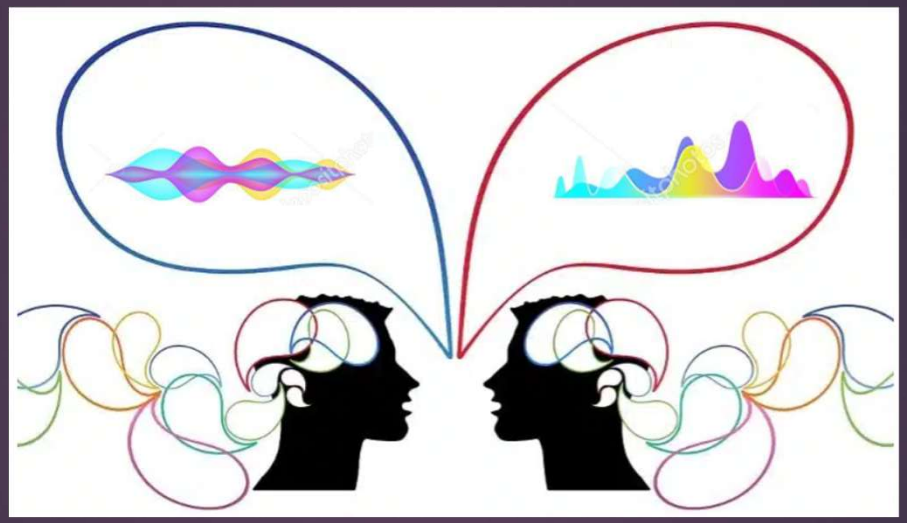
ΕΠΕΞΕΡΓΑΣΙΑ ΦΩΝΗΣ



ΔΡ. ΑΝΑΣΤΑΣΙΑ ΠΕΝΤΑΡΗ

ΗΛΕΚΤΡΟΝΙΚΟΣ ΜΗΧΑΝΙΚΟΣ & ΜΗΧΑΝΙΚΟΣ ΥΠΟΛΟΓΙΣΤΩΝ

14/1/2026

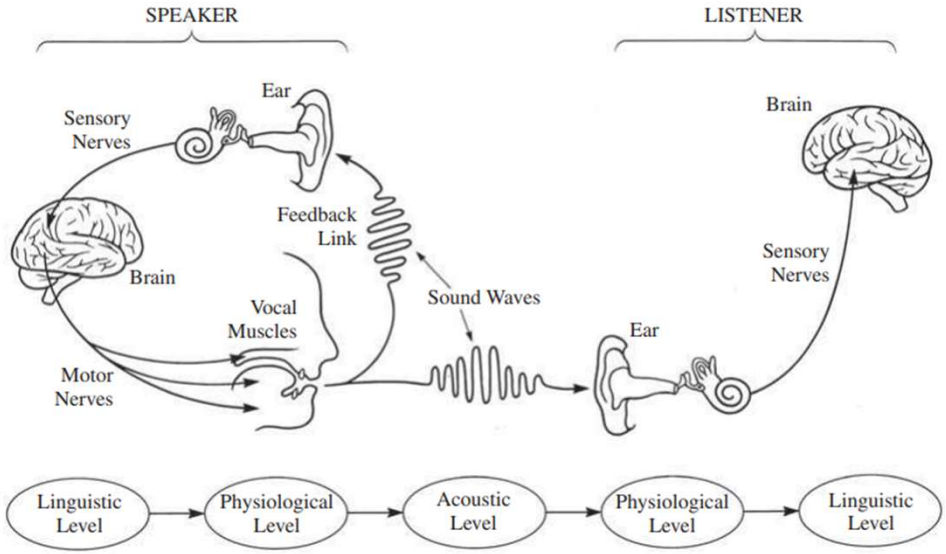
•Φωτογραφία εξωφύλλου: https://neurosciencenews.com/neuroscience-terms/speech-analysis/

ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

- Σπουδές:
  - 2013: Δίπλωμα, ΗΜΜΥ, Πολυτεχνείο Κρήτης
  - 2015: Μεταπτυχιακό Δίπλωμα Ειδίκευσης, ΗΜΜΥ, Πολυτεχνείο Κρήτης
  - 2022: Διδακτορικό Δίπλωμα, Επιστήμη Υπολογιστών, Πανεπιστήμιο Κρήτης
  - 2022: Μεταδιδακτορική Ερευνήτρια, CBML, ITE
- Κύρια Ερευνητικά Ενδιαφέροντα:
  - Επεξεργασία Φωνής και Λόγου (ιατρικών δεδομένων)
  - Επεξεργασία Σήματος για Αναγνώριση Συναισθηματικών Καταστάσεων
  - Ανάπτυξη Αλγορίθμων Βασισμένων σε Μηχανική και Βαθιά Μάθηση
  - Θεωρία Γράφων προσαρμοσμένη σε Σήματα (όπως φωνής, ΗΕΓ)

Τι είναι ο λόγος

- Ο λόγος αποτελεί το κύριο μέσο επικοινωνίας μεταξύ των ανθρώπων
- Χαρακτηρίζεται από εναλλαγές στον τόνο, το ύφος, την άρθρωση και άλλα στοιχεία



Μεταφορά Λόγου από Ομιλητή σε Ακροατή

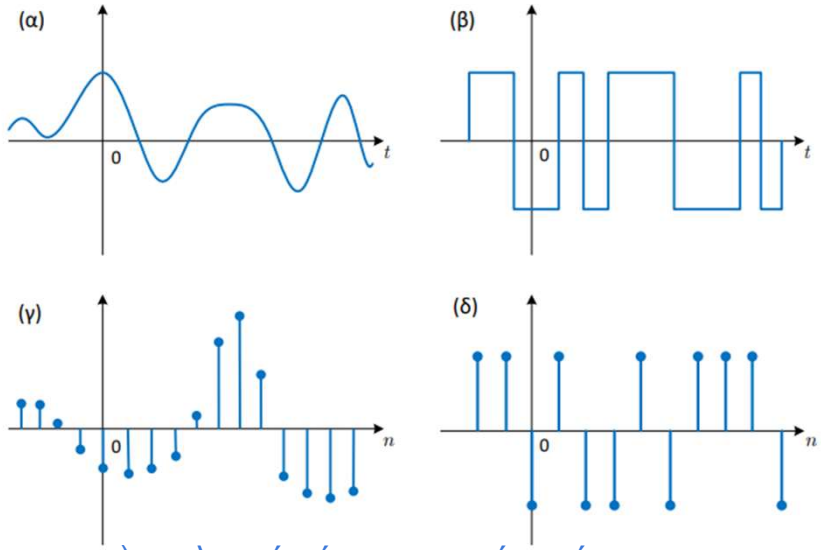
- Κατά την γλωσσολογία, αρχικά σχηματίζεται μία ιδέα στο μυαλό του ομιλητή
- Η ιδέα αυτή μετατρέπεται σε **λέξεις**, **φράσεις** και **προτάσεις**, σύμφωνα με τους γραμματικούς κανόνες **κάθε** γλώσσας
- Κατά την φυσιολογική επικοινωνία, ο εγκέφαλος αρχίζει και δημιουργεί ηλεκτρικά σήματα που κινούνται κατά μήκος των κινητικών νεύρων
- Τα ηλεκτρικά σήματα αυτά ενεργοποιούν μύες στην φωνητική οδό και στις φωνητικές χορδές

Μεταφορά Λόγου από Ομιλητή σε Ακροατή

- Η κίνηση της φωνητικής οδού και των φωνητικών χορδών καταλήγει στην δημιουργία **ηχητικού κύματος** που σχηματίζεται από τα χείλη
- Το ηχητικό κύμα διαδίδεται στον χώρο (ως αλυσιδωτή αντίδραση μεταξύ των σωματιδίων του αέρα) και δονεί το τύμπανο του αυτιού του ακροατή
- Η δόνηση στο τύμπανο του αυτιού του ακροατή προκαλεί ηλεκτρικά σήματα που κινούνται κατά μήκος των αισθητήριων νεύρων προς τον εγκέφαλο, σε φυσιολογικές καταστάσεις
- Τέλος, ο εγκέφαλος του ακροατή εκτελεί αναγνώριση και κατανόηση της ομιλίας, σε γλωσσικό επίπεδο

Ο Ψηφιακός Λόγος και η Ψηφιακή Επεξεργασία του

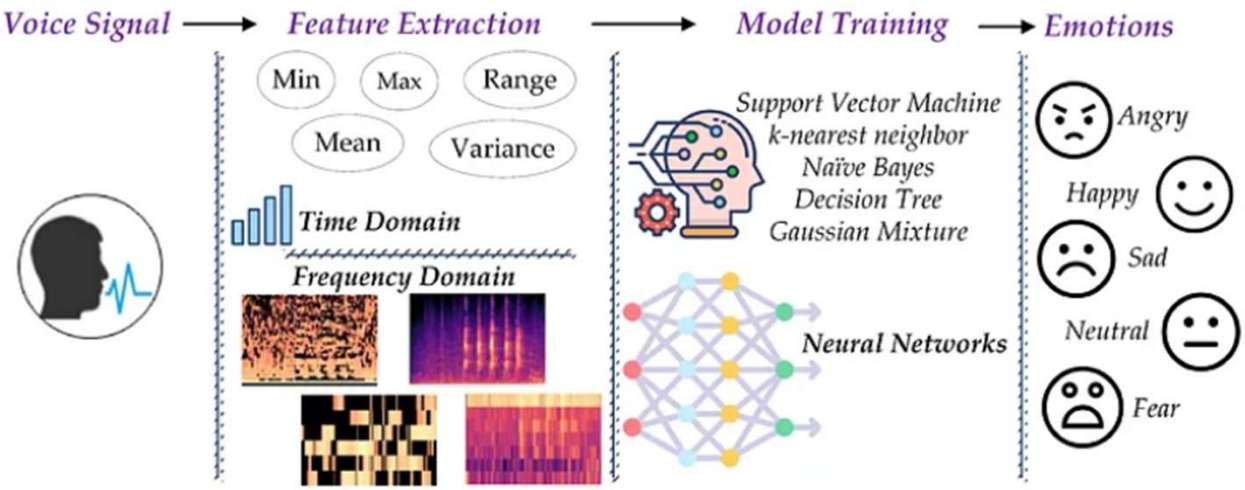
- Μέσω της ψηφιακής τεχνολογίας και περαιτέρω μέσω της μετατροπής αναλογικών ποσοτήτων σε ψηφιακές ο λόγος μπορεί να πάρει την μορφή ***σήματος***
- Το σήμα λόγου μπορεί να αναπαρασταθεί ως ένα *σήμα διακριτού χρόνου*, δηλαδή ως ένα ***σύνολο δειγμάτων***



α) αναλογικό σήμα, συνεχούς χρόνου
β) ψηφιακό σήμα, συνεχούς χρόνου
γ) αναλογικό σήμα, διακριτού χρόνου
δ) ψηφιακό σήμα, διακριτού χρόνου

Κύριες Εφαρμογές της Επεξεργασίας Λόγου-Σήματος

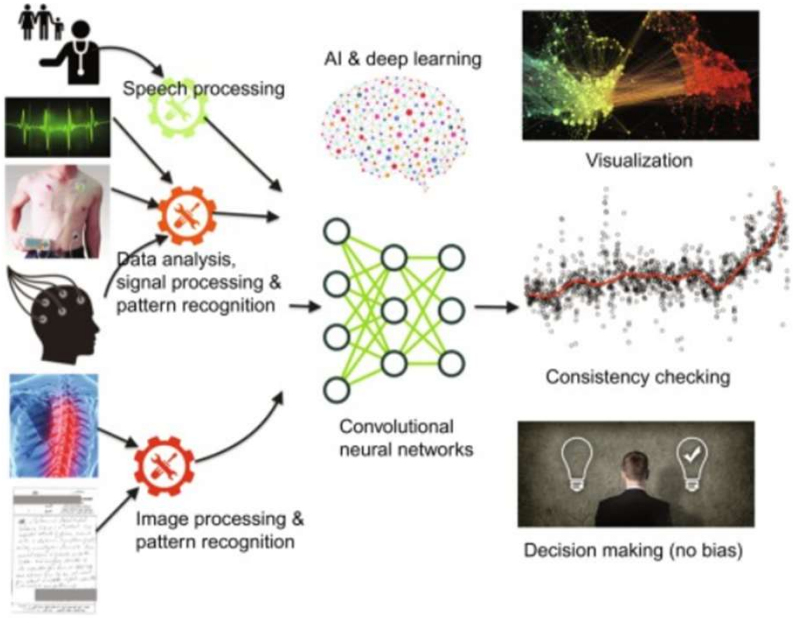
- Εργαλεία ανάλυσης ομιλίας: Μέσω της ανάλυσης του λόγου γίνεται επεξεργασία με στόχο την **αναγνώριση συναισθημάτων**



https://medium.com/@evertongomede/speech-emotion-recognition-unveiling-the-emotional-spectrum-through-sound-2dde24273616

Κύριες Εφαρμογές της Επεξεργασίας Λόγου-Σήματος

- Η **ανάλυση ομιλίας** βοηθά στη διάγνωση και την παρακολούθηση καταστάσεων ψυχικής υγείας, όπως η κατάθλιψη, το άγχος, αναλύοντας τις αλλαγές στον τόνο της φωνής, τα μοτίβα ομιλίας και το λεξιλόγιο



https://www.sciencedirect.com/topics/computer-science/speech-analysis

Κύριες Εφαρμογές της Επεξεργασίας Λόγου-Σήματος

- **Διαχωρισμός ασθενών-υγιών**
- Επεξεργασία σημάτων λόγου → Εξαγωγή χαρακτηριστικών
- Χαρακτηριστικά-Δείκτες-Βιοδείκτες: ποσοτικοποίηση χαρακτηριστικών του λόγου με στόχο αναγνώριση ασθενειών/πορεία ασθένειας/καθορισμός φαρμακευτικής αγωγής
- Τηλεϊατρική: Πολύ συχνά η συλλογή δεδομένων λόγου γίνεται μέσω εφαρμογής→ Επεξεργασία → Κλινική αξιολόγηση → Κλινική απάντηση (όπως σε απομακρυσμένες περιοχές)

Κύριες Εφαρμογές της Επεξεργασίας Λόγου-Σήματος

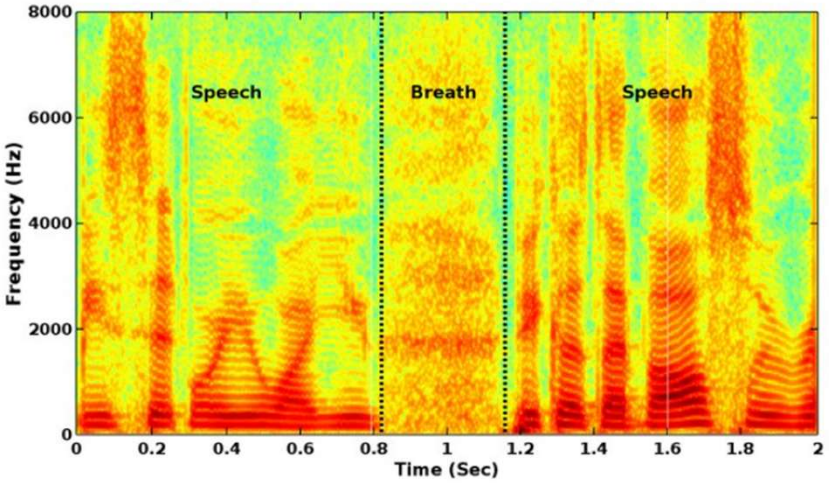
- **Βιομετρία φωνής:** Χρησιμοποιείται για λόγους ασφάλειας και ελέγχου ταυτότητας. Αναλύοντας μοναδικά χαρακτηριστικά φωνής, η ανάλυση ομιλίας μπορεί να επαληθεύσει την ταυτότητα ενός ατόμου, παρέχοντας ένα επιπλέον επίπεδο ασφάλειας στις τραπεζικές ή ευαίσθητες λειτουργίες
- **Έξυπνοι βοηθοί:** Εικονικοί βοηθοί (όπως η Alexa) χρησιμοποιούν ανάλυση ομιλίας για να κατανοούν τις εντολές των χρηστών, να αναγνωρίζουν συγκεκριμένες φωνές και να παρέχουν εξατομικευμένες απαντήσεις
- **Πολυμέσα & Ψυχαγωγία:** Χρησιμοποιείται στον υποτιτλισμό, τις υπηρεσίες ομιλίας σε κείμενο και τη βελτίωση της αλληλεπίδρασης των χρηστών σε παιχνίδια ή άλλες εφαρμογές ψυχαγωγίας

Διαδικασία Επεξεργασίας Σημάτων Λόγου

- Καταγραφή του λόγου
- Πολύ συνηθισμένη διαδικασία είναι η **αποθορυβοποίησή** του:
  - Διαχωρισμός σήματος ομιλητή ενδιαφέροντος από άλλους ομιλητές
  - Μείωση του θορύβου από τον περιβάλλοντα χώρο καταγραφής
- Η **εξαγωγή βέλτιστης πληροφορίας** από το σήμα φωνής:
  - Εξαρτάται από την εφαρμογή μας
- **Εξαγωγή κατάλληλων χαρακτηριστικών**
- Αναπαράσταση σε **γράφο** ή **φασματογράφημα**
- Κατηγοριοποίηση/Αναγνώριση/Διαχωρισμών Ομιλητών

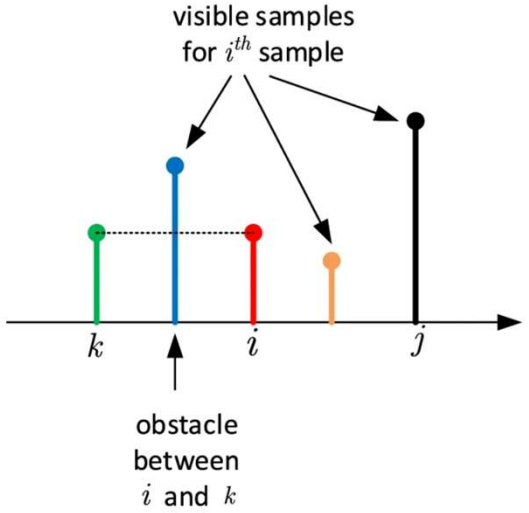
Παραδείγματα Αναπαραστάσεων Σημάτων Λόγου

Φασματογράφημα



https://link.springer.com/chapter/10.1007/978-3-319-66429-3_9

Γράφος



Adjacency Matrix

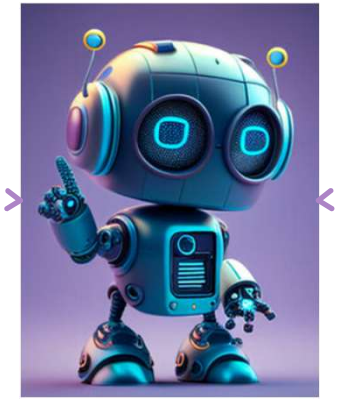
0	1	0	0	0
1	0	1	1	0
0	1	0	1	1
0	1	1	0	1
0	1	1	1	0

https://www.nature.com/articles/s41598-024-52989-2

Διαδικασίες Επεξεργασίας Σημάτων Λόγου

Αποθορυβοποίηση: Βελτίωση ποιότητας ήχου

Αναγνώριση Λόγου: Εξαγωγή Χαρακτηριστικών



Αναγνώριση Ομιλητή: Ταυτοποίηση ομιλητή μέσω της φωνής του

Κωδικοποίηση Λόγου: Συμπίεση σημάτων λόγου-Τηλεπικοινωνίες

https://stock.adobe.com/gr_en/search?k=robot+cartoon&asset_id=559871209

Τι είναι τα χαρακτηριστικά (features)

Υγιείς

- Κανονική-Δομημένη Ροή Λόγου
- Ευρύ εύρος Φωνής
- Υψηλή συχνότητα Ομιλίας

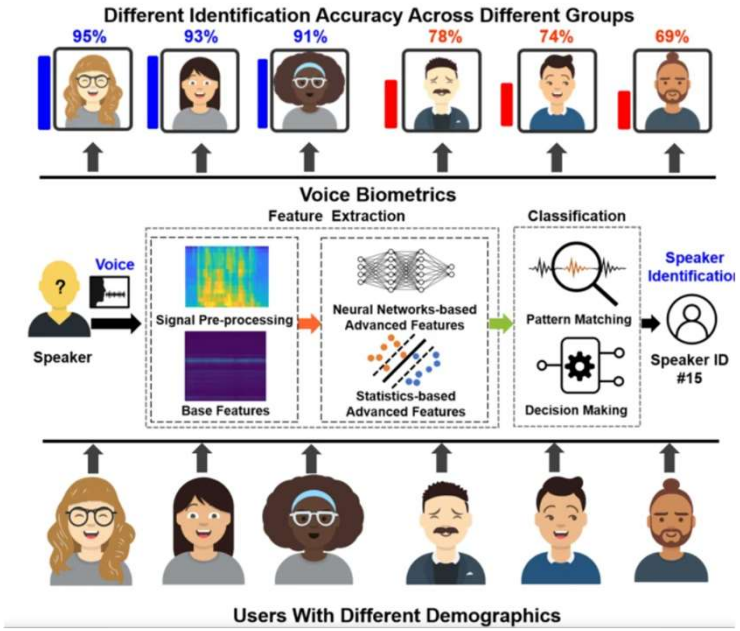
Ασθενείς με Νόσο Parkinson's

- Λόγος με συχνές παύσεις
- Περιορισμένο εύρος Φωνής
- Χαμηλή/Αργή συχνότητα Ομιλίας

Έστω ότι θέλουμε να διαχωρίσουμε υγιείς από ασθενείς με ΝΠ, βάσει της ομιλίας τους.
Θα ψάξουμε να βρούμε διαφορές μεταξύ των δύο ομάδων.
Η ποσοτικοποίηση αυτών των διαφοροποιήσεων → Χαρακτηριστικά

Τι είναι τα χαρακτηριστικά (features)

- Τα features είναι αντιπροσωπευτικές ποσότητες ενός γενικού γνωρίσματος ενός εξεταζόμενου συνόλου
- Ή αλλιώς, είναι **αναπαραστάσεις της χρήσιμης πληροφορίας των σημάτων σε διαφορετικό χώρο ή και πεδία που μας βοηθάνε να ποσοτικοποιούμε τα σήματα φωνής/λόγου**



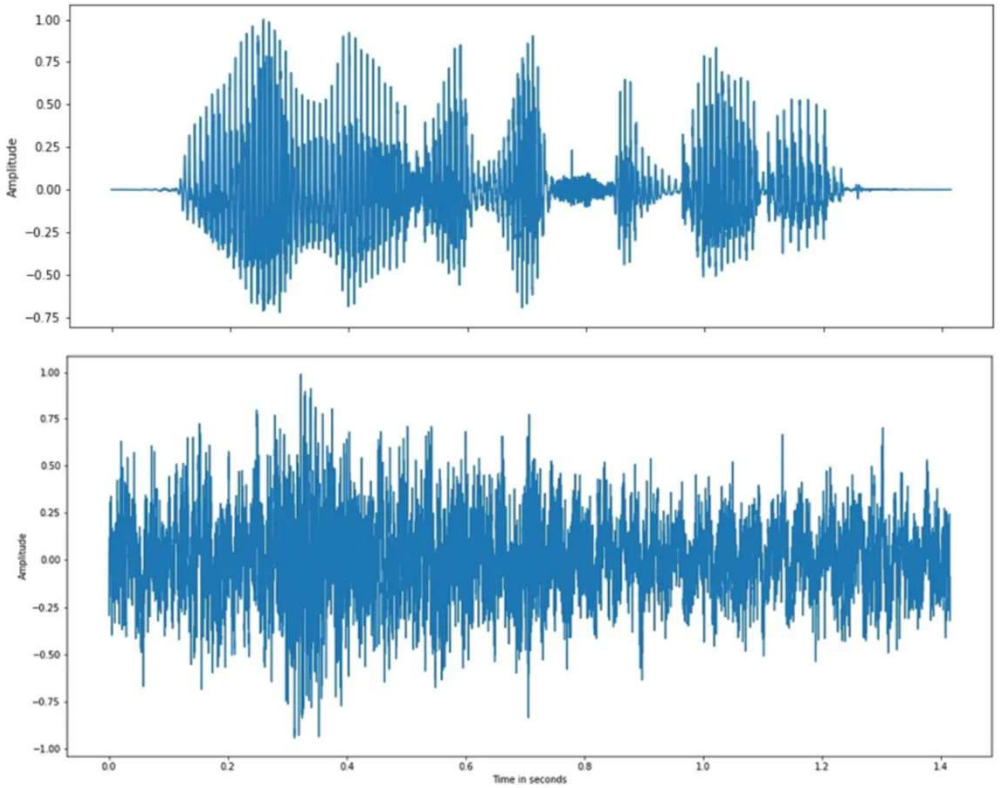
https://www.nature.com/articles/s41598-022-06673-y

Χρησιμότητα των Χαρακτηριστικών Λόγου

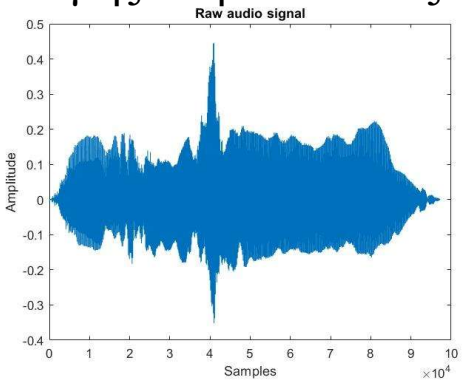
- **Τα σήματα λόγου συνήθως έχουν αρκετή διάρκεια → long duration → many samples**
- Επίσης, ενώ η αυθεντική καταγραφή του λόγου δίνει σημαντικά στοιχεία στον ακροατή (όπως συναισθηματική κατάσταση)
- Οι υπολογιστές δεν αντιλαμβάνονται τα στοιχεία του λόγου, όπως εμείς
- Έτσι: (Συνήθως) **Γίνεται εξαγωγή/υπολογισμός χαρακτηριστικών για να βοηθήσουμε τις υπολογιστικές μεθόδους να κατανοήσουν αυτό που αντιλαμβανόμαστε κι εμείς**

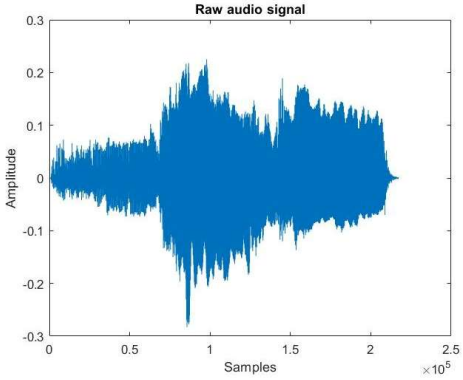
Παραδείγματα Σημάτων Λόγου

Γυναίκα-Άνδρας



Υγιής-Παρκινσονικός





Χαρακτηριστικά του Λόγου

Μπορούμε να εξάγουμε χαρακτηριστικά της φωνής και του λόγου αναπαριστώντας τα σήματα σε διάφορα πεδία ή μέσω διάφορων μαθηματικών διαδικασιών:

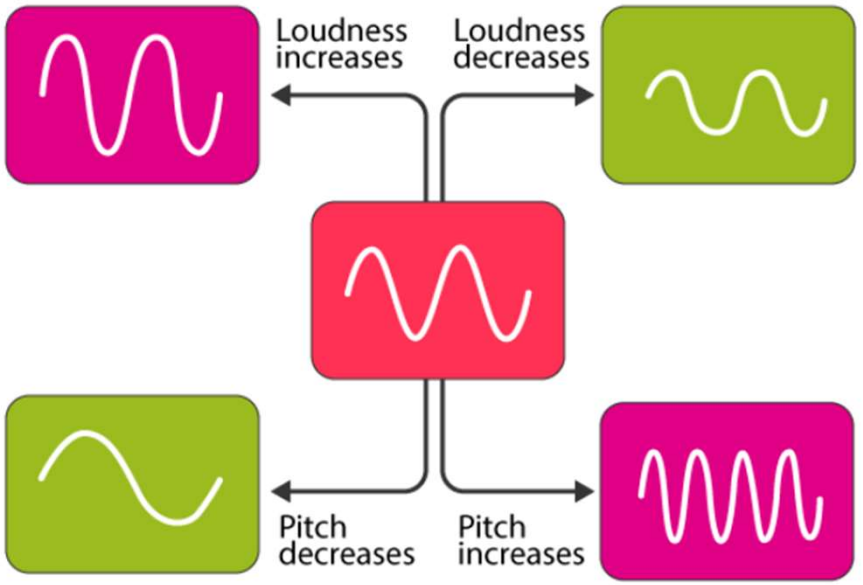
Ενδεικτικά, **οι κύριες επεξεργασίες τους γίνονται α) στο πεδίο του χρόνου, β) στο πεδίο των συχνοτήτων, συμπεριλαμβανομένων και των φασματικών αναπαραστάσεων**

Προσωδία

- Στην γλωσσολογία η προσωδία (prosody) είναι:
  - Ο ρυθμός
  - Ο τόνος
  - Και ο επιτονισμός της ομιλίας
- Μέσω της προσωδίας μπορούμε να κατανοήσουμε ποικίλα χαρακτηριστικά του ομιλητή, όπως:
  - Την συναισθηματική του κατάσταση
  - Την μορφή του εκφωνήματος (όπως ερώτηση)
  - Την παρουσία ειρωνείας ή σαρκασμού

Προσωδικά Χρονικά Χαρακτηριστικά

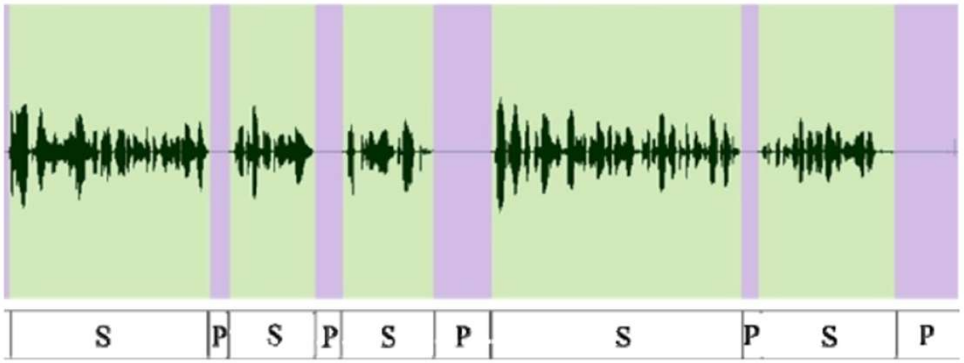
- **Pitch:** ο τόνος της φωνής/η συχνότητα με την οποία ο ακροατής αντιλαμβάνεται τον ομιλητή
- Το μήκος των φωνημάτων: μικρής ή μεγάλης **διάρκειας**
- **Ένταση** της φωνής: σχετίζεται με το πλάτος του σήματος λόγου
- **Ποιότητα** της φωνής



https://byjus.com/physics/loudness-of-sound/

Παραδείγματα Χρονικών Χαρακτηριστικών

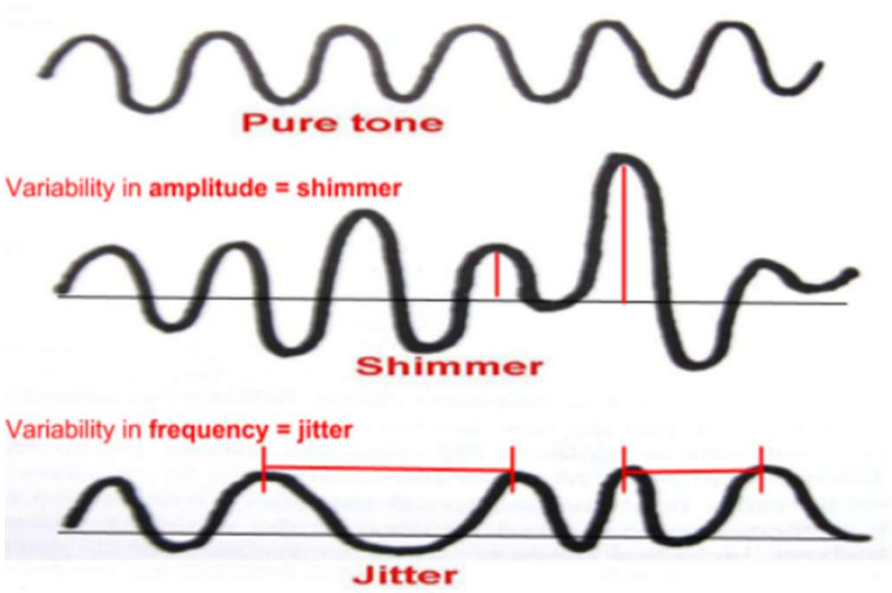
- Ποσοστό των **ισχυρότερων κορυφών** ανά δευτερόλεπτο
- Μέση τιμή και τυπική απόκλιση των **συνεχόμενων τμημάτων** του λόγου (δηλαδή, ποια τμήματα δεν περιέχουν παύσεις)
- Αξιοποίηση των **παύσεων** του λόγου



https://ebrary.net/93224/language_literature/segmentation_speech_signal

Παραδείγματα Χρονικών Χαρακτηριστικών

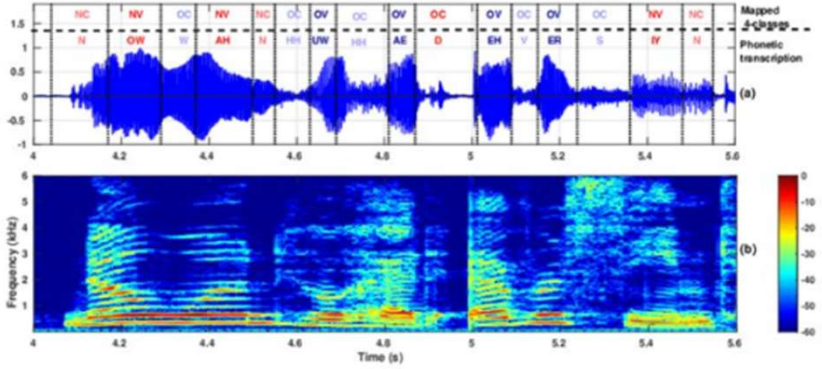
- **Shimmer:** οι διαφορές του μέγιστου πλάτους της ομιλίας
- **Jitter:** διακυμάνσεις της συχνότητας της ομιλίας



https://quizlet.com/790202689/ch67-speech-and-hearing-science-flash-cards/

Χαρακτηριστικά στο πεδίο Συχνοτήτων

- Φάσμα (spectrum): Μετατροπή του σήματος στο πεδίο συχνοτήτων
- Η κύρια συχνότητα μίας φωνής συμβολίζεται ως Fo και δεν είναι formant
- Οι δευτερεύουσες συχνότητες ονομάζονται Formants, που ποσοτικοποιούν ουσιαστικά, την ενέργεια που φέρει ένα ηχητικό σήμα

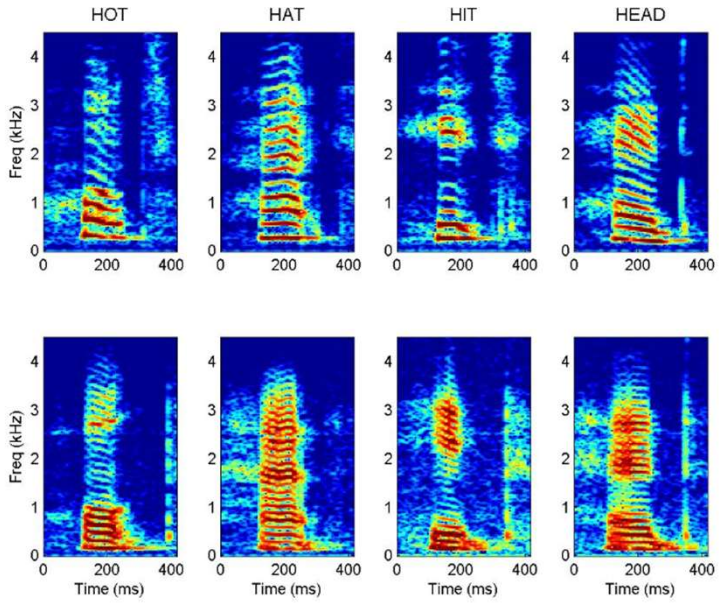


https://www.researchgate.net/publication/341084647_Deep_Learning_Based_Prediction_of_Hypernasality_for_Clinical_Applications/figures?lo=1

Παραδείγματα Φασμάτων

High Pitch Voice

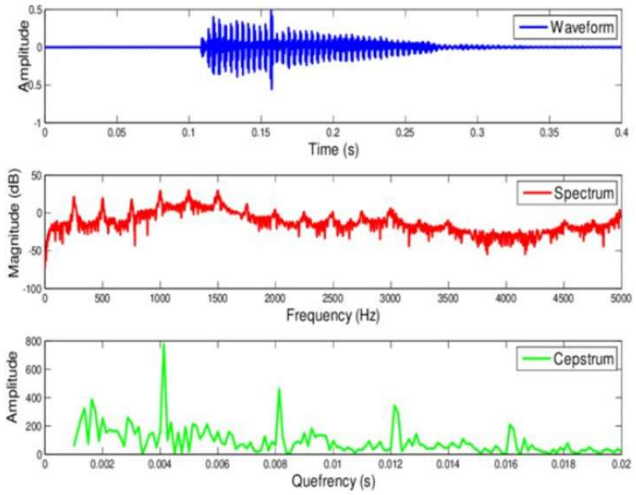
Low Pitch Voice



https://auditoryneuroscience.com/vocalizations-speech/formants-harmonics

Spectrum Vs. Cepstrum

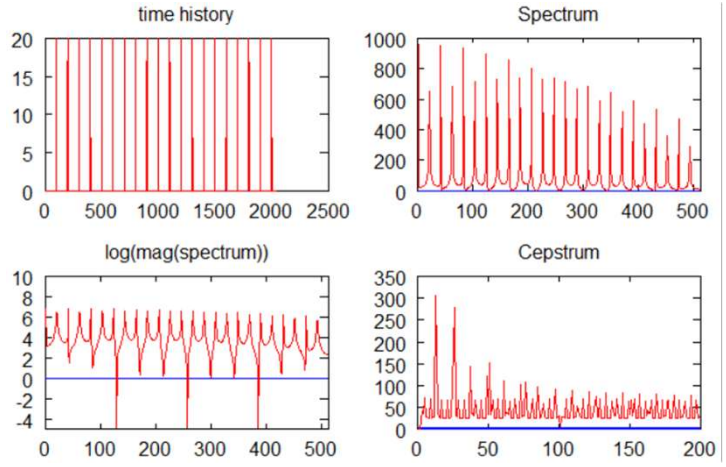
- **Cepstrum:** Αναγραμματισμός του spectrum→Σχετίζεται με αυτό
- Η ανάλυση μέσω cepstrum χρησιμοποιείται ευρέως στην επεξεργασία του λόγου
- Είναι ένα «εργαλείο» μέσω του οποίου μπορούμε να ανιχνεύσουμε **περιοδικές δομές στο φάσμα συχνοτήτων**
- **Βασικό πλεονέκτημα:** Η εξαγωγή cepstrum-based παραμέτρων ποσοτικοποιούν την περιοδικότητα των σημάτων λόγου και έχουν αποδειχθεί πιο αποδοτικές ποσότητες έναντι των spectrum-based



https://www.researchgate.net/publication/330015428_Text_Dependent_Speaker_Recognition_Using_HTK/figures?lo=1

Υπολογισμός του Cepstrum και MFCCs

- α) Μετατροπή σήματος από το πεδίο του **χρόνου** στο πεδίο της **συχνότητας**, β) Υπολογισμός του **λογαρίθμου** του πλάτους του φάσματος (spectrum), γ) Εφαρμογή Inverse Fourier Transform



Πηγή πληροφοριών: wikipedia

Mel-Frequency Cepstrum Coefficients (MFCCs): Από το cepstrum μπορούν να εξαχθούν 14 χαρακτηριστικά, τα οποία **επικεντρώνονται** στον **τρόπο** με τον οποίο οι **άνθρωποι αντιλαμβάνονται τον ήχο**, ειδικά την ομιλία. Μειώνουν τον θόρυβο και τονίζουν σημαντικά χαρακτηριστικά ήχου

eGeMAPS

- Από τα πιο γνωστά σύνολα χαρακτηριστικών του λόγου είναι τα Geneva Minimalistic Acoustic Parameter Set (GeMAPS) καθώς και η extended version του τα extended-GeMAPS (eGeMAPS)
- Περιλαμβάνουν 88 χαρακτηριστικά του λόγου
- Έχουν αποδειχθεί πολύ αποδοτικά σε πληθώρα δεδομένων του λόγου, κυρίως στην αναγνώριση του συναισθήματος
- Περισσότερες πληροφορίες: F. Eyben et al., "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing," in IEEE Transactions on Affective Computing, vol. 7, no. 2, pp. 190-202, 1 April-June 2016

Παράδειγμα Αποδοτικότητας Χαρακτηριστικών του Λόγου

[%]		ResNet						eGeMAPS	
		best		mean					
Dataset	chance	devel	test	devel		test		devel	test
FAU Aibo	20.0	55.5	39.3	**54.4**	(±0.8)	37.2	(±1.3)	46.2	**38.3**
MELD	16.7	24.4	21.9	22.6	(±1.1)	19.8	(±1.4)	**23.7**	**23.1**
DEMoS	14.3	90.8	77.8	**90.3**	(±0.3)	**76.6**	(±0.9)	38.4	43.2
SIMIS	20.0	46.2	34.4	**42.0**	(±2.4)	31.1	(±2.9)	38.5	**31.7**
IEMOCAP	25.0	60.9	56.0	**59.4**	(±0.9)	52.9	(±1.8)	56.2	**58.4**
SmartKom	14.3	26.6	25.6	24.2	(±1.6)	21.6	(±3.5)	**25.2**	**28.6**
SUSAS	25.0	66.5	59.6	**62.0**	(±2.7)	45.6	(±10.0)	56.6	**56.5**
PPMMK	25.0	76.5	48.9	**73.8**	(±2.0)	**44.8**	(±2.8)	48.5	40.4
EU-EV	5.6	16.8	11.7	**15.0**	(±1.5)	9.8	(±1.3)	7.8	**10.9**
EA-ACT	14.3	42.9	42.9	**37.0**	(±3.6)	31.6	(±6.6)	22.9	**57.1**
EA-BMW	33.3	93.0	56.7	**90.0**	(±1.9)	38.8	(±7.9)	79.8	**56.3**
GVEESS	7.7	38.9	27.9	34.0	(±2.5)	**22.5**	(±3.2)	**38.5**	21.2
eNTERFACE	16.7	73.7	80.5	**70.1**	(±1.8)	**75.8**	(±2.5)	44.8	49.1
GEMEP	5.9	52.7	28.1	**47.3**	(±2.9)	25.4	(±1.5)	41.1	**35.4**
CASIA	16.7	38.7	35.0	**35.2**	(±2.5)	**28.8**	(±5.6)	33.7	24.7
EmoFilm	20.0	48.2	49.5	46.3	(±1.3)	47.4	(±1.8)	**46.9**	**54.6**
EmotiW-2014	14.3	32.8	27.4	29.1	(±2.1)	23.7	(±1.7)	**30.5**	**30.6**
CVE	14.3	49.1	53.7	40.8	(±5.3)	45.4	(±4.8)	**58.9**	**60.4**
AD	50.0	89.2	76.3	87.6	(±1.4)	73.3	(±2.3)	**88.5**	**78.7**
EA-WSJ	50.0	100.0	98.1	**100.0**	(±0.0)	97.3	(±0.9)	**100.0**	**100.0**
EMO-DB	14.3	80.8	71.0	**73.4**	(±15.4)	**59.8**	(±11.8)	71.9	48.4
TurkishEmo	25.0	80.7	67.0	**76.1**	(±2.3)	**61.6**	(±2.6)	64.8	56.8
DES	20.0	63.6	38.9	**40.8**	(±16.5)	27.0	(±8.8)	31.0	**46.5**
BES	16.7	85.0	70.0	**77.5**	(±4.5)	**66.5**	(±2.8)	53.3	58.3
ABC	25.0	54.8	48.0	40.0	(±9.5)	32.7	(±9.2)	**47.0**	**54.4**
MES	16.7	30.0	26.7	18.8	(±3.9)	18.2	(±3.0)	**61.7**	**61.7**

Γενικά **τα χαρακτηριστικά του λόγου μπορεί να αποδειχθούν αποδοτικά σε μία εφαρμογή αλλά όχι το ίδιο σε μία άλλη.** Κυριότερες αιτίες:

1. Διαφορετικές γλώσσες
2. Διαφορετικοί γραμματικοί κανόνες
3. Διαφορετικές κουλτούρες
4. Διαφορετικότητα μεταξύ των ομιλητών

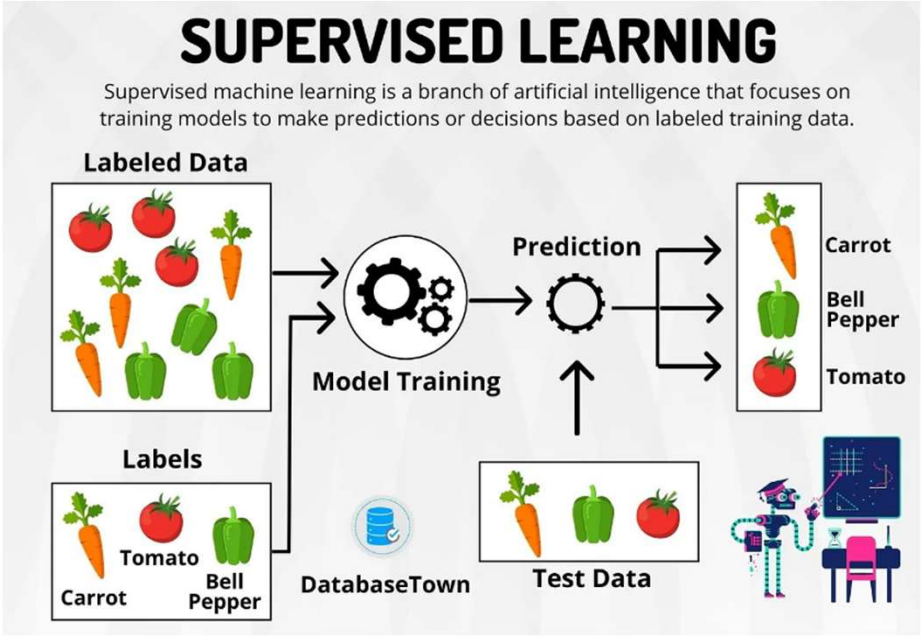
Μηχανική Μάθηση

- Η Μηχανική Μάθηση (machine learning) είναι το πεδίο της επιστήμης υπολογιστών που μελετά την υλοποίηση αλγορίθμων, οι οποίοι **μαθαίνουν** χωρίς να έχουν καθοριστεί συγκεκριμένοι κανόνες
- Ουσιαστικά, αυτοί οι αλγόριθμοι αυτοί **χρησιμοποιούν δεδομένα με στόχο να ανακαλύψουν μοτίβα και σχέσεις μεταξύ τους**, για να κάνουν προβλέψεις ή να πάρουν αποφάσεις

Τεχνικές Μηχανικής Μάθησης

- **Επιβλεπόμενη Μάθηση (supervised learning):** Στην τεχνική αυτή το πρόγραμμα εκπαιδεύεται για να κατανοήσει την σχέση μεταξύ των δεδομένων που του δίνονται και ενός επιθυμητού αποτελέσματος-Δίνονται οι ταυτότητες των δεδομένων (labels)
- **Μη-επιβλεπόμενη Μάθηση (unsupervised learning):** Στην τεχνική αυτή δίνονται δεδομένα στον αλγόριθμο και προσπαθεί να ανακαλύψει πιθανά μοτίβα μεταξύ τους, συνήθως δημιουργώντας ομάδες (clusters)-Δεν δίνονται οι πραγματικές ταυτότητες των δεδομένων (no labels)

Επιβλεπόμενη Μάθηση-Supervised Learning



https://medium.com/@gerzson.boros/a-simple-introduction-into-supervised-learning-dcce83ee3ada

Διαδικασία Supervised Learning

- Η **βασική λογική** ακολουθεί την εξής διαδικασία:
- Από τα διαθέσιμα δεδομένα εξάγουμε την χρήσιμη πληροφορία, δημιουργώντας έτσι τον πίνακα **χαρακτηριστικών-features**
- Χωρίζουμε τα διαθέσιμα επεξεργασμένα δεδομένα (δηλαδή τα features) σε training set (80%) και testing set (20%)→ **Ανεξάρτητα μεταξύ τους σύνολα**
- **Training procedure:** Επιλέγουμε τον αλγόριθμο εκπαίδευσης και του δίνουμε ως είσοδο το training set & τις αντίστοιχες ταυτότητες των δεδομένων (labels)

Διαδικασία Supervised Learning

- **Testing procedure:** Αφού εκπαιδευτεί ο αλγόριθμος, περνάμε στην **πρόβλεψη** του testing set, δηλαδή εξετάζουμε αν μπορεί να αναγνωρίσει τις ταυτότητες των δεδομένων (υπό την μορφή επίσης features) που **δεν έχει ξαναδεί**
- Κάποιες φορές χρειάζεται να χρησιμοποιήσουμε πχ. 10% των δεδομένων για cross-validation, δηλαδή για να βρούμε ποιος αλγόριθμος αποδίδει καλύτερα στα δεδομένα ή βέλτιστες τιμές παραμέτρων τους
- Εξίσου σημαντική διαδικασία είναι το **hyperparameter tuning:** Χρησιμοποιώντας ένα υποσύνολο των δεδομένων εξετάζουμε τις τιμές των υπερ-παραμέτρων που δίνουν την βέλτιστη απόδοση

Πλεονεκτήματα & Μειονεκτήματα

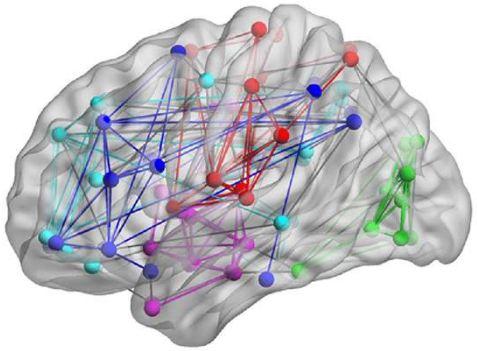
Πλεονεκτήματα

- Υψηλή Ακρίβεια
- Προσαρμοστικότητα
- Αυτοματοποίηση
- Ανθεκτικότητα στο θόρυβο
- Μπορούν να αποδώσουν σε (τεράστιο) όγκο δεδομένων

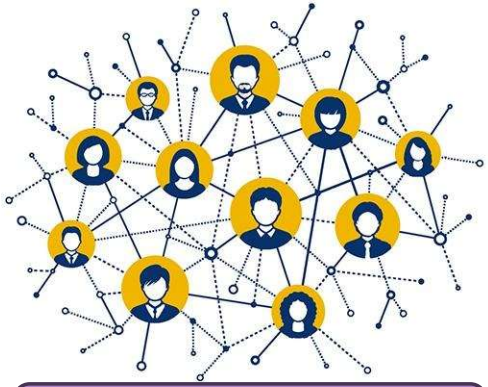
Μειονεκτήματα

- Ανάγκη για μεγάλα σύνολα δεδομένων
- Υψηλό υπολογιστικό κόστος
- Τα μοντέλα μπορεί να αποδίδουν καλύτερα σε ένα σύνολο ομιλητών από ότι σε κάποιο άλλο
- Δυσκολία ερμηνευσιμότητας χαρακτηριστικών

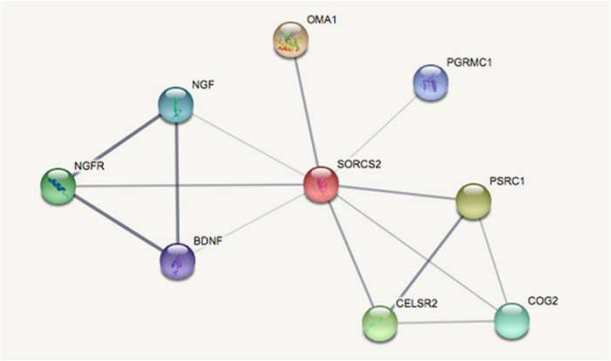
Θεωρία των Γράφων



Neuroscience



Social Networks



Molecular Biology

Θεωρία των Γράφων

- Ο γράφος αποτελείται από κόμβους και ακμές
- Κάθε κόμβος θεωρείται ένα «αντικείμενο»
- Ενώ, κάθε ακμή συνδέει 2 κόμβους που έχουν κάποιο «κοινό» χαρακτηριστικό

Πλεονεκτήματα & Μειονεκτήματα Γράφων

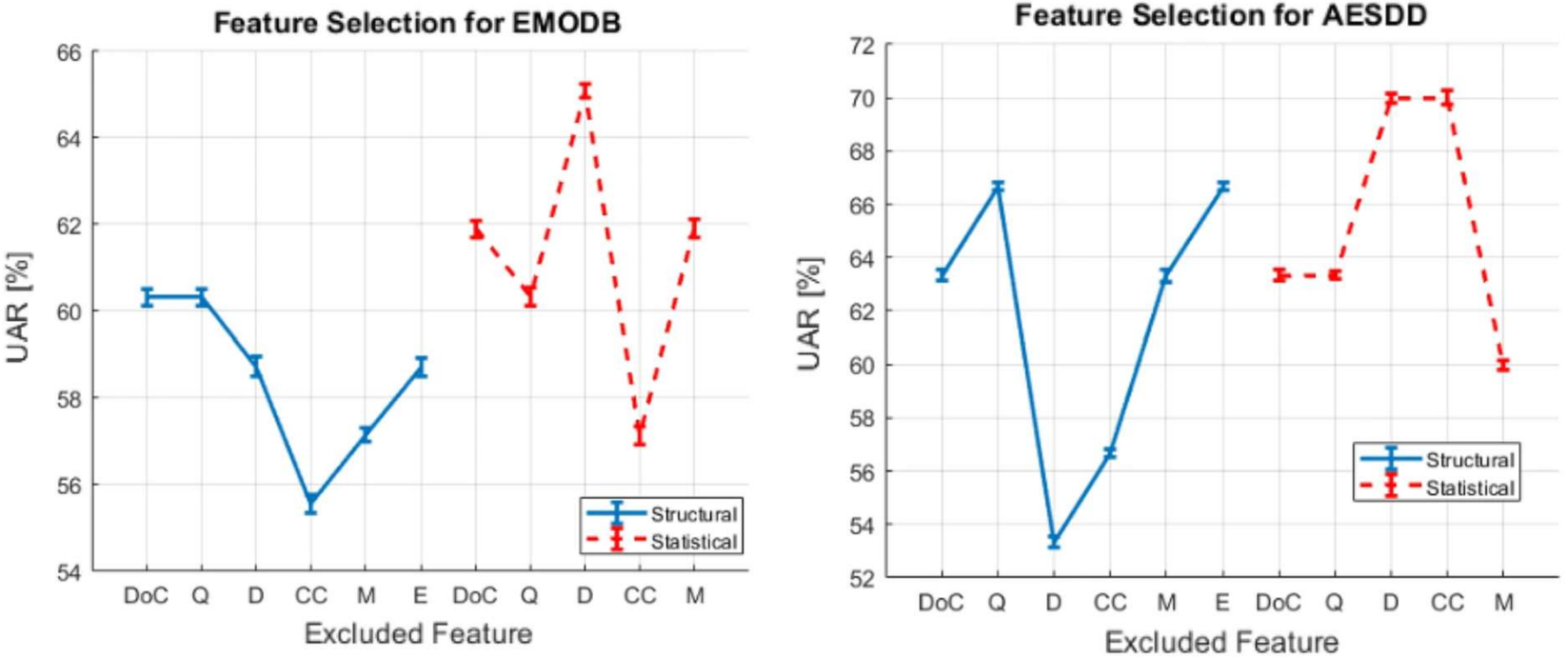
Πλεονεκτήματα

- Οι γράφοι εκμεταλλεύονται τις επιμέρους συσχετίσεις μεταξύ 2 αντικειμένων
- Έχουν αποδειχθεί πιο αποδοτικοί από την ατομική επεξεργασία των δεδομένων
- Μπορούν να μας δώσουν στοιχεία ενός συστήματος που δεν θα βρίσκαμε διαφορετικά
- Εξαγωγή γραφο-χαρακτηριστικών

Μειονεκτήματα

- Για όγκο δεδομένων → Αυξάνεται η υπολογιστική πολυπλοκότητα
- Εύρεση «καλύτερων» τιμών των υπερ-παραμέτρων που χρησιμοποιούνται σε κάθε εφαρμογή
- Δύσκολη η ερμηνεία των γραφο-χαρακτηριστικών σε κάποιες εφαρμογές

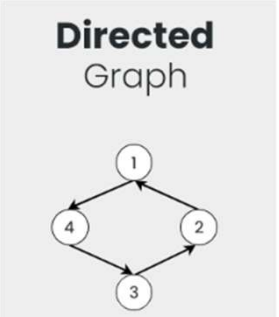
Παράδειγμα Χαρακτηριστικών σε Επεξεργασία Ομιλίας

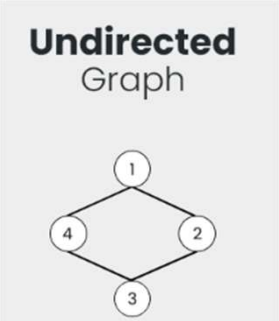


https://www.nature.com/articles/s41598-024-52989-2

Ορισμός του Γράφου

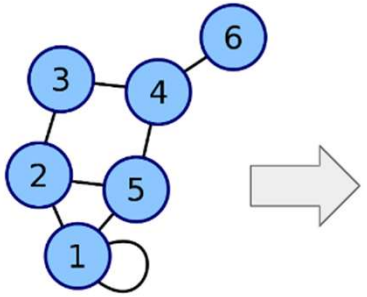
- Ένας γράφος $G= (V,E)$ είναι ένα σύνολο από κόμβους V και συνδέσεις/ακμές E
- Ο αριθμός των ακμών είναι υποσύνολο των κόμβων ανά ζεύγη
- Ο μέγιστος αριθμός ακμών είναι: $N = |V||V|$ σε έναν πλήρως συνδεδεμένο γράφο
- Οι ακμές μπορούν να είναι κατευθυνόμενες ή μη





Βασικά στοιχεία των Γράφων

- Η βασικότερη ποσότητα ενός γράφου, ουσιαστικά, είναι ο πίνακας που τον ορίζει, ο επονομαζόμενος πίνακας γειτνίασης ή αλλιώς ***adjacency matrix*** *(A)*
- Μπορεί να είναι δυαδικός (**binary**-{0,1}) ή με βάρη (**weighted**-[0,1])
- Το μέγεθος του ισούται με $|V|x|V|$. δηλαδή είναι ένας **τετραγωνικός** πίνακας μεγέθους ανάλογο των κόμβων
- Για μη-κατευθυνόμενους γράφους, ο *A* συνήθως είναι συμμετρικός



1	1	0	0	1	0
1	0	1	0	1	0
0	1	0	1	0	0
0	0	1	0	1	1
1	1	0	1	0	0
0	0	0	1	0	0

Χαρακτηριστικά (features) των Γράφων

- Degree of Connectivity (DoC)
- Density (D)
- Modularity (Q)
- Clustering Coefficient (CC)
- Shortest Path Length (L)
- Small-world Phenomenon-Coefficient

Degree of Connectivity (DoC)

- Δοθέντος του πίνακα γειτνίασης, ο βαθμός συνδεσιμότητας υπολογίζεται ως εξής:

$$DoC = \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{A}_{i,j}$$

- Ο οποίος μας δείχνει πόσες ακμές *(i,j)* ξεκινούν από τον κόμβο *i*
- **Όσο μεγαλύτερες τιμές *DoC* έχουμε, τόσο πυκνότερος είναι ο γράφος**

Density (D)

- Δοθέντος του πίνακα γειτνίασης, η πυκνότητα υπολογίζεται ως εξής:

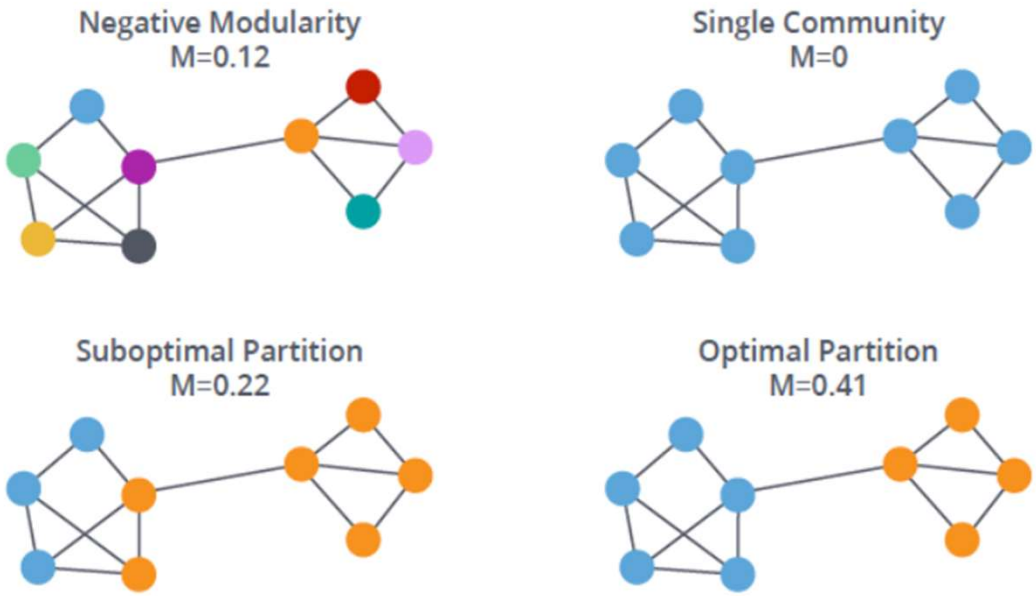
$$D = \frac{2|E|}{N(N-1)}$$

- Όπου ως $|E|$ συμβολίζουμε τον αριθμό των ακμών, ενώ ως N τον αριθμό των κόμβων
- Είναι μία ποσότητα που μας δείχνει πόσο αραιός ή πυκνός είναι ένας γράφος

Modularity (Q)

- Περιγράφει την τάση ενός γράφου να διασπάται σε υπο-γράφους, τα λεγόμενα *modules*
- Εντός του module η ισχύς της συνδεσιμότητας είναι υψηλή
- Ενώ, μεταξύ δύο modules η ισχύς συνδεσιμότητας είναι χαμηλή
- Μία υψηλή τιμή του Q υποδεικνύει στενά συνδεδεμένο γράφο

Modularity (Q)-Examples



Clustering Coefficient (CC)

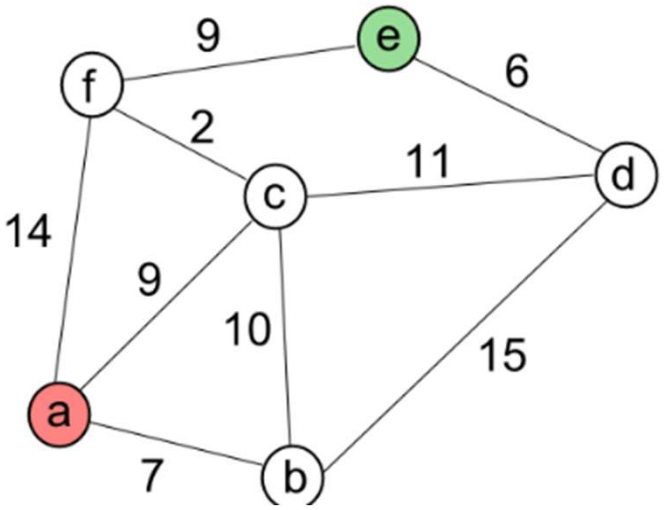
- Δείχνει την τάση του γράφου να δημιουργεί κλίκες/clusters
- Δοθέντος του πίνακα γειτνίασης, υπολογίζεται ως εξής:

$$CC_i = \frac{1}{k_i(k_i - 1)} \sum_{j=1,l=1}^{N} \mathbf{A}_{i,j}\mathbf{A}_{j,l}\mathbf{A}_{l,i}$$

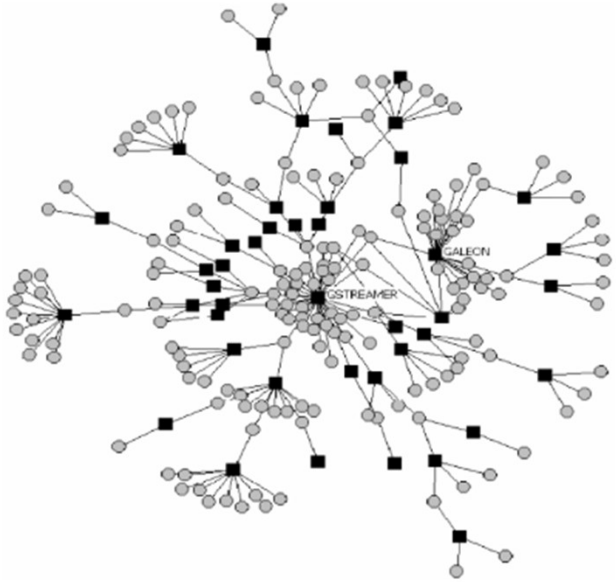
- Όπου: $k_i = \sum_{j=1}^{N} \mathbf{A}_{i,j}$
- Ενώ ο συνολικός *CC = mean(CCi)*

Shortest Path Length (L)

- Εκφράζει το ελάχιστο δυνατό μονοπάτι που πρέπει να «διασχίσουμε» ώστε να μεταφερθούμε από τον κόμβο *i* προς τον κόμβο *j* (χωρίς επανάληψη)
- Ποιο είναι το κόστος μεταφοράς από τον κόμβο a→e?

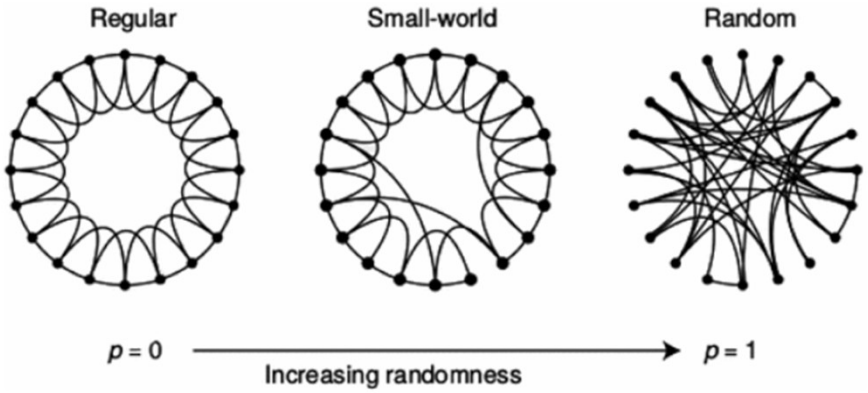


Small-World Phenomenon



Η τάση ενός γράφου να δημιουργεί «hubs»

Small-world Coefficient (S)



- Αρχικά χρειάζεται να δημιουργήσουμε «τυχαίους» γράφους/random graphs βάσει των αλγόριθμων:
  - Erdos-Renyi
  - Watts-Strogatz
- Αν S>1 υπάρνει small-world phenomenon

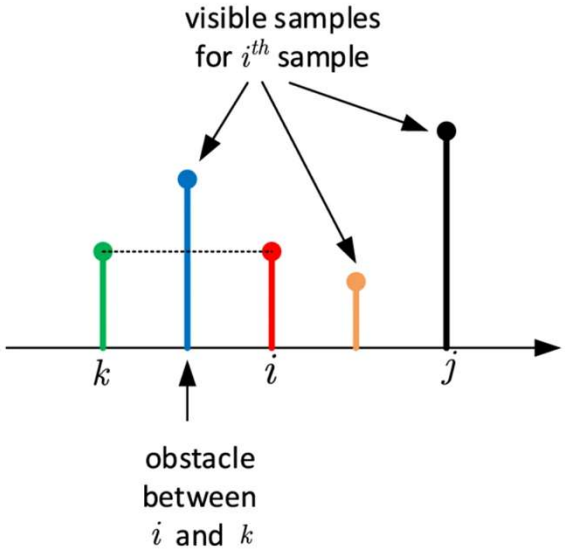
$$S_i = \frac{\frac{CC_i}{CC_{i,random}}}{\frac{L_i}{L_{i,random}}}$$

Δημιουργία Adjacency Matrix: Μέθοδος 1

- Ο πιο συνηθισμένος τρόπος δημιουργίας του πίνακα γειτνίασης είναι μέσω του στατιστικού μέτρου Pearson's correlation
- Όμως έχει αποδειχθεί ότι και η γεωμετρική πληροφορία μίας ποσότητας μπορεί να δώσει σημαντικά στοιχεία στην ανάλυσή μας!
- Συγκεκριμένα, υπάρχει η θεωρία της «ορατότητας των γράφων»/visibility graph theory

Visibility Graph Theory

- Βάσει αυτής της θεωρίας εκμεταλλευόμαστε την γεωμετρική σχέση μεταξύ των δειγμάτων μίας χρονοσειράς
- Απαραίτητη προϋπόθεση της χρονοσειράς είναι να αποτελείται από *θετικές* τιμές
  - Mean value?
  - **Standard deviation?**
  - Absolute distance?
  - Other ideas?
- Ο εξαγόμενος πίνακας γειτνίασης μπορεί να είναι binary/weighted

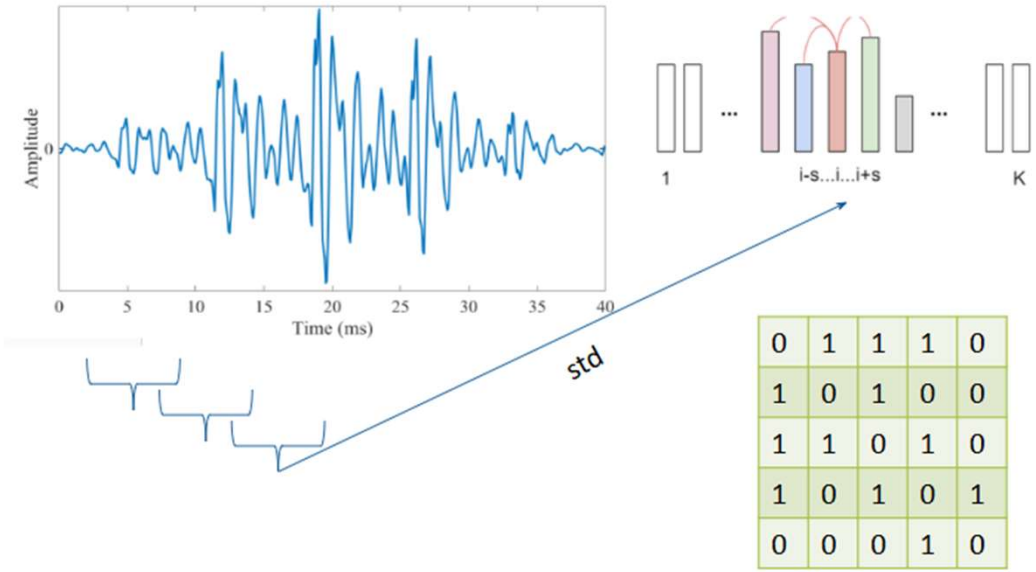


Adjacency Matrix

0	1	0	0	0
1	0	1	1	0
0	1	0	1	1
0	1	1	0	1
0	1	1	1	0

Visibility Graph Theory based on Std

1. Παίρνουμε ένα σήμα λόγου και το «σπάμε» σε επικαλυπτόμενα τμήματα
2. Από κάθε τμήμα υπολογίζουμε την std των τιμών των δειγμάτων του
3. Δημιουργείται μία νέα χρονοσειρά, μικρότερου μήκους
4. Εφαρμόζουμε visibility graph theory
5. Εξάγεται ο πίνακας γειτνίασης
6. Μπορούμε να εξάγουμε τα γραφο-χαρακτηριστικά



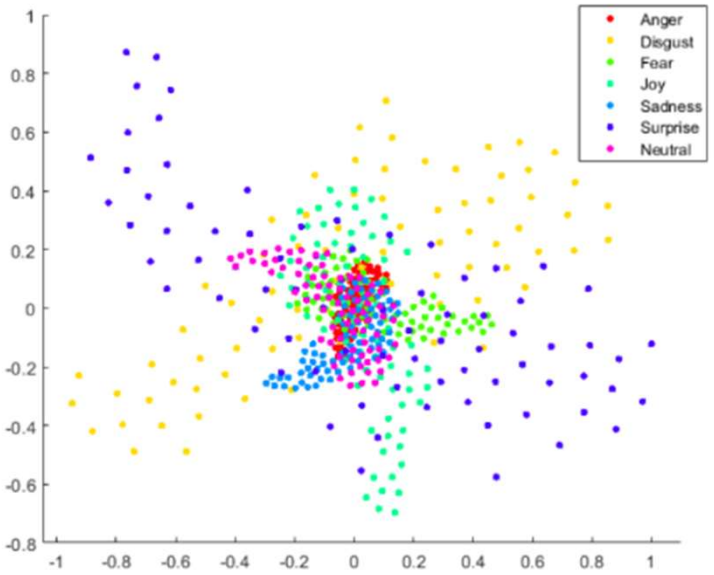
0	1	1	1	0
1	0	1	0	0
1	1	0	1	0
1	0	1	0	1
0	0	0	1	0

Παράδειγμα Εφαρμογής Γράφων σε Ομιλία

	SAVEE	EMOVO
eGeMAPS	40.8	37.4
Graph-based Features	**70**	**98**

Unweighted Average Recall-**UAR** (%) Classification Accuracy, multi-class SVM, Leave-One-Speaker-Out (LOSO) Method

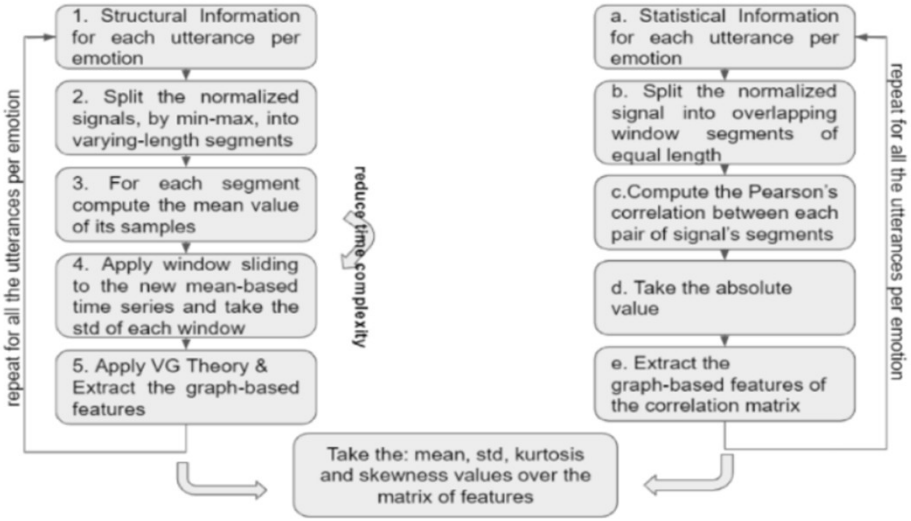
- **SAVEE**: British Speech Database of 4 male actors and 480 utterance expressing 7 emotions, i.e., anger, disgust, fear, happiness, sadness, surprise and neutral.
- **EMOVO**: Italian Speech Database of 6 actors and 7 emotional states, consisting of 588 utterances



Αναγνώριση Συναισθηματικής Κατάστασης μέσω Γραφοθεωρίας

Δημιουργία Adjacency Matrix: Μέθοδος 2

- Χρήση Visibility Graph Theory για δημιουργία πίνακα γειτνίασης γεωμετρικής πληροφορίας σημάτων λόγου
- Χρήση Pearson's Correlation για δημιουργία πίνακα γειτνίασης στατιστικής πληροφορίας σημάτων λόγου
- Εξαγωγή γραφο-χαρακτηριστικών
- Χρήση στατιστικών μετρικών



Παράδειγμα Εφαρμογής Γράφων σε Ομιλία

Classification LOSOCV Accuracy [%]

Methodology	Metric	EMODB	DEMoS	AESDD
eGeMAPS	UAR	48.4[9]	41.19	57.1[9]
Mel-Spec. and ResNet	UAR	59.8[9]	72.76	37[9]
Graph-based	Speaker-based	77.8	79.1	70

Αναγνώριση Συναισθηματικής Κατάστασης μέσω Γραφοθεωρίας

Διαχωρισμός Υγιών-Ασθενών ΝΠ βάσει Ομιλίας

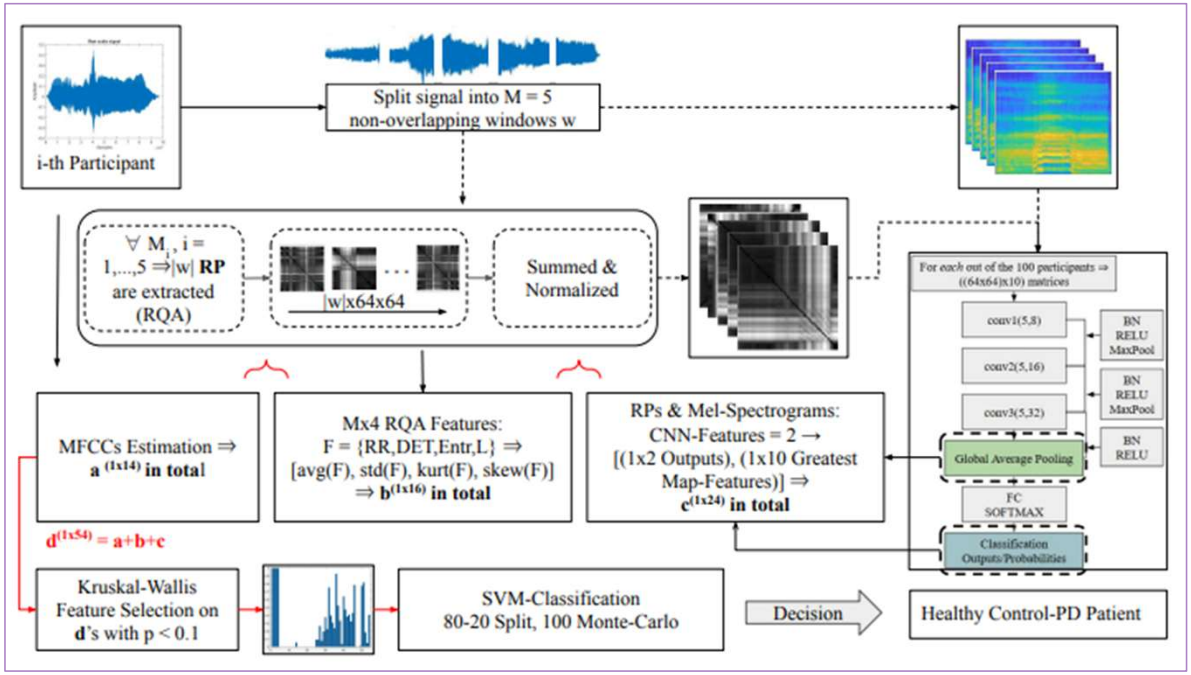


TABLE I
CLASSIFICATION ACCURACY (STD) OF PD PATIENTS AND CONTROL SUBJECTS DIFFERENTIATION-SPANISH DATASET

Task	Exercise	Validation	Testing
Vowels	/a/	0.88 (±0.03)	0.90 (±0.06)
	/e/	0.86 (±0.03)	0.88 (±0.08)
	/i/	0.89 (±0.03)	0.89 (±0.05)
	/o/	0.94 (±0.02)	0.93 (±0.04)
	/u/	0.96 (±0.02)	0.97 (±0.03)
	Averaged Features	0.99(±0.01)	0.99 (±0.01)
Syllables	/ka/	0.94 (±0.03)	0.94 (±0.05)
	/pa/	0.96 (±0.01)	0.97 (±0.04)
	/ta/	0.91 (±0.02)	0.94 (±0.05)
	Averaged Features	0.98 (±0.01)	0.99 (±0.02)
Words	coco	0.91 (±0.03)	0.94 (±0.05)
	pato	0.93 (±0.03)	0.94 (±0.05)
	petaka	0.92 (±0.02)	0.92 (±0.06)
	Averaged Features	0.97 (±0.02)	0.97 (±0.04)
Sentences	laura	0.94 (±0.02)	0.95 (±0.05)
	luisa	0.89 (±0.03)	0.91 (±0.07)
	rosita	0.94 (±0.02)	0.94 (±0.05)
	Averaged Features	0.94 (±0.02)	0.96 (±0.04)

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=11250588