5 για ταξινόμηση. Κατηγοριοποιήθηκαν τέσσερις κλάσεις συναισθημάτων οι οποίες ήταν ο πανικός, η λύπη, η χαρά, και ο θυμός καθώς και η ουδετερότητα. Δημιουργήθηκε ένα μοντέλο ταξινόμησης για κάθε ηθοποιό με την χρήση δεδομένων εκπαίδευσης από μια βάση και κάθε πρόταση χρησιμοποιήθηκε για την εξέταση της ακρίβειας του συστήματος. Για όλα τα μοντέλα που δημιούργησαν, πέτυχαν μια μέση ακρίβεια της τάξεως του 80,46%, η οποία δηλώνει την ποιότητα των μοντέλων που δημιουργήθηκαν. Ωστόσο η δομή που ακολούθησαν πραγματοποιούσε μια ανάλυση σε επίπεδο πρότασης για την οποία τα ποσοστά επιτυχίας ήταν στο 85,40%. Η βάση δεδομένων που χρησιμοποιήθηκε ήταν η Linguistic Data Consortium 2002S28 [149]. Άλλα συμπεράσματα στα οποία κατέληξαν είναι ότι ο θυμός και η ουδετερότητα έχουν τα υψηλότερα ποσοστά αναγνώρισης.

Οι Wang και λοιποί, προσπάθησαν μέσα από την έρευνά τους να βελτιώσουν την αυτόματη αντίληψη των φωνητικών συναισθημάτων με δύο τρόπους [155]. Πρώτον, σύγκριναν δύο πηγές συναισθηματικών δεδομένων: την φυσικότητα, τον αυθόρμητο συναισθηματικό λόγο και τον προσποιούμενο συναισθηματικό λόγο. Αυτή η σύγκριση δείχνει τα πλεονεκτήματα και τα μειονεκτήματα και των δύο μεθόδων και το πώς αυτές οι μέθοδοι

επιδρούν στην αναγνώριση συναισθημάτων στην τελική εφαρμογή. Δεύτερον, ασχολήθηκαν με δύο μεθόδους ταξινόμησης οι οποίες δεν έχουν εφαρμοστεί σε αυτό το πεδίο. Την επιτηδευμένη γενικοποίηση (stacked generalization) και την ψήφο χωρίς βαρύτητα (unweighted vote). Έδειξαν πώς αυτές οι τεχνικές μπορούν να βοηθήσουν στην εξέλιξη των παραδοσιακών μεθόδων ταξινόμησης. Η βάση δεδομένων που χρησιμοποιήθηκε ήταν η ESMBS (Emotional Speech of Mandarin and Burmese Speakers) και η βάση NATURAL της οποίας η συλλογή έγινε από τηλεφωνικά κέντρα. Οι ταξινομητές που χρησιμοποιήθηκαν ήταν μέσα από το WEKA [150]. Τα αποτελέσματα της ταξινόμησης έδειξαν ότι με την δεύτερη μέθοδο ταξινόμησης τα αποτελέσματα ήταν καλύτερα στην NATURAL βάση δεδομένων ενώ η πρώτη μέθοδος απέδωσε περισσότερα στην ESMBS. Στην βάση NATURAL τα ποσοστά επιτυχίας για την ουδετερότητα ήταν στο 60% ενώ για τον θυμό ήταν στο 40%. Το πρόβλημα που αντιμετώπισαν στην χρήση του αυθόρμητου λόγου ήταν η έλλειψη ελέγχου η οποία οδηγεί σε μη ισορροπημένα δεδομένα, που συνήθως έχουν έναν παρασκηνακό θόρυβο πράγμα που κάνει την αναγνώριση κάποιων συναισθημάτων δύσκολη. Τέλος, κατέληξαν ότι ο προσποιούμενος λόγος μπορεί να δίνει μεγαλύτερα ποσοστά επιτυχίας σε σχέση με τον αυθόρμητο, αλλά ο αυθόρμητος είναι αυτός που δίνει την εικόνα που είναι πιο κοντά στην πραγματικότητα.

Οι Talieh Seyed Tabatabaei, Sridhar Krishnan και Aziz Guergachi πρότειναν την χρήση πρωτότυπων ηχητικών χαρακτηριστικών και Μηχανών Διανυσματικής Υποστήριξης και Ελάχιστης Ευθυγράμμισης (Least Square-Support Vector Machine) στην υλοποίηση ενός ανεξάρτητου και αυτόματου συστήματος αναγνώρισης ανθρώπινων συναισθημάτων [1]. Έξι διακριτές συναισθηματικές καταστάσεις ταξινομηθήκαν μέσα από την έρευνά τους, η ευτυχία, η λύπη, ο θυμός, ο φόβος, η έκπληξη και η αηδία. Η βάση δεδομένων που χρησιμοποιήθηκε ήταν η «The Enterface'05 audio-visual emotion database». Εφαρμόστηκαν διαφορετικές μέθοδοι SVM για την επίτευξη του καλύτερου αποτελέσματος. Τα αποτελέσματα του ταξινομητή LS-SVM είναι συγκρίσιμα με τα αποτελέσματα κάποιου Γραμμικού Ταξινομητή, αφού τα ποσοστά επιτυχίας έφτασαν το 81,3%.

Οι Kostoulas και λοιποί στην έρευνά τους εν μέρη υποστηρίχτηκαν από την έρευνα LOGOS η οποία χρηματοδοτήθηκε από την Γενική Γραμματεία Έρευνας και Τεχνολογίας του Ελληνικού Υπουργείου Ανάπτυξης [177]. Σε αυτή την έρευνα αξιολόγησαν ένα ανιχνευτή αρνητικών συναισθηματικών καταστάσεων με σκοπό να προβάλλουν ένα σύστημα διαλόγου δια ομιλίας που να λειτουργεί σε περιβάλλον έξυπνου σπιτιού. Μεικτά μοντέλα μονάδων μαγνητικής επαγωγής και μια σειρά από συχνούς παράγοντες που χρησιμοποιούνται στην ομιλία, χρησιμοποιήθηκαν και στην προτεινόμενη σύνθεση. Αξιολόγησαν το σύστημα σε

στοιχεία πραγματικών καταστάσεων που συλλέχθηκαν από την διάδραση χρηστών που δεν γνώριζαν, με το σύστημα αυτό διαλόγου ομιλίας στο έξυπνο σπίτι. Για την διεξαγωγή του πειράματος χρησιμοποιήθηκε και μια βάση δεδομένων προσποιούμενου λόγου (The Linguistic Data Consortium 2002S28, Emotional Prosody Speech and Transcripts) [149]. Τα πειραματικά αποτελέσματα τους έδειξαν ότι η ακρίβεια αναγνώρισης αρνητικών συναισθημάτων στα πραγματικά δεδομένα είναι χαμηλότερη από εκείνη που αναφέρεται όταν εξετάζεται σε βάση προσποιούμενων δεδομένων. Παρά τα χαμηλά αποτελέσματα, η πειραματική αυτή μέθοδος είναι πολλά υποσχόμενη αφού συχνά ούτε οι άνθρωποι δεν είναι σε θέση να διαχωρίσουν τα συναισθήματά κρίνοντας μόνο από την ομιλία.

Πέντε βασικά συναισθήματα, όπως η ευτυχία, ο θυμός, η πλήξη, ο φόβος και η θλίψη συζητούνται και αναγνωρίζονται μέσω ενός πλαισίου που συνδυάζει την ανάλυση κύριων τμημάτων (Principal Component Analysis PCA) και το νευρωνικό δίκτυο οπισθοδιάδοσης (Back Propagation Neural Network) στην εργασία των Wang και λοιπών [151]. Όσον αφορά την PCA, οι συντάκτες έκαναν την υπόθεση ότι υπάρχει γραμμική σχέση μεταξύ των μεταβλητών, και έτσι μειώθηκαν από 43 σε 11 οι μεταβλητές που αντιπροσωπεύουν μια ορισμένη συναισθηματική κατάσταση. Δύο μοντέλα νευρωνικών δικτύων χρησιμοποιήθηκαν και συγκρίθηκαν το one class one Network(OCON)και All class one Network (ACON). Το αποτέλεσμα κυμαίνεται 52%-62% και είναι υποσχόμενο για την αναγνώριση συναισθημάτων από την ομιλία.

Ο Zhou και λοιποί παρουσιάζουν μία προσέγγιση χρησιμοποιώντας χαρακτηριστικά άρθρωσης Articulatory Features(AFs), τα οποία πηγάζουν από φασματικές μετρήσεις για την αναγνώριση συναισθημάτων [152]. Τα αποτελέσματα των πειραμάτων τα οποία βασίζονται στην CASIA Mandarin συναισθηματική βάση δεδομένων δείχνουν ότι τα AFs μόνα τους δεν είναι κατάλληλα για αναγνώριση συναισθημάτων(ER). Επιπροσθέτως, ο συνδυασμός των AFS και των φασματικών χαρακτηριστικών δείχνει ότι δεν είναι αρκετά ώστε να βελτιώσουν την αποδοχή του συστήματος. Αν και τα AFs έχουν υιοθετηθεί ως εναλλακτικά ή συμπληρωματικά χαρακτηριστικά για αναγνώριση φωνής, αναγνώριση γλώσσας και μέτρηση εμπιστοσύνης, δεν είναι αποτελεσματικά για αναγνώριση συναισθημάτων.

Ο Shah και λοιποί έχουν αναλύσει την αποτελεσματικότητα μιας μηχανής αυτόματης ER [153]. Οι συγγραφείς έχουν χρησιμοποιήσει και φασματική ανάλυση και τεχνικές διαμορφώσης για την εξαγωγή χαρακτηριστικών καθώς και ταξινομητή πολυεπίπεδου δικτύου αισθητήρων (MLP). Κατασκεύασαν ένα σετ δεδομένων ανεξάρτητο από ομιλητή και φύλο για το πείραμα τους, το οποίο αποτελείται από εφτακόσιες εκφράσεις οι οποίες χωρίζονται σε τέσσερα διαφορετικά συναισθήματα (θυμός, χαρά, λύπη και ουδέτερο), χρησιμοποιώντας και

διακριτό μετασχηματισμό κύματος (DWT) και MFCCs για ταξινόμηση και αναγνώριση. Η συνολική απόδοση αναγνώρισης φτάνει το 68.5% και δείχνει ότι το DWT δίνει πάρα πολύ καλά ποσοστά αναγνώρισης για το θυμό και τα MFCCs δίνουν πολύ καλά ποσοστά αναγνώρισης για τη χαρά και το θυμό επίσης. παρόμοια με αυτή την μέθοδο, στην εργασία του Yang και λοιπών η αναπαράσταση των χαρακτηριστικών βασίστηκε σε μη αρνητική μίτρα παραγοντοποίησης (nonnegative matrix factorization (NMF)) και τα αποτελέσματα είναι 90% και 77% για εξαρτώμενο από ομιλητή και ανεξάρτητα από ομιλητή αντίστοιχα [154].

Ο Yang και οι λοιποί μελέτησαν το πρόβλημα χρησιμοποιώντας Twins support vectrol Machines (TWINs SVM) [154]. Διάφορα ακουστικά χαρακτηριστικά όπως ο ρυθμός ομιλίας, προφορά, προσωδία, τόνος, ενέργεια, συμφραζόμενο και περιεχόμενο από διαφορετικές συναισθηματικές εκφράσεις ομιλίας χρησιμοποιήθηκαν για να διακρίνουν επτά βασικά ανθρώπινα συναισθήματα όπως θυμός, πλήξη, αηδία, φόβος/αγωνία, χαρά, ουδέτερο, λύπη. Συγκρίσεις μεταξύ των δύο αλγόριθμων ταξινόμησης των TWINs SVM και του Standartd SVM δείχνουν ότι ο πρώτος έχει οριακά καλύτερη απόδοση. Επιπροσθέτως, πειράματα που έγιναν για να αξιολογήσουν την ορθότητα των συναισθηματικών χαρακτηριστικών σε διαφορετική γλώσσα δείχνουν ότι τα χαρακτηριστικά της ομιλίας που έχουν επιλεγεί δεν είναι τα πλέον κατάλληλα για ER σε διαφορετική γλώσσα.

Από τους Zhou και λοιπούς έγινε μία έρευνα για την αναγνώριση και μετατροπή συναισθημάτων από την ομιλία [155]. Η συναισθηματική Mandarin βάση δεδομένων η οποία προέρχεται από το Κινέζικο LDC (Linguistic Data Consortium) χρησιμοποιήθηκε και περιέχει μικρές εκφράσεις που περιλαμβάνουν τα πέντε βασικά συναισθήματα: θυμός, χαρά, έκπληξη, ουδέτερο και λύπη. Οι συγγραφείς ερευνούν τη μάντα συχνότητων της κανονικής κατανομής ενός ατόμου χρησιμοποιώντας την F-Αναλογία. Σύμφωνα με την έρευνα τους, σχεδίασαν μη ομοιόμορφα φίλτρα ζώνης για να εξάγουν δύο ακόμη χαρακτηριστικά από τα MFCCs και από τη μέθοδο γραμμικής πρόβλεψης(Linear Predictive method). Πειράματα αναγνώρισης συναισθημάτων χρησιμοποιώντας GMM ταξινομητή, έδειξαν ότι τα χαρακτηριστικά που εξήχθησαν με την προτεινόμενη μέθοδο υπερέχουν από τα παραδοσιακά MFCC και PLP χαρακτηριστικά. Η συνολική απόδοση ήταν 86.5%.

Μαζί με τις κλασικές μετρήσεις όπως η θεμελιώδης συχνότητα, η ενέργεια και τα Formants, ο Yang και λοιποί υπολόγισαν την αποδοτικότητα του ρυθμού ομιλίας σαν χαρακτηριστικό ταξινόμησης [156]. Παρόλο που χρησιμοποιήθηκε η συναισθηματική βάση δεδομένων του Βερολίνου (EMO-DB) για τα πειράματα, μόνο πέντε συναισθήματα από αντρική ομιλία μελετήθηκαν. Τα διαισθητικά ασαφή σετ Intuitionistic Fuzzy Sets (IFS) χρησιμοποιήθηκαν για να λύσουν το πρόβλημα της ER και η μέθοδος αυτή έδειξε ότι είναι

πολύ αποδοτικά στην επεξεργασία φωνής. Σύμφωνα με τα αποτελέσματα της ταξινόμησης η μέση απόδοση ER είναι μόλις 69% για τα πέντε συναισθήματα που μελετήθηκαν.

Μια συναισθηματική βάση δεδομένων ψιθυριστής ομιλίας με πέντε συναισθήματα όπως χαρά, θυμός, έκπληξη, λύπη και ηρεμία δημιουργήθηκε από τον Jin και λοιπούς [157]. Με την μέθοδο της συνάρτησης συσχέτισης και της διαδικασίας αναλυτικής ιεραρχίας (Analytic Hierarchy Process) των ασαφών μαθηματικών (Fuzzy), δοκίμασαν τις ηχογραφημένες συναισθηματικές προτάσεις έτσι ώστε να είναι ποιοτική η συγκεκριμένη βάση δεδομένων ψιθύρων. Συγκρίνοντας την διαφορά των κυματομορφών και των φασματογραφήματων κανονικής και ψιθυριστής ομιλίας και συγκρίνοντας τα φασματογραφήματα της ίδια πρότασης με τέσσερα διαφορετικά συναισθήματα αυτή η μελέτη καταλήγει στο συμπέρασμα ότι η διάρκεια και η ενέργεια της πρότασης, που είναι σημαντικοί παράμετροι στην κανονική ομιλία, μπορούν επίσης να χρησιμοποιηθούν και στην ομιλία ψιθύρων. Επιπροσθέτως, και κάποια άλλα χαρακτηριστικά που σχετίζονται με την θεμελιώδη συχνότητα μπορούν να τροποποιηθούν ή να αντικατασταθούν.

Ένα σύστημα SER δημιουργήθηκε βασισμένο σε ένα κβαντικού διανύσματος-τεχνητό νευρωτικό δίκτυο Vector Quantization-Artificial Neural Network (VQ-ANN) από τον Wenjing και λοιπούς [158]. Μια συναισθηματική βάση δεδομένων που περιέχει θυμό, χαρά, λύπη και έκπληξη ηχογραφήθηκε και χρησιμοποιήθηκε. Τα κείμενα για κάθε συναίσθημα δεν είναι όλα ίδια και όλα περιέχουν διαφορετικούς τύπους προτάσεων, τόνους συλλαβών και προφοράς. Το σύστημα πρώτα εξάγει τις βασικές παραμέτρους προσωδίας και τα MFCCs. Ένας συνδυασμός από VQ-based MFCCs μετρικές και οι στατιστικές μετρήσεις των παραμέτρων προσωδίας χρησιμοποιήθηκαν για την δημιουργία διανύσματος χαρακτηριστικών σαν είσοδος στο σύστημα. Το ANN χρησιμοποιήθηκε για να επεξεργαστεί το συνδυασμό χαρακτηριστικών και των στατιστικών μετρήσεων όλων των εξαγόμενων παραμέτρων αντίστοιχα. Τα αποτελέσματα του πειράματος δείχνουν ότι ο συνδυασμός των χαρακτηριστικών υπερτερεί των στατιστικών μετρήσεων. Πιο λεπτομερής ανάλυση δείχνει ότι αυτά τα χαρακτηριστικά μπορούν να χαρακτηρίσουν τον χώρο των συναισθημάτων καλύτερα από τα στατιστικά χαρακτηριστικά.

Οι Yun και Yoo παρουσιάζουν μια μέθοδο για την SER με ένα μεγάλο περιθωρίου πλαίσιο ενσωματώνοντας μια loss function βασισμένη στο πρότυπο των Tellegen και Watson [159][160]. Τα χαρακτηριστικά του συναισθήματος αποτελούνται από 39 διαστάσεις βασισμένες στα 12 MFCCs, την ενέργεια και στους αντίστοιχους συντελεστές Δέλτα και επιτάχυνσής τους. Κάθε συναίσθημα διαμορφώνεται από ένα single state HMM που εκπαιδεύεται με την μεγιστοποίηση του ελάχιστου διαχωριστικού περιθωρίου μεταξύ των συναισθημάτων, ενώ το περιθώριο κλιμακώνεται με μια loss function. Οι Yun και Yoo

καθορίζουν ένα μέτρο απόστασης μεταξύ των συναισθημάτων, τη συνάρτηση απώλειας loss function, η οποία υπολογίζεται από τον πρώτο κανόνα του μέτρου απόστασης [159]. Η συναισθηματική βάση δεδομένων του Βερολίνου χρησιμοποιήθηκε προκειμένου να εφαρμοστούν τα πειράματα για αυτό το πλαίσιο. Τα αποτελέσματα αναφέρουν ότι το πλαίσιο μέγιστου περιθωρίου που ενσωματώνει τη συνάρτηση απώλειας, ονομάζεται margin scaling (MS), αποδίδουν καλύτερα από άλλα συμβατικά κριτήρια εκπαίδευσης για HMM, όπως η μέγιστη πιθανότητα (maximum likelihood (ML)) και η εκτίμηση της μέγιστης αμοιβαίας πληροφορίας (maximum mutual information (MMI)). Η συνολική απόδοση φτάνει το 88.9%.

Από τον Giannakopoulos και λοιπούς 1500 ακουστικά δείγματα φωνής από περισσότερες από 30 ταινίες διαφόρων ειδών όπως κωμωδία, θρίλερ κλπ [161]. Ο στόχος της έρευνας ήταν διπλός. Ο πρώτος στόχος ήταν να ερευνηθούν εάν ο τροχός συναισθημάτων (Emotion Wheel (EW)) προσφέρει μια καλή αναπαράσταση για τις συναισθήματα που συνδέονται με τα σήματα ομιλίας. Για αυτόν τον λόγο, διάφοροι άνθρωποι έχουν σχολιάσει τα δεδομένα φωνής από τις ταινίες χρησιμοποιώντας τον EW και το επίπεδο διαφωνίας έχει υπολογιστεί ως μέτρο της ποιότητας αναπαράστασης. Η πειραματική αξιολόγηση δείχνει ότι το μέσο λάθος χρηστών είναι μικρό, που σημαίνει ότι η διαφωνία των χρηστών είναι επίσης χαμηλή, και συνεπώς το EW μπορεί να χρησιμοποιηθεί στα πλαίσια του προσδιορισμού των συναισθημάτων στα σήματα ομιλίας. Μετά οι συγγραφείς επέλεξαν τον K-NN ταξινομητή για να προβλέψουν τη θέση ενός άγνωστου σήματος ομιλίας στον EW. Κάθε τμήμα φωνής αναπαρίσταται με ένα διάνυσμα δέκα ακουστικών χαρακτηριστικών όπως MFCC, Fast Fourier Transform of the signal, zero crossing rate, pitch, and spectral measurements. Το ομαλοποιημένο λάθος (normalized error) αυτής της προσέγγισης παλινδρόμησης βρέθηκε για να είναι ίσο με 0.97, το οποίο είναι πολύ κοντά στο μέσο ανθρώπινο λάθος. Τα αποτελέσματα δείχνουν ότι αυτή η αρχιτεκτονική μπορεί να υπολογίσει τις καταστάσεις συναισθημάτων από τις φωνές των ταινιών με ικανοποιητική ακρίβεια.

Ο Mao και λοιποί προτείνουν έναν υβριδικό ταξινομητή SER ο οποίος χρησιμοποιεί HMMs και ANN για πληροφορία στο επίπεδο έκφρασης και τμηματοποίησης από την ομιλία [162]. Για να αξιολογήσουν την απόδοση της προτεινόμενης προσέγγισης, δύο βάσεις συναισθημάτων ομιλίας χρησιμοποιήθηκαν η Berlin Speech Emotional Database (EMO-DB) και η Beihang University Database of Emotional Speech (BHUEDS), οι οποίες περιέχουν πέντε συναισθήματα.

Η δομή αυτού του υβριδικού συστήματος περιλαμβάνει έξι μέρη: εξαγωγή ακουστικών χαρακτηριστικών, μοντελοποίηση του συναισθήματος της φωνής με HMMs (όμοια με το [163]), την κατηγοριοποίηση φωνημάτων, την κανονικοποίηση ακολουθίας, διόρθωση

παραμόρφωσης κι αναγνώριση συναισθήματος με ANN. Τα πειράματα έχουν δείξει ότι η βέλτιστη απόδοση (71.6% και 78.9% για το emo-DB και BHUDES αντίστοιχα) των απομονωμένων HMMs επιτεύχθηκε από ένα σετ ακουστικών χαρακτηριστικών που περιέχει την πρώτη και δεύτερη παράγωγο της θεμελιώδους συχνότητας, πρώτη και δεύτερη παράγωγο του πλάτους ενέργειας, τα δέκα πρώτα LPCC και δώδεκα πρώτα MFCC. Όταν το υβριδικό HMM/ANN μοντέλο δοκιμάστηκε η συνολική απόδοση βελτιώθηκε κατά 3% και για τις δύο βάσεις δεδομένων. Σε παρόμοια εργασία ο Fu και λοιποί, παρόμοια διανύσματα χαρακτηριστικών που εισήχθησαν φάνηκαν να είναι πιο αποδοτικά έναντι των απόλυτων διανυσμάτων χαρακτηριστικών γνωρισμάτων καθώς ο θόρυβος από τη διαφορά της ακουστικής κάθε απόμου μειώνεται [164].

Στην εργασία των Mishra και Sekhar παρουσιάζεται η δυνατότητα εφαρμογής των μεταβαλλόμενων Gaussian Mixture Models (GMMs) στη SER [165]. Τα μεταβαλλόμενα GMMs επιλέγουν τον αριθμό των mixture στοιχείων αυτόματα κόβοντας τα περιττά στοιχεία. Τα μεταβαλλόμενα GMMs είναι επίσης απαλλαγμένα από τα προβλήματα ιδιομορφίας που προκύπτουν συχνά σε GMMs.

Σε αυτήν την μελέτη οι συντάκτες χρησιμοποίησαν δύο προσεγγίσεις και τις σύγκριναν με GMM. Στην πρώτη προσέγγιση, υπέθεσαν τα priors για τις παραμέτρους mixture βάρη, μέσες τιμές και μήτρες ακρίβειας και στη δεύτερη προσέγγιση τα priors υπολογίστηκαν μόνο για τη μέση τιμή και για τον μήτρα ακρίβειας. 34 διαστάσεων διανύσματα ακουστικών χαρακτηριστικών αποτελούμενα από 16 MFCCs, 16 delta-MFCCs, ενέργεια and delta-energy μετρήσεις, δοκιμάστηκαν στην Berlin Emotional Speech database (EMO-DB). Αναφέρεται ότι τα μεταβαλλόμενα GMMs αποδίδουν καλύτερα από τις μεθόδους μέγιστης πιθανότητας (ML) και ότι η σύγκλιση μεταβαλλόμενων GMM είναι γρηγορότερη από GMM. Επίσης αποδεικνύεται ότι τα μεταβαλλόμενα GMMs αποδίδουν καλύτερα ακόμα και για μικρή ποσότητα δεδομένων.

Μια μέθοδος για την αναγνώριση των συναισθημάτων βάσει του σθένους παρουσιάζεται στο [166]. Ο Yao συγκρίνει τρεις διαφορετικούς αλγορίθμους ταξινόμησης για πέντε συναισθήματα από δύο σετ δεδομένων ομιλίας Μανδαρινών. Τα αποτελέσματα δείχνουν ότι τα HMMs υπερτερούν από τις δύο άλλες τεχνικές ταξινόμησης (KNN and Linear discriminate analysis) από 3-6%. Επιπλέον, τα τεσσάρων διαστάσεων διακριτά HMMs με το επιλεγμένο σετ χαρακτηριστικών προσωδίας, πέτυχαν υψηλή ακρίβεια στο να διακρίνει συναισθήματα θυμού/χαράς ή πλήξης/λύπης τα οποία έχουν παρόμοια προσωδία. Τα πειράματα αναγνώρισης συστάδων δείχνουν ότι τα επιλεγμένα χαρακτηριστικά γνωρίσματα είναι σημαντικά για να διακρίνουν τις συστάδες θυμού/χαράς, ουδέτερου και πλήξης/λύπης.

Η επιλογή ακουστικών χαρακτηριστικών από σήματα φωνής παρουσιάζεται αναλυτικά από τον Cen και λοιπούς [167]. Η προτεινόμενη μέθοδος επιλογής ακουστικών χαρακτηριστικών χρησιμοποιεί την ανάλυση κανονικής συσχέτισης Canonical Correlation Analysis (CCA) για να αφαιρέσει τις άσχετες πληροφορίες. Μετά την CCA το σετ των ακουστικών χαρακτηριστικών μειώνεται κατά 72.89% (46 χαρακτηριστικά από τα αρχικά 166), και φαίνεται να είναι το πιο αποδοτικό σύμφωνα με αυτή την έρευνα. Το σετ αυτό περιλαμβάνει στατιστικά της θεμελιώδους συχνότητας, της έντασης και των MFCCs παρόμοια με την εργασία των Cheng και Zhao [168]. Τα πειράματα έχουν πραγματοποιηθεί χρησιμοποιώντας ένα πιθανοτικό νευρικό δίκτυο (PNN), το οποίο ταξινομεί τα συναισθήματα των εκφράσεων που είναι ηχογραφημένες στην LDC βάση δεδομένων. Η μέγιστη απόδοση φτάνει περίπου το 60% και στα εξαρτώμενα/ανεξάρτητα πλαίσια ομιλητών και φύλου. Παρόμοιες προσεγγίσεις παρουσιάζονται στα papers [169][170][171].

Οι Atassi και Esposito προτείνουν μια ενδιαφέρουσα προσέγγιση για την αυτόματη αναγνώριση συναισθημάτων στο πλαίσιο ανεξάρτητου ομιλητή [172]. Η κύρια ιδέα είναι να χωριστεί η διαδικασία αναγνώρισης συναισθημάτων σε δύο βήματα. Στο πρώτο βήμα, ένας συνδυασμός επιλεγμένων ακουστικών χαρακτηριστικών γνωρισμάτων χρησιμοποιείται για να ταξινομήσει έξι συναισθήματα όπως πλήξη, φόβος, χαρά, λύπη και το ουδέτερο από την Berlin Speech Emotional Βάση δεδομένων μέσω ενός GMM ταξινομητή. Γι' αυτό το σκοπό, ακουστικά χαρακτηριστικά προσωδίας και ποιότητας φωνής όπως η θεμελιώδης συχνότητα, η ενέργεια, η zero crossing αναλογία και η αρμονικότητα μελετήθηκαν.

Κατόπιν, τα δύο συναισθήματα με τη μεγαλύτερη πιθανότητα επιλέγονται για περαιτέρω επεξεργασία στο δεύτερο βήμα για μια ανταγωνιστική ταξινόμηση μεταξύ τους. Για κάθε πιθανό ζευγάρι(π.χ. χαρά/λύπη, θυμός/ουδέτερο κτλ.) το καλύτερο σετ ακουστικών χαρακτηριστικών που μπορεί να διακρίνει ένα από τα δύο συναισθήματα προσδιορίζεται μέσω του αλγόριθμου Sequential Floating Forward Selection (SFFS), ακολουθούμενο από ένα ξεχωριστό GMM, για την τελική απόφαση. Το μέσο αποτέλεσμα ταξινόμησης φτάνει το 80.7%, το οποίο είναι καλύτερο από την αντίστοιχη απόδοση (74.5%) μιας παρόμοιας εργασίας Lugger και Yang στην ανεξάρτητη από τον ομιλητή αναγνώριση των συναισθηματικών φωνητικών εκφράσεων [173].

Η κυβική παρεμβολή (cubic spline) χρησιμοποιήθηκε για να ταιριάζει τις καμπύλες των χαρακτηριστικών της προσωδίας τα οποία εξήχθησαν από τα σήματα ομιλίας στην εργασία των Wang και λοιπών [174]. Κατόπιν, ελήφθησαν οι μετρήσεις των παραγώγων αυτών των καμπυλών, συλλέγοντας συνολικά 23 χαρακτηριστικά γνωρίσματα βασισμένα στο πλάτος, την μικρής χρονικής διάρκειας ενέργεια και τη θεμελιώδη συχνότητα. Η διαδικασία

έχει 2 στάδια: το στάδιο της επιλογής χαρακτηριστικών γνωρισμάτων μέσω ενός γενετικού αλγορίθμου (GA) και το στάδιο της ταξινόμησης που χρησιμοποιεί έναν SVM. Με βάση τη βέλτιστη αναζήτηση του γενετικού αλγορίθμου και τη διάκριση των συναισθημάτων οι συγγραφείς ήταν σε θέση να ανακαλύψουν τα κατάλληλα χαρακτηριστικά για κάθε συναίσθημα. Αυτό το σύστημα αναγνώρισης συναισθημάτων επέτυχε ένα μέσο ποσοστό 82.8% σε ένα σετ δεδομένων το οποίο περιέχει εκφράσεις αντρών και γυναικών με έξι συναισθήματα.

Ο Pao και λοιποί χρησιμοποίησαν ένα σύστημα πολλαπλών ταξινομητών χρησιμοποιήθηκε για να πετύχει τη μέγιστη δυνατή ταξινόμηση συναισθημάτων στη Mandarin βάση δεδομένων η οποία περιέχει πέντε συναισθήματα [175]. Οι ταξινομητές που ερευνήθηκαν από τους συγγραφείς περιλαμβάνουν τον K- Nearest Neighbor (KNN), Weighted KNN (WKNN), Weighted Average Patterns of Categorical KNN (WCAP), Weighted Discrete KNN (W-DKNN) και SVM. Η συνολική απόδοση σωστής ταξινόμησης υπερτερεί της μεθόδου ταξινόμησης με ένα ταξινομητή και φτάνει το 83.8%.

Ο Pao και λοιποί παρουσιάζουν ένα σύστημα αναγνώρισης συναισθημάτων για να συγκρίνουν διάφορους ταξινομητές από την καθαρή και θορυβώδη ομιλία [176]. Πέντε συναισθήματα όπως θυμός, χαρά, λύπη, ουδέτερο και πλήξη ερευνήθηκαν από τις Mandarin βάσεις δεδομένων. Οι ταξινομητές που χρησιμοποιήθηκαν είναι οι KNN, WCAP, GMM, HMM και τα W-DKNN. Μετά την επιλογή των ακουστικών χαρακτηριστικών το βέλτιστο σετ χαρακτηριστικών περιλαμβάνει κυρίως MFCCs, Linear Prediction Cepstral Coefficients (LPCC) και Linear Predictive Coefficients (LPC). Τα αποτελέσματα της έρευνας έδειξαν ότι το προτεινόμενο W-DKNN υπερτερεί για κάθε SNR ομιλία μεταξύ των τριών KNN-βασισμένων ταξινομητών και επέτυχε την υψηλότερη ακρίβεια από την καθαρή ομιλία την 20dB θορυβώδη ομιλία.

Στην εργασία του Kostoulas και λοιποί ένας ανιχνευτής αρνητικών συναισθημάτων negative emotional states (DNES) αξιολογείται, αυτό υπηρετεί το σκοπό της βελτίωσης ενός συστήματος διαλόγου που λειτουργεί σε ένα περιβάλλον έξυπνου σπιτιού [177]. Η αρνητική κατηγορία αποτελείται από τις εκφράσεις που χαρακτηρίστηκαν ως θυμός ή έντονος θυμός και μια ερευνητική κατηγορία από τις εκφράσεις που χαρακτηρίζονται ως ουδέτερο, χαρά, ευτυχία, ευχαρίστηση. Το DNES βασίζεται στα Gaussian mixture models και ένα σύνολο 16 προσωδικών χαρακτηριστικών όπως θεμελιώδης συχνότητα, ενέργεια, αρμονικότητα (harmonics-to-noise ratio) και τα δεκατρία πρώτα MFCCs. μετά την απομάκρυνση των τμημάτων που δεν περιέχουν ομιλία, μόνο τα διανύσματα ακουστικών χαρακτηριστικών που αντιπροσωπεύουν τα τμήματα που περιέχουν ομιλία χρησιμοποιήθηκαν

για τη δημιουργία ενός GMM για τις δύο κατηγορίες. Τα αποτελέσματα του πειράματος δείχνουν ότι τα ποσοστά αναγνώρισης αρνητικών συναισθημάτων στον πραγματικό κόσμο είναι χαμηλότερα από αυτά στην προσποιούμενη ομιλία. Αντί της επιλογής χαρακτηριστικών γνωρισμάτων και της μείωσης του διαστήματος χαρακτηριστικών γνωρισμάτων από γνωστές μεθόδους.

Ο Schuller και λοιποί προτείνουν τη διεύρυνση του διανύσματος ακουστικών χαρακτηριστικών παράγοντας καινούργια χαρακτηριστικά βασισμένα στα ήδη υπάρχοντα [178]. Σύμφωνα με αυτή την έρευνα, η μετατροπή των χαρακτηριστικών με μαθηματικές πράξεις μπορεί να οδηγήσει σε καλύτερη αναπαράσταση αυτών. Το σετ ακουστικών χαρακτηριστικών ξεκινάει με δυναμικούς χαμηλού επιπέδου περιγραφείς όπως ο τόνος, η ένταση, τα formants και η φασματική πληροφορία. Κατόπιν, η συστηματική παραγωγή της προσωδίας, της άρθρωσης και της ποιότητας φωνής πραγματοποιήθηκε με περιγραφική στατιστική ανάλυση. Μετά την περιγραφική στατιστική ανάλυση το σετ χαρακτηριστικών ενισχύεται με αυτόματες μετατροπές των χαρακτηριστικών για να βρεθεί η βέλτιστη αναπαράσταση στο χώρο των χαρακτηριστικών με σκοπό την ταξινόμηση. Η αναζήτηση αυτή εκτελείται μετά από τις αρχές του εξελικτικού προγραμματισμού. Αυτή η απολύτως διαφορετική προσέγγιση αναφέρεται ότι βελτιώνει την απόδοση ταξινόμησης από τις προηγούμενες εργασίες των συγγραφέων χρησιμοποιώντας δύο δημόσιες βάσεις δεδομένων.

Στην εργασία των Shami και Kamel κάθε έκφραση θεωρείται ως μια σειρά από ξεχωριστά τμήματα φωνής κι όχι ως μια οντότητα [179]. Τα μέρη της φωνής τμηματοποιούνται, ακολουθούμενα από τις στατιστικές μετρήσεις του φασματικού σχήματος, της έντασης και της θεμελιώδους συχνότητας. Η τελική ταξινόμηση γίνεται συνθέτοντας τις αποφάσεις ταξινόμησης κάθε τμήματος φωνής χρησιμοποιώντας ένα σταθερό σύστημα σύνθεσης.

Τέλος, υπάρχει μια πρόσθετη υποκατηγορία στις παραγλωσσολογικές μεθόδους που αφορά τις ανθρώπινες φωνήσεις (που αναφέρονται συχνά ως μη-γλωσσικές φωνήσεις). Τα γέλια, οι κραυγές, οι στεναγμοί, τα χασμουρητά, και άλλα τέτοια φωνητικά ξεσπάσματα φαίνονται καταρχάς να είναι καλά παραδείγματα των εκφράσεων κάποιου συναισθήματος. Ένα αστείο προκαλεί διασκέδαση, η οποία παράγει ένα γέλιο. Μια απώλεια προκαλεί θλίψη, η οποία παράγει ένα κλάμα. Μια μονότονη διάλεξη προκαλεί πλήξη, η οποία παράγει ένα χασμουρητό. Είναι ενδιαφέρον να σημειωθεί ότι μερικά συναισθήματα όπως η απογοήτευση και η πλήξη φαίνονται να είναι ευπροσδιόριστα από τις μη-λεκτικές εκφράσεις όπως τους στεναγμούς και τα χασμουρητά. Λίγες προσπάθειες προς την αυτόματη αναγνώριση των μη λεκτικών εκφράσεων όπως τα γέλια [180][181] και κλάματα [182][183][184] έχουν επίσης

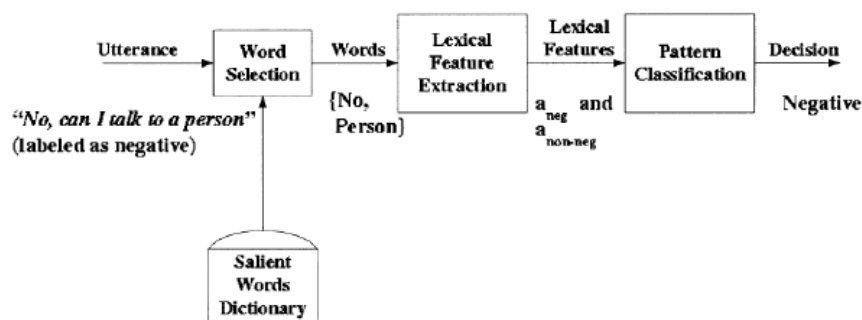
αναφερθεί πρόσφατα. Επιπλέον, συγκεκριμένες φωνητικές εκφράσεις και εκφράσεις προσωδίας έχουν συνδεθεί με συγκεκριμένες διανοητικές ασθένειες όπως η ψυχαναγκαστική αναταραχή (OCD) [185], Tourette Σύνδρομο (TS) [186] ή κατάθλιψη [187]. Αυτές οι εκφράσεις φανερώνονται συνήθως ως φωνητικές νύξεις του φόβου και της θλίψης, που συνοδεύονται συνήθως με τις εκφράσεις του προσώπου. Εντούτοις, καμία προσπάθεια προς την αποτελεσματική αυτόματη ανάλυση αναγνώρισης συναισθημάτων βασισμένη στα φωνητικά ξεσπάσματα δεν έχει αναφερθεί μέχρι τώρα.

5.1.2 Γλωσσολογική- Λεκτική Πληροφορία

Σε μια έρευνα τους ο Batliner και οι λοιποί πρότειναν ότι, εκτός από τα προσωδιακά χαρακτηριστικά της ομιλίας και διάφορα συντακτικά και συμπεριφοριστικά χαρακτηριστικά, όπως οι επαναλήψεις σε ένα διάλογο και το τι μέρος του λόγου είναι μια λέξη (part-of-speech features (POS)) πρέπει να μελετηθούν (εικόνα 3) [188]. Σύμφωνα με αυτούς οι επαναλήψεις κι οι αναδιατυπώσεις (χρήση των διαφορετικών λέξεων για να αποδώσουν το ίδιο νόημα) έχουν βρεθεί για να είναι δείκτες αλλαγής της συναισθηματικής κατάστασης των ομιλητών, όπως η αυξανόμενη απογοήτευση και ο θυμός. Επιπλέον, το POS κάθε λέξης αναπαριστά μια συντακτική δομή η οποία αποτελείται από έξι κατηγορίες όπως ουσιαστικό, ρήμα, συνδετικό ρήμα, αντωνυμία, κλιτό και μη κλιτό επίθετο. Το POS κάθε λέξης έχει γραφτεί σε ένα λεξικό που περιέχει όλους τους τύπους λέξεων που βρέθηκαν στη βάση δεδομένων. Η έρευνα αποκαλύπτει ότι τα ουσιαστικά και τα επίθετα είναι πιο χρήσιμα στην κατηγοριοποίηση των συναισθημάτων υποστηρίζοντας την άποψη ότι γενικά, οι λέξεις με περιεχόμενο (content words) είναι πιο σημαντικές και χαρακτηρίζουν πιο εύκολα το συναίσθημα από τις λέξεις λειτουργίας (function words), όπως τα ρήματα.

Ο Lee και Narayanan χρησιμοποίησαν ένα συνδυασμό τριών χαρακτηριστικών (ακουστικών, λεκτικών και ομιλίας) για την αναγνώριση συναισθημάτων [189]. Για τη σύλληψη της συναισθηματικής πληροφορίας στο επίπεδο της γλώσσας, χώρισαν τα συναισθήματα σε δύο υπερκλάσεις στα αρνητικά συναισθήματα και στα μη αρνητικά. Η σημαντικότητα μιας λέξης στην αναγνώριση συναισθημάτων μπορεί να οριστεί ως η αμοιβαία πληροφορία μεταξύ μιας συγκεκριμένης λέξης και συναισθήματος. Για παράδειγμα, η λέξη wrong (λάθος) μας οδηγεί στην κλάση των αρνητικών συναισθημάτων, ενώ η λέξη exactly (ακριβώς) δείχνει ένα μη αρνητικό συναίσθημα. Μία λίστα τέτοιων λέξεων δημιουργήθηκε κι αν κάποια λέξη σε μια δοσμένη έκφραση ταιριάζει με τις λέξεις στη λίστα σημαίνει ότι η έκφραση αυτή ανήκει στην αντίστοιχη υπερκλάση. Η βελτιστοποίηση της αναγνώρισης συναισθημάτων από ακουστικά χαρακτηριστικά επιτεύχθηκε διερευνώντας διαφορετικά σετ

χαρακτηριστικών τα οποία επιλέχθηκαν από διαφορετικές μεθόδους επιλογής, ακολουθούμενα από principal component ανάλυση (PCA). Τα αποτελέσματα δείχνουν ότι συνδυάζοντας όλη την πληροφορία και όχι μόνο την ακουστική πληροφορία βελτιώνει την ταξινόμηση συναισθημάτων κατά 40.7% για τους άνδρες και 36.4% για τις γυναίκες. Ωστόσο, η έρευνα συμπεραίνει ότι η κατηγοριοποίηση των συναισθημάτων πρέπει να επεκταθεί από μεμονωμένες λέξεις σε προτάσεις δεδομένου ότι οι άνθρωποι συχνά λαμβάνουν υπόψη ακολουθίες λέξεων(προτάσεις) για να αναγνωρίσουν ένα συναίσθημα.



Εικόνα 3: Τυπική διαδικασία ER με χρήση λεκτικών χαρακτηριστικών

Όπως και οι Lee και Narayanan, οι Litman και Forbes εξετάζουν τη χρησιμότητα των ακουστικών και λεκτικών χαρακτηριστικών για να προβλέψουν τα συναισθήματα των φοιτητών σε διαλόγους που έγιναν κατά τη διάρκεια αλληλεπιδραστικής διδασκαλίας μεταξύ ανθρώπου και υπολογιστή [189][190]. Τα συναισθήματα που εξετάστηκαν ήταν το ουδέτερο, θετικά, αρνητικά κι ανάμικτα συναισθήματα. Γι' αυτό το σκοπό, εξήχθησαν προσωδιακά χαρακτηριστικά από την ομιλία και λεκτικά, όπως οι λέξεις. Τα αποτελέσματα δείχνουν μια σχετική βελτίωση κατά 19-36% από προηγούμενες έρευνες. Όπως και οι Litman και Forbes, οι συγγραφείς ερεύνησαν επίσης και το ρόλο του νοήματος των λέξεων(π.χ. υποκείμενο, φύλο κλπ) στην αναγνώριση συναισθημάτων από την ομιλία [191]. Συζητάνε για ένα σχήμα αναγνώρισης των συναισθημάτων των φοιτητών από διαλόγους κατά τη διάρκεια μιας διδασκαλίας για τρεις κατηγορίες συναισθημάτων(αρνητικά, μη αρνητικά και ουδέτερα). Επίσης, εκτός από τα λεκτικά στοιχεία(λέξεις), η εκδήλωση ενός συναισθήματος επηρεάζεται κι από άλλα χαρακτηριστικά όπως το φύλο του φοιτητή, μια μεγάλη παύση πριν την απάντηση σε μια ερώτηση ή μια ξαφνική αλλαγή στην ένταση και στο ρυθμό της ομιλίας.

Οι διαδικασίες ανάλυσης του στρες από τη φωνή στηρίζονται σε χαμηλού επιπέδου δείκτες του άγχους ως έμμεσοι δείκτες της απάτης. Με βάση τα παραπάνω, ο Graciarena και

λοιποί, προτείνουν μια σύνθετη προσέγγιση χρησιμοποιώντας λεκτική πληροφορία μαζί με προσωδιακά χαρακτηριστικά για την ανίχνευση της απάτης [192]. Τα αποτελέσματα από ένα SVM σύστημα βασισμένο σε λεκτικά χαρακτηριστικά συνδυασμένο με τα αποτελέσματα ενός Gaussian mixture μοντέλου βασισμένο σε ακουστικά χαρακτηριστικά δίνουν καλύτερη απόδοση απ' ό,τι μόνα τους. Τα λεκτικά χαρακτηριστικά υπολογίστηκαν αυτόματα χρησιμοποιώντας αληθινές λέξεις οι οποίες συλλέχθηκαν από υπηρεσίες επιβολής του νόμου και περιλαμβάνουν κατηγορίες με πολλές παύσεις, συντακτικά χαρακτηριστικά, ετικέτες ενεργού διαλόγου όπως συγκεκριμένες αρνήσεις, λέξεις που εκφράζουν θετικά κι αρνητικά συναισθήματα και κωδικοποίηση χαρακτηριστικών εάν το υποκείμενο απάντησε στην ερώτηση του ανακριτή με ερώτηση.

Στην εργασία του Schuller και λοιπών, 381 ακουστικά χαρακτηριστικά εξήχθησαν και η σημαντικότητά- συνάφειά τους υπολογίστηκε με Sequential Forward Floating Αναζήτηση σε σύγκριση με τη μείωση των χαρακτηριστικών με Principal Component Ανάλυση [193]. Μερικές παραλλαγές για τον υπολογισμό των λεκτικών χαρακτηριστικών περιγράφονται και περιέχουν δέσμες λέξεων (bunch-of-words (BoW)), n-grams, σημαντικότητα κι αμοιβαία πληροφορία. Επιπλέον, μείωση των χαρακτηριστικών με μεθόδους επιλογής βασισμένες σε φίλτρα αξιολογήθηκαν μειώνοντας 2334 λεκτικά χαρακτηριστικά. Η έρευνα καταλήγει σ' ένα σετ από 93 λέξεις- όρους προκειμένου να αυξήσει την απόδοση.

Επιπροσθέτως, σε μια πρόσφατη εργασία ο Schuller και λοιποί βελτιώνουν την αυτόματη αναγνώριση συναισθημάτων από ομιλούμενες λέξεις με τη μοντελοποίηση διανυσματικού χώρου έναντι χαρακτήρων [194]. Εκτός από το προφορικό περιεχόμενο οι συγγραφείς συμπεριέλαβαν και το POS και την επισήμανση της υψηλότερης σημαντικότητας στην ανάλυση τους. Σε αντίθεση με τις περισσότερες εργασίες στο πεδίο αυτό, οι οποίες υποθέτουν 100% σωστή αναγνώριση ομιλίας, αξιολόγησαν την απόδοση χρησιμοποιώντας ένα αυτόματο σύστημα αναγνώρισης φωνής (ASR) βασισμένο σε συνεχόμενα HMM χρησιμοποιώντας ακουστικά χαρακτηριστικά όπως οι δεκατρείς πρώτοι συντελεστές MFCC. Εκτεταμένα πειράματα έγιναν στην FAU Aibo Emotion Βάση Δεδομένων σε περισσότερες από τέσσερις χιλιάδες πραγματικές συναισθηματικές αλληλεπιδράσεις μεταξύ παιδιών και ρομπότ κι έδειξαν εξαιρετικά χαμηλή απόδοση σε σχέση με άλλα συστήματα αναγνώρισης φωνής από ηχογραφημένες προσποιούμενες βάσεις δεδομένων. Στα αποτελέσματα του πειράματος ο σάκος των λέξεων Bag of Words κυριάρχησε έναντι όλων των άλλων μορφών μοντελοποίησης με βάση το προφορικό περιεχόμενο.

Ο BoW είναι μια μορφή μοντελοποίησης του διανυσματικού χώρου, είναι μια πασίγνωστη μορφή αριθμητικής αναπαράστασης του κειμένου στην αυτόματη

κατηγοριοποίηση εγγράφων, ανάκτησης πληροφοριών και επεξεργασίας φυσικής γλώσσας natural language processing (NLP) [195]. Κάθε λέξη στο λεξιλόγιο προσθέτει μια διάσταση σε ένα λεκτικό διάνυσμα που αναπαριστά τη συχνότητα του όρου(term) μέσα στην ίδια έκφραση. Αυτή η μέθοδος με επιτυχία μεταφέρθηκε και στην αναγνώριση συναισθημάτων [196].

Ο Ijima και λοιποί παρουσίασαν μια τεχνική για την αναγνώριση συναισθημάτων από την ομιλία η οποία μπορεί να πάρει λεκτική και μη λεκτική πληροφορία [197]. Η φωνητική βάση δεδομένων περιέχει τρία είδη συναισθημάτων ουδέτερο, λύπη και χαρά κι αποτελείται από 503 προτάσεις τις οποίες πήραν από την ATR Ιαπωνική Βάση Δεδομένων, η οποία ηχογραφήθηκε από δύο άντρες και δύο γυναίκες. Η τεχνική αυτή χρησιμοποιεί την πολλαπλής παλινδρόμησης HMM (MRHMM) και βασίζεται στην εκτίμηση του συναισθήματος. Χρησιμοποιώντας ένα μοντέλο ανεξαρτήτου ομιλητή(speaker-independent), το MRHMM εκπαιδεύτηκε με ένα μικρό ποσό δεδομένων. Επιπλέον, το ακουστικό μοντέλο για την αναγνώριση φωνής προσαρμόστηκε ώστε να εισάγει το στυλ ομιλίας από το εκπαιδευμένο MRHMM. Έδειξαν ότι η απόδοση της προτεινόμενης τεχνικής είναι υποσχόμενη και για την αναγνώριση συναισθήματος και για την αναγνώριση φωνής.

5.3 ΣΥΜΠΕΡΑΣΜΑΤΑ

Οι παραπάνω μελέτες δείχνουν ότι οι ερευνητές έχουν στραφεί σε διάφορους τομείς για την αξιοποίηση της αναγνώρισης συναισθημάτων από τους υπολογιστές. Παρατηρώντας την εξέλιξη των ερευνών χρονολογικά, βλέπουμε ότι κάποια πράγματα που το 1999 ήταν υπό διερεύνηση τώρα χρησιμοποιούνται ως παραδοχές, πράγμα που μας οδηγεί στην εξέλιξη των συστημάτων και στην χρήση μιας κοινής βάσης. Αυτό έχει ιδιαίτερο ενδιαφέρον αφού κατά το ξεκίνημα αυτών των ερευνών η έλλειψη μιας κοινής βάσης οδηγούσε σε λανθασμένα συμπεράσματα και αποτελέσματα. Τα τελευταία χρόνια, γίνεται προσπάθεια χρήσης αυτών των συστημάτων σε πραγματικές καταστάσεις της καθημερινής ζωής κάτι που δίνει στους ερευνητές την ώθηση να προχωρήσουν τις μελέτες τους, έτσι ώστε να φτάσουν σε ένα τελειοποιημένο σύστημα αναγνώρισης συναισθημάτων.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Emotion Recognition Using Novel Speech Signal Features Talieh Seyed Tabatabaei, Sridhar Krishnan, Aziz Guergachi, Ryerson University, Toronto Canada.
- [2] http://en.wikipedia.org/wiki/Human-computer_interaction; Last access date [September 8, 2011].
- [3] Darwin, Ch. The expression of the emotions in man and animals. Chicago: University of Chicago Press, 1965 (Original Work Published in 1872).
- [4] Emotion Recognition in speech signal: Experimental study, development, end application, Valery A. Petrushin, ICSLP 2000.
- [5] Picard, R. Affective Computing. The MIT Press. 1997.
- [6] Benzooijen, R. van The characteristics and recognizability of vocal expression of emotions, Dordrecht, The Netherlands: Foris, 1984 και Scherer, K.R., Basse, R., Wallboot, H.G., Goldbeck T. Vocal clues in emotion encoding and decoding *Motiv Emotion* 1991;15:123-148, 1991.
- [7] Murray, I.R. and Arnott, J.L. Toward the Simulation of emotion in synthetic speech: A review of the literature on human vocal emotions. *Journal Acoustical society of America*; 93(2): 1079-1108, 1993.
- [8] Dallaert, F., Polzin, Th., and Waibel, A. Recognizing emotions in speech ICSLP 96.
- [9] Tosa, N. and Nakatsu, R. Life-like communication agent – emotion sensing character “MIC” and feeling session character “MUSE”. *Proceedings of IEEE Conference on Multimedia 96*, 12-19, 1999.
- [10] Floyd, C, Zullighovem. H. Budde R. Keil-Slawik, R (eds) *Software Development and Reality Construction*. Heidelberg: Springer, 1992.
- [11] B. Laurel, *Computers as Theatre*. Reading, MA: Addison-Wesley Publishing Co., 1991.
- [12] Frank Thissen, University Karlsruhe, Germany “Emotional and Meta Communicational Aspects of Human-Computer Interaction” 2002.
- [13] Doring, N. *Sozialpsychologie des Internet*. Gottingen: Hogrefe, 1999.
- [14] Reeves, B. Nass, C. *The Media Equation*. Cambridge University Press, 1996.

- [15] Weizenbaum J. Die Macht der Computer und die Ohnmacht der Vernunft. Frankfurt/M. Suhrkamp [orig. Weizenbaum J,1976] Computer Power and Human Reason. New York: Freeman & Company, 1978.
- [16] Bateson, G Okologie des Geistes. Frankfurt Suhrkamp 1985.
- [17] Watzlawick P Beavin, J H Jackson D.D. Menschliche Kommunikation Bern Huber 1996. Pragmatics of Human Communication. New York 1967.
- [18] Maturana, Humberto R./Varela, Francisco J. (1980): Autopoiesis and Cognition. The Realization of the Living. Dordrecht: Reidel, p. 13.
- [19] Ciompi, Die Emotionalen Grundlagen des Denkens. Göttingen: Vandenhoeck, 1997.
- [20] Damasio, Descartes Error, New York 1995.
- [21] Goleman. Emotional Intelligence. New York 1997.
- [22] Markowitsch Dem Gedächtnis auf der Spur. Darmstadt: Wiss Buchgesellschaft, 2002.
- [23] Roth, Das Gehirn und seine Wirklichkeit, Frankfurt 1997.
- [24] Spitzer M, Lernen Heidelberg: Spektrum 2002.
- [25] Prensky Digital Game-Based Learning. New York: McGraw Hill, 2001.
- [26] Forester H.v. Wissen und Gewissen Frankfurt 1993.
- [27] Maturana & Varela F.J Der Baum der Erkenntnis. Bern 1987.
- [28] Siebert H. Pädagogischer Konstruktivismus. Neuwied 1999.
- [29] Luhmann. Soziale Systeme 2001 . Frankfurt Social Systems, Stanford Univ Press 1995.
- [30] Tin Lay New, Say Wei Foo, Liyanage C De Silva, Speech Emotion Recognition Using Hidden Markov models 2003.
- [31] Otaley K Jenkis JM Understanding Emotions, Blackwell, Oxford U K 1996.
- [32] Nass C, Steuer J, Tauber E "Computers as social actors, " In Proceedings of CHI'94 Human factors in Computing Systems pp 72-78 New York ACM Press 1994.
- [33] Anne Aula, The effects of Emotional Feedback in Human-Computer Interaction. University of Tampere Dep. Psychology, 2001.
- [34] Cahn JE The Generation of Affect in Synthesized Speech. Journal of the American Voice I/O Society, 8, 1-19, 1990.
- [35] Bachorowski JA & Owren MJ Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context, Psychological science 6, 219-224, 1995.
- [36] Scherer K.R. Vocal measurement of emotion. In Plutchik, R. & Kellerman, H. Theory, Research and Experience pp 233-259, 1989.

- [37] Scherer, Banse, Wallbott HG & Goldbeck T Vocal cues in emotion encoding and decoding. *Motivation and emotion* 15, 123-148, 1991.
- [38] Pittam J & Scherer KR Vocal expression and communication of emotion. In Lewis M & Heciland JM. *Handbook of emotions* pp 185-197. New York, The Guilford Press, 1993.
- [39] Murray, I. R. & Arnott. J.L. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of acoustic society of America*, 93 1097-1108, 1993.
- [40] Kappas A Hess U & Scherer KR Voice and emotion. In Feldman RS & Rime B. *Fundamentals of nonverbal behavior*. New York: Cambridge University Press, 1991.
- [41] Scherer, K.R., Banse, & Silverman K.E.A. Vocal cues to speaker affect: Testing two models. *Journal of the acoustical society of America* 76, 1346-1356, 1984.
- [42] Frijda N.H. *The emotions*. New York Cambridge University Press 1986.
- [43] Levenson R.W. Human emotion: A functional view. In Ekman P & Davidson RJ. *The nature of emotion : Fundamental questions* pp 123-126 New York: Oxford Press, 1994.
- [44] Schwartz G.E. Emotion and psychophysiological organization. In Coles, M.G.H. Donchin E., & Proges S. *Psychophysiology: Systems processes and applications* pp 378-406 New York: Cuilford Press, 1986.
- [45] Surakka, Kasvonilmeet ja emootioiden tutkimus. *Psykologia* 31, 412-420, 1996.
- [46] Surakka, Contagion and modulation of human emotions. A doctoral dissertation. *Acta Universitatis Tamperensis*, 627. Vammalan Kirjapaino OY: Vammala, 1998.
- [47] Davidson RJ & Irwin. The functional neuroanatomy of emotion and affective style. *Trends in cognitive sciences* 3 (1) 11-21, 1999.
- [48] LeDoux JE Emotion and the amygdale. In Aggleton JP. *The amygdala: Neurobiological aspects of emotion, memory and mental dysfunction*, pp 339-351. New York: Wiley-Liss, 1992.
- [49] LeDoux JE Emotional networks in the brain. In Lewis M & Heciland JM. *Handbook of emotions* pp 109-118. NY The Guilford Press ,1993.
- [50] Panksepp J. *Affective neuroscience: The foundations of human and animal emotions*. NY: Oxford Univ Press, 1998.
- [51] Camras, LA Holland EA & Patterson MJ *Facial Expression*. 1993.
- [52] Lang PJ Bradley MM & Guthbert BN *Motivated attention: affect, activation and attention* 1997.

- [53] Lang PJ, Greenwald MK, Bradley MM & Hamm AO, Looking at pictures: Affective facial visceral and behavioural reactions. *Psychophysiology*, 30, 261-273, 1993.
- [54] Ekman P. An argument for basic emotions. *Cognition and emotion*, 6(3/4) 169-200, 1992.
- [55] Izard, C.E. Organizational and motivational functions of discrete emotions, 1993.
- [56] Bradley & Lang Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavioral Therapy & Experimental Psychiatry* 25 (1), 49-59,1994.
- [57] Lang P. The Emotion probe: Studies of motivation and attention. *American Psychologist* 50, 372-385, 1995.
- [58] Bradley MM Greenwald MK, Petry MC & Lang PJ. Remembering pictures: Pleasure and arousal in memory. *Journal of experimental psychology*, 18 (2), 379-390, 1992.
- [59] Ekman P. All emotions are basic. *The nature of emotion : Fundamental questions*, 1994.
- [60] A.Batliner, K.Fischer., R.Huber, J. Silker, E.Noeth, How to find trouble in communication, *Speech Communication*, 2003.
- [61] Ekman,P.,Friesen, W.V., 1969. The repertoire of nonverbal behaviour: Categories, Origin,usage and coding, *Semiotica* I 49-98.
- [62] Scherer, K., Ceschi,G 1997 Lost Luggage: a field study of emotion-antecedent appraisal. *Motivation and Emotion* 21, 211-235.
- [63] Scherer, K., Ceschi,G 2000 Criteria for emotion recognition from verbal and nonverbal expression: studying baggage loss in the airport. *Personality and Social Psychology Bulletin* 26 327-339.
- [64] Bechara A, Damasio H, Tranel D & Damasio A.R. Deciding advantageously before knowing the advantageous strategy, *Science* 275, 1293-1295.
- [65] Isen AM Positive affect and decision making. 1993.
- [66] Martin & Jones GV. Integral bias in the cognitive functioning of emotionally linked pictures. *British journal of psychology*, 86, 419-432, 1995.
- [67] Tischer,B. 1993 Die vokale Kommunikation von Gefuhlen, *Fortschritte der psychologischen Forschung*, Vol.18. Psychologie Verlags Union, Weinheim.
- [68] Li Y Zhao Y, Recognizing emotions in speech using short-term and long-term features, 1998.
- [69] Paeschke, A, Kinast M, Sendlmeier W.F. F0 contours in emotional speech, 1999.
- [70] Scherer K., How emotion is expressed in speech and singing, 1995.

- [71] Cowie R, Douglas-Cowie E., Schroder M, A Conceptual Framework for Research, Proc. Of the ISCA Workshop On Speech Emotion, 2000.
- [72] Batliner A, Prosody focus and focal structure: some remarks on methodology, 1994.
- [73] Fischer K. What is a situation? Gothenburg Papers in Computational Linguistics 85-92, 2000.
- [74] Selting, Emphatic speech style –With special focus on the prosodic signaling of heightened emotive involvement in conversation, 1994.
- [75] Fiehler R, Kommunikation und Emotion. Theoretische und empirische Untersuchungen zur Rolle von Emotionen in der verbalen Interaktion. De Gruyter, Berlin, 1990.
- [76] B.M.De Paulo, D.A.Kashy, S E Kirkendol, M.M. Wyer and J.A. Epstein: Lying in Everyday Life. Journal of Personality and Socia Psychology, 1996.
- [78] Oliviero Stock and WP8 members, HUMAINE, Report on Basic Cues and Open Research Topics in Communication and Emotions.
- [79] W.Bosma & E. Andre: Exploting Emotions to Disambiguate Dialogue Acts. Proceedings of the 9th International Conferencs on Intelligent User Interfaces, ACM Press, 85-92, 2004.
- [80] Anoli & R.Ciceri: The voice of deception: vocal strategies of naive and able liars. Journal of Nonverbal Behavior, 1997.
- [81] Mc Kenzie F, Scerbo, Catanzaro J & Phillips M. Nonverbal indicators of malicious intent: affective components of interrogative virtual reality training. International Journal of Human Computer Studies, 2003.
- [82] Castelfranch & Poggi Lying as pretending to give information. Pretending to communicate, 1993.
- [83] Carofiglio, Rosis and Castelfranchi: Ascribing and Weighting Beliefs in Deceptive Information Exchanges. M.Bauer, P.J. Gmytrasiewicz, B.A. Juliano, R.S.Renner. ICCS, Berlin 2001.
- [84] Lee & Wilks: Eliminating deception and mistaken belief to infer conversational implicature. IJCAI Workshop on “Cooperation, Collaboration and Conflict in Dialogue Systems”, 1997.
- [85] Ward & Hexmoor: Deception as a Means for Power Among Collaborative Agents, 2003.
- [86] Castelfranchi & Falcone & Rosis, Deceiving in GOLEM: hot to strategically pilfer help, 1998.

- [87] Hirschberg :Emotional Speech (course notes)
<http://www1.cs.columbia.edu/~julia/cs6998/emotion.ppt>, 2004.
- [88] Ekman P. Universals and cultural differences in facial expressions of emotion, 1972.
- [89] Fiske & Kitayama & Markus & Nisbett. The cultural matrix of social psychology, 1998.
- [90] Markus & Kitayama. Culture and the self: Implications for cognition, emotion and motivation, 1991.
- [91] Mesquita & Frijda Cultural variations in emotions. A review, 1992.
- [92] Mesquita & Frijda & Scherer. Culture and emotion, 1997.
- [93] Scherer, The role of culture in emotion antecedent appraisal. *Journal of Personality and Social Psychology*, 1997.
- [94] Scherer & Wallbott . Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Psychology*, 1994.
- [95] Matsumoto D. American Japanese cultural differences in the recognition of universal facial expressions. *Journal of Cross-Cultural Psychology*, 1992.
- [96] Russell, Negative results on a reported facial expression of contempt. *Motivation and Emotion*, 1991.
- [97] Ekman.,Strong Evidence for universals in facial expressions: A reply to Russell's mistaken critique, 1994.
- [98] Izard CE Innate and universal facial expressions. Evidence from developmental and cross cultural research, 1994.
- [99] Yimngam, S., W. Premchaisawadi, and W. Kreesuradej. Thai emotion words analysis. in *Natural Language Processing, 2009. SNLP '09. Eighth International Symposium on*. 2009.
- [100] Binali, H., W. Chen, and V. Potdar. *Computational approaches for emotion detection in text*. in *Digital Ecosystems and Technologies (DEST), 2010 4th IEEE International Conference on*.
- [101] Trabelsi, A. and C. Frasson. *The Emotional Machine: A Machine Learning Approach to Online Prediction of User's Emotion and Intensity*. in *Advanced Learning Technologies (ICALT), 2010 IEEE 10th International Conference on*.
- [102] Kenneth Ward Church and Patrick Hanks. Word association norms, mutual information and lexicography. In proceedings of ACL 27, pp.76-83, Vancouver, Canada, 1989.

- [103] Lewis D.D. and M. Ringuette. Comparison of two learning algorithms for text categorization. In Proceedings of the Third Annual Symposium on Document Analysis and Information Retrieval (SDAIR'94), 1994
- [104] C. Apte, F. Damerau, and S. Weiss. Automated learning of decision rules for text categorization. *ACM Transactions on Information Systems*, 12(3): pp.233-251, 1994.
- [105] E. C.-C. Kao, L. Chun-Chieh, Y. Ting-Hao, H. Chang-Tai, and S. Von-Wun, "Towards Text-based Emotion Detection," in International Conference on Information Management and Engineering, 2009, pp. 70-74.
- [106] S. Aman, "Recognizing Emotions in Text," Master of Computer Science, University of Ottawa, 2007.
- [107] F. Sebastiani, "Machine Learning in Automated Text Categorization," *ACM Computing Surveys*, vol. 34, No.1, pp. 1-47, 2002.
- [108] J. Read, "Recognising Affect in Text using Pointwise-Mutual Information ": Master of Science, University of Sussex, 2004.
- [109] A. Esuli and F. Sebastiani, "Determining Term Subjectivity and Term Orientation for Opinion Mining," in Proceedings of EACL 2006.
- [110] Chung-Hsien Wu, Ze-Jing Chuang and Yu-Chung Lin. 2006. Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models, *ACM Transactions on Asian Language Information Processing*, Vol. 5, No. 2, Pages 165-182.
- [111] T. Mitchell, *Machine Learning*. New York : McGraw Hill, 1997.
- [112] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques", *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 79-86, 2002.
- [113] C.M. Lee, R. Pieraccini: "Combining acoustic and language information for emotion recognition," *Proceedings ICSLP 2002* , Denver, CO, USA, 2002
- [114] Y. Yang. Expert Network: Effective and efficient learning from human decisions in text categorization and retrieval. In 17th Ann Int ACM SIGIR Conference on Research and Development in information retrieval (SIGIR'94), pp 13-22, 1994
- [115] Yiming Yang, Xin Liu. A re-examination of text categorization methods. In 22nd Ann Int ACM SIGIR Conference on Research and Development in information retrieval (SIGIR'99), pp. 424-431, 1999
- [116] Y. Yang and W.J. Wilbur. Using corpus statistics to remove redundant words in text categorization. In *J Amer Soc Inf*, 1996

- [117] Y. Yang, and J. Pedersen, "A comparative study on feature selection in text categorization", Proceedings of ICML-97, the 14th International Conference on Machine Learning, 1997.
- [118] K. Nigam, A. Mccallum, S. Thrun, and T. Mitchell, "Text classification from labeled and unlabeled documents using EM", Machine Learning, Vol 39, No. 2/3, pp. 103-134, 2000.
- [119] S. Li, R. Xia, C. Zong, and C. Huang, "A framework of feature selection methods for text categorization", Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, pp. 692-700, Singapore, 2009.
- [120] Guo, Y. Shao, Z. Hua N. "Automatic text categorization based on content analysis with cognitive situation models", Information Sciences, Volume 180, Issue 5, 2010, pp. 613-630.
- [121] Ye Wu; Fuji Ren; , "Improving emotion recognition from text with fractionation training," Natural Language Processing and Knowledge Engineering (NLP-KE), 2010 International Conference on , vol., no., pp.1-7, 21-23 Aug. 2010.
- [122] Shaikh Mostafa Al Masum, Helmut Prendinger, Mitsuru Ishizuka, A cognitively based approach to affect sensing from text. Intelligent User Interfaces 2006_T Sydney, Australia_T 303-305.
- [123] Chunling Ma, Helmut Prendinger, Mitsuru Ishizuka, Emotion Estimation and Reasoning Based on Affective Textual Interaction, ACII 2005, 622-628.
- [124] Kuhn, R., (1988). Speech recognition and the frequency of recently used words: A modified markov model for natural language. In 12th International Conference on Computational Linguistics, pp. 348–350.
- [125] Kuhn, R., and De Mori R., (1990). A cache-based natural language model for speech reproduction. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-12(6):570–583.
- [126] Kuhn, R., and De Mori, R., (1992). Roland Kuhn and Renato De Mori. Correction to: A cache-based natural language model for speech reproduction. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-14(6):691–692.
- [127] Kupiec, J., (1989). Probabilistic models of short and long distance word dependencies in running text. In Proceedings of the DARPA Workshop on Speech and Natural Language, pages 290–295.
- [128] Rosenfeld, R., (1996). A Maximum Entropy Approach to Adaptive Statistical Language Modeling, Computer, Speech and Language, vol. 10, 187—228.

- [129] Beeferman, D., Berger, A., and Lafferty, J. (1997). A model of lexical attraction and repulsion. In *ACL/EACL '97*.
- [130] Iyer, R.M., and Ostendorf, M., (1999). Modeling long distance dependence in language: topic mixtures versus dynamic cache models *Speech and Audio Processing, IEEE Transactions on* , Volume: 7 Issue: 1 , 30 -39.
- [131] Jelinek, F. (1995). The language modeling summer workshop at Johns Hopkins University. Closing remarks.
- [132] Kuhn, R., (1988). Speech recognition and the frequency of recently used words: A modified markov model for natural language. In *12th International Conference on Computational Linguistics*, pp. 348–350.
- [133] Anna Aula *The effects of Emotional Feedback in Human Computer Interaction*, 2001.
- [134] Cowie, Douglas Cowie, Tsapatsoulis G.Votsis, S Kollias W Fellenz and Taylor *Emotion Recognition in Human Computer Interaction*.
- [135] Ekman & Friesen, *Unmasking the face*, 1975.
- [136] Cabeza & Nyburg, *Cognitive Neuroscience*, 1997.
- [137] Ekman & Friesen *Facial Action Coding System: A technique for the measurement of facial action*, 1978.
- [138] Cacioppo Petty Losch & Kim, *Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions*, 1986.
- [139] Tassinari & Cacioppo. *Unobservable facial actions and emotions*, 1992.
- [140] Fridtjof A.J. & Cacioppo *Guidelines for human electromyographic research*, 1986.
- [141] Cacioppo Klein Bernston & Hatfield, *The psychophysiology of emotion*, 1993.
- [142] Camras L.A. Holland & Patterson. *Facial Expression*, 1993.
- [143] Greenwald M K Cook III & Lang P J. *Affective judgment and psychophysiological response: dimensional covariation in the evaluation of pictorial stimuli*, 1989.
- [144] Jancke & Vogt & Musial & Lutz & Kalveram, *Facial EMG responses to auditory stimuli. International journal of psychophysiology*, 1996.
- [145] McKenna and S Gong “Tracking faces” 1996.
- [146] Growley & Berard “Multimodal tracking of faces for video communications”, 1997.
- [147] H. Graf, E. Cosatto, Gibbon Kocheisen Petajan “Multi modal system for locating heads & faces”, 1996.
- [148] Young & Ellis *Handbook of Research on Face Processing*. Amsterdam, The Netherlands, 1989.

- [149] Tom Johnstone, Carien M, van Reekum, Terrence R.Oakes, and Richard J.Davidson
The voice of emotion: an FMRI study of neural responses to angry and happy vocal expressions.
- [150] A. Fernald. Intonation and communicative intent in mothers speech to infants: is the melody the message? *Child Development*, 1989; 60: 1497-510]. [A.Fernald,; H., Morikawa. Common themes and cultural variations in Japanese and American mothers speech to infants. *Child Development*. 1993;64:637-56.
- [151] Shackman JE, Pollak SD. Experiential Influences on multimodal perception of emotion. *Child Development* 2005; 76: 1116-26.
- [152] J. Hughlings Jackson. On affectations of speech from diseases of the brain. *Brain*.1915;38:107-74.
- [153] Tucker DM, Watson RT, Heilman KM. Discrimination and evocation of affectively intoned speech in patients with right parietal disease. *Neurology*. 1977;27:947-50.
- [154] Williams & Stevens, *Vocal correlates of emotional states*, 1981.
- [155] Cloria J Borden Katherine S. Harris, Lawrence J.Raphael, *Speech Science Primer* 3rd edd., Williams & Wilkins press, 2000.
- [156] De Silva, *Bimodal emotion recognition. Automatic Face and Gesture Recognition*, 2000.
- [157] Chen Huang Myiasato & Nakatsu *Multimodal human emotion /expression recognition*, 1998.
- [158] De Silva & Miyasato & Nakatsu, *Facial Emotion Recognition Using Multimodal Information*, 1997.
- [159] Yoshitomi Sung Kim, Kawano, Kilazoe, *The effect of sensor fusion for recognition of emotional states using voice, face image and thermal image of face*, 2000.
- [160] Huang, Chen, Tao, Miyasato & Nakatsu, *Bimodal Emotion Recognition by Man and Machine*.
- [161] Chen, Huang, *Emotional expressions in audiovisual human computer interaction*, 2000.
- [162] Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, Shrikanth Narayanan, *Analysis of emotion recognition using Facial Expressions, Speech and Multimodal Information*, 1995.
- [163] Camurri & Coglio. *An architecture for emotional agents*. IEEE, *Multimedia*, 1998.
- [164] Cornelius, *The Science of Emotion*, 1996.

- [165] Cornelius Theoretical approaches to emotion. Workshop on Speech and Emotions, 2000.
- [166] Vladimir Hozjan and Adravko Kacic, Context-Independent Multilingual Emotion Recognition from Speech Signals,2000.
- [167] Tomkins, Affect Theory, 1982.
- [168] James W. What is an emotion?, 1884.
- [169] Arnold M B , Emotion and Personality, Physiological aspect, NY, 1980.
- [170] Zajonc R.B. Feeling and thinking: Preferences need no inferences, 1980.
- [171] Lutz, Unnatural emotions: Everyday sentiments on an Micronesian atoll and their challenge to western theory. Chicago, 1988.
- [172] Kramer, Judgment of personal characteristics and emotions from nonverbal properties of speech, 1963.
- [173] Johnstone, T. Emotional speech elicited using computer games. in Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on. 1996.
- [174] Demitrios Ververidis and Constantine Kotropoulos, A State of the Art Review on Emotional Speech Databases, Aristotele University of Thessaloniki, 2004.
- [175] Ambrus D.C. "Collecting and recording of an emotional Speech Database" Technical Report, faculty of Electrical Engineering and Computer Science, Institute of Electronics, University of Maribor.
- [176] Ostermann J "Face animation in MPEG-4" in MPEG-4 facial animation, 2002.
- [177] Cowie R , Douglas-Cowie, Savvidou S "Feeltrace:An instrument for recording perceived emotion in real time" ,Belfast 2000.
- [178] Schroeder M. Cowie , Douglas-Cowie "Acoustic correlates of emotion dimensions in view of speech synthesis" 2001.
- [179] Mc Gilloway Cowie-Douglas "Approaching automatic recognition of emotion from voice: A rough benchmark" Newcastle, 2000.
- [180] Douglas Cowie, Cowie and Schroeder "A New emotion Database: Considerations, Sources and Scope" Belfast 2000.
- [181] The Center of Spoken Language Research (CSLR) CU Kid's speech corpus, http://cslr.colorado.edu/beginweb/reading /data_collection.html.
- [182] Linguistic Data Consortium LDC <http://www.ldu.upenn.edu/>. Last access date [September 8, 2011].

- [183] G. Holmes; A. Donkin and I.H. Witten (1994). "Weka: A machine learning workbench". Proc Second Australia and New Zealand Conference on Intelligent Information Systems, Brisbane, Australia. Retrieved 2007-06-25.
- [184] S. Wang , X. Ling , F. Zhang , Jianing Tong, — Speech Emotion Recognition Based on Principal Component Analysis and Back Propagation Neural Network, Proc. of International Conference on Measuring Technology and Mechatronics Automation, 2010, pp. 437-440.
- [185] Y. Zhou , Y. Sun , L. Yang , Y. Yan, —Applying Articulatory Features to Speech Emotion Recognition, Proc. of International Conference on Research Challenges in Computer Science, 2009, pp. 73-76.
- [186] A. Firoz Shah, V.R. Vimal Krishnan , A. Raji Sukumar , Athulya Jayakumar , P. Babu Anto, —Speaker Independent Automatic Emotion Recognition from Speech: A Comparison of MFCCs and Discrete Wavelet Transforms, Proc. of International Conference on Advances in Recent Technologies in Communication and Computing, 2009, pp. 528-531.
- [187] C. Yang, L. Ji, G. Liu, —Study to Speech Emotion Recognition Based on TWINsSVM, Proc. of 5th International Conference on Natural Computation, 2009, pp. 312-316.
- [188] Y. Zhou, J. Zhang, L. Wang, Y. Yan, —Emotion Recognition and Conversion for Mandarin Speech, Proc. of 6th International Conference on Fuzzy Systems and Knowledge Discovery, 2009, pp. 179-183.
- [189] T. Yang, J. Yang , F. Bi, —Emotion Statuses Recognition of Speech Signal Using Intuitionistic Fuzzy Set, Proc. of World Congress on Software Engineering, 2009, pp. 204-207.
- [190] Y. Jin, Y. Zhao , C. Huang , L. Zhao, —Study on the Emotion Recognition of Whispered Speech, Proc. of Global Congress on Intelligent Systems, 2009, pp. 242-246.
- [191] H. Wenjing, L. Haifeng , G. Chunyu, —A Hybrid Speech Emotion Perception Method of VQ-based Feature Processing and ANN Recognition, Proc. of Global Congress on Intelligent Systems, 2009, pp. 145-149.
- [192] S. Yun , C. D. Yoo, —Speech emotion recognition via a max-margin framework incorporating a loss function based on the Watson and Tellegen's emotion model, Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 4169-4172.

- [193] A. Tellegen, D. Watson, L.A. Clark, —On The Dimensional and Hierarchical Structure of Affect, *Psychological Science*, vol. 10, no. 4, pp. 297–303, 1999.
- [194] T. Giannakopoulos, A. Pikrakis, S. Theodoridis, —A dimensional approach to emotion recognition of speech from movies, *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 65-68.
- [195] X. Mao, L. Chen, L. Fu, —Multi-level Speech Emotion Recognition Based on HMM and ANN, *World Congress on Computer Science and Information Engineering*, 2009, pp. 225-229.
- [196] B. Schuller, G. Rigoll, M. Lang, —Hidden Markov model-based speech emotion recognition, *Proc. of International Conference on Multimedia and Expo*, vol. 1, 2003, pp. 401-404.
- [197] L. Fu, X. Mao, L. Chen, —Relative Speech Emotion Recognition Based Artificial Neural Network, *Proc. of IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, 2008, pp. 140-144.
- [198] H. K. Mishra, C. C. Sekhar, —Variational Gaussian Mixture Models for Speech Emotion Recognition, *Proc. of 7th International Conference on Advances in Pattern Recognition*, 2009, pp. 183-186.
- [199] W. Yu, —Research and Implementation of Emotional Feature Classification and Recognition in Speech Signal, *Proc. of International Symposium on Intelligent Information Technology Application Workshops*, 2008, pp. 471-474.
- [200] L. Cen, W. Ser, Z. L. Yu, —Speech Emotion Recognition Using Canonical Correlation Analysis and Probabilistic Neural Network, *Proc. of 7th International Conference on Machine Learning and Applications*, 2008, pp. 859-862.
- [201] X. M. Cheng, P. Y. Cheng, L. Zhao, —A Study on Emotional Feature Analysis and Recognition in Speech Signal, *Proc. of International Conference on Measuring Technology and Mechatronics Automation*, 2009, pp. 418-420.
- [202] J. Rong, Y.P. P. Chen, M. Chowdhury, G. Li, —Acoustic Features Extraction for Emotion Recognition, *Proc. of 6th IEEE/ACIS International Conference on Computer and Information Science*, 2007.
- [203] K.H. Hyun, E.H. Kim, Y.K. Kwak, Robust Speech Emotion Recognition Using Log Frequency Power Ratio, *Proc. of 2006 SICE-ICASE International Joint Conference*, 2006, pp. 2586-2589.
- [204] M. You, C. Chen, J. Bu, J. Liu, J. Tao, —Emotional Speech Analysis on Nonlinear Manifold, *Proc. of 18th International Conference on Pattern Recognition*, vol. 3, 2006.

- [205] H. Atassi, A. Esposito, —A Speaker Independent Approach to the Classification of Emotional Vocal Expressions, Proc. of 20th IEEE International Conference on Tools with Artificial Intelligence, 2008, pp. 147-152.
- [206] M. Lugger, B. Yang, —The Relevance of Voice Quality Features in Speaker Independent Emotion Recognition, Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Honolulu, Hawaii, 2007, vol. IV, pp. 17-20.
- [207] Y. Wang, S. Du, Y. Zhan, —Adaptive and Optimal Classification of Speech Emotion Recognition, Proc. of 4th International Conference on Natural Computation, 2008, pp. 407-411.
- [208] T.L. Pao, C.S. Chien, Y.T. Chen, J.H. Yeh, Y. M. Cheng, W. Y. Liao, —Combination of Multiple Classifiers for Improving Emotion Recognition in Mandarin Speech, Proc. of 3rd International Conference on International Information Hiding and Multimedia Signal Processing, 2007.
- [209] T.L. Pao, W.Y. Liao, Y.T. Chen, J.H. Yeh, Y.M. Cheng, C. S. Chien, —Comparison of Several Classifiers for Emotion Recognition from Noisy Mandarin Speech, Proc. of 3rd International Conference on International Information Hiding and Multimedia Signal Processing, 2007.
- [210] T. Kostoulas, T. Ganchev, I. Mporas, N. Fakotakis, —Detection of Negative Emotional States in Real-World Scenarios, Proc. of 19th IEEE International Conference on Tools with Artificial Intelligence, vol.2, 2007, pp. 502-509.
- [211] B. Schuller, S. Reiter, G. Rigoll, —Evolutionary Feature Generation in Speech Emotion Recognition, Proc. of 2006 IEEE International Conference on Multimedia and Expo, 2006, pp. 5-8.
- [212] M.T. Shami, M.S. Kamel, —Segment-based approach to the recognition of emotions in speech, Proc. of IEEE International Conference on Multimedia and Expo, 2005.
- [213] J.A. Russell, J. Bachorowski, and J. Fernandez-Dols, —Facial and Vocal Expressions of Emotion, Annual Revisions of Psychology, vol. 54, 2003, pp. 329-349.
- [214] S. Petridis, M. Pantic, —Audiovisual Discrimination between Laughter and Speech, Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2008, pp. 5117-5120.
- [215] P. Pal, A.N. Iyer, R.E. Yantorno, —Emotion Detection from Infant Facial Expressions and Cries, Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2006, vol. 2, pp. 721-724, 2006.

- [216] M. Schroder, —Experimental study of affect bursts, Speech Communication, Special Issue on Speech and Emotion, vol. 40, Issue 1-2, 2003, pp. 99–116.
- [217] S. Matos, S.S. Birring, I.D. Pavord, D.H. Evans, —Detection of Cough Signals in Continuous Audio Recordings Using HMM, IEEE Trans. on Biomedical Engineering, vol. 53, no. 6, 2006, pp. 1078-1083.
- [218] M. Aigner, G. Sachs, E. Bruckmüller, B. Winklbaaur, W. Zitterl, I. Kryspin-Exner, R. Gur, H.Katschnig, —Cognitive and emotion recognition deficits in obsessive–compulsive disorder, Psychiatry research, Elsevier, Vol. 149, Issue 1, 2007, pp. 121-128.
- [219] J. Calder, A. D. Lawrence and A. W. Young, —Neuropsychology of Fear and Loathing, Nature Reviews Neuroscience, vol. 2, 2001, pp. 352-363.
- [220] France, D.J., Shivavi, R.G., Silverman, S., Silverman, M., Wilkes, M., "Acoustical properties of speech as indicators of depression and suicidalrisk," IEEE Trans. on Biomedical Engineering, vol. 7, 2000, pp. 829-837.
- [221] A. Batliner, K. Fischer, R. Huber, J. Spilker, E. Nolth, —How to find trouble in communication, Speech Communication, vol. 40, 2003, pp. 117–143.
- [222] C.M. Lee, S. S. Narayanan, —Toward Detecting Emotions in Spoken Dialogs, IEEE Trans. on Speech and Audio Processing, vol. 13, no. 2, 2005, pp. 293-303.
- [223] D. J. Litman, K. Forbes-Riley, —Predicting Student Emotions in Computer-Human Tutoring Dialogues, Proc. of the 42nd Annual Meeting on Association for Computational Linguistics, 2004.
- [224] K. Forbes-Riley and D. Litman, —Predicting Emotion in Spoken Dialogue from Multiple Knowledge Sources, Proc. of Human Language Technology Conference, North American Chapter of the Assoc. Computational Linguistics (HLT/NAACL), 2004.
- [225] M. Graciarena, E. Shriberg, A. Stolcke, F. Enos, J. Hirschberg, S. Kajarekar, Combining Prosodical Lexical and Cepstral Systems for Deceptive Speech Detection, Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, 2006.
- [226] B. Schuller, R. J. Villar, G. Rigoll, M. Lang, —Meta-classifiers in Acoustic and Linguistic Feature Fusion-based Affect Recognition, Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, vol. I, 2005, pp. 325-328.

- [227] B. Schuller , A. Batliner , S. Steidl , D. Seppi, —Emotion recognition from speech: Putting ASR in the loop|| Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 4585-4588.
- [228] D. D. Lewis , —Naive (Bayes) at forty: The independence assumption in information retrieval||, Machine Learning: ECML-98, Lecture Notes in Computer Science, Springer, vol. 1398, 1998, pp. 4-15.
- [229] B. Schuller, R. Muller, M. Lang, and G. Rigoll, —Speaker Independent Emotion Recognition by Early Fusion of Acoustic and Linguistic Features within Ensembles||, Proc. of 9th Eurospeech - Interspeech, 2005, pp. 805–809.
- [230] Y. Ijima, M. Tachibana, T. Nose, T. Kobayashi, —Emotional speech recognition based on style estimation and adaptation with multiple-regression HMM||, Proc of 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 4157-4160.