Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου
και Ρομποτικής

Αναγνώριση Προσωπικότητας από Σπεκτογράμματα Φωνής σε διαφορετικές Χρονικές Κλίμακες

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΝΙΚΟΛΑΟΣ ΠΑΝΤΕΛΑΙΟΣ

Επιβλέπων : Αλέξανδρος Ποταμιάνος
Αναπληρωτής Καθηγητής

Αθήνα, Ιούλιος 2018

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου
και Ρομποτικής

Αναγνώριση Προσωπικότητας από Σπεκτογράμματα Φωνής σε διαφορετικές Χρονικές Κλίμακες

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΝΙΚΟΛΑΟΣ ΠΑΝΤΕΛΑΙΟΣ

Επιβλέπων : Αλέξανδρος Ποταμιάνος
Αναπληρωτής Καθηγητής

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 9η Ιουλίου 2018.

Αλέξανδρος Ποταμιάνος Αναπληρωτής Καθηγητής	Κωνσταντίνος Τζαφέστας Αναπληρωτής Καθηγητής	Γιώργος Στάμου Αναπληρωτής Καθηγητής

Αθήνα, Ιούλιος 2018

Νικόλαος Παντελαίος

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

Περίληψη

Ο κλάδος της ψυχολογίας έχει για πολλές δεκαετίες ασχοληθεί με την εύρεση ενός μοντέλου περιγραφής της Προσωπικότητας. Με τον όρο Προσωπικότητα αναφερόμαστε σε συγκεκριμένες διαφορές σε χαρακτηριστικά πρότυπα σκέψης, αισθημάτων και συμπεριφοράς. Η προσωπικότητα καθορίζει τους τρόπους που επικοινωνεί και αλληλεπιδρά το άτομο. Δεδομένου της σημασίας της για την επικοινωνία και την ανάπτυξη των διαπροσωπικών σχέσεων στη σημερινή εποχή της ραγδαίας τεχνολογικής ανάπτυξης, η μελέτη της Αναγνώρισης Προσωπικότητας κρίνεται πιο σημαντική από ποτέ.

Υπάρχουν διάφοροι τρόποι αναπαράστασης της προσωπικότητας και οι πιο διαδεδομένοι από αυτούς δίνουν μια ικανή αναπαράσταση για την αποκωδικοποίηση της προσωπικότητας. Η μελέτη της Προσωπικότητας και ο συνολικός κλάδος που αυτή εντάσσεται, επικεντρώνεται στην εξαγωγή χαρακτηριστικών μέσα από προγραμματιστικά μοντέλα και κατα συνέπεια στην κατηγοριοποίηση τους στην αντίστοιχη κλάση. Οι διάφοροι αλγόριθμοι κατηγοριοποίησης έχουν εξελιχτεί στην πάροδο των χρόνων και οι πιο πρόσφατες τεχνικές συμπεριλαμβάνουν νευρωνικά δίκτυα τόσο απλούστερα, όσο και πολύπλοκα νευρωνικά δίκτυα με ανάδραση.

Η μελέτη της Προσωπικότητας γίνεται με κάθε δυνατό μέσο, συγκεκριμένα από την ανάλυση ήχου, κειμένου, εικόνας, βίντεο και κάθε άλλου διαθέσιμου μέσου καθώς και συνδυασμού αυτών. Στην εργασία αυτή θα ασχοληθούμε με την Αναγνώριση Προσωπικότητας με την κατηγοριοποίηση αυτής να γίνεται μέσω ανάλυσης Σπεκτογραμμάτων που εξάγονται από το ηχητικό σήμα και με τη χρήση νευρωνικών δικτύων καταλήγουμε στην εκπαίδευση αυτών και στην τελική πρόβλεψη της Προσωπικότητας. Μετά την εξαγωγή Σπεκτογραμμάτων από τον ήχο, ακολουθεί η ανάλυση των Σπεκτογραμμάτων από Αυτόματους Κωδικοποιητές για την κατάλληλη μετατροπή τους που οδηγεί στα επόμενα στάδια ανάλυσης και μελέτης της εργασίας. Στην παρούσα εργασία μετά την εξαγωγή διαφορετικών χρονικών κλιμάκων από την αρχική είσοδο Σπεκτογραμμάτων , αυτές συνθέτονται, για την εξαγωγή περισσοτέρων χαρακτηριστικών. Με αυτή τη μέθοδο μπορούμε να κάνουμε μια προσέγγιση με καλά αποτελέσματα όσον αφορά την Αναγνώριση Προσωπικότητας.

Στη συνέχεια εφαρμόζουμε τις ίδιες αυτές αρχιτεκτονικές και μεθόδους για την Αναγνώριση Συναισθήματος, για την καλύτερη επαλήθευση αυτών των μεθόδων αλλά και για τη σύγκριση των δύο κλάδων στο βαθμό που αυτό καθίσταται δυνατό.

Συγκεκριμένα το μοντέλο που χρησιμοποιούμε διαχωρίζει τη διαδικασία σε στάδια. Στο πρώτο στάδιο εισάγουμε σαν είσοδο του δικτύου Σπεκτογράμματα που έχουν εξαχθεί από τη Φωνή και τα τροφοδοτούμε στον Αυτόματο Κωδικοποιητή. Εκπαιδεύοντας τον Κωδικοποιητή για διαφορετικές χρονικές Κλίμακες αποθηκεύουμε το εκπαιδευμένο δίκτυο για το επόμενο στάδιο. Διαχωρίζοντας τον Αυτόματο Κωδικοποιητή στο μεσαίο επίπεδο, καταλήγουμε στην εξαγωγή Χαρακτηριστικών από το αρχικό Σπεκτόγραμμα εισόδου. Αυτά τα Χαρακτηριστικά περνούν από τα επόμενα επίπεδα(από 3 - 5 επίπεδα) που είναι είτε Συνελικτικού Δικτύου επίπεδα, είτε επίπεδα πλήρως Συνδεδεμένου Δικτύου και στο τελευταίο στάδιο καταλήγουν στο επίπεδο εξόδου που καθορίζει τη δυαδική απόφαση για το κάθε δεδομένο εισόδου. Αν κάνουμε την διαδικασία εκπαίδευσης του Αυτόματου Κωδικοποιητή για διαφορετικές Κλίμακες, χρησιμοποιώντας διαφορετικά είδη πυρήνα, παίρνουμε εξαγωγή Χαρακτηριστικών σε διαφορετικές Χρονικές Κλίμακες και στη συνέχεια δοκιμάζουμε την συνένωση των διαφόρων κλιμάκων σε ένα είδος ιεραρχικού μοντέλου.

Τα αποτελέσματα που παίρνουμε είναι για τα βασικά πειράματα στο μέσο όρο των 5 αξόνων της Προσωπικότητας στο 58.59%, χρησιμοποιώντας Μηχανές Υποστήριξης Διανυσμάτων καθώς και Ανάλυση Κυρίων Συνιστωσών. Στη συνέχεια τα αποτελέσματα για ένα απλό νευρωνικό δίκτυο δύο

επιπέδων είναι στο 61.50% για το μέσο όρο των 5 αξόνων. Για διαφορετικές χρονικές Κλίμακες και Προεκπαίδευση στη βάση δεδομένων Αναγνώρισης Συναισθήματος, παίρνουμε μέσο όρο των 5 αξόνων της Προσωπικότητας 67.25%. Για εξαγωγή χαρακτηριστικών από Σπεκτογράμματα έχουμε μέσο όρο κλάσεων 63.15% και τελος, εφαρμόζοντας τις αρχιτεκτονικές συνένωσης διαφορετικών χρονικών Κλιμάκων παίρνουμε 68.51% για τους 5 άξονες της Προσωπικότητας.

Abstract

The field of phychology has for many decades studied Personality and the development of a model describing it. The term Personality refers to specific differences on characteristic patterns of thought, emotions and behaviour. Personality defines how a person communicates and interacts. Taking into account the significance of communication and the development of interpersonal relationships in today's age of rapid technology development, studying Personality Recognition is more crucial than ever.

There are many ways to represent Personality and the most important of them give the necessary tools to decode Personality. The study of Personality and the overall field it belongs to, focuses on feature extraction through programming models and consequently classifying it to the specified class. Various classifying algorithms have evolved over time kai most recent techniques include Neural Networks both simpler, as well as more complex neural networks with feedback.

Studying Personality includes every possible input, specifically from speech analysis, to text, image, video and every other available mean analysis, as well as combination of either of them. On this current Thesis we are gonna focus on Personality Recognition, where its classification is decided by Spectrogram Analysis, produced by speech with the help of neural networks and resulting to their training and the final Personality Classification. Immediately after Spectrogram Extraction from speech, the Spectrogram Analysis from Autoencoders follows, for their appropriate conversion which leads to next analysis stages of this project. Then, after extracting different time scales from initial Spectrogram input, they are combined together, for further feature extraction. This method leads to good results in Personality Recognition.

Moreover, we apply the same architectures and methods for Emotion Recognition, for a further technique verification as well as a comparison between the two fields, as far as this is possible.

Particularly, the model we are using distills separates the process in stages. In the first stage, we feed our Autoencoder with Spectograms, extracted from Speech data. By training the Autoencoder for different time scales, we save our network for the next stages. By cutting the Autoencoder in the middle layer, we manage to extract features from the initial Spectogram input. These features come through the next layers (3 - 5 layers), which are either Convolutional or Fully-Connected Layers and in the last stage they connect to the output layer which defines the binary classification for every input data. Following the Autoencoder training process, we extract Features in different time scales and consequently we apply concatenation of the different time scales in a way that a hierarchical model is constructed.

The results we end up with for our basic experiments are 58.59% , using SVM and PCA. Afterwards, using a simple Neural Network of 2 hidden layers we have for the mean value of the 5 Personality Axes a 61.50% classification result. Using Transfer Learning and pretraining our network on IEMOCAP dataset, we have 67.25% for the OCEAN values of Personality. Using Feature Extraction and a single timescale, our classification results are 63.15% and finally, by using different timescales concatenation the classification results rise to 68.51% for the 5 Personality axes.

Ευχαριστίες

Με την εργασία αυτή ολοκληρώνεται ο κύκλος σπουδών μου στη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών στο Εθνικό Μετσόβιο Πολυτεχνείο. Νιώθω την ανάγκη και την υποχρέωση να αποδώσω τις ευχαριστίες σε όσους με βοήθησαν και με οδήγησαν, σε αυτούς που το αξίζουν.

Αρχικά θα ήθελα να ευχαριστήσω τον κ. Αλέξανδρο Ποταμιάνο, επιβλέποντα καθηγητή της διπλωματικής μου εργασίας, για την προσφορά του στην περάτωση αυτής καθώς και για τη συνολική αλληλεπίδραση που είχαμε σε ερευνητικό επίπεδο. Μου προσέφερε καθοδήγηση, εμπιστοσύνη και μία ερευνητική ομάδα υψηλού επιπέδου, παράγοντες καταλυτικοί για την δική μου εξέλιξη και της έρευνας μου. Πάνω απ'όλα τον ευχαριστώ για το ενδιαφέρον του, την κατανόηση που μου έδειξε και το συνολικό ερευνητικό του έργο που με ενέπνευσε να συνεχίζω την έρευνα μου και να προσπαθω συνεχώς περισσότερο.

Έπειτα θα ήθελα να ευχαριστήσω την οικογένεια μου γιατί χωρίς εκείνους δεν θα βρισκόμουν εδώ που είμαι σήμερα. Το υγιές περιβάλλον που δημιούργησαν με βοήθησε να κάνω τις επιλογές μου χωρίς άγχος και με γνώμονα το καλύτερο για εμένα. Η καθοδήγηση και η στήριξη τους εκτείνεται σε όλες τις βαθμίδες εκπαίδευσης και κατ' επέκταση στην ολοκλήρωση μιας σχολής που μου αρέσει και επέλεξα προσωπικά. Ο πατέρας μου, η μητέρα μου, τα τρία μου αδέρφια, ο παππούς μου, οι γιαγιάδες μου · όλοι συνέβαλλαν στο να είμαι ευτυχισμένος, να κυνηγάω τα όνειρα μου και να είμαι όλα όσα είμαι σήμερα.

Τέλος ευχαριστώ τους φίλους μου για τις στιγμές που περάσαμε μαζί, τις εμπειρίες που μοιραστήκαμε και τις αναμνήσεις που δημιουργήσαμε σε αυτό το στάδιο της ζωής μας ως φοιτητές.

Νικόλαος Παντελαίος,

Αθήνα, 9η Ιουλίου 2018

Περιεχόμενα

Κατάλογος πινάκων

Κατάλογος σχημάτων

Κεφάλαιο 1

Εισαγωγή

1.1 Προσωπικότητα: Ορισμός και Γνωρίσματα

Οι θεωρίες σχετικά με το τι είναι η προσωπικότητα και πως αναπτύσσεται είναι πολλές. Στη δημιουργία και διαμόρφωση μια προσωπικότητας εμπλέκονται διάφοροι παράγοντες. Σύμφωνα με μια εκδοχή η προσωπικότητα αναφέρεται σε συγκεκριμένες διαφορές σε χαρακτηριστικά πρότυπα σκέψης, αισθημάτων και συμπεριφοράς. Η μελέτη της προσωπικότητας επικεντρώνεται σε δύο διαφορετικούς τομείς. Ο πρώτος αφορά την κατανόηση των διαφορών σε συγκεκριμένο χαρακτηριστικό της προσωπικότητας όπως η κοινωνικότητα. Ο δεύτερος είναι η κατανόηση των διαφορετικών πλευρών του ανθρώπου και πως αυτές συνθέτους συνολικά την προσωπικότητα του [4].

Ένα από τα βασικά σημεία της προσωπικότητας είναι η συνοχή. Υπάρχει μια αναγνωρίσιμη τάξη και κανονικότητα στις συμπεριφορές του ατόμου ανάλογα με την προσωπικότητα του. Αυτό σημαίνει ότι ένας άνθρωπος, μια προσωπικότητα, θα αντιδρά με τον ίδιο ή πολύ παρόμοιο τρόπο σε μια ποικιλία καταστάσεων. Η προσωπικότητα εκφράζεται επίσης με πολλές μορφές συμπεριφοράς. Αυτή επηρεάζει όχι μόνο το πως ένας άνθρωπος ανταποκρίνεται ή προσαρμόζεται στο περιβάλλον του αλλά και το πώς ενεργεί. Η προσωπικότητα ενός ατόμους εκδηλωνεται ή εμφανίζεται , στις σκέψεις , στα αισθήματα , στις στενές σχέσεις και σε άλλες κοινωνικές αλληλεπιδράσεις. Οι σχολές σκέψης , ανάλυσης και θεωρίας για τη γένεση και την εξέλιξη της προσωπικότητας έχουν κάποτε αντικρουόμενες απόψεις. Στις θεωρίες για την προσωπικότητα περιλαμβάνονται οι ψυχοδυναμικές, οι συμπεριφορικές , οι ανθρωπιστικές και άλλες.

1.2 Μελέτη Προσωπικότητας στο Σήμερα

Η προσωπικότητα, με την έννοια του συνόλου του χαρακτηριστικών του κάθε ατόμου , είναι σήμερα δείκτης για διάφορα θέματα, κυρίως στις διαπροσωπικές σχέσεις. Η προσωπικότητα έχει συνδεθεί με την καλλιέργεια των φιλικών σχέσεων, το πόσο εύκολα αποδέχεται τις αλλαγές της τεχνολογίας και τα νέα επιτεύγματα ακόμα και με τη μακροζωία. Συγκεκριμένα, έρευνες εξετάζουν πόσο η προσωπικότητα του ατόμου επηρεάζει το βαθμό στον οποίο αποδέχεται την τεχνολογία [5].

Οι πιο πρόσφατες προσπάθειες για αναγνώριση της προσωπικότητας καταλήγουν σε εξειδικευμένα διαγωνίσματα προσωπικότητας *personality tests* που καθορίζουν την προσωπικότητα των εξεταζόμενων ανάλογα με το εκάστοτε ψυχολογικό μοντέλο που ακολουθεί το καθένα. Όσων αφορά την αυτοματοποιημένη πρόβλεψη της προσωπικότητας, οι τελευταίες προσπάθειες εστιάζουν στην κατασκευή υπολογιστικών μοντέλων που εκμεταλλεύονται ήδη υπάρχοντα δεδομένα για τους χρήστες και αναλύοντας τα προσπαθούν να εξάγουν πρότυπα που αυτά ακολουθούν. Στην παρούσα φάση στόχος είναι η δημιουργία ενός τέτοιου μοντέλου που να κάνει επιτυχημένη πρόβλεψη της προσωπικότητας δεχόμενο ως είσοδο δεδομένα σε κάθε μορφή, είτε είναι ήχος , είτε εικόνα, είτε κείμενο, είτε συνδυασμός.

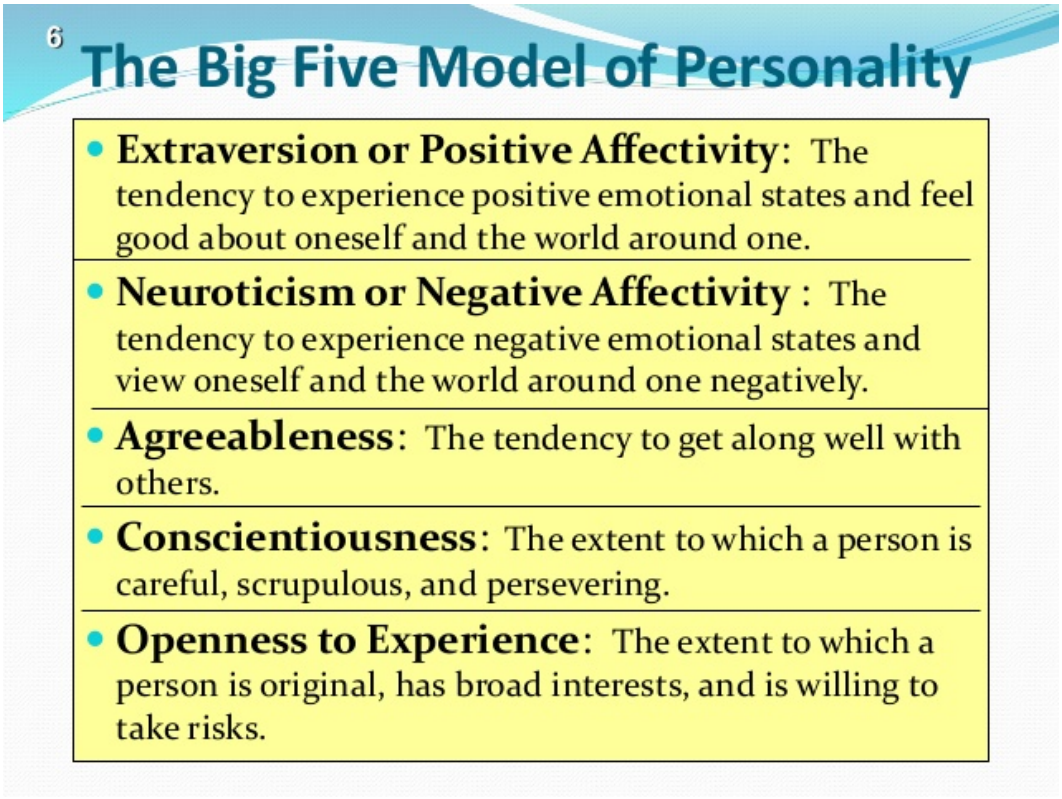
1.3 Αναπαράσταση Προσωπικότητας

Οι τρόποι με τους οποίους μπορεί να αναπαρασταθεί η προσωπικότητα για καλύτερη οπτική αναγνώριση ποικίλλουν και οι περισσότεροι εξ αυτών, αν όχι όλοι, προέρχονται από μοντέλα ψυχολογίας που έχουν δοκιμαστεί σαν θεωρίες για να παράξουν τα αποτελέσματα που χρησιμοποιουν πλέον όλοι οι κλάδοι για την αναπαράσταση της προσωπικότητας. Φυσικά δεν υπάρχει καθολικά σωστός τρόπος αναπαράστασης καθώς κάθε φορά εξαρτάται από το είδος του συμπεράσματος που θέλουμε να εξαχθεί, την ποσότητα και την ποιότητα των δεδομένων και σε εξίσου μεγάλο βαθμό , από την διαθεσιμότητα των δεδομένων. Συγκεκριμένα εδώ θα πρέπει να τονιστεί ότι ο κλάδος της αυτόματης πρόβλεψης της προσωπικότητας πλήττεται από τον περιορισμό των δεδομένων καθώς αυτά βρίσκονται είτε σε πολύ μικρή κλίμακα, είτε είναι πολύ κοστοβόρα διαδικασία, αλλά αυτό θα αναλυθεί στη συνέχεια.

Η εστίαση θα γίνει στην παρουσίαση συνοπτικά των πιο διαδεδομένων μοντέλων που χρησιμοποιούνται στην αναπαράσταση προσωπικότητας, πώς έχουν δοκιμαστεί αυτά τα μοντέλα και που αναφέρεται η βιβλιογραφία στη σημερινή εποχή ως προς τις προτιμήσεις των μοντέλων.

1.3.1 Big-5

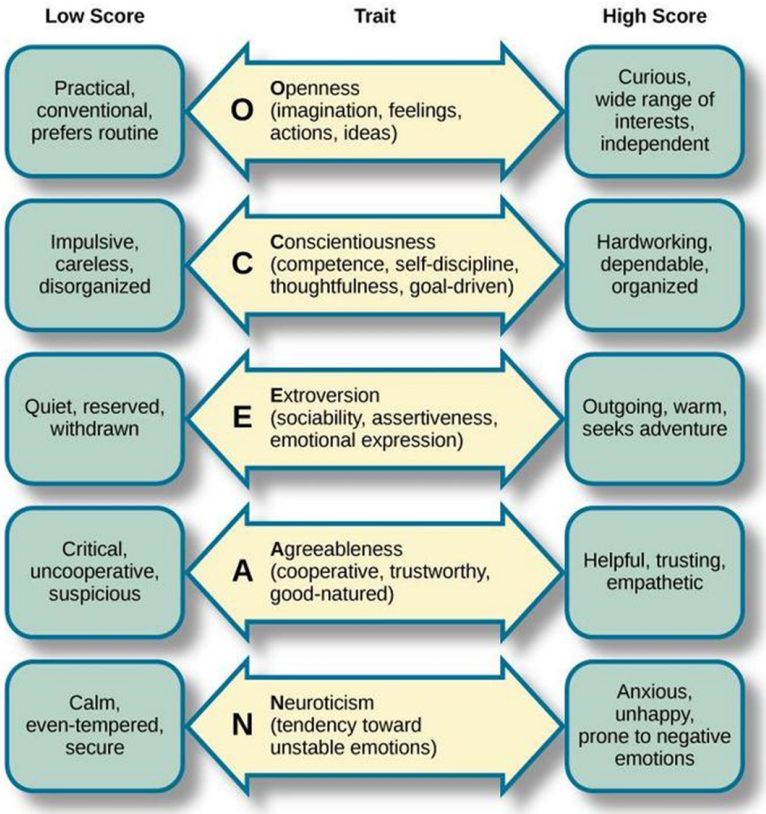
Μία από τις πιο γνωστές κατηγοριοποιήσεις είναι αυτή του Goldberg , που δημοσιεύτηκε το 1990. [2]. Σύμφωνα με αυτή χωρίζει το χώρο της προσωπικότητας σε 5 υποχώρους δυαδικής υπόστασης στο σχήμα 1.1.



Σχήμα 1.1: Οι 5 άξονες της προσωπικότητας κατά Goldberg [2].

Οι πέντε αυτοί υπόχωροι είναι η Εξωστρέφεια (*Extraversion*), η Νευρικότητα (*Neuroticism*), η Ανοικτότητα (*Openness*), η Ευσυνειδησία (*Conscientiousness*) και η Προθυμία (*Agreeableness*). Αναλύοντας έτσι τους πέντε ξεχωριστούς και ανεξάρτητους δυαδικούς άξονες, μπορούμε να βγάλουμε συμπεράσματα για το προφίλ της προσωπικότητας ενός ατόμου. Σαν συντομογραφία αυτοί οι πέντε άξονες αποκαλούνται Big-5 άξονες της προσωπικότητας ή O.C.E.A.N. , από τα αρχικά των ονομάτων

τους στα αγγλικά. Ένας απλός μετασχηματισμός των πέντε αυτών εννοιών σε επίθετα που χαρακτηρίζουν τον άνθρωπο και είναι πιο κοντά στην κατανόηση μας παρουσιάζεται στο σχήμα 1.2



Σχήμα 1.2: Αντιστοίχιση του O.C.E.A.N. σε χαρακτηριστικά [2].

Έτσι αυτό το μοντέλο γίνεται πιο κατανοητό για την απόδοση προσωπικότητας . Το συγκεκριμένο μοντέλο μάλιστα είναι και εκείνο που θα χρησιμοποιήσουμε παρακάτω στην συλλογή, ανάλυση και αξιοποίηση των δεδομένων μας και γι' αυτό παρουσιάζεται και πρώτο.

1.3.2 Myers-Brigg

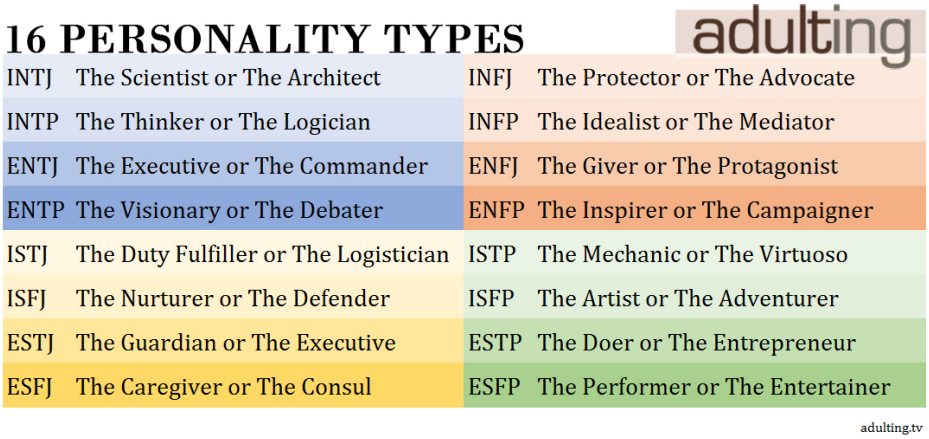
Μία διαφορετική αναπαράσταση είναι αυτή που προτείνει η Myers-Briggs [3]. Συγκεκριμένα χωρίζει την προσωπικότητα σε 4 διαφορετικούς τομείς και ο συνδυασμός αυτών καθορίζει τη συνολική προσωπικότητα. Οι 4 αυτοί τομείς είναι 1) ο τρόπος στον οποίο επικεντρώνεσαι (εξωστρεφής ή εσωστρεφής), 2) ο τροπος με τον οποίο εξετάζεις τις πληροφορίες (δεχόμενος την πραγματικότητα ή φανταζόμενος τις πιθανότητες), 3) ο τρόπος με τον οποίο προτιμάς να παίρνεις τις αποφάσεις (σκεπτόμενος ή με το συναίσθημα) και 4) πως αντιλαμβάνεσαι το περιβάλλον σου (κρίνοντας με βάση τους κανόνες ή με το συναίσθημα) . Ο συνδυασμός των δυαδικών αυτών αξόνων καθορίζει την αντιστοιχη προσωπικότητα, όπως φαίνεται και στο σχήμα 1.3.

Συνολικά έχουμε 4 δυαδικούς άξονες και όλοι οι πιθανοί συνδυασμοί αυτών μας δίνουν 16 διαφορετικά είδη προσωπικοτήτων. Οι 4 άξονες χωρίζονται ως εξής :

I-E : Εξωστρεφής και Εσωστρεφής
S-I : Προμελετημένος και Ενστικτώδης
T-F : Σκεπτόμενος και Αισθανόμενος
J-P : Κριτικός και Αντιληπτικός

και λίγα λόγια για κάθε έναν από τους 16 συνδυασμούς :

ISTJ : ειλικρινής, αναλυτικός, συγκρατημένος, ρεαλιστικός, συστηματικός
ISFJ : θερμός, διακριτικός, ευγενικός, υπεύθυνος, πραγματιστής
INFJ : ιδεαλιστής, οργανωτικός, διορατικός, αξιόπιστος, συμπονετικός
INTJ : καινοτόμος, ανεξάρτητος, στρατηγικός, λογικός, συγκρατημένος
ISTP : δραστικός, λογικός, αναλυτικός, αυθόρμητος, ανεξάρτητος
ISFP : πράος, ευαίσθητος, βοηθητικός, ευέλικτος, ρεαλιστής
INFP : δημιουργικός, ιδεαλιστής, αντιληπτικός, πιστός, ευαίσθητος
INTP : διανοούμενος, λογικός, ακριβής, συγκρατημένος, ευέλικτος
ESTP : κοινωνικός, ρεαλιστής, δραστικός, προσαρμοστικός, αυθόρμητος
ESFP : ενθουσιώδης, φιλικός, αυθόρμητος, στρατηγικός, ευέλικτος
ENFP : ενθουσιώδης, δημιουργικός, αυθόρμητος, αισιόδοξος, υποστηρικτικός
ENTP : εφευρετικός, ενθουσιώδης, στρατηγικός, εξεταστικός, προσαρμοστικός
ESTJ : αποδοτικός, κοινωνικός, συστηματικός, ρεαλιστής, αξιόπιστος
ESFJ : φιλικός, κοινωνικός, αξιόπιστος, ευσυνείδητος, οργανωτικός
ENFJ : καλόκαρδος, ενθουσιώδης, ιδεαλιστής, οργανωτικός, διπλωματικός
ENTJ : στρατηγικός, λογικός, αποδοτικός, κοινωνικός, φιλόδοξος



Σχήμα 1.3: Briggs-Myers κλίμακα προσωπικότητας [3].

1.3.3 Περαιτέρω Αναπαραστάσεις Προσωπικότητας

Υπάρχουν και άλλες αναπαραστάσεις της προσωπικότητας, λιγότερο ίσως χρησιμοποιούμενες όπως η αναπαράσταση 6 αξόνων του σχήματος 1.4

Σύμφωνα με το [6] η σύνθεση της ανθρώπινης προσωπικότητας είναι το αντικείμενο πολλών θεωριών. Κάθε μία από αυτές εστιάζει σε μία ή περισσότερες πλευρές των διαστάσεων της προσωπικότητας, αλλά καμία από αυτές λαμβάνουν υπόψη την πλήρη αναπαράσταση , το βάθος και την αντικειμενικότητα των διαστάσεων της προσωπικότητας. Μερικές θεωρίες εστιάζουν περισσότερο στις διαφορές σε τύπους , ενώ άλλες στη σχετική ανάπτυξη των συγκεκριμένων χαρακτηριστικών και αρμοδιοτήτων και κάποιες άλλες στη δυναμικότητα της αλληλεπίδρασης μεταξύ διάφορων ψυχολογικών κατασκευών όπως το υπερεγώ και η ταυτότητα.

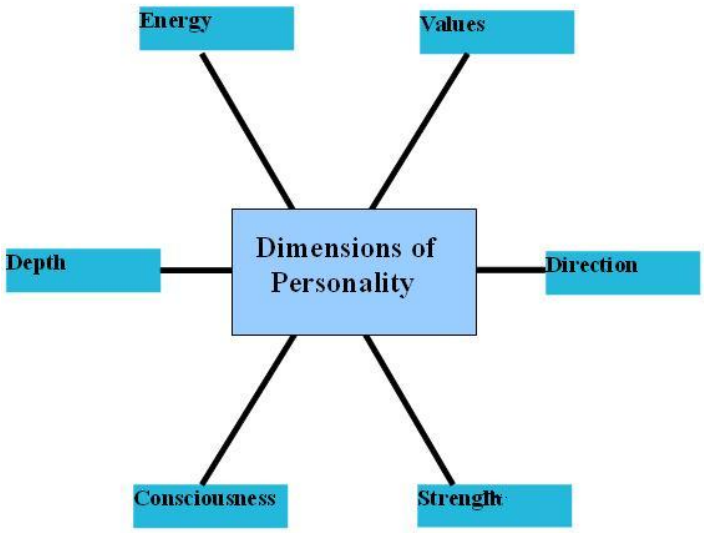
Μια τυπολογική προσέγγιση για την κατηγοριοποίηση των προσωπικοτήτων αποτυγχάνει να λάβει υπόψη τις ποικιλομορφίες της έντασης που θα μπορούσαν επαρκώς να διαχωρίσουν μια προσωπικότητα ανθρώπου τύπου Ναπολέων από έναν τύπο προσωπικότητα κυρίαρχου τοπικού ηγέτη. Οι περιγραφές δεν ξεχωρίζονται επαρκώς μεταξύ τους σε σχέση με τις εξωτερικές εκφράσεις της προσωπικότητας και της εσωτερικής ώθησης. Σε μια προσπάθεια προς επιστημονική αντικειμενικότητα και αμεροληψία, οι περισσότερες θεωρίες παραλείπουν να δίνουν τιμές στις διαφορές στην κατεύθυνση προσωπικότητας των άλλεων ανθρώπων και του γύρω περιβάλλοντος, κι όμως παρ' όλα αυτά

η διάκριση μεταξύ ενός καλού, ευγενικού και καλόκαρδου ατόμου που είναι θετικός με όλο τον περίγυρο του και ενός ζηλιάρη , μοχθηρού με αμφιλεγόμενα κίνητρα ανθρώπου δεν μπορεί απλώς να απορριφθεί ως διαφορά τύπων.

Μια κατανοητή θεωρία της προσωπικότητας θα πρέπει να λαμβάνει υπόψη και να ενσωματώνει όλες αυτές τις διαστάσεις , τους τύπους χαρακτήρων, τα επίπεδα διαφορετικότητας και της ανάπτυξη της δυναμικής. Το πρώτο βήμα σε αυτή την κατεύθυνση θα ήταν μια θεωρία που ξεκάθαρα αναγνωρίζει τις δομικές διαστάσεις στις οποίες οι διαφορετικές προσωπικότητες διαφέρουν.

Οι προσωπικότητες των ανθρώπων μπορούν να εξεταστούν και να διαχωριστούν σε έξι βασικές διαστάσεις :

1. Ενέργεια
2. Κατεύθυνση
3. Αρχές
4. Βάθος - Μοναδικότητα
5. Ευσυνειδησία
6. Δύναμη προσωπικότητας



Σχήμα 1.4: Αναπαράσταση Προσωπικότητας 6 διαστάσεων.

1.4 Συνεισφορά Εργασίας

Ο κύριος σκοπός της παρούσας εργασίας είναι η Αναγνώριση της Προσωπικότητας του Ομιλητή από Σήματα Φωνής. Σύμφωνα με το σχήμα απόδοσης προσωπικότητας του Goldberg [2], η αναπαράσταση Προσωπικότητας 5 αξόνων είναι αυτή που θα χρησιμοποιήσουμε για να ταξινομήσουμε την προσωπικότητα των δειγμάτων που έχουμε.

Η προσωπικότητα ενός ατόμου θα θεωρηθεί σαν ένα σύνολο από χαρακτηριστικά που αφορούν την δυαδική ταξινόμηση στους 5 διαφορετικούς άξονες (Εξωστρέφεια (*Extraversion*), η Νευρικότητα (*Neuroticism*, η Ανοικτότητα (*Openness*), η Ευσυνειδησία (textitConscientiousness) και η Προθυμία (*Agreeableness*). Κατά συνέπεια στο σύνολο τους συνθέτουν τη συνολική προσωπικότητα, κάτι το οποίο στην παρούσα εργασία θα θεωρηθεί ως δεδομένο.

Προηγούμενες δουλειές πάνω στην Αναγνώριση Προσωπικότητας έχουν γίνει τόσο σε θεωρητικό επίπεδο [7], όσο και σε επίπεδο Αναγνώρισης Φωνής [8] και [9] , τομέας που θα αναλύσουμε και εμείς

στη συνέχεια.

Η οργάνωση της εργασίας ακολουθεί την ακόλουθη δομή. Στο πρώτο κεφάλαιο αναλύουμε την έννοια της Προσωπικότητας, τη σημαντικότητα της σαν τομέας έρευνας καθώς και συνήθεις αναπαραστάσεις της. Στο δεύτερο κεφάλαιο αναλύουμε την ανθρώπινη φωνή, που είναι η βάση για την Αναγνώριση Προσωπικότητας που προέρχεται από Φωνή και τα διάφορα ακουστικά Χαρακτηριστικά που μπορούν να εξαχθούν από αυτή. Στο τρίτο κεφάλαιο της εργασίας αναλύουμε διάφορα μαθηματικά μοντέλα, που αποτελούν και το υπόβαθρο αυτής της εργασίας, ξεκινώντας από τη θεωρία πιθανοτήτων, συνεχίζοντας στις Μηχανές Υποστήριξης Διανυσμάτων και κλείνοντας με τα διάφορα είδη Νευρωνικών Δικτύων που θα χρησιμοποιηθούν στη συνέχεια. Στο τέταρτο κεφάλαιο της εργασίας θα αναλυθούν οι Αυτόματοι Κωδικοποιητές, τα Προεκπαιδευμένα Νευρωνικά Δίκτυα και η Μεταφορά Μάθησης, έννοιες και αρχιτεκτονικές που χρησιμοποιήθηκαν σε όλα σχεδόν τα πειράματα που παρουσιάζονται στο κεφάλαιο πέντε. Ξεκινώντας από μια αναδρομή σε σχετική έρευνα στον συγκεκριμένο τομέα, γίνεται ανάλυση των βάσεων δεδομένων που χρησιμοποιήθηκαν και στη συνέχεια παρουσιάζονται τα πειράματα που έγιναν, οπτικές αναλύσεις των πειραμάτων και σύγκριση τους. Τέλος στο έκτο και τελευταίο κεφάλαιο της εργασίας αναφέρονται τα συμπεράσματα και οι προεκτάσεις της εργασίας.

Κεφάλαιο 2

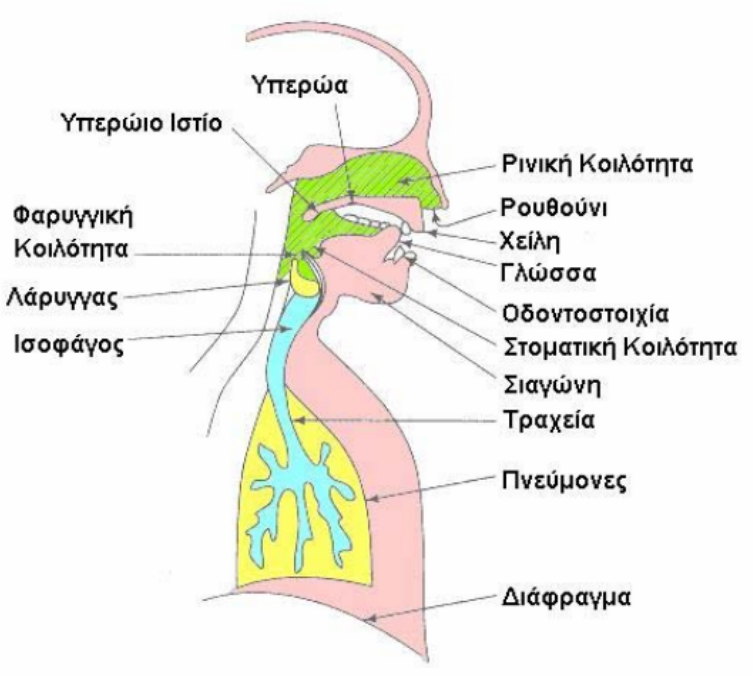
Η ανθρώπινη Φωνή

2.1 Εισαγωγή

Η φωνή, μέσα από την εξελικτική διεργασία, έγινε το μέσο επικοινωνίας των ανθρώπων μεταξύ τους και συνέβαλλε σε μεγάλο βαθμό στην εξέλιξη του πολιτισμού παίζοντας κυρίαρχο ρόλο στην ιστορία του πλανήτη. Αυτή η ξεχωριστή ικανότητα του ανθρώπου, προσέδωσε πλεονέκτημα σε σχέση με τους υπόλοιπους ζωντανούς οργανισμούς που συμβίωναν με αυτόν. Επιπροσθέτως, η φωνή είναι αυτή που βοηθάει τον άνθρωπο να εκφραστεί όσον αφορά τα συναισθήματα του και να δείξει ή να προσποιηθεί πλευρές της προσωπικότητας του. Συνολικά λοιπόν η φωνή οδηγεί στην εξέλιξη της ανθρωπότητας αλλά και στην καλύτερη κατανόηση στις διαπροσωπικές σχέσεις [10].

2.2 Παραγωγή φωνής

Η φωνή εξετάζεται ως αντικείμενο μελέτης τόσο για τον μηχανισμό της παραγωγής της όσο και για τα ξεχωριστά της χαρακτηριστικά. Η φωνή ελέγχεται, αναπτύσσεται και συντηρείται από τη συνεχή ροή πληροφοριών μεταξύ του των μυών υπεύθυνων για την παραγωγή της και του εσωτερικού μηχανισμού ακοής. Εν συνεχεία , σε συνεργασία με τμήματα του εγκεφάλου συντονίζονται τα επιμέρους μέλη για την παραγωγή της φωνής. Συνολικά δηλαδη, το σύστημα παραγωγής της θεωρείται ως ένα από τα πιο περίπλοκα συστήματα του ανθρώπινου οργανισμού και μία τυπική αναπαράσταση του βρίσκευται στο σχήμα 2.1.

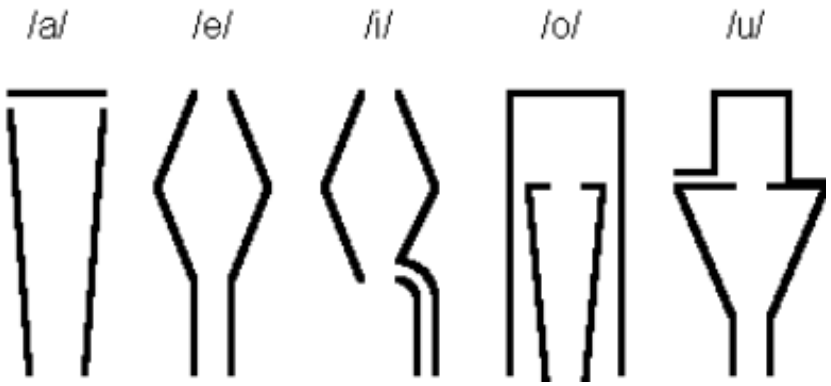


Σχήμα 2.1: Σύστημα Παραγωγής Φωνής.

Η παραγωγή της φωνής είναι μια δράση σύνθετη που περιλαμβάνει πληθώρα οργάνων και συστημάτων του σώματος. Αρχικά μέσω του αναπνευστικού συστήματος ο οργανισμός μας εκμεταλλεύεται τον αέρα σε συνδυασμό με τις φωνητικές χορδές και τις διάφορες αντανακλάσεις του αέρα στα φυσικά ηχεία του σώματος παράγεται η φωνή. Αναπνοή είναι η εισπνοή και εκπνοή του αέρα. Με την εισπνοή εισάγεται αέρας με κατεύθυνση και τελικό προορισμό τους πνεύμονες, ενώ εκπνοή είναι η εξαγωγή του αέρα από τον οργανισμό μας. Κατά τη διάρκεια της εισπνοής ενεργοποιείται μεγάλος οριζόντιος μύς που ονομάζεται διάφραγμα που βρίσκεται κάτω από τον πνεύμονες , ο οποίος χαμηλώνει με σκοπό τη δημιουργία χώρου. Με αυτήν την πράξη ενεργοποιείται ο μηχανισμός διαστολής του όγκου των πνευμόνων για να καλύψει το χώρο. Έτσι ο οργανισμός δημιουργεί έναν επιπλέον χώρο για να γεμίσει με αέρα. Ουσιαστικά το διάφραγμα σε συνδυασμό με τους πλευρικούς μύες δημιουργούν έναν κλειστό χώρο συμπιεσμένου αέρα. Η παραγωγή του ήχου αρχίζει να ενεργοποιείται κατά τη διάρκεια της εκπνοής αέρα και βρίσκει αντίσταση στους πνεύμονες με σκοπό τη δημιουργία κραδασμού. Η δύναμη αυτή στη συνέχεια ενεργοποιεί τις φωνητικές χορδές που δημιουργούν τον ήχο. Αυτή η δράση πραγματοποιείται χιλιάδες φορές ανά δευτερόλεπτο. Η τελική μορφή του ήχου καθορίζεται από το φάρυγγα, τη στοματική κοιλότητα συμπεριλαμβανοντας τα χείλη, τη γνάθο, τον ουρανίσκο , τη γλώσσα καθώς και τη ρινική κοιλότητα. Στον λάρυγγα υπάρχει ένας επιπλέον χόνδρος σε σχήμα πυραμίδας ο οποίος κοινά ονομάζεται μήλο του Αδάμ. Ο χόνδρος αυτός είναι η ασπίδα των φωνητικών χορδών που βρίσκονται μέσα του. Πίσω από το χόνδρο αυτό βρίσκεται η επιγλωττίδα που καθορίζει την κατάποση και την αντίστροφη δράση με σκοπό να προστατεύσει τις φωνητικές χορδές κατά τη διάρκεια της κατάποσης τροφίμων και υγρών.

2.3 Κωδικοποίηση Ομιλίας

Η πρώτη μηχανή που προσπάθησε να παράγει ανθρώπινους ήχους κατασκευάσηκε το 1779 από τον Christian Kratzenstein και αποτελούταν από ένα σύστημα αντηχείων 2.2.



Σχήμα 2.2: Αντηχεία Kratzenstein

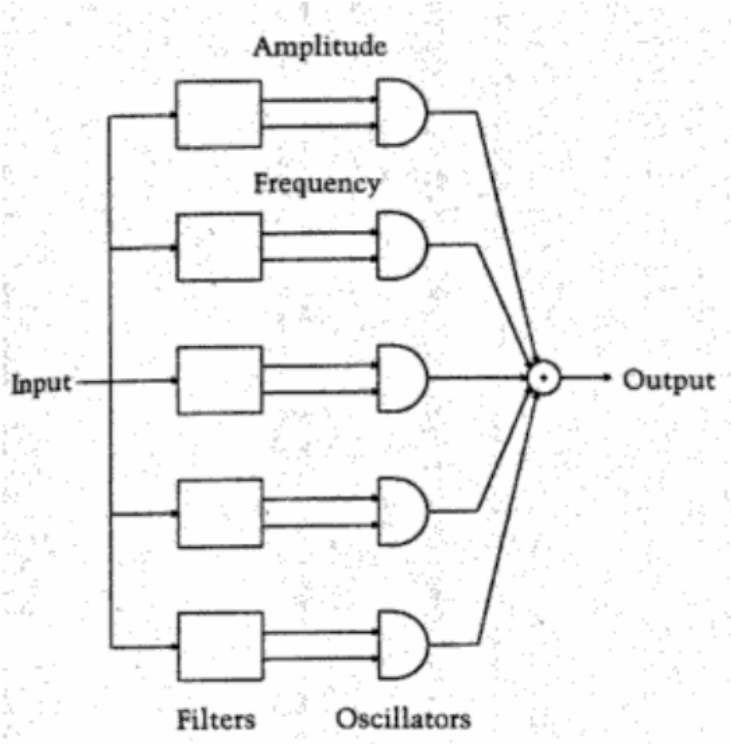
΄Οταν κάποιος φυσούσε απ' την μία άκρη , ακουγόταν από την άλλη τα 5 φωνήεντα α,ε,ι,ο,ου. Λίγα χρόνια αργότερα το 1791 ο Wolfgang von Kempelen εισήγαγε την ομιλούσα μηχανή του που έχοντας εξομοιωτές της γλώσσας και των χειλιών κατάφερνε να προφέρει και κάποια από τα σύμφωνα. Την δεκαετία του 1810 ο Charler Wheatstone με μια παρόμοια μηχανή παρήγαγε μερικές ολόκληρες λέξεις. Τα πειράματα με μηχανικα και ημιηλεκτρικά αναλογικά ηχητικά συστήματα συνεχίστηκαν μέχρι το 1960 περίπου χωρίς αξιοσημείωτες επιτυχίες.

Ηλεκτρική συσκεύη που πετύχαινε σύνθεση κάποιων φωνηέντων εισήχθηκε το 1922 από τον Stewart. Παρ'όλα αυτά η πρώτη ηλεκτρική μηχανή που μπορεί να θεωρηθεί συνθέτης φωνής είναι ο Vocoder του H. Dudley. Το όνομα Vocoder προήλθε από τη σύντμηση των λέξεων Voice και Encoder και δημιουργήθηκε τη δεκαετία του 1930 για να μειώσει το εύρος ζώνης που καταλαμβάνει η ομιλία όταν μεταδίδεται. Αυτό επετεύχθη ανακτώντας την πληροφορία από το ηχητικό σήμα και μεταδίδοντας την με μειωμένο ρυθμό . Μια σειρά από ζωνοδιαβατά φίλτρα χρησιμοποιούνταν για να χωρίσουν

το αρχικό σήμα σε ζώνες συχνοτήτων, τα βάρη των οποίων, αφού μεταδοθούν αρκούν για να επανασυνθέσουν αρκετά καλά το ηχητικό σήμα. Παρουσιάστηκε πρώτη φορα το 1939, στη Διεθνή 'Εκθεση της Νέας Υόρκης. Σήμερα χρησιμοποιούνται κυρίως κωδικοποιητές δύο κατηγοριών, γραμμικής πρόβλεψης και κωδικοποιητές φάσης.

Εξέλιξη αυτών των κωδικοποιητών έχουμε από phase vocoders, σε τράπεζα φίλτρων, σε κωδικοποίηση γραμμικής πρόβλεψης.

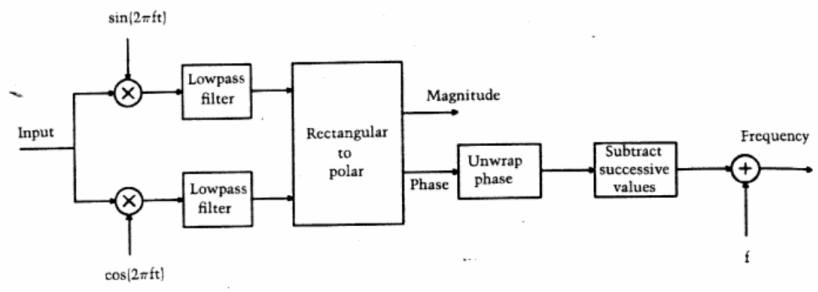
Στους Κωδικοποιητές Φάσης (*(*Phase Vocoders) το σήμα θεωρείται ότι αποτελείται από ένα άθροισμα ημιτονοειδών κυμάτων, το πλάτος και την συχνότητα των οποίων προσπαθούμε να βρούμε. Για να γίνει αυτό περνάμε το σήμα από μία τράπεζα φίλτρων, με την έξοδο του καθενός να εκφράζεται σαν ένα μεταβλητό με τον χρόνο πλάτος στην συγκεκριμένη κεντρική συχνότητα 2.3.



Σχήμα 2.3: Σύστημα Κωδικοποιητή

Η τράπεζα φίλτρων πρέπει να ικανοποιεί 3 περιορισμούς: 1. Η κρουστική απόκριση των ζωνοδιαβατών φίλτρων πρέπει να διαφέρει μόνο στην κεντρική συχνότητα της ζώνης διέλευσης τους. 2. Οι κεντρικές συχνότητες να είναι ισομερώς κατανεμημένες σε όλο το φάσμα από 0 ως το μισό της συχνότητας δειγματοληψίας. 3. Η συνισταμένη κρουστική απόκριση να προσεγγίζει ικανοποιητικά μία σταθερή συνάρτηση σε όλη την έκταση του φάσματος. Η τελευταία προϋπόθεση εξασφαλίζει πως σε καμία συχνοτική συνιστώσα δεν δίνεται δυσανάλογο βάρος. Εξαιτίας των προδιαγραφών τα μοναδικά ζητούμενα στη σχεδίαση μιας τράπεζας φίλτρων είναι ο αριθμός αυτών και η ανεξάρτητη απόκριση τους. Ο αριθμός των φίλτρων πρεέπει να είναι τέτοιος ώστε να μην υπάρχει πάνω από ένα μέρος του σήματος μέσα στη ζώνη διέλευσης κάποιου φίλτρου. Η ανάλυση της λειτουργίας αυτών γίνεται με βάση το σχήμα 2.4.

Αρχικά το σήμα οδηγείται σε δύο παράλληλους δρόμους και διασπάται μέσω ενός μείκτη συχνότητας ίση με την κεντρική του φίλτρου και ενός χαμηλοπερατού φίλτρου. Μόνο τα μέρη του σήματος που είναι κοντά στην κεντρική συχνότητα διέρχονται από το φίλτρο. Έτσι παράγονται δύο στενά ίδια σήματα με διαφορά φάσης $\frac{\pi}{2}$. Τα διαχωρισμένα σήματα οδηγούνται σε ένα μετατροπέα καρτεσιανών σε κυλινδρικές συντεταγμένες. Το αποτέλεσμα προφανώς θα έχει σταθερό με τον χρόνο πλάτος. Η κεντρική συνχότητα υπολογίζεται μετρώντας τη φάση σε δύο χρονικές στιγμές διαιρώντας με το χρόνο. Για να μπορεί να γίνει αυτό η φάση πρέπει να αλλάξει ώστε να μην παίρνει τιμές μόνο στο διάστημα



Σχήμα 2.4: Κωδικοποίηση Φωνής

[0, 360] αλλά σε όλο το R+. Τέλος, προσθέτοντας την κεντρική συχνότητα του φίλτρου λαμβάνουμε το επιθυμητό επεξεργασμένο σήμα.

Η κωδικοποίηση γραμμικής πρόβλεψης είναι μία από τις πιο διαδεδομένες τεχνικές κωδικοποίησης καλής ποιότητας ομιλίας σε χαμηλό ρυθμό μετάδοσης. Αυτή προσπαθεί να προσεγγίσει τον τρόπο παραγωγής της ανθρώπινης φωνής , υποθέτοντας πως παράγεται από ένα βομβητή στο τέλος ενός σωλήνα. Οι LPC αναλύουν το ηχητικό σήμα εκτιμώντας τις αντηχήσεις, αφαιρούν την επίδραση τους στο σήμα και υπολογίζουν την ένταση και τη συχνότητα του εναπομείναντος τόνου. Η διαδικασία αυτή ονομάζεται αντίστροφο φιλτράρισμα και το εναπομείνον σήμα υπόλειμμα. Οι αριθμοί που εκφράζουν τις αντηχήσεις και το υπόλειμμα μπορούν να αποθηκευτούν ή να μεταδοθούν κάπου αλλού. Η σύνθεση γίνεται αντιστρέφοντας την επεξεργασία: με το υπόλειμμα παράγεται ένας αρχικός τόνος, ο οποίος οδηγείται σ'ένα φίλτρο με απόκριση που ορίζουν οι αποθηκευμένες τιμές των αντηχήσεων. Επειδή τα σήματα ομιλίας μεταβάλλονται πολύ γρήγορα με το χρόνο, αυτή η διαδικασία γίνεται σε μικρα κομμάτια του λογου που ονομάζονται frames. Συνήθως, 30-50 frames το δευτερόλεπτο δίνουν επαρκή λόγο με καλή συμπίεση. Το βασικότερο πρόβλημα στους LPC Vocoders έχει να κάνει με τον καθορισμό των αντηχήσεων του ηχητικού σήματος. Η βασική λύση είναι μια εξίσωση διαφορών που εκφράζει κάθε δείγμα σαν γραμμικό συνδυασμό των προηγούμενων. Αυτή η εξίσωση ονομάζεται γραμμικής πρόβλεψης κι από κει προέρχτεεαι το όνομα αυτών των κωδικοποιητών. Οι συντελεστές της εξίσωσης καθορίζουν τις αντηχήσεις, έτσι το LPC σύστημα πρέπει να τους προσεγγίζει όσο γίνεται καλύτερα. Η βελτιστοποίηση γίνεται ελαχιστοποιώντας το μέσο τετραγωνικό σφάλμα μεταξύ του προβλεπόμενου και του πραγματικού σήματος. Η μέθοδος αυτή δίνει ικανοποιητικά αποτελέσματα για τους περισσότερους φθόγγους, όμως σε "ρινικούς" ηχούς λόγω της εισαγωγής ενός ακόμα κλάδου στον ηχητικο σωλήνα ο αλγόριθμος οφείλει να γίνει πιο πολύπλοκος.

Εάν οι συντελεστές πρόβλεψης είναι ακριβείς, τότε μετά από αντίστροφο φιλτράρισμα καταλήγουμε σε έναν καθαρό τόνο. Σε ένα τέτοιο σήμα μπορούμε αρκετά εύκολα να υπολογίσουμε το πλάτος και τη συχνότητα του και να τα κωδικοποιήσουμε. Δυστυχώς, όμως, υπάρχουν κάποια σύμφωνα τα οποία παράγονται με αρκετά στοχαστική ροή αέρα και ακούγονται σα σφύριγμα. Γι'αυτό ο LPC κωδικοποιητής πρέπει να αποφασίζει για κάθε frame αν η ηχητική πηγή είναι τόνος ή σφύριγμα , να αποθηκεύει αυτή την πληροφοριά και στην πρώτη περίπτωση να υπολογίζει τη συχνότητα, ενώ στη δεύτερη την ένταση αυτής. Βεβαιώς εξαιτίας αυτής της προσέγγισης, σύμφωνα που παράγονται από συνδυασμό τόνου και στοχαστικής ροής αέρα δε θα ακούγονται φυσιολογικά (π.χ. δέλτα).

Τα παραπάνω πρόβλήματα δεν θα υπήρχαν αν αποθηκεύαμε ολόκληρο το υπόλειμμα. Κάτι τέτοιο θα είχε σαν αποτέλεσμα να μην υπάρχει καμία συμπίεση, αφού το υπόλειμμα χρειάζεται τον ίδιο αριθμό bits με το αρχικό σήμα. Αυτό είναι απαγορευτικό αφού ο αρχικός μας στόχος ήταν η συμπίεση του σήματος. Έτσι ,διάφορες τροποποιήσεις προσπάθησαν να βελτιώσουν την απόδοση του απλού LPC χωρίς να μειώσουν τη συμπίεση αισθητά. Οι πιο πετυχημένες χρησιμοποιούν έναν πίνακα τυπικών υπολειμμάτων (*codebook*) αποθηκευμένο στο σύστημα. Κατά τη λειτουργία ο αναλυτής λαμβάνει ένα σήμα, το οποίο συγκρίνει με όλα όσα βρίσκονται στον πίνακα υπολογίζοντας ποιο είναι το κοντινότερο και στη συνέχεια στέλνει τον κωδικό αυτού. Ο συνθέτης παίρνει τον κωδικό , αποκαθιστά το υπόλειμμα και αντιστοιχεί αυτόν και το χρησιμοποιεί για να διεγείρει το καθορισμένο φίλτρο. Τέτοιες μέθοδοι ονομάζονται CELP (*Code Excited Linear Prediction*).

Πρόκειται για μια σχετικά καινούρια τεχνολογία που αρχικά αναπτύχθηκε στο πανεπιστήμιο του Stanford και τα τελευταία χρόνια χρησιμοποιείται όλο και περισσότερο. Σ'αυτήν τη μέθοδο ο ήχος χωρίζεται σε μια περιοδική συνιστώσα που αναλύεται με την ίδια περίπου μέθοδο που ακολουθείται και στους κωδικοποιητές φάσης και μια στοχαστική συνιστώσα που αναλύεται σαν φιλτραρισμένος λευκός θόρυβος. Κατά την ανασύνθεση διάφορες τεχνικές δίνουν αρκετά μεγάλες δυνατότητες για επεξεργασία, συμπεριλαμβανομένης και τη μορφοποίηση της χροιάς, αλλά δυστυχώς φαίνεται ότι λειτουργουν πολύ καλύτερα με συγκεκριμένο τύπο ήχων απ' ότι με άλλους. Επειδή με τα σήματα ομιλίας ο συγκεκριμένος τύπος κωδικοποιητή παρουσιάζει σημαντικά προβλήματα προς το παρόν η χρήση του περιορίζεται σε πειραματικές διατάξεις.

2.4 Ανθρωπομορφικές βαθμίδες διασύνδεσης χρήστη

Σύγκριση των σημαντικότερων τύπων κωδικοποιητών ομιλίας

Κάθε μία από αυτές τις τρεις μεθόδους κωδικοποίησης ομιλίας έχει κάποιες κοινές ικανότητες, όπως η δυνατότητα ξεχωριστού χειρισμού του τόνου και της διάρκειας του σήματος , που τις έκαναν τις πιο δημοφιλείς στον τομέα τους. Εν γένει, όμως η καθεμιά έχει τα ιδιαίτερα χαρακτηριστικά της. Αναλυτικότερα, οι κωδικοποιητές φάσης έχουν το πλεονεκτημα να είναι εύκολοι στο χειρισμό, να αποδώσουν αρκετά καλά σε ένα μεγάλο εύρος αρμονικών και μη αρμονικών ήχων. Συνήθως είναι η καλυτερη ή ευκολότερη μέθοδος για χρήση όταν έχουμε να κανουμε με πολύ χαμηλής ή πολύ υψηλής τονικότητας και όταν ο στόχος μας είναι η πιστή αναπαραγωγή του αρχικού σήματος ομιλίας. Όμως , είναι αρκετά περιορισμένοι όσον αφορά την επεξεργασία της ομιλίας. Αν αυξήσουμε τη διάρκεια, πολλές φορές εισάγονται ανεπιθύμητες αλλοιώσεις (μεταλλική χροιά, ηχώ, θόρυβος). Επίσης μετατόπιση του τόνου προκαλεί μετατόπιση και των αρμονικών κάτι που οφηγεί πολλές φορες σε αλλαγή του φασματικού περιεχομένου του σήματος. Τέλος, συχνά η ανάλυση οφείλει να είναι ειδικά κατασκευασμένη για συγκεκριμένο τύπο επανασύνθεσης με αποτέλεσμα να μην λειτουργεί το ίδιο καλά σε άλλους. Οι κωδικοποιητές γραμμικής πρόβλεψης δουλεύουν αρκετά καλά και για τονικούς και για στοχαστικούς ήχους. Η μετατροπή του τόνου δεν προκαλέι μετατόπιση των αρμονικών αντίθετα με τις προαναφερθείσες τεχνικές. Επίσης παρέχει μεγάλες δυνατότητες για τροποποιήσεις. Για τους παραπάνω λόγους οι κωδικοποιητές LPC χρησιμοποιούνται ευρέως παρόλη την σαφώς μεγάλη πολυπλοκότητα και την αρκετά απρόβλεπτη συμπεριφορά. Ένα άλλο μειονέκτημα τους είναι ότι κατά την ανασύνθεση παραγουν ένα απλοποιημένο υπερβολικά αρμονικό μοντέλο ομιλίας, κάτι που υπό συνθήκες μπορεί να γίνει αντιληπτό από το ανθρώπινο αυτί.

Εφαρμογές των Vocoders

Οι Vocoders αρχικά εισήχθησαν κατά τη διάρκεια του δεύτερου παγκοσμίου πολέμου για να ασφαλίσουν τις ραδιοεπικοινωνίες μεταξύ των συμμάχων από τις δύο μεριές του Ατλαντικού. Άρχισαν να χρησιμοποιούνται πιο συχνά κατά τη διάρκεια της δεκαετίας του 1970 , μετά δηλαδή την παρουσία των phase vocoders το 1966. Σήμερα βρίσκουν πολλές εφαρμογές σε διάφορους τομείς, οι κυριότερες των οποίων είναι:

1. Συμπίεση της πληροφορίας που μεταφέρει η φωνή , με κυριότερη τη χρήση στην κινητή τηλεφωνία. Για να καταγραφεί η ανθρώπινη φωνή , η συχνότητα της οποίας κυμαίνεται απο 500Hz έως 8kHz, απαιτείται εύρος ζώνης 64kbit/s ενώ ένας κωδικοποιητής ομιλίας παρέχει καλή εξομοίωση της φωνής με εύρος ζώνης που κυμαίνεται απο 2400bit/s ως 32kbit/s, δηλαδή συμπίεση από 2 έως 26 φορές. Εκτός του βαθμού συμπίεσης άλλοι παράμετροι που χαρακτηρίζουν έναν Vocoder είναι : α) η ποιότητα φωνής στην έξοδο του συστήματος , β) η πολυπλοκότητα του αλγοριθμου και γ) η ανθεκτικότητα σε θόρυβο και σφάλματα μετάδοσης. 2. Κρυπτογράφηση φωνής σε περιπτώσεις απόρρητων και ευαίσθητων συνδιαλέξεων. 3. Καλλιτεχνικές εφαρμογές. Στον κινηματογράφο χρησιμοποιήθηκε Vocoder για πρώτη φορα στην τανία "Η οδύσσεια του διαστήματος" , όπου η φωνή του υπολογιστή HAL παράγεται από κωδικοποιητή ομιλίας. Από τότε , αρκετές φορές οι φωνές ρομπότ ή φανταστικών μηχανών προέκυψαν από τη χρήση Vocoders. 4. Συστήματα αναγνώρισης ή παραγωγής ομιλίας (speech-to-text, synthetic speech). Τα συστήματα αναγνώρισης ομιλίας μπορούν να χρησιμοποιηθούν σε μια σειρά εφαρμογών όπως υπαγόρευση κειμένου , φωνητική αναγνώριση για λόγους ασφαλείας,

έλεγχος φωνής για άτομα με αναπηρία κ.α. Οι συνθέτες ομιλίας βρίσκουν επίσης πολλές εφαρμογές οι κυριότερες από τις οποίες είναι : τεχνητή ομιλία για άτομα φωνητικά ανάπηρα, οι μηχανές διαβάσματος για τους τυφλούς και οι μηχανές διδασκαλίας για δυσλεξικούς.

2.5 Ακουστικά Χαρακτηριστικά

Ένα από τα πιο κρίσιμα συστατικά για την επιτυχημένη αναγνώριση προτύπων είναι το σύνολο των χαρακτηριστικών που εξάγονται. Μετά από δεκαετίες προόδου στην Αναγνώριση Φωνής, πολλές μέθοδοι εμπνευσμένοι από εκεί βρίσκουν εφαρμογή και στην Αναγνώριση Προσωπικότητας, όπως θα δούμε στη συνέχεια. Αυτό οφείλεται στο γεγονός ότι η φωνή του καθενός είναι ξεχωριστή σε μεγάλο βαθμό σε συνδυασμό με το ότι η φωνή του ατόμου οδηγεί τους γύρω του στην εξαγωγή συμπερασμάτων για την Προσωπικότητα του. Οι κοινές συνισταμένες λοιπόν του γενικότερου τομέα της φωνής και της Αναγνώρισης Προσωπικότητας μας οδηγούν στην εξαγωγή χαρακτηριστικών από τη φωνή. Τα χαρακτηριστικά αυτά μας οδηγούν στην τελική κατηγοριοποίηση της Προσωπικότητας έπειτα από την ανάλογη επεξεργασία των διάφορων αρχιτεκτονικών. Τα χαρακτηριστικά αυτά διαχωρίζονται σε διάφορες κατηγορίες ανάλογα με τον τρόπο που έχουν εξαχθεί.

Για την Αναγνώριση Φωνής, τα βασικά χαρακτηριστικά που χρησιμοποιούνται είναι τα χαρακτηριστικά που αφορούν τον τόνο της φωνής, χαρακτηριστικά που αφορούν την ένταση και χαρακτηριστικά που αφορούν τη διάρκεια του φωνητικού σήματος.

Τα χαρακτηριστικά έντασης συνήθως θεωρούνται τα πιο σημαντικά χαρακτηριστικά φωνής που αποτυπώνουν μονότονη ομιλίας ή συλλαβές με διακριτή προφορά και επιτυγχάνουν καλά αποτελέσματα σε μεγάλο εύρος εφαρμογών. Τα χαρακτηριστικά έντασης αφορούν την ενέργεια της φωνής και χρησιμοποιούνται ξεχωριστά. Τέλος, τα χαρακτηριστικά διάρκειας χρησιμοποιούν μετρικές για αθροίσματα των επιμέρους στατιστικών σε δεδομένα χρονικά διαστήματα.

Τόνου Φωνής	μέση τιμή, ελάχιστο, μέγιστο, μαθηματικός μέσος, τυπική απόκλιση, εύρος, απόκλιση τιμής, ρυθμός εναλλαγής
Έντασης	κανονική ένταση, σχετική ένταση, δύναμη της φωνής, ενέργεια, πλάτος ενέργειας
Διάρκειας	ρυθμός διάσχισης μηδενικών, ρυθμός εναλλαγής

Πίνακας 2.1: Τύποι Χαρακτηριστικών Φωνής

Ο κύριος σκοπός των συστημάτων Αναγνώρισης Φωνής είναι η ικανότητα να ακούν, να καταλαβαίνουν και μετά να δρουν με βάση την πληροφορία αυτή. Όσον αφορά τα χαρακτηριστικά Φωνής, αποτελούν τα δύο πρώτα στάδια Αναγνώρισης Προσωπικότητας.

Το πρώτο στάδιο είναι η ανάλυση, δηλαδή όταν ο ομιλητής εκφράζεται, έχει κάποια συγκεκριμένα χαρακτηριστικά που βοηθούν στην αναγνώριση του. Η πληροφορία είναι διαφορετική εξαιτίας της ιδιαιτερότητας της φωνής του και των χαρακτηριστικών συμπεριφοράς. Αυτό το στάδιο της ανάλυσης μπορεί να χωριστεί σε τρεις επιμέρους κατηγορίες.Τα χαρακτηριστικά αναλόγως με τη διάρκεια του σήματος που αντιπροσωπεύουν είναι σε πλαίσιο(*frame-level*) , ομάδας πλαισίων (*segment-level*) και σε ολόκληρης της εκφώνησης (*utterance*).

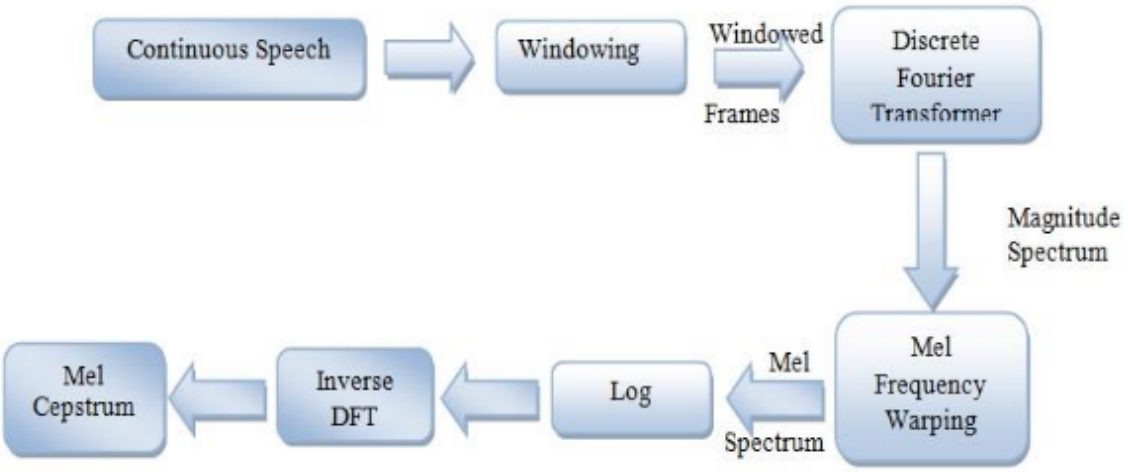
α. Στην ανάλυση χαρακτηριστικών με πλαίσιο, η προσπάθεια εξαγωγής της πληροφορίας του ομιλητή γίνεται με την υλοποίηση ενός μεταβλητού πλαισίου μεγέθους ανάμεσα σε 10 έως $30msec$.

β. Στην ανάλυση υπο-πλαισίων(*sub-segmental*) , η εξαγωγή της πληροφορίας του ομιλητή γίνεται με την υλοποίηση μεταβλητών πλαισίων μεγέθους από 3 έως $5msec$. Έτσι εξάγονται και αναλύονται τα χαρακτηριστικά παραγωγής φωνής.

γ. Στην ανάλυση υπερ-πλαισίων(*supra-segmental*), η προσπάθεια εξαγωγής της πληροφορίας του ομιλητή γίνεται με την υλοποίηση ενός μεταβλητού πλαισίου μεγέθους ανάμεσα σε 50 και $200msec$.

Το δεύτερο στάδιο είναι η επιμέρους εξαγωγή των χαρακτηριστικών. Αυτό το στάδιο της διαδικασίας θεωρείται το πιο σημαντικό για τη γενικότερη Αναγνώριση Προσωπικότητας. Η λειτουργία του

είναι να εξάγει αυτά τα χαρακτηριστικά από το φωνητικό σήμα εισόδου που βοηθούν στην αναγνώριση του ομιλητή. Η εξαγωγή χαρακτηριστικών συμπιέζει το πλάτος του σήματος εισόδου χωρίς να προκαλεί μείωση στην ισχύ του σήματος φωνής. Υπάρχουν πολλές τεχνικές εξαγωγής [11].



Σχήμα 2.5: features1

Στο παραπάνω σχεδιάγραμμα, από τη μία πλευρά εισάγουμε το σήμα συνεχούς ομιλίας για την επεξεργασία της παραθυροποίησης (*windowing*). Στη διαδικασία της παραθυροποίησης οι διακοπές που υπάρχουτν στην αρχή καθώς και στο τέλος του κάθε πλαισίου ελαχιστοποιούνται. Μετά από αυτή τη διαδικασία, το σήμα συνεχούς φωνής μετατρέπεται σε παραθυροποιημένα πλαίσια. Αυτά τα πλαίσια περνάνε στον διακριτό μετασχηματισμό Fourier που τα μετατρέπει στο φάσμα ισχύος των παραθυροποιημένων πλαισίων. Στο επόμενο βήμα, η φασματική ανάλυση γίνεται με μία υποκειμενική κλίμακα συχνότητας σταθερής ανάλυσης που ουσιαστικά είναι η συχνότητα Mel (*Mel-frequency*) που παράγει το φάσμα Mel (*Mel-spectrum*). Αυτό το φάσμα περνάει έπειτα από τον αντίστροφο διακριτό μετασχηματισμό Fourier και παράγει το τελικό αποτέλεσμα σαν φάσμα Mel. Το φάσμα Mel αποτελείται από τα χαρακτηριστικά που είναι αναγκαία για την αναγνώριση του ομιλητή. Κάποιες από τις τεχνικές εξαγωγής χαρακτηριστικών περιλαμβάνουν:

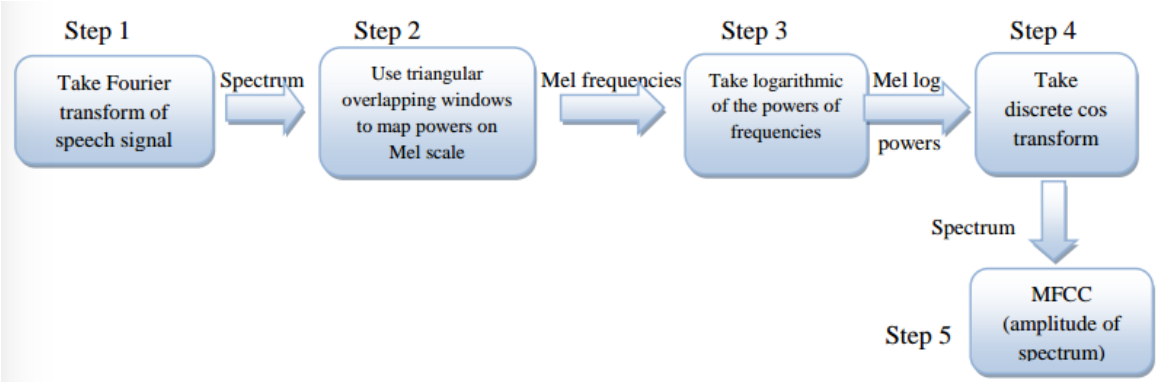
α. Ο Κώδικοποιητής Γραμμικής Πρόβλεψης (*Linear Predictive coding*) είναι ένα εργαλείο που χρησιμοποιείται για την επεξεργασία φωνής. Βασίζετααι στην αξίωση ότι σε μια σειρά δειγμάτων ομιλίας, μπορούμε να προβλέψουμε το ν-οστό δείγμα από το άθροισμα των προηγούμενων κ δειγμάτων του σήματος. Η παραγωγή ενός αντίστροφου φίλτρου γίνεται με τέτοιο τρόπο που ανταποκρίνεται στις περιοχές μορφοποίησης (*formants*).

β. Το φάσμα συχνότητας Mel (*Mel-frequency cepstrum / MFCCs*) βασίζεται στις γνωστές παραλλαγές της συχνότητας του ανθρώπινου αυτιού, για εύρος κάτω απο $1000Hz$. Ο κύριος σκοπός του επεξεργαστή MFCC είναι να προσπαθεί να αντιγράψει τη συμπεριφορά των ανθρώπινων αυτιών. Η διαδικασία της εξαγωγής των MFCCs γίνεται ως εξής :

Οι συντελεστές του φίλτρου $w(n)$ προκύπτουν μετά από την παραθυροποίηση με το παράθυρο Hamming ,

$$W(n) = 0.54 - 0.46\cos(\frac{2\pi n}{N-1}), \quad 0 \leq n \leq N-1 \tag{2.1}$$

Όπου N είναι ο συνολικός αριθμός του δείγματος και n είναι το παρών σήμα. Μετά την παραθυροποίηση , εφαρμόζουμε FFT (*Fast Fourier Transform*) που υπολογίζεται για κάθε πλαίσιο για την εξαγωγη των στοιχείων συχνότητας του σήματος στο πεδίο του χρόνου. Ο FFT χρησιμοποιείται για



Σχήμα 2.6: Εξαγωγή MFCCs

την επιτάγχυση της διαδικασίας. Η λογαριθμική Mel (*Mel-scaled*) τράπεζα φίλτρων εφαρμόζεται στο μετασχηματισμένο πλαίσιο.

Τα MFCCs χρησιμοποιούν Mel τράπεζα φίλτρων όπου τα φίλτρα συχνότητας έχουν μεγάλο εύρος από τα φίλτρα χαμηλής συχνότητας, κρατώντας τις χρονικές αναλύσεις σταθερές. Το τελευταίο στάδιο είναι να προσδιοριστεί ο διακριτός Μετασχηματισμός Συνημιτόνου (*Discrete Cosine Transform*) των εξόδων της τράπεζας φίλτρων. Ο Μετασχηματισμός Συνημιτόνου αντιστοιχεί τους συντελεστές σύμφωνα με τη σημαντικότητα τους όπου ο μηδενικός συντελεστής δεν θεωρείται σημαντικός και αναιρείται.

Για κάθε πλαίσιο φωνής, ένα συνολο συντελεστών MFCCs υπολογίζεται. Αυτό το σύνολο των συντελεστών ονομάζεται ακουστικό διάνυσμα (*acoustic vector*) και αναπαριστά τα φωνητικά σημαντικά χαρακτηριστικά της ομιλίας και είναι πολύ χρήσιμο για περαιτέρω ανάλυση στην Αναγνώριση Φωνής και ειδικότερα εδώ στην Αναγνώριση Προσωπικότητας.

Κεφάλαιο 3

Θεωρητικό Υπόβαθρο

3.1 Εισαγωγή

Υπάρχουν πολλές ξεχωριστές ερμηνείες του όρου πιθανότητα. Για την πλήρη ανάλυση της έννοιας θα πρέπει να εξερευνήσουμε διάφορες περιοχές γνώσης όπως η φιλοσοφία, η θεωρία αλγορίθμων , η τυχαιότητα. Γι'αυτό θα εστιάσουμε σε δύο μόνο αναλύσεις. Η πρώτη εκδοχή εξαρτάται από τη λεγόμενη αντικειμενική σκοπιά και η δεύτερη εκδοχή από την υποκειμενική σκοπιά.

Η υποκειμενική σκοπία ορίζει τις πιθανότητες ως υποκειμενικές έννοιες που βασίζονται στη λογική σκέψη και ανάλυση των διαθέσιμων πληροφοριών. Κάποιοι ερευνητές της υποκειμενικής σχολής ερμηνεύουν τις πιθανότητες ως το βαθμό βεβαιότητας. Αυτός είναι και ο λόγος που είναι δύσκολο να ερμηνευθούν οι πιθανότητες ως γεγονός.

Η αντικειμενική σχολή από την άλλη μερία ορίζει τις πιθανότητες ως μακρινά συνδεδεμένες σχετικές συχνότητες. Αυτό πρακτικά σημαίνει ότι η πιθανότητα ενός γεγονότος υπολογίζεται ως ο αριθμός των επιτυχημένων προσπαθειών που εμφανίζεται ένα γεγονός προς τον αριθμό των συνολικών προσπαθειών και επεκτείνοντας υπολογίζουμε το όριο αυτής της σχέσης καθώς ο αριθμός των προσπαθειών είναι επαρκώς μεγάλος. Κάποιοι στατιστικολόγοι ασκούν βέτο στον όρο "μακρινά συνδεδεμένες". Η αντικειμενική σχολή σκέψης για τον όρο της πιθανότητας αναπτύχθηκε από τους Von Mises(1928) και Kolmogorov(1965). Ο ρώσος μαθηματικός Kolmogorov έδωσε έναν σταθερό ορισμό για τη δημιουργία της θεωρίας των πιθανοτήτων χρησιμοποιώντας τη θεωρία μέτρου. Το πλεονέκτημα της θεωρίας του Kolmogorov είναι ότι είναι πλέον εφικτή η δημιουργία πιθανοτήτων ακολουθώντας τους κανόνες , υπολογίζοντας διαφορετικές πιθανότητες βασιζόμενος στα αξιώματα και έπειτα η ερμήνευση αυτών των πιθανοτήτων.

Στην πραγματικότητα όλος ο κλάδος της θεωρίας Πιθανοτήτων βασίζεται στην θεωρία που ανέπτυξε ο Kolmogorov. Υπάρχουν φυσικά πολλές εφαρμογές της θεωρίας των πιθανοτήτων. Μελετάμε τη θεωρία των πιθανοτήτων επειδή θα θέλαμε να μετελήσουμε τα μαθηματικά στατιστικά. Η στατιστική ασχολείται με την ανάπτυξη μεθόδων και τις εφαρμογές τους για τη συλλογή, ανάλυση και κατανόηση των ποσοτικών δεδομένων με τρόπο τέτοιο ώστε η αξιοπιστία ενός συμπεράσματος βασισμένο στα δεδομένα να μπορεί να αξιολογηθεί αντικειμενικά μέσω των κανόνων της στατιστικής. Η πιθανοτική θεωρία χρησιμοποιείται για την αξιολόγηση της αξιοπιστίας των συμπερασμάτων και των υποθέσεων βασισμένων σε δεδομένα. Για αυτόν τον λόγο η θεωρία των πιθανοτήτων είναι θεμελιώδης για τα μαθηματικά στατιστικά.

Για ένα γεγονός , έστω A, ενός διακριτού χώρου δειγμάτων S, η πιθανότητα του A μπορεί να υπολογιστεί από τον μαθηματικό τύπο :

$$P(A) = \frac{N(A)}{N(S)}$$

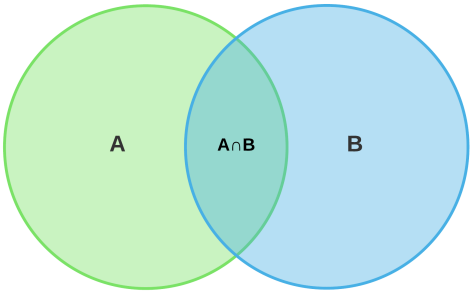
όπου N(A) υποδηλώνει τον αριθμό των στοιχείων του A και N(S) των αριθμό των συνολικών στοιχείων του χώρου δειγμάτων S. Σε διαφορετική περίπτωση, η πιθανότητα ενός γεγονότος A μπορεί

να υπολογιστεί αριθμώντας τον αριθμό των στοιχείων στο A και διαιρώντας το με τον αριθμό των στοιχείων του χώρου δειγμάτων S.

3.2 Πιθανότητες

3.2.1 Δεσμευμένες Πιθανότητες

Συνεχίζοντας στην κρισιμότητα των πιθανοτήτων, θα αναφέρουμε πόσο σημαντικές είναι οι δεσμευμένες πιθανότητες. Για να ορισουμε την δεσμευμένη πιθανότητα θεωρούμε τυχαίο πείραμα του οποίου ο χώρος δειγμάτων είναι ο χώρος S. Έστω $\subset C$ σε πολλές περιπτώσεις μας ενδιαφέρει μόνο τα γεγονό τα που είναι στοιχεία του B. Αυτό σημαίνει ότι θεωρούμε το B σαν το νέο χώρο δειγμάτων μας.



Σχήμα 3.1: Δεσμευμένη Πιθανότητα στα Σύνολα

Θεωρούμε ότι το σύνολο S είναι ένας μη κενός, πεπερασμένος χώρος δειγμάτων και ότι το B είναι ένα μη κενό υποσύνολου του S. Για να ορίσουμε την πιθανότητα του γεγονότος B, ως προς το νέο χώρο δειγμάτων B ως εξης :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

δεδομένου ότι ο αριθμός δειγμάτων του S είναι διάφορος του μηδενός.

Συνεπώς, αν ο χώρος δειγμάτων είναι πεπερασμένος, τότε ο αποπάνω ορισμός της πιθανότητας του γεγονότος A δεδομένου του γεγονότος B βγάζει νόημα διαισθητικά. Σε κάθε χώρο δειγμάτων η δεσμευμένη πιθανότητα ορίζεται ως εξής.

Αυτή η δεσμευμένη πιθανότητα P(A|B) ικανοποιεί και τα τρία αξιώματα του μέτρου πιθανότητας. Δηλαδή,

$$P(A|B) \geq 0$$
$$P(B|B) = 1$$

Αν A_1 , A_2 ,..., A_k ,...είναι αμοιβαίως αποκλειόμενα, τότε

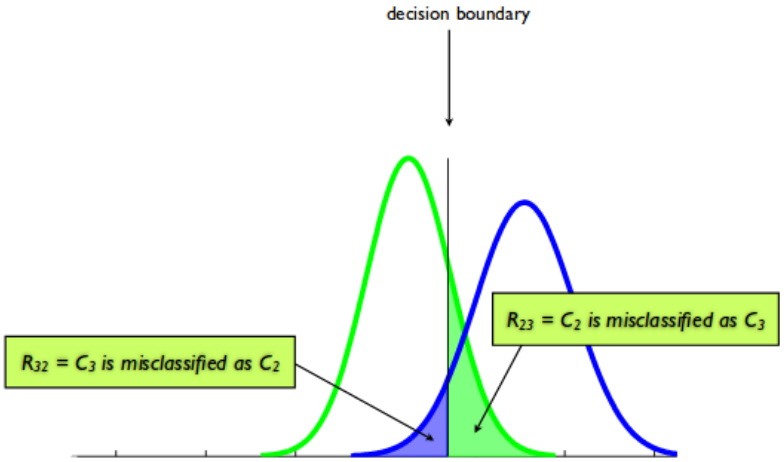
$$P(\bigcup_{i=1}^{\infty} A_k|B) = \sum_{k=1}^{\infty} P(A_k|B)$$

Έτσι λοιπόν, είναι το μέτρο πιθανότητας ως προς το νέο χώρο δειγμάτων B.

Καταλήγουμε λοιπόν σε μια πιθανότητα για κάθε κλάσση και μια συνολική πρόβλεψη για την κλάση που είναι πιο πιθανή.

$$p(error) = \int_{R32} p(C_3|x)dx + \int_{R23} p(C_2|x)dx$$

όπου C_2 και C_3 είναι οι 2 κλάσεις ταξινόμησης καθώς και η γραφική του απεικόνιση στο Σχήμα 3.2



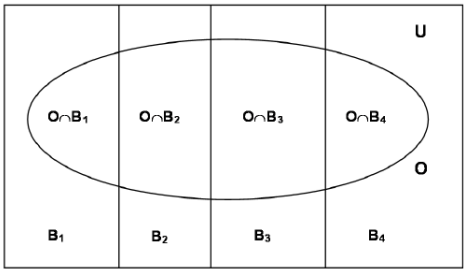
Σχήμα 3.2: Πιθανοτικό Σφάλμα 2 κλάσεων

3.2.2 Θεώρημα Bayes

Υπάρχουν πολλές περιπτώσεις όπου το τελικό αποτέλεσμα ενός πειράματος εξαρτάται από το τι έχει συμβεί στα ενδιάμεσα στάδια. Αυτό το πρόβλημα λύνεται με το θεώρημα του Bayes.

'Εστω S ένα σύνολο και $P = \{A_i\}_{i=1}^{m}$ είναι μια συλλογή υποσυνόλων του S. Η συλλογή P ονομάζεται διαμέριση του S αν:

(α) $S = \bigcup_{i=1}^{\infty} A_i$
(β) $A_i \cap A_j = 0, \quad i \neq j$



Σχήμα 3.3: Θεώρημα Bayes

Αν τα γεγονότα $\{B_i\}_{i=1}^m$ ορίζουν μία διαμέριση του χωρου δειγματων S και $P(B_i) \neq 0$ για κάθε $i = 1, 2, \ldots, m$, τότε για κάθε γεγονός A στο S

$$P(A) = \sum_{i=1}^{m} P(B_i)P(A|B_i)$$

Επίσης αν τα γεγονότα $\{B_i\}_{i=1}^m$ ορίζουν μία διαμέριση του χωρου δειγματων S και $P(B_i) \neq 0$ για κάθε $i = 1, 2, \ldots, m$, τότε για κάθε γεγονός A στο S όπου $P(A) \neq 0$ τότε

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_{i=1}^{m} P(B_i)P(A|B_i)} \quad k = 1, 2, \ldots, m.$$

3.3 Συναρτήσεις Κόστους

Στους περισσότερους ταξινομητές , χρησιμοποιείται κάποια συνάρτηση κόστους για την μεταφορά στα επόμενα στάδια καθώς και για τον υπλογισμό της εξόδου. Μάλιστα σε κάποιους περίπλοκους ταξινομητές , όπως τα νευρωνικά δίκτυα, χρησιμοποιείται συνάρτηση κόστους σε κάθε επίπεδο και μάλιστα αυτές μπορεί να αλλάζουν κιόλας κατά το μήκος της αρχιτεκτονικής του ταξινομητή.

Γενικότερα, το σφάλμα υπολογίζεται ως η διαφορά της επιθυμητής εξόδου και της προβλεπόμενης εξόδου του συστήματος.

$$J(w) = p - p_1$$

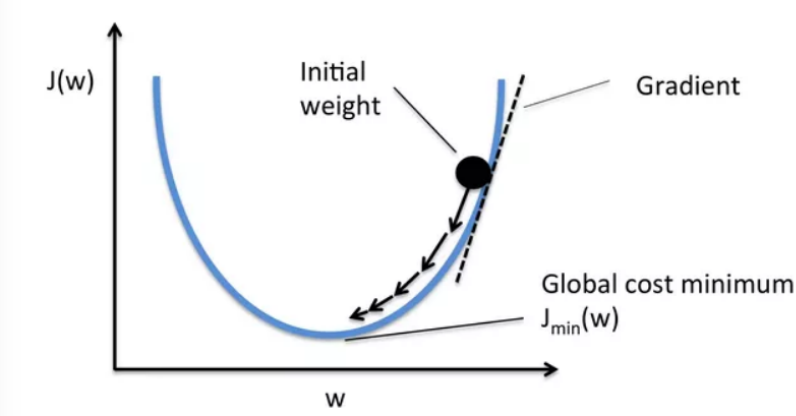
Η συνάρτηση που χρησιμοποιείται για να υπολογιστεί το σφάλμα είναι γνωστή και ως συνάρτηση κόστους J(.). Διαφορετικές συναρτήσεις κόστους θα δώσουν διαφορετικά σφάλματα για την ίδια προβλεπόμενη έξοδο και γι'αυτό έχουν σημαντική λειτουργία στην απόδοση του μοντέλου. Μία από τις πιο διαδεδομένες συναρτήσεις κόστους είναι το μέσο τετραγωνικό σφάλμα, που υπολογίζει το τετράγωνο της διαφοράς μεταξύ της πραγματικής τιμής και της προβλεπόμενης τιμής . Για διαφορετικές λειτουργίες χρησιμοποιούμε και διαφορετικές συναρτήσεις κόστους όπως για οπισθοδρόμηση, ταξινόμηση κλπ.

Το σφάλμα J(w) είναι μία συνάρτηση εσωτερικών παραμέτρων του μοντέλου όπως για παράδειγμα τα βάρη λαο τα υπό μεροληψία βάρη. Για ακριβείς προβλέψεις χρειαζόμαστε να ελαχιστοποιήσουμε το τελικό σφάλμα. Σε ένα νευρωνικό πχ, αυτό γίνεται με την πίσω διάδοση. Το παρών σφάλμα είναι διαδίδεται τυπικά προς τα πίσω στο προηγούμενο επίπεδο, όπου χρησιμοποιείται για την μετατροπή των βαρών και της μεροληψίας με τρόπο τέτοιο ώστε το σφάλμα να ελαχιστοποιείται. τα βάρη που αλλάζουν χρησιμοποιούν μια συνάρτηση που ονομάζεται συνάρτηση βελτιστοποίησης 3.4.

Οι συναρτήσεις βελτιστοποίησης χρησιμοποιούνται συνήθως για να υπολογίσουν την κλίσης, πχ τη μερική παράγωγο της συνάρτησης κόστους σε σχέση με τα βάρη και τα βάρη μετατρέπονται με αντίστροφο ρυθμό. Συνεχίζουμε την ίδια διαδικασία μέχρι το κόσμος να ελαχιστοποιηθεί σύμφωνα με τον τύπο για κάθε κύκλο:

$$W^{(k+1)} = W^{(k)} - \frac{\partial}{\partial W^{(k)}} J(W)$$

Γι'αυτό τα στοιχεία ενός νευρωνικού δικτύου , πχ η συνάρτηση ενεργοποίησης, η συνάρτηση κόστους και ο αλγόριθμος βελτιστοποίησης παίζουν πολύ σημαντικό ρόλο στη λειτουργία του μοντέλου και στην παραγωγή σωστών αποτελεσμάτων. Διαφορετικές εργασίες απαιτούν διαφορετικές συναρτήσεις κόστους.



Σχήμα 3.4: Συνάρτηση Βελτιστοποίησης

Συνεπώς, οι συναρτήσεις κόστους είναι βοηθητικές στην απόδοση των ταξινομητών. Δεδομένης μιας εισόδου και ενός στόχου, υπολογίζουν το κόστους , δηλαδή τη διαφορά ανάμεσα σε αυτά τα δύο και επίσης αυτές μπορούν να καταταχθούν σε τέσσερις κυρίαρχες κατηγορίες.

Οι συναρτήσεις κόστους για Οπισθοδρόμηση χρησιμοποιούνται όταν ο στόχος είναι συνεχής μεταβλητή. Κυρίως χρησιμοποιούμενη συνάρτηση κόστους σε αυτήν την κατηγορία είναι το μέσο τετραγωνικό σφάλμα. Άλλες συναρτήσεις κόστους εδώ είναι

1. Απόλυτου σφάλματος - υπολογίζει τη μέση απόλυτη τιμή ανά στοιχείο διαφορά μεταξύ της εισόδου.

2. ομοιόμορφου απόλυτου σφάλματος - μια παραλλαγή του κριτηρίου απόλυτης τιμής.

Οι συναρτήσεις κόστους για ταξινόμηση είναι αυτές οι συναρτήσεις κόστους όπου η μεταβλητή εξόδου σε ένα πρόβλημα ταξινόμησης είναι μια τιμή πιθανότητας f(x), που ονομάζεται σκορ για την είσοδο x. Γενικώς, το πλάτος της τιμής αυτής αντιπροσωπεύει το βαθμό εμπιστοσύνης ως προς την πρόβλεψη μας. Ο στόχος y, είναι μια δυαδική τιμή. Σε ένα παράδειγμα (x,y), το περιθώριο ορίζεται ως $y * f(x)$. Το περιθώριο είναι ένα μέτρο του πόσο σωστό είναι το αποτέλεσμα μας. Τα κόστη των περισσότερων ταξινομητών στοχεύουν στη μεγιστοποίηση αυτού του περιθωρίου. Καποιες συναρτήσεις κόστους για ταξινόμηση είναι

1. Δυαδική διασταυρούμενη εντροπία

2. Αρνητική λογαριθμική πιθανότητα

3. Ταξινομητής Περιθωρίου

4. Απαλός ταξινομητής Περιθωρίου

Οι συναρτήσεις κόστους για ενσωμάτωση χειρίζονται προβήματα που πρέπει να δούμε αν δύο είσοδοι είναι παρόμοιες ή όχι. Για παράδειγμα

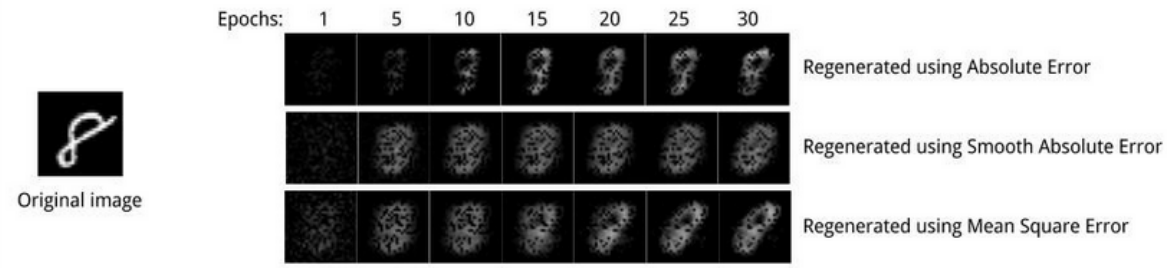
1. L1 σφάλμα Hinge - Υπολογίζει την L1 νόρμα μεταξύ των δύο εισόδων.

2. Σφάλμα συνημιτόνου - η απόσταση συνημιτόνου ανάμεσα στις δύο εισόδους.

Για την οπτικοποίηση των συναρτήσεων κόστους προσπαθούμε να ανακατασκευάσουμε την εικόνα χρησιμοποιώντας αυτόματους Κωδικοποιητές για δεδομένα της βάσης δεδομένων MNIST με τρεις διαφορετικές συναρτήσεις κόστους.

1. Συνάρτηση Κόστους Απόλυτης Τιμής

2. Συνάρτηση Κόστους μέσου τετραγώνου



Loss function for Mean Square Error $loss(x, y) = \frac{1}{n}\sum |x_i - y_i|^2$

Loss function for Absolute Error $loss(x, y) = \frac{1}{n}\sum |x_i - y_i|$

Loss function for Smooth Absolute Error $loss(x, y) = \frac{1}{n}\sum \begin{cases} 0.5 * (x_i - y_i)^2, \; if \; |x_i - y_i| < 1 \\ |x_i - y_i| - 0.5, \; otherwise \end{cases}$

Σχήμα 3.5: Σύγκριση Συναρτήσεων Κόστους

3. Συνάρτηση Κόστους ομοιόμορφης Απόλυτης Τιμής

Ενώ η συνάρτηση του απλώς υπολογίζει τη διαφορά της μέσης απόλυτης τιμής ανά εικονοστοιχείο(*pixel*), η συνάρτηση κόστους μέσου τετραγωνικού σφάλματος χρησιμοποιεί το μέσο σφαλμα στο τετράγωνο. Γι' αυτό είναι και πιο ευαίσθητη σε μακρινές τιμές και τα μετατρέπει σε 0 ή 1. Αντίθετα η ομοιόμορφη L1 νόρμα σφάλματος είναι λιγότερο ευαίσθητη σε αυτές τις μεταβολές και αποτρέπει ακραίες μεταβολές των εικονοστοιχείων.

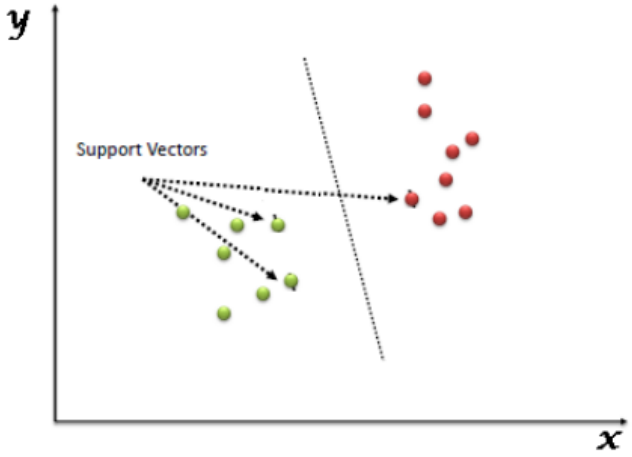
3.4 Μηχανές Υποστήριξης Διανυσμάτων

Οι μηχανές Υποστήριξης Διανυσμάτων είναι ένας αλγόριμος μηχανικής μάθησης με επίβλεψη που μπορεί να χρησιμοποιηθεί τόσο σε προβλήματα οπισθοδρόμησης όσο και σε προβλήματα ταξινόμησης. Ωστόσο, χρησιμοποιείται κυρίως σε εφαρμογές ταξινόμησης. Σε αυτόν τον αλγόριμο, σχεδιάζουμε κάθε δεδομένο σαν ένα σημείο στον N-διάστατο χώρο (όπου N είναι ο αριθμός των χαρακτηριστικών που έχουμε εξάγει και τροφοδοτήσει την είσοδο με αυτά) με την κάθε τιμή να είναι η αντίστοιχη συντεταγμένη στο χώρο. Έπειτα, εφαρμόζουμε ταξινόμηση βρίσκοντας το υπερ-επίπεδο που διαχωρίζει τις δύο κλάσεις 3.6.

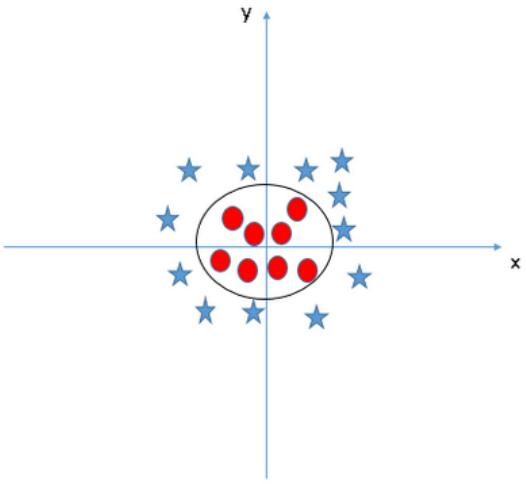
Τα διανύσματα υποστήριξης είναι απλώς συντεταγμένες των επιμέρους παρατηρήσεων. Η μηχανή των διανυσμάτων υποστήριξης είναι το υπερ-επίπεδο που διαχωρίζει καλύτερα τις δύο κλάσσεις.

Ο τρόπος που δουλεύουν είναι να μεγιστοποιήσουμε τις αποστάσεις μεταξύ των κοντινότερων (από κάθε κλάση) σημείων . Αυτή η απόσταση ονομάζεται περιθώριο (*margin*). Όσο μεγαλύτερο το περιθώριο, τόσο καλύτερος ο διαχωρισμός των κλάσεων. Οι μηχανές διανυσμάτων υποστήριξης πρεέπει να αγνοούν πολύ ακραίες τιμές ώστε να βρεί το μέγιστο περιθώριο που θα δημιουργήσει το υπερ-επίπεδο διαχωρισμού. Άρα είναι εύρωστο ως προς τις ακραίες τιμές.

Στις μηχανές υποστήριξης διανυσμάτων , είναι εύκολο να έχουμε ένα γραμμικό υπερ-επίπεδο μεταξύ των δύο κλάσεων. Αλλά, μερικές φορές χρειάζεται να χρησιμοποιήσουμε την τεχνική του πυρήνα. Αυτή είναι κάποιες συναρτήσεις που παίρνουν δεδομένα εισόδου χαμηλής διάστασης και τα μετασχηματίζουν σε υψηλότερη διάσταση, συναρτήσεις που ονομάζονται πυρήνες. Κυρίως χρησιμοποιείται σε προβληματα μη-γραμμικά διαχωρίσιμα 3.7. Χρησιμοποιεί περίπλοκους μετασχηματισμούς δεδομένων και βρίσκει τη λύση για τον διαχωρισμό των κλάσεων ανάλογα με τους στόχους και τις εξόδους που προβλέπουν οι μηχανές υποστήριξης διανυσμάτων.



Σχήμα 3.6: Δυαδική Διαχώριση διδιάστατου χώρου



Σχήμα 3.7: Ταξινόμηση μη-γραμμικών κλάσεων με τεχνική πυρήνα

3.5 Νευρωνικά Δίκτυα

3.5.1 perceptron

Ο όρος perceptron αναφέρεται σε συγκεκριμένο μοντέλο μάθησης με επίβλεψη, αναφέρθηκε το 1957 από τον Rosenblatt. Η αρχιτεκτονική και η συμπεριφορά του perceptron είναι παρόμοια με αυτή των βιολογικων νευρώνων και συχνά θεωρείται ως η πιο απλά μορφή νευρωνικού δικτύου. Συγκεκριμένα το perceptron είναι ένα νευρωνικό δίκτυο ενός επιπέδου και αντίστοιχα ένα perceptron πολλών επιπέδων ονομάζεται νευρωνικό δίκτυο.

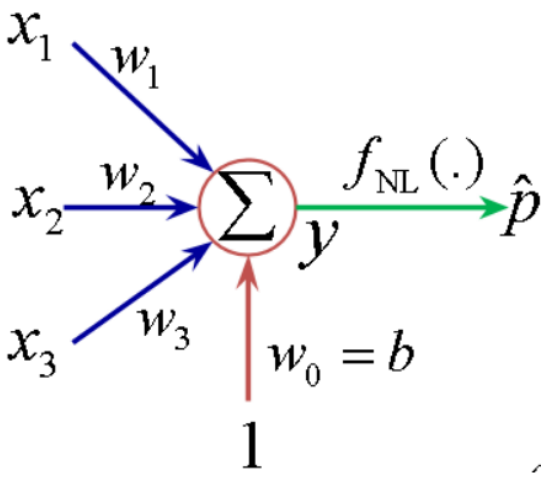
Το perceptron είναι ένας γραμμικός ταξινομητής και συγκεκριμένα δυαδικός. Επίσης, χρησιμοποιείται στην μάθηση με επίβλεψη. Βοηάει στην ταξινόμηση των δοσμένων δεδομένων εισόδου. Το perceptron αποτελείται από τέσσερα μέρη.

1. Τιμές εισόδου ή ένα επίπεδο εισόδου.
2. Τα βάρη και η μεροληψία των βαρών.
3. Το δίκτυο των αθροισμάτων
4. Μια συνάρτηση ενεργοποίησης

Για να καταλάβουμε πως δουλεύει το νευρωνικό δίκτυο, θα πρέπει πρωτίστως να κατανοήσουμε πως λειτουργεί το perceptron 3.8.

Ο τρόπος λειτουργίας του perceptron μπορεί να αναλυθεί σε μια σειρά από βήματα

α. Όλες οι είσοδοι x πολλαπλασιάζονται με τα αντίστοιχα βάρη w. Έστω k το γινόμενο τους 3.8.



Σχήμα 3.8: Το perceptron

β. Προσθέτω ταα γινόμενα και το άθροισμα αυτό το ονομάζουμε Άθροισμα με Βάρη.

$$y = w_0 + w_1 x_1 + w_2 x_2 + ..$$
$$y = [\mathbf{w} \quad \mathbf{b}][\mathbf{x} \quad \mathbf{1}]^T$$

γ. Εφαρμόζουμε το Άθροισμα με βάρη στην Σύνάρτηση Ενεργοποίησης. Το αποτέλεσμα των προηγούμενων βημάτων περνάει από τον μετασχηματίζεται ώστε να δημιουργεί η επιθυητή έξοδος μέσω της εκάστοτε συνάρτησης ενεργοποίησης , μιας μη γραμμικής συνάρτησης. Εφόσον η έξοδος είναι μια πιθανότητα στο διάστημα [0,1], μπορούμε να χρησιμοποιήσουμε τη συνάρτηση sigmoid. Άλλες συναρτήσεις που χρησιμοποιούνται είναι η ReLu , Tan, Tanh και Identity.

Να επισημάνουμε επίσης πως χρειαζόμαστε τα βάρη και την μεροληψία των βαρών ώστε να μπορούμε να μετακινήσουμε την συνάρτηση ενεργοποίησης προς τα πάνω ή προς τα κάτω. Επιπλέον, ο λόγος που χρειαζόμαστε τη συνάρτηση ενεργοποίησης για την ταξινόμηση δεδομένων που δεν είναι γραμμικώς διαχωρισμένα.

3.5.2 Πλήρως Ενωμένα Δίκτυα

Τα Πλήρως Ενωμένα Δίκτυα, παρ' ότι είναι η απλούστερη μορφή νευρωνικού δικτύου, καταλαβαινοντας τις βασικές αρχές που τα διακατέχουν καθώς και τα μαθηματικά μοντέλα πίσω από αυτά, μας βοηθάει για τη γενικότερη κατανόηση των νευρωνικών δικτύων, ακόμα και των πιο πολύπλοκων. Στα νευρωνικά δίκτυα αυτού του τύπου, όταν κάνουμε εκπαίδευση με επίβλεψη, υπάρχουν δύο βασικοί παράγοντες. Το προς τα εμπρός πέρασμα (*forward-pass*) και το προς τα πίσω πέρασμα (*backward-pass*), που αποτελούν τους βασικούς άξονες όλης της εξέλιξης της Τεχνητής Νοημοσύνης στο σήμερα.

Τα βασικά στοιχεία για την κατανόηση που θα χρειαστούμε για ένα πλήρως συνδεδεμένα δίκτυο με L επίπεδα είναι τρία :

1. Ένα επίπεδο εισόδου (με δομικές μονάδες (*units*) τα u_i^0) των οποίων οι τιμές καθορίζονται από τα δεδομένα εισόδου.

2. Τα κρυφά επίπεδα (με δομικές μονάδες τα u_i^l) των οποίων οι τιμές προκύπτουν από τα προηγούμενα επίπεδα.

3. Το επίπεδο εξόδου (με δομικές μονάδες τα u_i^L) του οποίου οι τιμές προκύπτουν από το τελευταίο κρυφό επίπεδο.

Το νευρωνικό δίκτυο μαθαίνει προσαρμόζοντας τα βάρη w_{ij}^l όπου w_{ij}^l είναι τα βάρη από τις εξόδους κάποιων δομικών μονάδων u_i^l προς κάποιες δομικές μονάδες u_i^{l+1}

Προχωράμε τώρα στην ανάλυση των δύο σταδίων. Το προς τα εμπρος πέρασμα στάδιο είναι βασικά ένα σύνολο από λειτουργίες που μετασχηματίζει την είσοδο του δικτύου στο χώρου εξόδου. Κατά το στάδιο αυτό η έξοδος του νευρωνικού είναι η έξοδος του τελευταίο επιπέδο u^L. Ορίζουμε ως :

x_i^l την συνολική είσοδο του u_i^l και
y_i^l την συνολική έξοδο του u_i^l

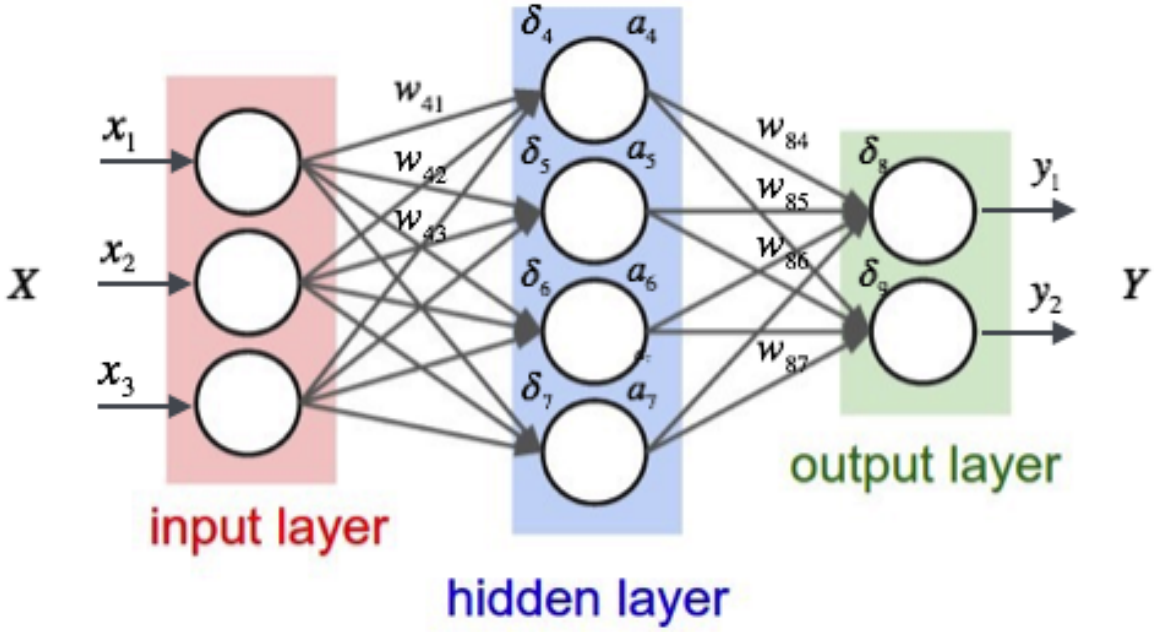
Για να υπολογίσουμε την έξοδο από την αντίστοιχη είσοδο , εφαρμόζουμε κάποια μη-γραμμικότητα σ(x) στην είσοδο. Η είσοδος στο πρώτο επίπεδο είναι εξαρχής ορισμένη ενώ οι είσοδοι των επόμενων επιπέδων υπολογίζονται σαν το άθροισμα των βαρών του εκάστοτε προηγούμενου επιπέδου.

Προς τα εμπρός πέρασμα :

1. Υπολογίζω τις ενεργοποιήσεις για τα επίπεδα με γνωστές εισόδους:
   $y_i^l = (x_i^l) + I_i^l$

2. Υπολογίζω τις εισόδους για το επόμενο επίπεδο με αυτές τις ενεργοποιήσεις :
   $x_i^l = \sum_j w_{ij}^{l-1} y_j l - 1$

3. Επαναλαμβάνουμε τα βήματα 1 και 2 μέχρι να φτάσουμε στο επίπεδο εξόδου με γνωστές τις τιμές του y^L



Σχήμα 3.9: Πλήρως Συνδεδεμένο Νευρωνικό Δίκτυο

Όπως ορίσαμε προηγουμένως, το x_i^l είναι η είσοδος στη δεδομένη δομική μονάδα , ενώ το y_i^l είναι η ενεργοποίηση ή η έξοδος αυτής της δομικής μονάδας. Σε αυτές τις εξισώσεις, όταν μια δομική μονάδα δεν έχει καθόλου εισόδους, ο όρος σ(x) στο y_i^l γίνεται σταθερά, επιτρέποντας στη δομική μονάδα

να ορίζεται εξωτερικά από την τιμή I_i^l. Αυτό αντιστοιχεί στο επίπεδο εισόδου του νευρωνικού. Γι' αυτό ξεκινάμε θέτωντας κάποιες ενεργοποισεις στο επίπεδο εισόδου , υπολογίζουμε τις εισόδους των νευρώνων για το επόμενο επίπεδο, χρησιμοποιούμε τη μη-γραμμικότητα για να πάρουμε τις ενεργοποιήσεις τους και συνεχίζουμε να προωθούμε τιμές μέχρι να φτάσουμε στο επίπεδο εξόδου. Σε αυτό το σημείο υπολογίζουμε τις τελικές ενεργοποιήσεις y^L.

Στο προς τα πίσω πέρασμα ο σκοπός είναι ο υπολογισμός του σφάλματος , δηλαδή να μπορούμε να βελτιστοποιήσουμε τα βάρη προς την ελαχιστοποίηση του, διαδικασία που ονομάζεται μάθηση. Ο αλγόριθμος που χρησιμοποιούμε ονομάζεται προώθηση προς τα πίσω (*backpropagation*), όπου προκύπτει με τρόπο ανάλογο με τον αλγόριθμο της προώθησης προς τα εμπρός. Προκειμένου να χρησιμοποιήσουμε τον συντελεστή κλίσης (*gradient descent*) ή καποιον άλλον αντίστοιχο αλγόριθμο για την εκπαίδευση του δικτύου μας, χρειάζεται να υπολογίζουμε την παράγωγο του σφάλματος σε σχέση με το κάθε βάρος. Χρησιμοποιώντας τον κανόνα την αλυσίδας παίρνουμε:

$\frac{\partial E}{\partial w_{ij}^l} = \frac{\partial E}{\partial x_j^{l+1}} \frac{\partial x_j^{l+1}}{\partial w_{ij}^l}$

Συνεπώς για τη συνολική ανάλυση του αλγορίθμου προώθηση προς τα πίσω έχουμε τα εξής βήματα :

1. Υπολογίζουμε τα σφάλματα του επιπέδου εξόδου L:

   $\frac{\partial E}{\partial y_i^l} = \frac{\partial E(y^L)}{\partial y_i^L}$

2. Υπολογίζουμε το σφάλμα της μερικής παραγώγου σε σχέση με την είσοδο του νευρωνικού στο πρώτο επίπεδο l που έχει γνωστά σφάλματα:

   $\frac{\partial E}{\partial x_{jj}^l} = (x_j^l) \frac{\partial E}{\partial y_j^l}$

3. Υπολογίζουμε τα σφάλματα στο προηγούμενο επίπεδο:

   $\frac{\partial E}{\partial y_{jj}^l} = \sum w_{ij}^l \frac{\partial E}{\partial x_j^l}$

4. Επαναλαμβάνουμε τα βήματα 2 και 3 μέχρι να φτάσουμε στο στάδιο εισόδου.

5. Υπολογίζουμε το τελικό σφάλμα

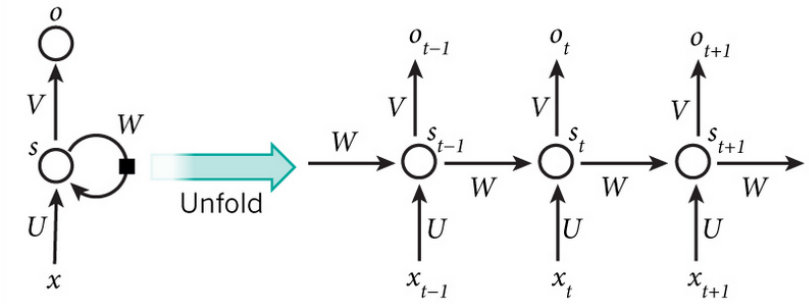
3.5.3 Αναδρομικά Νευρωνικά Δίκτυα

Η ιδέα πίσω από τα Αναδρομικά νευρωνικά δίκτυα (*Recurrent Neural Networks - RNN*) είναι η χρήση της πληροφορίας σαν αλληλουχία. Σε ένα παραδοσιακό νευρωνικό δκτυο υποθέτουμε ότι όλες οι είσοδοι(και έξοδοι) είναι ανεξάρτητες μεταξύ του. Σε πολλές όμως εφαρμογές αυτό δεν αποτελεί τη βέλτιστη λύση. Αν θέλουμε να προβλέψουμε την επόμενη λέξη σε μια πρόταση, είναι πολύ χρήσιμο να ξέρουμε ποιες λέξεις έχουν προηγηθεί και αντίστοιχα ισχύει το ίδιο και στην ομιλία. Τα νευρωνικά αυτά δίκτυα ονομάζονται αναδρομικά γιατί εκτελούν την ίδια λειτουργία για κάθε στοιχείο μιας ακολουθίας , με την έξοδο να είναι εξαρτώμενη από τους προηγούμενους υπολογισμούς. Ένας διαφορετικός τρόπος να κατανοήσουμε τα Αναδρομικά Νευρωνικά Δίκτυα είναι σαν να έχουν μνήμη που εμπεριέχει αποθηκευμένη την πληροφορία που έχει υπολογιστεί έως τώρα. Στην θεωρία τα RNNs χρησιμοποιούν μνήμη για αυθαίρετα μεγάλες αλληλουχίες, αλλά στην πράξη μπορούν να κοιτάξουν πίσω πεπερασμένο αριθμό βημάτων. Ένα τυπικό RNN μοιάζει όπως στην εικόνα 3.10.

Το διάγραμμα 3.10 δείχνει ένα RNN να ξετυλίγεται σε ένα πλήρες δίκτυο. Δηλαδή το αναλύουμε για το σύνολο της ακολουθίας. Οι μαθηματικοί τύποι που διέπουν το Αναδρομικό Νευρωνικό Δίκτυο είναι:

x_t είναι η είσοδος για ένα βήμα χρόνου t. Για παράδειγμα το x_1 θα μπορούσε να είναι ένα δυαδικό διάνυσμα που αντιστοιχεί στη δεύτερη λέξη της πρότασης.

s_t είναι η κρυφή κατάσταση για ένα βήμα χρόνου t. Είναι σαν μνήμη του δικτύου. Το s_t υπολογίζεται βασιζόμενο στην προηγούμενη κατάσταση και την είσοδο του παρόντος βήματος χρόνο :



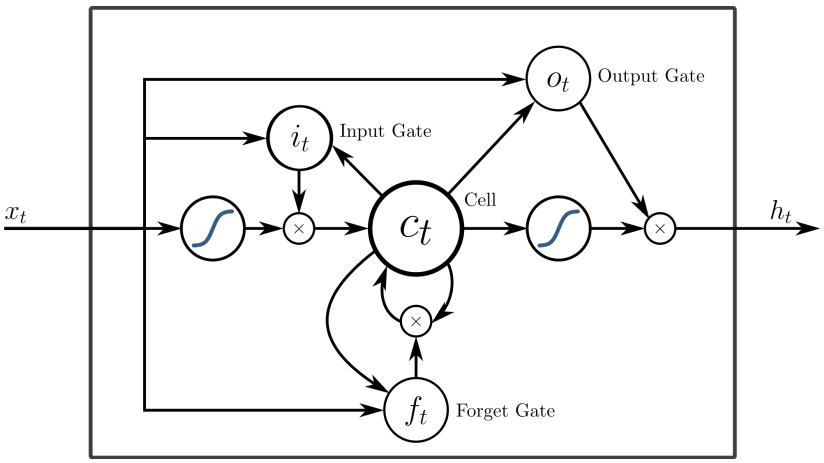
Σχήμα 3.10: Αναδρομικό Νευρωνικό Δίκτυο - RNN

$s_t = f(Ux_t + Ws_t - 1)$. Η συνάρτηση f συνήθως είναι μη γραμμικότητα όπως η tanh και η ReLu . Το s_{-1} που χρειάζεται για την αρχική κρυφή κατάσταση, το αρχικοποιούμε στο μηδέν.

o_t είναι η έξοδος του βήματος t. Για παράδειγμα αν θέλαμε να προβλέψουμε την επόμενη λεη σε μια πρόταση θα ήταν το διάνυσμα με τις πιθανότητες κατά μήκος του λεξικού μας. $o_t = softmax(Vs_t)$.

3.5.4 Νευρωνικά Δίκτυα Μακράς-Βραχέας Μνήμης

Τα δίκτυα Μακράς-Βραχέας Μνήμης (*LSTM*) χρησιμοποιούνται σε όλο το φάσμα των εφαρμογών για την Αναγνώριση Προσωπικότητας, τόσο από είσοδο Φωνής, όσο και από τα υπόλοιπα μέσα. Κάποια από τα δίκτυα της προηγούμενης κατηγορίας (Αναδρομικά Νευρωνικά Δίκτυα) είναι και Δίκτυα μακράς Βραχέας Μνήμης. Τα LSTMs είναι ικανά να πιάνουν εξαρτήσεις χρονικές από την αλληλουχία εισόδου σε μεγαλύτερο βαθμό από τα προηγούμενα. Η κύρια διαφορά τους είναι στο ότι έχουν μια επιπλέον κατάσταση Μνήμης και αλλάζει ο τρόπος που υπολογίζεται η κρυφή κατάσταση. Μία ενδεικτική αρχιτεκτονική τους είναι 3.11. Για να αναλύσουμε ουσιαστικά τη λειτουργία τους, χωρίζεται η είσοδος σε επιμέρους σήματα ώστε κάθε σήμα να έχει σταθερό μήκος και για κάθε ένα από αυτά υπολογίζεται η έξοδος. Για το επόμενο λαμβάνεται υπόψη τόσο τα βάρη όσο και η κατάσταση , η οποία έχει αποθηκευτεί στη μνήμη από το προηγούμενο στάδιο. Τελικώς δηλαδή φτάνουμε στον υπολογισμό της εξόδου , λαμβάνοντας υπόψη μας την αλληλουχία.



Σχήμα 3.11: Δίκτυο Μακράς - Βραχέας Μνήμης

Κεφάλαιο 4

Μελέτη Εργασίας

4.1 Εισαγωγή

Σε αυτή την ενότητα θα παρουσιαστούν διάφορες τεχνικές που χρησιμοποιήθηκαν, όπως η Ανάλυση Κύριων Συνιστωσών, οι Αυτόματοι Κωδικοποιητές καθώς και θα γίνει ανάλυση των Προεκπαιδευμένων Νευρωνικών Δικτύων και της Μεταφοράς Μάθησης μέσω Νευρωνικών Δικτύων που εκπαιδεύονται σε κάποια δεδομένα και χρησιμοποιούνται τα εκπαιδευμένα αυτά δίκτυα για την κατηγοριοποίηση άλλων παρόμοιων δεδομένων.

4.2 Μέθοδος Ανάλυσης Κύριων Συνιστωσών

Η μέθοδος Ανάλυσης Κύριως Συνιστωσών(*PCA*) είναι ένα εργαλείο για τη μείωση των διαστάσεων που μπορεί να χρησιμοποιηθεί για τη μείωση ενός μεγάλου συνόλου μεταβλητών σε ένα μικρότερο σύνολο που παρ' όλα αυτά περιέχει την ίδια ποσότητα πληροφορίας με το αρχικό μεγάλο σύνολο.

Η μέθοδος Ανάλυσης Κύριων Συνιστωσών είναι μια μαθηματική διεργασία που μετασχηματίζει έναν αριθμό πιθανώς συσχετιζόμενων μεταβλητών σε ένα μικρότερο αριθμό ασυσχέτιστων μεταβλητών που ονομάζονται Κύριες Συνιστώσες (*principal components*). Η πρώτη κύρια συνιστώσα αντιπροσωπεύει όσο μεγαλύτερη ποικιλία στα δεδομένα όσο είναι δυνατόν και κάθε διαδοχική συνιστώσα που ακολουθεί περιέχει πάλι τη μέγιστη ποικιλομορφία που είναι εφικτό. Η Μέθοδος Ανάλυσης Κύριως Συνιστωσών είναι παρόμοια με μια άλλη διαδικασία πολυμεταβλητών που ονομάζεται ανάλυση παραγόντων.

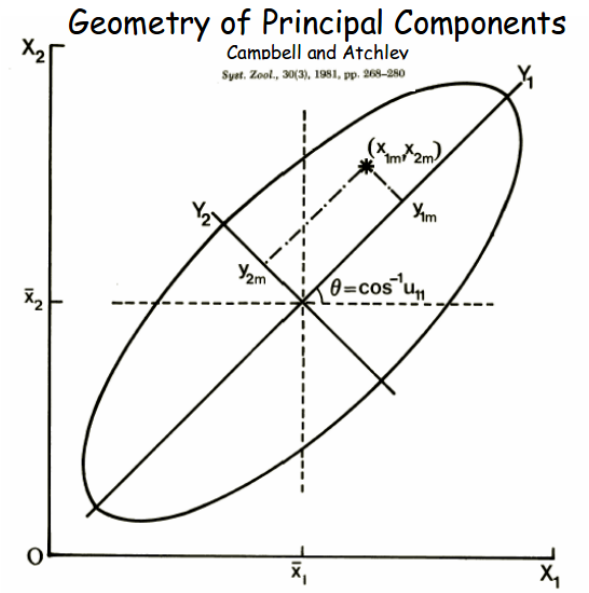
Συνηθέστερα, η μέθοδος Κύριων συνιστωσών πραγματοποιείται σε τετραγωνικό συμμετρικό πίνακα. Μπορεί να είναι ένας πίνακας αθροισμάτων και γινομένων, ένας πίνακας συσχέτισης ή ένας πίνακας συνδιακύμανσης. Η μέθοδος δεν αλλάζει σημαντικά , αλλάζει μόνο ο συνολικός παράγοντας κλιμάκωσης. Ένας πίνακας συσχέτισης χρησιμοποιείται αν οι μεταβλητές διαφέρουν πολύ ή αν οι επιμέρους δομικές μονάδες διαφέρουν σημαντικά.

Όπως βλέπουμε στο διάγραμμα 4.1 για δύο μεταβλητές κάθε μία περιέχει μια έλλειψη 95 συγκεντρωτικότητας και κύριους άξονες $_1$ και $_2$. Τα σημεία y_{1m} και y_{2m} δίνουν τα σκορ των κύριων συνιστωσών για τις παρατηρήσεις $x_1 = (x_{1m}, x_{2m})$. Το συνημίτονο της γωνίας θ μεταξύ του Y_1 και του X_1 δίνει τον πρώτο παράγοντα u_{11} του αντίστοιχου στο Y_1 ιδιοδιάνυσματος.

Ο συνολικός σκοπός είναι η δημιουργία μοντέλου του οποίου οι παράγοντες βασίζονται στη σύνοψη της συνολικής διακύμανσης. Με την Μέθοδο Ανάλυσης Κύριων Συνιστωσών, χρησιμοποιούνται πίνακας όπου η διαγώνιος του πίνακας συσχέτισης υποδηλώνει ότι η διακύμανση είτε είναι κοινή είτε μοιράζεται ανάμεσα στα στοιχεία.

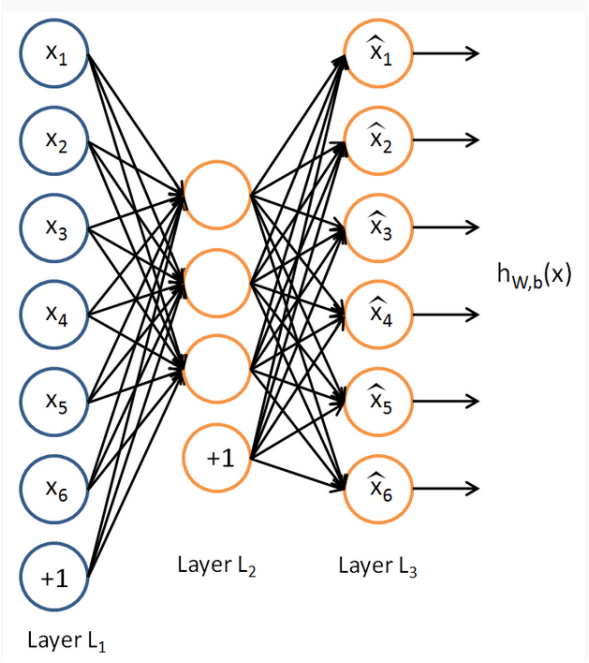
4.3 Αυτόματοι Κωδικοποιητές - Autoencoders

Ας υποθέσουμε ότι έχουμε δεδομένα για εκπαίδευση μαζί με τις ετικέτες τους. Τότε εφαρμόζουμε κανονικά μηχανική μάθηση με επίβλεψη. Αν όμως δεν είχαμε τις ετικέτες των δειγμάτων, έστω



Σχήμα 4.1: Μέθοδος Ανάλυσης Κύριως Συνιστωσών

$\{x^{(1)}, x^{(2)}, x^{(3)}, \dots\}$ όπου $x^(i)\epsilon R^n$ [12]. Ένας αυτόματος κωδικοποιητής (*Autoencoder*) ως νευρωνικό δίκτυο, είναι ένας αλγόριθμος μάθησης χωρίς επίβλεψη που εφαρμόζει οπισθοδρόμηση προς τα πίσω, καθορίζοντας τις τιμές του στόχου να είναι ίσες με την είσοδο, δηλαδή προσπαθεί $y^{(i)} = x^{(i)}$ 4.2.



Σχήμα 4.2: Αυτόματος Κωδικοποιητής

Ο αυτόματος κωδικοποιητής προσπαθέι να μάθει μια συνάρτηση $H_{W,b}(x) = x$. Δηλαδή, προσπαθέι να μάθει μια προσέγγιση της ταυτο-συνάρτησης. Προσπαθεί δηλαδή να μάθει την προσέγγιση ^{x} ώστε να είναι παρόμοια με το x. Αν και η ταυτο-συνάρτηση είναι κάτι σχετικά απλοϊκό , είναι ένα βασικό παράδειγμα για να καταλαβουμε την βασική λειτουργία του Αυτόματου Κωδικοποιητή [13],[12].

4.4 Προεκπαιδευμένα Νευρωνικά Δίκτυα

Μία από τις μεθόδους που χρησιμοποιούμε για την εκπαίδευση των μοντέλων είναι τα Προεκπαιδευμένα Νευρωνικά Δίκτυα. Αυτά τα χρησιμοποιούμε κυρίως όταν θέλουμε να γλυτώσουμε χρόνο στην εκπαίδευση του δικτύου για τα δεδομένα μας, ειδικά όταν η δουλειά έχει ξαναγίνει σε πιο εξειδικευμένα δεδομένα [14],[15]. Ας υποθέσουμε ότι έχουμε να εκπαιδεύσουμε ένα νευρωνικό δίκτυο για να πρόβλημα αναγνώρισης Φωνής ή Εικόνας και τα αντίστοιχα δεδομένα. Για να εφαρμόσουμε τα προεκπαιδευμένα δίκτυα αρχικοποιούμε τα βάρη τυχαία. Όταν αρχίσει η εκπαίδευση, τα βάρη αρχίζουν να αλλάζουν έτσι ώστε να γίνει η βελτιστοποίηση για το συγκεκριμένο σετ δεδομένων. Μετά από κάποιο διάστημα εποχών, όταν οι συναρτήσεις κόστους έχουν αρχίσει να συγκλίνουν, αποθηκεύουμε τα αποτελέσματα καθώς και τα βάρη του νευρωνικού δικτύου.

Τώρα θα πρέπει να εφαρμόσουμε τα προηγούμενα αποθηκευμένα δεδομένα για ένα διαφορετικό πρόβλημα, ενδεχομένως σε ένα διαφορετικό σύνολο δεδομένων από το προηγούμενο. Αντί λοιπόν να επαναλάβουμε τη διαδικασία ξανά, κάτι που θα περιλάμβανε την εκ νέου αρχικοποίηση των βαρών και την ίδια συνέχεια, μπορουμε να χρησιμοποιήσουμε τα αποθηκευμένα βάρη για να αρχικοποιήσουμε τα βάρη στο νέο μας δίκτυο.

Η ιδέα πίσω από τα προεκπαιδευμένα δίκτυα είναι ότι η τυχαία αρχικοποίηση, δυσκολεύει τη σύγκλιση των συναρτήσεων κόστους, ενώ αν πάρουμε τα ήδη εκπαιδευμένα βάρη , γλυτώνουμε και πόρους και χρόνο και έχουμε ήδη μια καλή προσέγγιση για το καινούριο σύνολο δεδομένων για την Αναγνώριση Φωνής [16],[17],[18].

4.5 Μεταφορά Μάθησης Προεκπαιδευμένων Δικτύων

Η μεταφορά μάθησης Νευρωνικών Δικτύων (*Transfer Learning*) είναι μια μέθοδος που χρησιμοποιείται στη μηχανική μάθηση έτσι ώστε ένα μοντέλο που αναπτύχθηκε για μια εφαρμογή να ξαναχρησιμοποιηθεί ως βάση για μια άλλη διαφορετική εφαρμογή. Είναι μια δημοφιλής μέθοδος στη μηχανική μάθηση όπου προεκπαιδευμένα μοντέλα χρησιμοποιούνται για αναγνώριση φωνης, κειμένου, εικόνας, δεδομένης τις τεράστιας υπολογιστικής ισχύς που χρειάζεται για να αναπτύξουμε τα νευρωνικά μας δίκτυα και επίσης ότι ένα εύρος εφαρμογών είναι παρόμοιες [19],[20] και αναγνώριση Ειδών Μουσικής από Δίκτυα Μεταφοράς Μάθησης [21],[22],[23].

Η μεταφορά μάθησης είναι μια βελτιστοποίηση που επιτρέπει τη ραγδαία πρόοδο ή την βελτιωμένη απόδοση όταν μοντελοποιούμε τη δεύτερη από τις εφαρμογές μας [19],[20].

Για την εφαρμογή της μεταφοράς μάθησης στα μοντέλα μας υπάρχουν δύο προσεγγίσεις

1. Η Προσέγγιση της ανάπτυξης μοντέλου
2. Η Προσέγγιση του προεκπαιδευμένου μοντέλου.

Στην προσέγγιση προεκπαιδευμένου μονλεου επιλέηουμε ένα από τα διαθέσιμα μοντέλα. Μπορούμε να βρούμε από μεγάλες εταιρείες που τα δημοσιεύουν και υπάρχει αρκετά μεγάλη ποικιλία. Αρχικοποιούμε το νέο μας μοντέλο με βάση το προεκπαιδευμένο και εξετάζουμε τα μέρη του μοντέλου που μας ενδιαφέρουν ανάλογα με την εφαρμογή μας. Μπορούμε στο τέλος, προαιρετικά, να προσαρμόσουμε ή να αλλάξουμε το μοντέλο για την είσοδο-έξοδο των δεδομένων μας.

Κεφάλαιο 5

Αναγνώριση Προσωπικότητας

5.1 Ερευνητικό Υπόβαθρο και Σχετική Έρευνα

Στο συγκεκριμένο κεφάλαιο θα ασχοληθούμε αναλυτικά με τα πειράματα που γίναν πάνω στον τομέα της Αναγνώρισης Προσωπικότητας από Φωνή.

Ο τομέας της Αναγνώρισης Προσωπικότητας έχει γνωρίσει ανάπτυξη κυρίως τον τελευταίο καιρό και αυτή τη στιγμή βρίσκεται σε άνοδο , καθώς είναι ένα αρκετά ενδιαφέρον ερευνητικό πεδίο. Προηγούμενες δουλειές αφορούν κυρίως τομείς στην Αναγνώριση Προσωπικότητας , αλλά από διαφορετικό μέσο, όπως από κείμενο [24] ή από βίντεο και εικόνα [25], παρόμοιες μέθοδοι χρησιμοποιούνται για αυτά τα δύο πεδία και εφαρμόζοντας αλγορίθμους που έχουν ήδη επιτυχία στην Αναγνώριση Συναισθήματος στην Αναγνώριση Προσωπικότητας , αλλά και αντίστροφα, οδηγεί συχνά σε καλά αποτελέσματα.

Σε αυτή την ενότητα αρχικά θα παρουσιάσουμε τα δεδομένα με τα οποία χρησιμοποιήσαμε για τις πειραματικές διατάξεις της παρούσας εργασίας. Έπειτα θα παρουσιάσουμε τις αρχιτεκτονικές των μοντέλων που χρησιμοποιήσαμε, από ποια μοντέλα εμπνευστήκαμε και τις αντίστοιχες εργασίες που χρησιμοποιήθηκαν για μελέτη. Συνεχίζοντας αυτό το κεφάλαιο θα παρουσιαστούν τα πειράματα κατά σειρά που εκτελέστηκαν, επισημαίνοντας κάθε φορά τις διαφορές στην αρχική και τελική διάταξη καθώς και τις διαφορές στα αποτελέσματα προσπαθώντας ταυτόχρονα να τις ερμηνεύσουμε. Στο τελευταίο κομμάτι θα παρουσιαστούν οι καλύτερες από αυτές τις διατάξεις , όταν εφαρμόζονται σε ένα σύνολο δεδομένων Αναγνώρισης Συναισθήματος , τόσο για επιβεβαίωση των αρχιτεκτονικών που χρησιμοποιήσαμε, όσο και για την μετεξέλιξη των πειραμάτων σε δεδομένα μεγαλύτερου όγκου καθώς και περισσότερων κλάσσεων για ταξινόμηση, κάτι που όπως θα δούμε δουλεύει εξίσου καλά. Να τονίσουμε επίσης ότι ο κύριος στόχος είναι και παραμένει η Αναγνώριση Συναισθήματος, ταυτόχρονα όμως η έλλειψη δεδομένων μας οδηγεί στη χρησιμοποίηση και άλλων παραπλήσιων τομέων για την εξαγωγή ασφαλών συμπερασμάτων.

Μεγάλο κομμάτι της έρευνας έχει γίνει σε έναν τομέα αρκετά παρόμοιο με την Αναγνώριση Προσωπικότητας, που είναι ο τομέας της Αναγνώρισης Συναισθήματος. Συγκεκριμένα στον τομέα της Αναγνώρισης Προσωπικότητας αλλά και γενικότερα στον τομέα της Αναγνώρισης Φωνής υπάρχει μεγάλο υπόβαθμο έρευνας όπως στο [26] όπου περιγράφει την ποσότητας πληροφορίας συναισθήματος στο χρόνο από κομμάτια ομιλίας, στο [27] που ερευνά ένα σύνολο χαρακτηριστικών για όλα τα πειράματα ώστε να υπάρχει αντικειμενικό κριτήριο σύγκρισης. Επίσης στο [28] αναφέρεται στα χαρακτηριστικά που εξάγονται για την Αναγνώριση Συναισθήματος. Στο [25] που κάνει αναγνώριση Κατάστασης Ομιλίας του χρήστη από Φωνή χρησιμοποιώντας και γενετικούς αλγορίθμους για την εξαγωγή χαρακτηριστικών. Στο [29] έχουμε την αναγνώριση ρεαλιστικού συναισθήματος αό φωνή και γενικότερη ανάλυση των έως τώρα καλύτερων αποτελεσμάτων της βιβλιογραφίας σε αυτόν τον τομέα. Επίσης έχει ασχοληθεί με την αλληλεπίδραση ανθωπου - μηχανής και πώς μπορεί ένα ρομπότ να αναπτύξει προσωπικότητα και να αποκτήσει χαρακτηριστικά όπως εξωστρεφής ή εσωστρεφής [30]. Υπάρχει ακόμα σχετική έρευνα για αναγνώριση Συναισθήματος σε πραγματικό χρόνο σε διαφορετικές χρονικές κλίμακες [31]. Εξαγωγή Υβριδικών Χαρκατηριστικών, τόσο ακουστικών όσο και γλωσσικών [32],[33],[34] καθώς και για άλλους τομείς όπως Γένος και Ηλικία [35].

Όσον αφορά την αναγνώριση Προσωπικότητας υπάρχει και μία επιπλέον βάση δεδομένων που ασχολείται με την προσωπικότητα και ονομάζεται youtube βάση δεδομένων, τα δεδομένα δίνονται

σαν κείμενο (*transcripts*) και έχουν εφαρμογή σε διάφορα κομμάτια της βιβλιογραφίας και γενικότερα στον τομέα της Αναγνώρισης Προσωπικότητας [36],[37].

Ακόμα υπάρχει το πρόβλημα των πολλαπλών διαστάσεων (*curse of dimensionality* στο οποίο πρέπει πρώτα να μειώσουμε τις διαστάσεις και έπειτα να συνεχίσουμε την ανάλυση και ταξινόμηση των δεδομένων [38]. Μπορούμε να κάνουμε προεκπαίδευση δικτύων χωρίς επίβλεψη, όπως στους Αυτόματους Κωδικοποιητές [39]. Έχουμε τα αμφίδρομα Νευρωνικά Δίκτυα με Ανάδραση [40] ,[41]. Νευρωνικά Μακράς - Βραχέας Μνήμης για την κατηγοριοποίηση ειδών μουσικής [42]. Επίσης εφαρμόζεται σε διαγωνισμό αναγνώρισης Προσωπικότητας, στο InterSpeech2013 [43]. Παρόμοιος διαγωνισμός Αναγνώρισης Προσωπικότητας το 2012 αναφέρει τα αποτελέσματα [43]. Η προσαρμογή Συνελικτικών δικτύων με Ταξινομητή Χρονικής Σύνδεση (*Connectionist Temporal Classifier - CTC*) στο στάδιο εξόδου [44]. Γενετικά μοντέλα παραγωγής νέων δεδομένων [45]. Περιορισμένες Μηχανές Boltzmann (*Restricted Boltzmann Machines*) για την παραγωγή των κυματομορφών φωνής [46]. Εξαγωγή Χαρακτηριστικών από Αυτόματους Κωδικοποιητές [47],[48] λαμβάνοντας ως είσοδο Σπεκτογράμματα Φωνής για την εξαγωγή Χαρακτηριστικών. Σύγκριση Αναπαράστασης δεδομένων για την Αναγνώριση προσωπικότητας όπου στόχος είναι η διάκριση των διαταραχών σε σχέση με τα υπόλοιπα δεδομένα [49]. Αναγνώριση Προσωπικότητας από Συνελικτικά Δίκτυα [50],[51]. Εξαγωγή Χαρακτηριστικών από κοινώς συμφωνημένα πρότυπα [52].

5.2 Ανάλυση Δεδομένων

Συνολικά θα ασχοληθούμε με 2 διαφορετικά σύνολα δεδομένων. Το πρώτο αφορά την Αναγνώριση Προσωπικότητας, είναι αυτό στο οποίο θα ρίξουμε το μεγαλύτερο βάρος και είναι το σύνολο Δεδομένων Προσωπικότητας του Mohammadi [1]. Το δεύτερο αφορά την Αναγνώριση συναισθήματος και είναι η βάση δεδομένων IEMOCAP [53].

5.2.1 Δεδομένα προσωπικότητας (*Personality Corpus*)

Η βάση δεδομένων αναγνώρισης προσωπικότητας [1] χρησιμοποιεί το σύστημα ταξινόμησης Προσωπικότητας με το μοντέλο των πέντε παραγόντων (*O.C.E.A.N. - Big-5*).

Το μοντέλο αξιολόγησης Προσωπικότητας χρησιμοποιώντας την μεγάλη Πεντάδα είναι μια κατασκευή για την κατασκευή μιας κλίμακας αξιολόγησης της προσωπικότητας όπως αναλύθηκε και στο κεφάλαιο 1 (Παράγραφος 1.3.1) για την εξαγωγή προτύπων σκέψης, συναισθήματος και συμπεριφοράς σε συνδυασμό με ψυχολογικούς μηχανισμούς - κρυφούς ή όχι - πίσω από αυτά τα πρότυπα [54]. Το μοντέλο της Μεγάλης Πεντάδας εκτιμάει την Προσωπικότητα μέσω μιας σειράς ερωτήσεων στη συγκεκριμένη βάση δεδομένων. Οι ερωτήσεις του ερωτηματολογίου είναι 10 και είναι οι εξής :

1	This person is reserved
2	This person is generally trusting
3	This person tends to be lazy
4	This person is relaxed, handles stress well
5	This person has few artistic interests
6	This person is outgoing, sociable
7	This person tends to find fault with others
8	This person does a thorough job
9	This person gets nervous easily
10	This person has an active imagination

Πίνακας 5.1: Personality Corpus [1]

Από αυτές τις ερωτήσεις μετράει το σκορ και υπολογίζει για κάθε μία από τις 5 κατηγορίες του μοντέλου ένα ακέραιο σκορ στο διάστημα [-4,4]. Το βασικό χαρακτηριστικό αυτής της αξιολόγησης είναι ότι ένα τέτοιο ερωτηματολόγιο 10 ερωτήσεων μπορεί να συμπληρωθεί από το συμμετέχοντα.

Όσον αφορά το σύνολο των δεδομένων, η αξιολόγηση έγινε σε 640 δείγματα των δέκα δευτερολέπτων το καθένα που είναι αποκόμματα από το Ελβετικό Ραδιόφωνο (*Radio Suisse Romande*) και πάρθηκαν τυχαία από 96 αποσπάσματα που εξέπεμψε ο σταθμός κατά το χρονικό διάστημα Φλεβάρης του 2005. Υπάρχει μόνο ένας ομιλητής ανά δείγμα και ο συνολικός αριθμός των ξεχωριστών υποψηφίων είναι 322. Το σύνολο των αξιολογητών είναι έντεκα συμμετέχοντες που ακούσαν το κάθε ένα από τα 640 δείγματα και αμέσως μετά το καθένα συμπλήρωναν το ερωτηματολόγιο των δέκα ερωτήσεων. Οι έντεκα δεν είχαν συναντηθεί ποτέ μεταξύ τους και δούλευαν ξεχωριστά χωρίς να βρίσκονται στον ίδιο χώρο. Οι αξιολογητές δεν δούλευαν πάνω από 60 λεπτά συνεχόμενα την ημέρα (χωρισμένα σε δύο συνεδρίες των 30 λεπτών η καθεμία) για την αποφυγή κούρασης. Τα δείγματα παρουσιάζονταν στους αξιολογητές σε τυχαία σειρά . Η γλώσσα είναι τα Γαλλικά και οι έντεκα αξιολογητές έχουν υπογράψει ένα έγγραφο που πιστοποιεί ότι δεν γνωρίζουν Γαλλικά, οπότε το περιεχόμενο των δειγμάτων επηρεάζει σε ελάχιστο βαθμό τις απαντήσεις των αξιολογητών.

Τέλος οι δέκα απαντήσεις των αξιολογητών για κάθε δείγμα αντιστοιχούν στους πέντε άξονες της Μεγάλης Πεντάδας ως εξής: Ερωτήσεις 1-6 στην Εξωστρέφεια, 2-7 στην Προθυμία , 3-8 στην Ευσυνειδησία , 4-9 στην Νευρικότητα και 5-10 στην Ανοικτότητα.

Modalities	Audio
Utterances	640
Subjects	322
Language	French
Annotation	O.C.E.A.N.
Type of Speech	Natural

Πίνακας 5.2: Personality Corpus [1]

5.2.2 Δεδομένα Συναισθήματος (*IEMOCAP*)

Η βάση δεδομένων IEMOCAP θα αναλυθεί σύμφωνα με τα και [55] και έχει χρησιμοποιηθεί σε αρκετές δημοσιεύσεις όπως στο [56] και στο [53].

Η συγκεκριμένη βάση δεδομένων ονομάζεται IEMOCAP (*Interactive Emotional Dyadic Motion Capture*) και περιέχει δεδομένα που έχουν παραχθεί από ελεγχόμενα σενάρια , είναι μια βάση που περιέχει πολλά είδη δεδομένων , όπως φωνή , εικόνα , βίντεο και κείμενο και επιπλέον περιέχει διαφορετικούς ομιλητές. Η συλλογή όλων αυτών των δεδομένων διοργανώθηκε και έγινε από το εργαστήριο SAIL του πανεπιστημίου USC(*University of Southern California*). Συνολικά περιέχει περίπου 12 ώρες οπτικοακουστικού υλικού , συμπεριλαμβανομένων βίντεο, ομιλίας, δεδομένα από τους αισθητήρες κίνησης που έχουν τοποθετηθεί στο πρόσωπο των υποψηφίων και επίσης το αντίστοιχο κείμενο. Η βάση αυτή δεδομένων αποτελείται από δυαδικές συνομιλίες όπου οι ηθοποιοί εκτελούν αυθορμητισμούς ή σενάρια που έχουν καταγραφεί εκ των προτέρων, συγκεκριμένα επιλεγμένα ώστε να επισημάνουν συγκεκριμένες εκφράσεις προσωπικότητας. Η βάση δεδομένων IEMOCAP αξιολογείται από πολλαπλούς αξιολογητές σε κατηγορική ταξινόμηση όπως θυμός, χαρά, λύπη, ουδετερότητα αλλά και σε δυαδική διαστασιακή ταξινόμηση όπως χαρά, ενεργοποίηση και κυριαρχία. Η λεπτομερής περιγραφή της πληροφορίας που ανακτάται από τους αισθητήρες κίνησης, η προσαρμοστικότητα στα αυθεντικά συναισθήματα καθώς και το μέγεθος της βάσης δεδομένων, καθιστούν αυτή τη βάση ως πολύτιμη πληροφορία για τα υπάρχοντα δεδομένα για την κοινότητα του τομέα Αναγνώρισης Συναισθήματος καθώς και για την μελέτη και μοντελοποίηση της πολύτροπης και εκφραστικής ανθρώπινης επικοινωνίας. Έχουμε και ένα παράδειγμα προκατασκευασμένων σεναρίων για το χρήστη που φοράει αισθητήρης εντοπισμού ομιλίας καθώς και για το χρήστη που δεν φοράει 5.1.

	Subject 1 (with markers)	Subject 2 (without markers)
1	(Fru) The subject is at the *Department of Motor Vehicles* (DMV) and he/she is being sent back after standing in line for an hour for not having the right form of IDs.	(Ang) The subject works at DMV. He/she rejects the application.
2	(Sad) The subject, a new parent, was called to enroll the army in a foreign country. He/she has to separate from his/her spouse for more than 1 year.	(Sad) The subject is his/her spouse and is extremely sad for the separation.
3	(Hap) The subject is telling his/her friend that he/she is getting married.	(Hap) The subject is very happy and wants to know all the details of the proposal. He/she also wants to know the date of the wedding.
4	(Fru) The subject is unemployed and he/she has spent last 3 years looking for work in his/her area. He/she is losing hope.	(Neu) The subject is trying to encourage his/her friend.
5	(Ang) The subject is furious, because the airline lost his/her baggage and he/she will receive only $50 (for a new bag that cost over $150 and has lots of important things).	(Neu) The subject works for the airline. He/she tries to calm the customer.
6	(Sad) The subject is sad because a close friend died. He had cancer that was detected a year before his death.	(Neu) The subject is trying to support his friend in this difficult moment.
7	(Hap) The subject has been accepted at USC. He/she is telling this to his/her best friend.	(Hap) The subject is very happy and wants to know the details (major, scholarship). He/she is also happy because he/she will stay in LA so they will be together.
8	(Neu) He/She is trying to change the mood of the customer and solve the problem.	(Ang) After 30 minutes talking with a machine, he/she is transferred to an operator. He/she expresses his/her frustration, but, finally, he/she changes his/her attitude.

Σχήμα 5.1: Σενάρια IEMOCAP

5.3 Βασικά Πειράματα

Χρησιμοποιώντας τη βάση δεδομένων για Αναγνώριση Προσωπικότητας, σε σύνολο 640 δειγμάτων έχουμε στον Πίνακα 5.3.

	Εξωστρέφεια	Προθυμία	Ευσυνειδησία	Νευρικότητα	Ανοικτότητα
Μηδενικά (0)	214	220	168	491	228
Μονάδες (1)	426	420	472	149	412

Πίνακας 5.3: Διαχωρισμός Κλάσεων των Δεδομένων

Χρησιμοποιούμε για όλα τα πειράματα διαχωρισμό (*K-fold*) και στη συνέχεια παίρνουμε το μέσο όρο των 4 τιμών.

Για την εξισορρόπηση των συνόλων εκπαίδευσης και επαλήθευσης έχουμε :

Στο σύνολο Επαλήθευσης για κάθε ένα απο τα 4 σύνολα διαχωρισμού , παίρνουμε το $\frac{1}{4}$ της κλάσης με τα λιγότερα δείγματα και τον ακριβώς ίδιο αριθμό από την άλλη κλάση. Επαναλαμβάνουμε τη διαδικασία άλλες τρεις φορές.

Στο σύνολο εκπαίδευσης για κάθε ένα από τα 4 σύνολα αφαιρούμε αντίστοιχα τα δείγματα του συνόλου εκπαίδευσης και επαναλαμβάνουμε τα δείγματα της κλάσης μειονότητας έως ότου οι δύο κλάσεις να έχουν ίδιο αριθμό δειγμάτων.

Για την εξαγωγή δεδομένων εφαρμόζουμε πρώτα VAD στο σύνολο των δεδομένων μας και παρατηρούμε ότι 97% των δειγμάτων δεν έχουν καθόλου θόρυβο, οπότε προχωράμε στη συνέχεια.

Εφαρμόζουμε τη Μέθοδο Ανάλυσης Κύριων Συνιστωσών έτσι ώστε να μειώσουμε τα χαρακτηριστικά που εξάγουμε από 1300 σε 239 . Τα χαρακτηριστικά αυτά τα εξάγουμε από το πρόγραμμα OpenSMILE και περιλαμβάνουν τόσο χαμηλού επιπέδου (*low-level*) χαρακτηριστικά, όσο και συναρτησιακά (*functional*) χαρακτηριστικά, όπως η ένταση της φωνής, ο τόνος , η κυρτότητα, το πλάτος του μετασχηματισμού fft , οι θέσεις μεγίστου και ελαχίστου, η ενέργεια RMS. Χρησιμοποιούμε για την εξαγωγή χαρατηριστικών την διαμόρφωση (*is*12_*speaker_trait.conf*) και ενδεικτικά τα χαρακτηριστικά που εξάγουμε είναι 5.2.

Με έναν αλγόριθμο όπως οι Μηχανές Υποστήριξης Διανυσμάτων (*SVM*) έχουμε στον Πίνακα 5.4.

Εξωστρέφεια	Προθυμία	Ευσυνειδησία	Νευρικότητα	Ανοικτότητα
0.5859	0.6015	0.625	0.5468	0.5703

Πίνακας 5.4: Πειράματα με Μηχανές Υποστήριξης Διανυσμάτων

'Επειτα στο πλαίσιο των βασικών πειραμάτων εισαγώγουμε τα χαρακτηριστικά που έχουμε εξάγει σε νευρωνικό δίκτυο δύο επιπέδων με αριθμό νευρώνων 128 και 64 αντίστοιχα 5.5.

Εξωστρέφεια	Προθυμία	Ευσυνειδησία	Νευρικότητα	Ανοικτότητα
0.607	0.61	0.623	0.618	0.567

Πίνακας 5.5: Νευρωνικά Με Ανάδραση

5.4 Προεκπαιδευμένος Αυτόματος κωδικοποιητης (*Autoencoder*)

Η αρχιτεκτονική που εφαρμόζουμε είναι αυτή του σχήματος 5.3.
Δηλαδή τροφοδοτούμε το δίκτυο με σπεκτογράμματα που έχουν προκύψει από τα δεδομένα ομιλίας που έχουμε, τα παιρνάμε από επίπεδα Συνελικτικού Νευρωνικού Δικτύου για την εξαγωγή χαρακτηριστικών , όπου αναλόγως τον πυρήνα του επιπέδου ορίζεται και μια διαφορετική κλίμακα και στη συνέχεια συνδυάζουμε τις διάφορες χρονικές κλίμακες για τη συνολική εξαγωγή χαρακτηριστικών και την ταξινόμηση των δεδομένων στο τελευταίο στάδιο του δικτύου. Συνολικά στο δίκτυο έχουμε εφαρμόσει πολλές διαφορετικές χρονικές κλίμακες και στη συνέχεια θα παρουσιάσουμε κάθε διαφορετικό στάδιο μέχρι την τελευταία αρχιτεκτονική που παρέχει και τα καλύτερα αποτελέσματα.

Statistical functionals (23)
(positive[2]) arithmetic mean, root quadratic mean, standard deviation, flatness, skewness, kurtosis, quartiles, inter-quartile ranges, 1 %, 99 % percentile, percentile range 1 %–99 %, percentage of frames contour is above: minimum + 25%, 50%, and 90 % of the range, percentage of frames contour is rising, maximum, mean, minimum segment length[1,3], standard deviation of segment length[1,3]
Regression functionals[1] (4)
linear regression slope, and corresponding approximation error (linear), quadratic regression coefficient a, and approximation error (linear)
Local minima/maxima related functionals[1] (9)
mean and standard deviation of rising and falling slopes (minimum to maximum), mean and standard deviation of inter maxima distances, amplitude mean of maxima, amplitude mean of minima, amplitude range of maxima
Other[1,3] (6)
LP gain, LPC 1 – 5

Σχήμα 5.2: Βασικά Χαρακτηριστικά openSMILE που εξάγονται

5.5 Επέκταση Αυτόματου Κωδικοποιητή

Στη συνέχεια αφού εκπαιδεύσουμε τον αυτόματο Κωδικοποιητή, παίρνουμε μικρότερης διάστασης σπεκτογράμματα όπως στο Σχήμα 5.4 με διαφορετικής διάστασης πυρήνες . Έπειτα αφού περάσουν από τα 3-5 κρυφά επίπεδα του Συνελικτικού νευρωνικού Δικτύου.

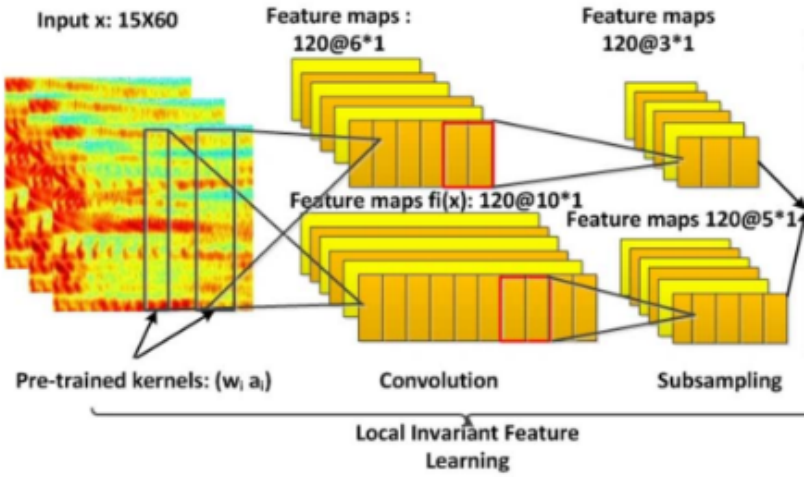
παρτίδα	Εξωστρέφεια	Προθυμία	Ευσυνειδησία	Νευρικότητα	Ανοικτότητα
512	0.6222	0.6207	0.6411	0.6631	0.6106
128	0.6113	0.6186	0.6367	0.6612	0.6074

Πίνακας 5.6: Εξαγωγή Χαρακτηριστικών μεγέθους (10*1)

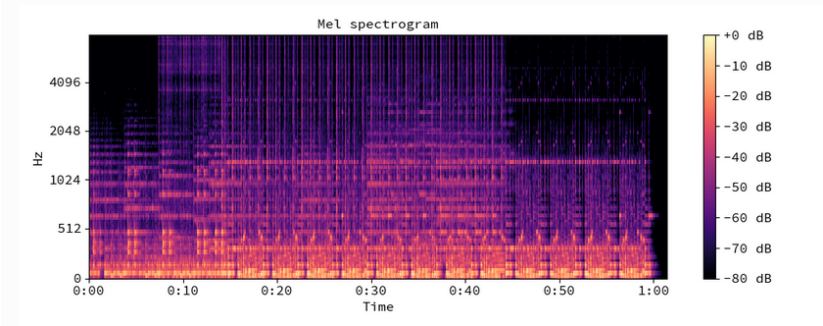
και για διαφορετικό μέγεθος χαρακτηριστικών

παρτίδα	Εξωστρέφεια	Προθυμία	Ευσυνειδησία	Νευρικότητα	Ανοικτότητα
512	0.6134	0.6197	0.6307	0.6589	0.6095

Πίνακας 5.7: Εξαγωγή Χαρακτηριστικών μεγέθους (6*1)



Σχήμα 5.3: Εξαγωγή Χαρακτηριστικών από Αυτόματο Κωδικοποιητή



Σχήμα 5.4: Παράδειγμα Σπεκτογράμματος

5.6 Διαφορετικές Χρονικές Κλίμακες

Ενώνουμε τα επιμέρους επίπεδα με τους διάφορους πυρήνες για ειδη πυρήνων 6,10,20,25,30,36,42 και έχουμε

Χαρακτηριστικά	παρτίδα	Εξωστρέφεια	Προθυμία	Ευσυνειδησία	Νευρικότητα	Ανοικτότητα
5*1	8	0.6589	0.6523	0.6657	0.6945	0.6501
5*1	16	0.6703	**0.6796**	0.6703	0.6879	0.6745
6*1	8	0.6689	0.6679	0.6703	0.6879	0.6673
6*1	16	0.6689	0.6757	**0.6894**	0.6911	0.6569
10*1	8	**0.6703**	0.6757	0.6813	**0.7024**	0.6673
10*1	16	0.6191	0.6589	0.6713	0.6493	0.6432
16*1	8	0.6756	0.6679	0.6813	0.6796	**0.6772**
16*1	16	0.6657	0.6601	0.6756	0.6879	0.6501
24*1	8	0.6713	**0.6796**	**0.6894**	0.6983	0.6673
24*1	16	0.6408	0.6269	0.6523	0.6632	0.6493
30*1	8	0.6601	0.6679	**0.6894**	0.6879	0.6673
30*1	16	0.6357	0.625	0.6673	0.6632	0.6441
48*1	8	0.6601	0.6601	0.6756	0.6796	0.6569
48*1	16	0.6532	0.6601	0.6703	0.6851	0.6493

Πίνακας 5.8: Διαφορετικές Χρονικές Κλίμακες

Σε αυτά τα αποτελέσματα παρατηρούμε ότι παρ' ότι αρκετές τιμές είναι κοντά, τα καλύτερα αποτελέσματα τα παιρνουμε για τις χρονικές κλίμακες 24 και 30, κοντά στο 67%

5.7 Συνένωση διαφορετικών Κλιμάκων

Συνενώνουμε τους διαφορετικούς συνδυασμούς των κλιμάκων και έχουμε Πίνακας 5.9.

Εξωστρέφεια	Προθυμία	Ευσυνειδησία	Νευρικότητα	Ανοικτότητα
0.6713	0.6796	**0.6911**	0.7024	**0.6813**

Πίνακας 5.9: Πλήρως Συνδεδεμένο στο Τελευταίο Σκέλος

και για διαφορετικό κομμάτι τελευταίου σκέλους, Συνελικτικό Δίκτυο αντί για Πλήρως Συνδεδεμένο 5.10

Εξωστρέφεια	Προθυμία	Ευσυνειδησία	Νευρικότητα	Ανοικτότητα
0.6703	0.6679	**0.6942**	0.6923	0.6772

Πίνακας 5.10: Συνελικτικό Δίκτυο στο Τελευταίο Σκέλος

Παρατηρούμε ότι για Πλήρως Συνδεδεμένο Δίκτυο στο τελευταίο στάδιο έχουμε καλύτερα αποτελέσματα απ' ότι με το Συνελικτικό Δίκτυο, κατά 0.4%.

Συνενώνοντας 2 διαφορετικές χρονικές Κλίμακες 5.11.

Εξωστρέφεια	Προθυμία	Ευσυνειδησία	Νευρικότητα	Ανοικτότητα
0.6813	0.6796	**0.6973**	0.7024	0.6772

Πίνακας 5.11: Συνένωση Χρονικών Κλιμάκων 10 και 20

Αυτά τα αποτελέσματα είναι και τα καλύτερα που πήραμε καθώς φτάνουν στο μέσο όρο των 5 κλάσεων στο 68.51%.

5.8 Επέκταση στην Αναγνώριση Συναισθήματος

Παρόμοιες μέθοδοι εφαρμόστηκαν και στην βάση Δεδομένων IEMOCAP με ρυθμίσεις :
Χρησιμοποιήσαμε για τα σπεκτογράμματα μας πλαίσιο $50ms$ και εξαγωγές χαρακτηριστικών από τον Αυτόματο Κωδικοποιητή μεγέθους $(10 * 1) + (20 * 1)$

WA	UA
40.1	39.3

Πίνακας 5.12: Αποτελέσματα μονής Χρονικής Κλίμακας IEMOCAP

WA	UA
48.3	49.5

Πίνακας 5.13: Αποτελέσματα συνδυασμού Χρονικών Κλιμάκων IEMOCAP

WA	UA
52.9	52.3

Πίνακας 5.14: Περαιτέρω τμηματοποίηση των σπεκτογραμμάτων και διαφορετικές Χρονικές Κλίμακες στην IEMOCAP

5.8.1 Αναγνώριση Προσωπικότητας από Προεκπαιδευμένα Δίκτυα στην IEMOCAP

Αφού λοιπόν εξετάσαμε τις δύο βάσεις δεδομένων ξεχωριστά , βάση Αναγνώρισης Προσωπικότητας 640 δειγμάτων και βάση IEMOCAP για Αναγνώριση Συναισθήματος, δοκιμάσαμε στην συνέχεια να προεκπαιδεύσουμε το δίκτυο στην IEMOCAP που είναι μεγαλύτερη σε μέγεθος βάση δεδομένων και να εφαρμόσουμε την καλύτερη αρχιτεκτονική μας ήδη προεκπαιδευμένη για την Αναγνώριση Προσωπικότητας στο δικό μας σύνολο δεδομένων και προέκυψαν τα εξής αποτελέσματα 5.15

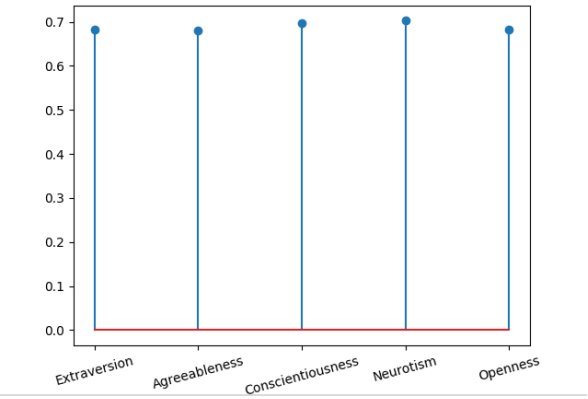
Εξωστρέφεια	Προθυμία	Ευσυνειδησία	Νευρικότητα	Ανοικτότητα
0.6589	0.6679	0.6703	0.6911	0.6745

Πίνακας 5.15: Αναγνώριση Προσωπικότητας από Προεκπαιδευμένο Δίκτυο στην IEMOCAP

5.9 Συγκεντρωτικά Αποτελέσματα

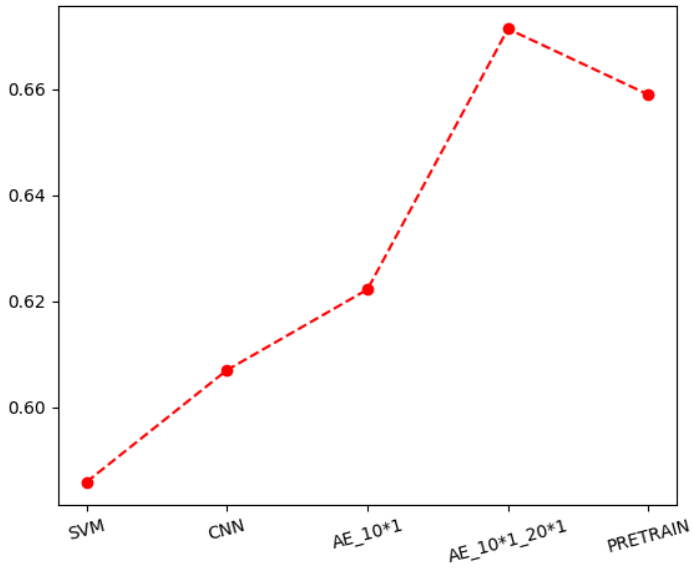
Αφού λοιπόν έχουμε αναλύσει όλο το εύρος των πειραμάτων που εκτελέσαμε και έχουμε παρουσιάσει τα αποτελέσματα, εδώ τα παρουσιάζουμε σε μια πιο οπτική μορφή για την καλύτερη μελέτη και σύγκριση τους.

Οι συνολικές τιμές που επιτύχαμε για τους 5 διαφορετικούς άξονες Προσωπικότητας άγγιξαν τα καλύτερα αποτελέσματα της βιβλιογραφίας που είναι 72% για την πιο μεγάλη τιμή στους 5 άξονες. Φτάσαμε έως 70.24% για την Νευροτικότητα συγκεκριμένα 5.5.

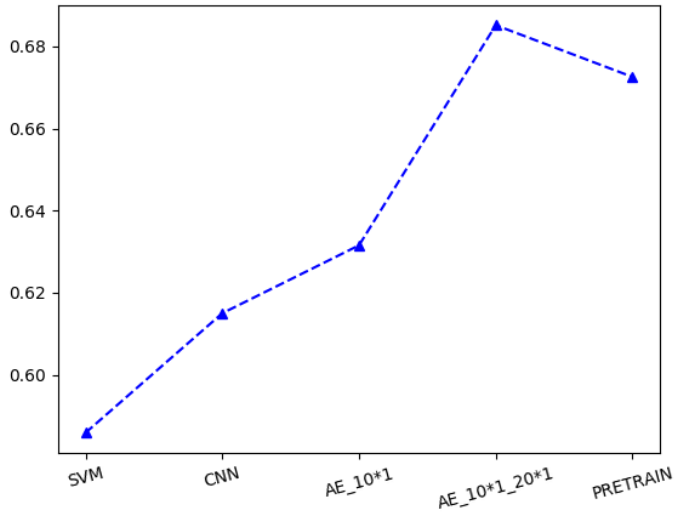


Σχήμα 5.5: Αποτύπωση των καλύτερων αποτελεσμάτων για κάθε μία από τις κλάσεις O.C.E.A.N.

Επίσης κάτι πολύ σημαντικό , βλέπουμε τη βελτίωση των αποτελεσμάτων όσο προχωράμε από τα Βασικά μας Πειράματα, στην απλή Χρονική Κλίμακα και κατά συνέπεια στην συνένωση διαφορετικών Χρονικών Κλιμάκων , τόσο για την κάθε Κλάση Αναγνώρισης Προσωπικότητας ξεχωριστά 5.6 όσο και συνολικά όταν παίρνουμε το μέσο όρο των 5 κλάσεων και βλέπουμε τα αποτελέσματα των πειραμάτων, όπως αυτά εξελίσσονται στην παρούσα εργασία 5.7.



Σχήμα 5.6: Συγκεντρωτικά πειράματα όλων των αρχιτεκτονικών για την ”Εξωστρέφεια”



Σχήμα 5.7: Συγκεντρωτικά πειράματα όλων των αρχιτεκτονικών για το μέσο όρο των 5 κλάσεων

Κεφάλαιο 6

Συμπεράσματα και Προεκτάσεις Εργασίας

6.1 Συμπεράσματα

Με την αναπαράσταση των πειραμάτων στο προηγούμενο κεφάλαιο, ακολουθεί η σύνοψη της διπλωματικής και η εξαγωγή των συμπερασμάτων. Η παρούσα εργασία εστίασε στην Αναγνώριση Προσωπικότητας από σήμα εισόδου ομιλίας. Εκπαιδεύσαμε Αυτόματους Κωδικοποιητές για την εξαγωγή χαρακτηριστικών από τα σπεκτογράμματα Φωνής που προωθούσαμε σαν είσοδο στα Νευρωνικά μας δίκτυα. Αναλόγως με τον μέγεθος του πυρήνα, φιλτράρουμε το σπεκτόγραμμα σε διαφορετικές χρονικές κλίμακες για την εξαγωγή χαρακτηριστικών διαφορετικού μεγέθους.

Στη συνέχεια κάνουμε ταξινόμηση των δειγμάτων για τις διάφορες χρονικές κλίμακες, εξετάζοντας το πρώτο κομμάτι του αυτόματου Κωδικοποιητή και δοκιμάζοντας στο δεύτερο μέρος διαφορετικά επίπεδα νευρωνικών δικτύων, έχοντας αρκετά καλά αποτελέσματα, μεχρι 66.12% . Έπειτα δοκιμάζουμε περισσότερα επίπεδα στο δεύτερο μέρος του νευρωνικού και έχουμε αποτελέσματα της τάξης του 67.2%.

Όταν δοκιμάζουμε τη μετατροπή του δικτύου έτσι ώστε διαφορετικές χρονικές κλίμακες να συνενώνονται και η εξαγωγή χαρακτηριστικών να γίνεται ταυτόχρονα σε 2 κλίμακες, τα αποτελέσματα μας βελτιώνονται ακόμα περισσότερο, φτάνοντας στο 68.51% στον μέσο όρο και έως 70.24% στις επιμέρους κλάσεις.

Τέλος χρησιμοποιούμε σαν δεδομένα για προεκπαίδευσης βάση δεδομένων που έχει δημιουργηθεί για Αναγνώριση Συναισθήματος και έπειτα ξαναεφαρμόζουμε την ίδια αρχιτεκτονική και παίρνουμε παρομοίως καλά αποτελέσματα καθώς το δίκτυο είναι επαρκώς εκπαιδεύσιμο και με το αρχικό σύνολο δεδομένων της βάσης για Αναγνώριση Προσωπικότητας.

Τα συμπεράσματα που καταλήξαμε είναι ότι το σπεκτόγραμμα χρησιμοποιείται ικανοποιητικά για την αναπαράσταση Φωνής και όχι μόνο εικόνας και έχει εφαρμογές και στην Αναγνώριση Προσωπικότητας. Επίσης η εξαγωγή χαρακτηριστικών από σπεκτόγραμμα οδηγεί σε καλής ποιότητας χαρακτηριστικά που μπορούν να χρησιμοποιηθούν στη συνέχεια. Τέλος, ο συνδυασμός εξαγωγής χαρακτηριστικών από διαφορετικές χρονικές κλίμακες οδηγεί σε μεγαλύτερη ποικιλία στα χαρακτηριστικά και κατά συνέπεια σε μεγαλύτερη απόδοση στην Αναγνώριση Φωνής. Όσον αφορά τα προεκπαιδευμένα δίκτυα, όταν το δίκτυο δεν περιέχει υπερβολικά μεγάλο αριθμό εκπαιδεύσιμων μεταβλητών, μπορεί να γίνει η σύγκλιση της εκπαίδευσης και με μικρότερη βάση δεδομένων.

6.2 Προεκτάσεις Εργασίας

Στην πορεία θα μπορούσαμε να δοκιμάσουμε και σε άλλα σύνολα δεδομένων τις παρούσες αρχιτεκτονικές για την εξαγωγή συμπερασμάτων όσων αφορά την Αναγνώριση Προσωπικότητας.

Συγκεκριμένα θα μπορούσαμε να δοκιμάσουμε δίκτυα με Ανάδραση και Νευρωνικά Δίκτυα Μακράς - Βραχέας Μνήμης αν είχαμε μεγαλύτερο σύνολο δεδομένων για την Προσωπικότητα. Επίσης θα μπορούσαμε να ενσωματώσουμε και ένα μοντέλο γλώσσας και να κάνουμε ταυτόχρονη αναγνώριση κειμένου από τα *transcripts* της βάσης, αλλά ο αρχικός καθορισμός των δεδομένων , δεν περιλάμβανε κατανόηση της γλώσσας (Γαλλικά) και άρα δεν θα είχε νόημα η αναγνώριση και από κείμενο όσον αφορά τους στόχους που έχουν ήδη τεθεί από τους αξιολογητές. Τέλος, θα μπορούσε να γινει

Αναγνώριση Προσωπικότητας είτε από εικόνα των ομιλητών, είτε από βιντεο, καθώς όμως αυτή τη στιγμή δεν υπάρχει διαθέσιμο κάποιο σύνολο δεδομένων αρκετά μεγάλο για την Αναγνώριση Προσωπικότητας, ευελπιστούμε στην κατασκευή του στο μέλλον για την καλύτερη δυνατή βοήθεια στην Αναγνώριση Προσωπικότητας.

Επιπροσθέτως, θα μπορούσαμε να εισάγουμε στα Νευρωνικά Δίκτυα Σπεκτογράμματα με διαφορετική Ανάλυση, ώστε η εξαγωγή Χαρακτηριστικών στη συνέχεια να αποτελεί διαφορετική χρονική κλίμακα, ακόμα και για τον ίδιο πυρήνα Συνελικτικού Δικτύου. Τέλος να αναφέρουμε ότι θα μπορούσε να χρησιμοποιηθεί Ιεραρχικό μοντέλο για τη συνολική αρχιτεκτονική του δικτύου, καθώς βλέπουμε ότι η συνένωση χρονικών Κλιμάκων αποδίδει τα καλύτερα αποτελέσματα, με επιπλέον επίπεδα ιεραρχίας από την τωρινή διάταξη.

Βιβλιογραφία

[1] G. Mohammadi, a. origlia, M. Pili, and A. Vinciarelli, “From speech to personality: Mapping voice quality and intonation into personality differences,” in *in Proceedings of ACM Multimedia 2012*, 2012.

[2] L. R. Goldberg, “An alternative ”description of personality”: The big-five factor structure,” vol. 59, pp. 1216–29, 01 1991.

[3] I. B. Myers, *The Myers-Briggs type indicator*. Palo Alto, Calif: Consulting Phychologists Press, 1962.

[4] A. E. Kazdin, *Encyclopedia of Phychology*, 2000.

[5] V. Arun, “Impact of personality on technology adoption: An empirical model,” *Journal of the American Society for Information Science and Technology*, vol. 56, no. 8, pp. 803–811. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20169

[6] http://humanscience.wikia.com/wiki/Dimensions_of_Personality.

[7] K. M. Lee, W. Peng, S.-A. Jin, and C. Yan, “Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction,” *Journal of Communication*, vol. 56, no. 4, pp. 754–772, 2006. [Online]. Available: http://dx.doi.org/10.1111/j.1460-2466.2006.00318.x

[8] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, “Using linguistic cues for the automatic recognition of personality in conversation and text,” *J. Artif. Int. Res.*, vol. 30, no. 1, pp. 457–500, Nov. 2007. [Online]. Available: http://dl.acm.org/citation.cfm?id=1622637.1622649

[9] C. X. Ivanov Alexei, “Modulation spectrum analysis for speaker personality trait recognition,” 2012.

[10] P. B. Dasgupta, “Detection and analysis of human emotions through voice and speech pattern processing,” 2017.

[11] S. Narang and M. D. Gupta, “Speech feature extraction techniques : A review,” 2015.

[12] E. M. Grais, H. Wierstorf, D. Ward, and M. D. Plumbley, “Multi-resolution fully convolutional neural networks for monaural audio source separation,” *CoRR*, vol. abs/1710.11473, 2017. [Online]. Available: http://arxiv.org/abs/1710.11473

[13] K. Y. Huang, C. H. Wu, M. H. Su, and H. C. Fu, “Mood detection from daily conversational speech using denoising autoencoder and lstm,” in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 5125–5129.

[14] O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, “Convolutional neural networks for speech recognition,” *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 22, no. 10, pp. 1533–1545, Oct. 2014. [Online]. Available: http://dx.doi.org/10.1109/TASLP.2014.2339736

[15] ——, "Convolutional neural networks for speech recognition," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 22, no. 10, pp. 1533–1545, Oct. 2014. [Online]. Available: http://dx.doi.org/10.1109/TASLP.2014.2339736

[16] Y. Miao, M. Gowayyed, and F. Metze, "EESEN: end-to-end speech recognition using deep RNN models and wfst-based decoding," *CoRR*, vol. abs/1507.08240, 2015. [Online]. Available: http://arxiv.org/abs/1507.08240

[17] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*, ser. ICML'14. JMLR.org, 2014, pp. II–1764–II–1772. [Online]. Available: http://dl.acm.org/citation.cfm?id=3044805.3045089

[18] Y. Zhang, M. Pezeshki, P. Brakel, S. Zhang, C. Laurent, Y. Bengio, and A. C. Courville, "Towards end-to-end speech recognition with deep convolutional neural networks," *CoRR*, vol. abs/1701.02720, 2017. [Online]. Available: http://arxiv.org/abs/1701.02720

[19] J. Gideon, S. Khorram, Z. Aldeneh, D. Dimitriadis, and E. M. Provost, "Progressive neural networks for transfer learning in emotion recognition," *CoRR*, vol. abs/1706.03256, 2017. [Online]. Available: http://arxiv.org/abs/1706.03256

[20] J. Deng, Z. Zhang, E. Marchi, and B. Schuller, "Sparse autoencoder-based feature transfer learning for speech emotion recognition," in *Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, ser. ACII '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 511–516. [Online]. Available: http://dx.doi.org/10.1109/ACII.2013.90

[21] K. Choi, G. Fazekas, M. B. Sandler, and K. Cho, "Transfer learning for music classification and regression tasks," *CoRR*, vol. abs/1703.09179, 2017. [Online]. Available: http://arxiv.org/abs/1703.09179

[22] D. Beaver, B. Zack Clark, E. Stanton Flemming, T. F. Jaeger, and M. Wolters, "When semantics meets phonetics: Acoustical studies of second-occurrence focus," vol. 83, pp. 245–276, 06 2007.

[23] J. Lee, J. Park, K. L. Kim, and J. Nam, "Sample-level deep convolutional neural networks for music auto-tagging using raw waveforms," *CoRR*, vol. abs/1703.01789, 2017. [Online]. Available: http://arxiv.org/abs/1703.01789

[24] M. Sambur, "Selection of acoustic features for speaker identification," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 23, no. 2, pp. 176–182, April 1975.

[25] M. Sidorov, C. Brester, E. Semenkin, and W. Minker, "Speaker state recognition with neural network-based classification and self-adaptive heuristic feature selection," *2014 11th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, vol. 01, pp. 699–703, 2014.

[26] A. Chorianopoulou, P. Koutsakis, and A. Potamianos, "Speech emotion recognition using affective saliency," pp. 500–504, 09 2016.

[27] B. M. . N. S. Busso, C., "Toward effective automatic recognition systems of emotion in speech. in social emotions in nature and artifact." 2013.

[28] C. Busso, S. Lee, and S. Narayanan, "Analysis of emotionally salient aspects of fundamental frequency for emotion detection," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 582–596, May 2009.

[29] B. Schuller, A. Batliner, S. Steidl, and D. Seppi, "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge," *Speech Commun.*, vol. 53, no. 9-10, pp. 1062–1087, Nov. 2011. [Online]. Available: http://dx.doi.org/10.1016/j.specom.2011.01.011

[30] A. Tapus and M. J. Mataric, "Socially assistive robots: The link between personality, empathy, physiological signals, and task performance," in *AAAI Spring Symposium: Emotion, Personality, and Social Behavior*, 2008.

[31] S. Kim, P. G. Georgiou, S. Lee, and S. Narayanan, "Real-time emotion detection system using speech: Multi-modal fusion of different timescale features," in *2007 IEEE 9th Workshop on Multimedia Signal Processing*, Oct 2007, pp. 48–51.

[32] B. Schuller, G. Rigoll, and M. Lang, "Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, May 2004, pp. I–577–80 vol.1.

[33] M. B. Mustafa, M. A. Yusoof, Z. M. Don, and M. Malekzadeh, "Speech emotion recognition research: An analysis of research focus," *Int. J. Speech Technol.*, vol. 21, no. 1, pp. 137–156, Mar. 2018. [Online]. Available: https://doi.org/10.1007/s10772-018-9493-x

[34] S. G. Karadoğan and J. Larsen, "Combining semantic and acoustic features for valence and arousal recognition in speech," in *2012 3rd International Workshop on Cognitive Information Processing (CIP)*, May 2012, pp. 1–6.

[35] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. P. Seligman, and L. H. Ungar, "Personality, gender, and age in the language of social media: The open-vocabulary approach," *PLOS ONE*, vol. 8, no. 9, pp. 1–16, 09 2013. [Online]. Available: https://doi.org/10.1371/journal.pone.0073791

[36] J.-I. Biel, D. Gatica-Perez, J. Dines, and V. Tsminiaki, "Hi youtube! personality impressions and verbal content in social video," *15th ACM International Conference on Multimodal Interaction, Sydney, Australia, ACM, 2013*, 2013.

[37] G. Mohammadi, A. Vinciarelli, and M. Mortillaro, "The voice of personality: Mapping nonverbal vocal behavior into trait attributions," in *Proceedings of the 2Nd International Workshop on Social Signal Processing*, ser. SSPW '10. New York, NY, USA: ACM, 2010, pp. 17–20. [Online]. Available: http://doi.acm.org/10.1145/1878116.1878123

[38] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed&uid=16873662&cmd=showdetailview&indexed=google

[39] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, "Why does unsupervised pre-training help deep learning?" *J. Mach. Learn. Res.*, vol. 11, pp. 625–660, Mar. 2010. [Online]. Available: http://dl.acm.org/citation.cfm?id=1756006.1756025

[40] M. Schuster and K. Paliwal, "Bidirectional recurrent neural networks," *Trans. Sig. Proc.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997. [Online]. Available: http://dx.doi.org/10.1109/78.650093

[41] A. Graves, N. Jaitly, and A. r. Mohamed, "Hybrid speech recognition with deep bidirectional lstm," in *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*, Dec 2013, pp. 273–278.

[42] J. Dai, S. Liang, W. Xue, C. Ni, and W. Liu, "Long short-term memory recurrent neural network based segment features for music genre classification," in *2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, Oct 2016, pp. 1–5.

[43] B. Schuller, S. Steidl, A. Batliner, E. Nöth, A. Vinciarelli, F. Burkhardt, f. Weninger, F. Eyben, T. Bocklet, G. Mohammadi, and B. Weiss, "A survey on perceived speaker traits: Personality, likability, pathology and the first challenge," *Computer Speech and Language*, vol. 19, no. 1, pp. 100–131, Jan. 2015, received 25 July 2013, Revised 26 June 2014, Accepted 15 August 2014, Available online 27 August 2014.

[44] K. Krishna, L. Lu, K. Gimpel, and K. Livescu, "A study of all-convolutional encoders for connectionist temporal classification," *CoRR*, vol. abs/1710.10398, 2017. [Online]. Available: http://arxiv.org/abs/1710.10398

[45] S. Ghaffarzadegan, H. Bořil, and J. H. L. Hansen, "Deep neural network training for whispered speech recognition using small databases and generative model sampling," *International Journal of Speech Technology*, vol. 20, no. 4, pp. 1063–1075, Dec 2017. [Online]. Available: https://doi.org/10.1007/s10772-017-9461-x

[46] N. Jaitly and G. Hinton, "Learning a better representation of speech sound waves using restricted boltzmann machines," pp. 5884 – 5887, 06 2011.

[47] Q. Mao, M. Dong, Z. Huang, and Y. Zhan, "Learning salient features for speech emotion recognition using convolutional neural networks," *IEEE Transactions on Multimedia*, vol. 16, no. 8, pp. 2203–2213, Dec 2014.

[48] Y. T. X. L. Szu-Wei Fu, Ting-yao Hu, "Complex spectrogram enhancement by convolutional neural network with multi-metrics learning."

[49] C. Fayet, A. Delhay, D. Lolive, and P.-F. Marteau, "Big Five vs. Prosodic Features as Cues to Detect Abnormality in SSPNET-Personality Corpus," in *Interspeech*, Stockholm, Sweden, Aug. 2017. [Online]. Available: https://hal.inria.fr/hal-01583510

[50] M. H. Su, C. H. Wu, K. Y. Huang, Q. B. Hong, and H. M. Wang, "Personality trait perception from speech signals using multiresolution analysis and convolutional neural networks," in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Dec 2017, pp. 1532–1536.

[51] S. Jothilakshmi and R. Brindha, "Speaker trait prediction for automatic personality perception using frequency domain linear prediction features," in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, March 2016, pp. 2129–2132.

[52] N. Takahashi, M. Gygli, and L. V. Gool, "Aenet: Learning deep audio features for video analysis," *CoRR*, vol. abs/1701.00599, 2017. [Online]. Available: http://arxiv.org/abs/1701.00599

[53] C. Busso, M. Bulut, C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. Chang, S. Lee, and S. Narayanan, "Iemocap: Interactive emotional dyadic motion capture database," *Language Resources and Evaluation*, vol. 42, no. 4, pp. 335–359, 12 2008.

[54] D. C. Funder, "Personality," *Annual Review of Psychology*, vol. 52, no. 1, pp. 197–221, 2001, pMID: 11148304. [Online]. Available: https://doi.org/10.1146/annurev.psych.52.1.197

[55] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "Iemocap: interactive emotional dyadic motion capture database,"

Language Resources and Evaluation, vol. 42, no. 4, p. 335, Nov 2008. [Online]. Available: https://doi.org/10.1007/s10579-008-9076-6

[56] S. Tripathi and H. S. M. Beigi, "Multi-modal emotion recognition on IEMOCAP dataset using deep learning," *CoRR*, vol. abs/1804.05788, 2018. [Online]. Available: http://arxiv.org/abs/1804.05788