Towards the use of speech as a health biomarker

PhD Advanced Course
Molecular Biomarkers and Technologies
18 September, 2020

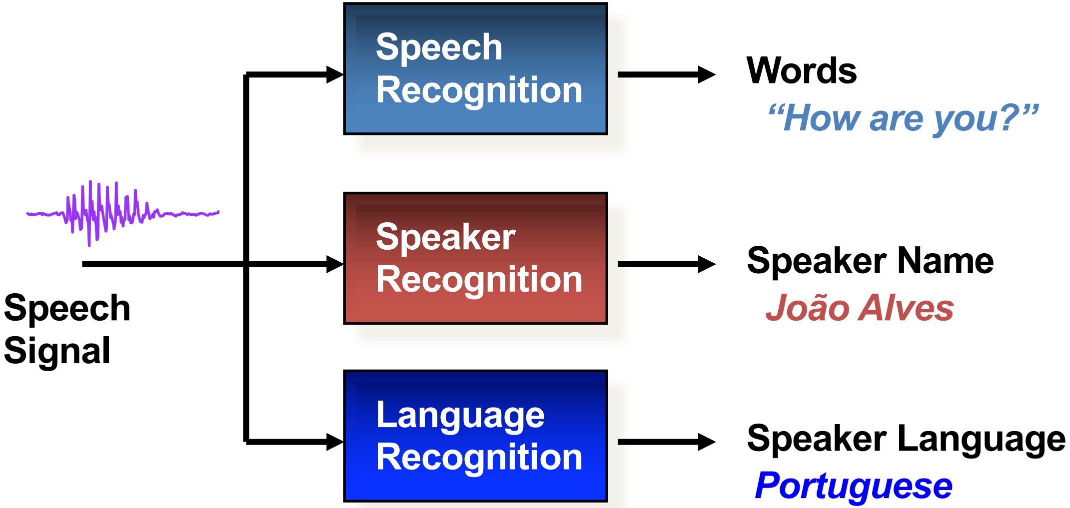
Alberto Abad
alberto.abad@inesc-id.pt

With contributions from: Anna Pompili, Catarina Botelho, Francisco Oliveira, Rubén Solera-Ureña, Isabel Trancoso

inesc id lisboa

TÉCNICO LISBOA

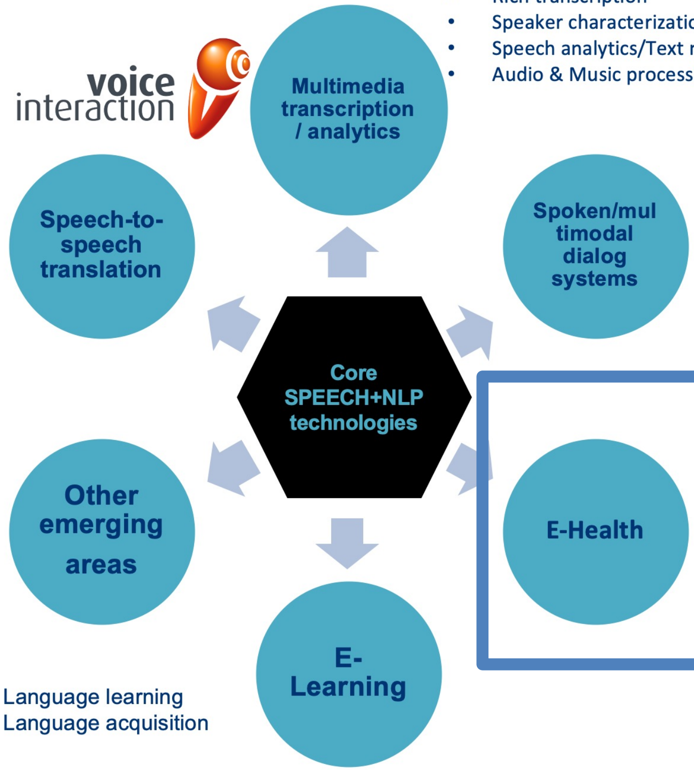
Human Language Technologies



Speech processing: Speech coding, Speech enhancement, Audio segmentation, Text-to-speech synthesis, Automatic speech recognition, Speaker and language identification; Other speech pattern classification tasks

Text processing: Morphological analysis, Syntactic analysis, Semantic analysis, Discourse analysis, Named entity extraction, NL Generation, Information retrieval, Summarization, Question answering, Machine translation, Text analytics, Recommendation

Core research & application areas @ INESC-ID



Multimedia transcription / analytics
- Rich transcription
- Speaker characterization
- Speech analytics/Text mining
- Audio & Music processing

Spoken/multimodal dialog systems
- Smart homes
- Cultural heritage
- Chat bots

E-Health
- Virtual therapists
- Monitoring neurodeg.
- Detection of pathological speech, depression, etc.
- Active ageing
- IE/IR for medical texts
- Sign language

E-Learning
- Language learning
- Language acquisition

Other emerging areas
- Security/Privacy
- Cyberbullying
- Brain interfaces

Speech-to-speech translation
- MT4 Microblogs
- Voice conversion
- Quality evaluation

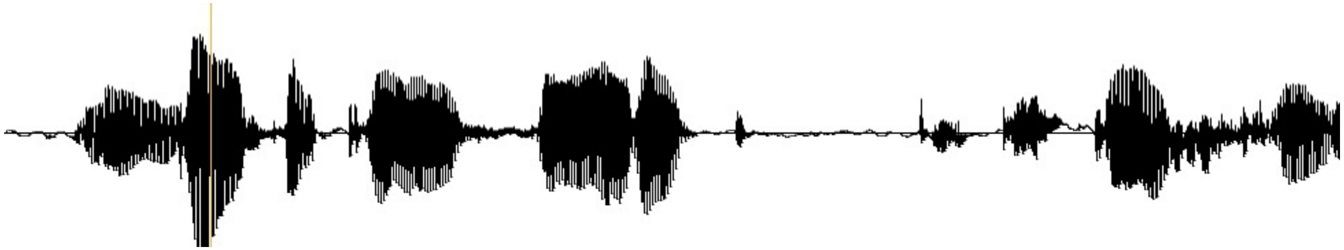
Outline

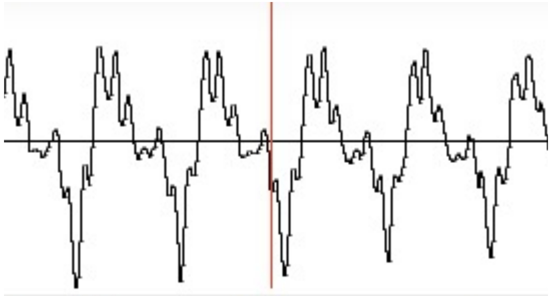
- Introduction to speech processing
- Three generations of HLT applications in e-health @INESC-ID:
  - 1G: Augmentative communication, assistive technologies and e-inclusion
  - 2G: Diagnosis and treatment of speech and language disorders
  - 3G: Speech as a health biomarker for speech-affecting diseases
- Challenges and open questions

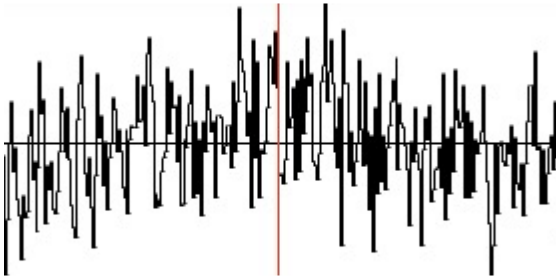
Outline

- Introduction to speech processing
- Three generations of HLT applications in e-health @INESC-ID:
  - 1G: Augmentative communication, assistive technologies and e-inclusion
  - 2G: Diagnosis and treatment of speech and language disorders
  - 3G: Speech as a health biomarker for speech-affecting diseases
- Challenges and open questions

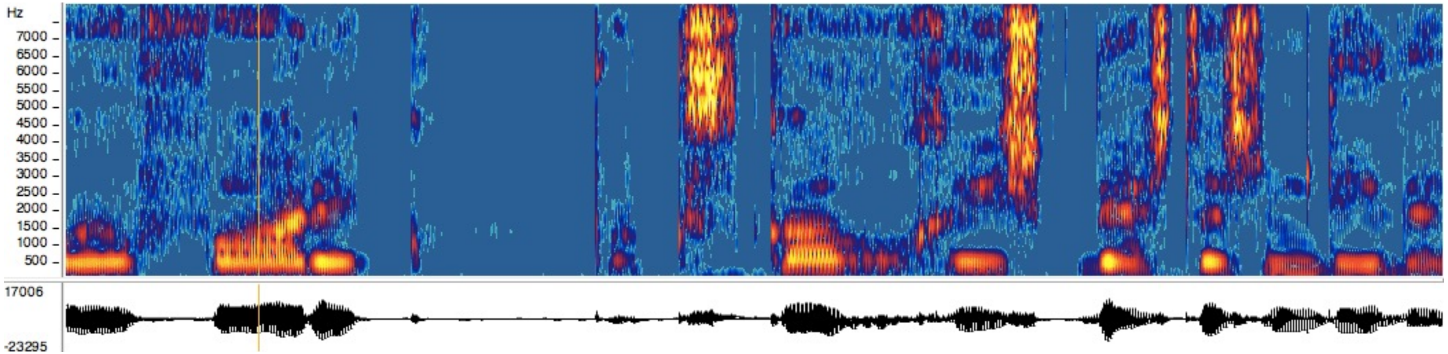
The speech signal in the time domain

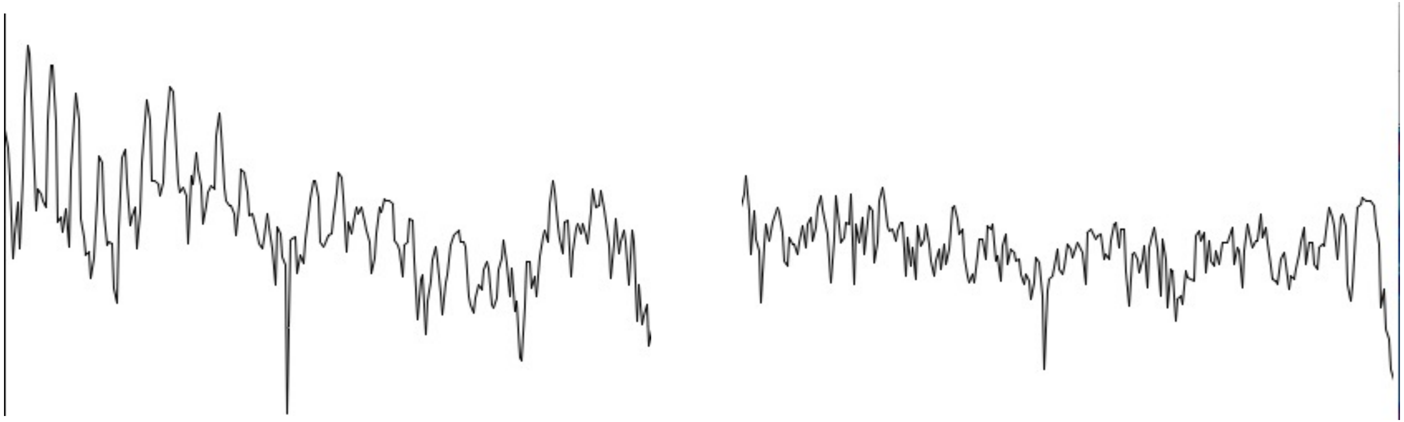




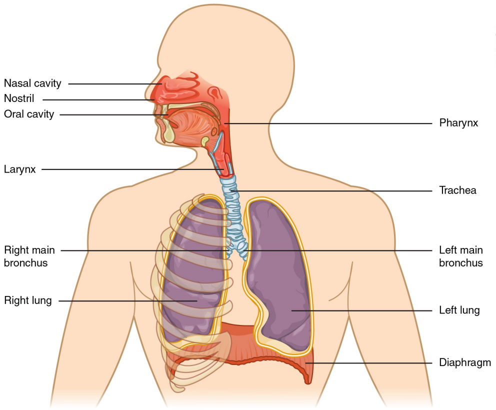


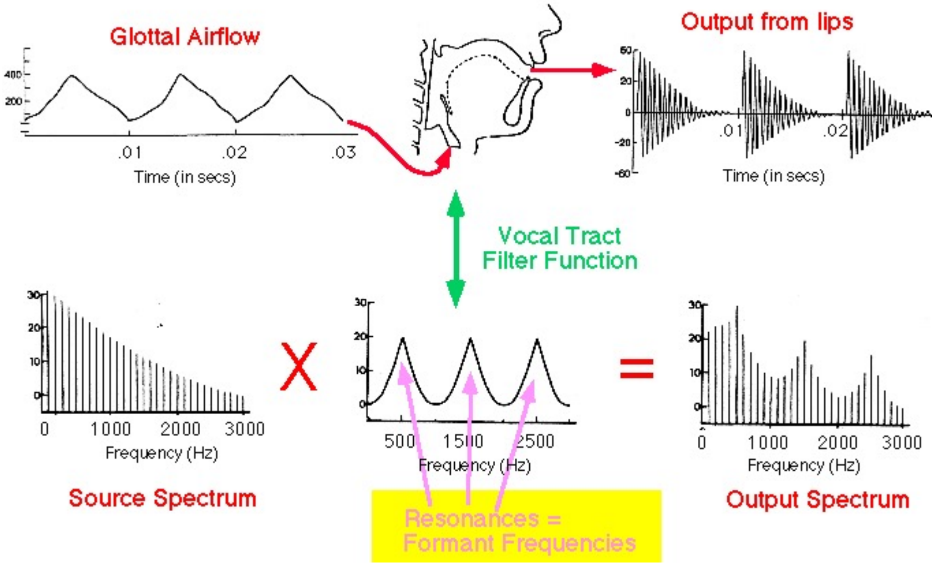
Speech signal time / frequency representation





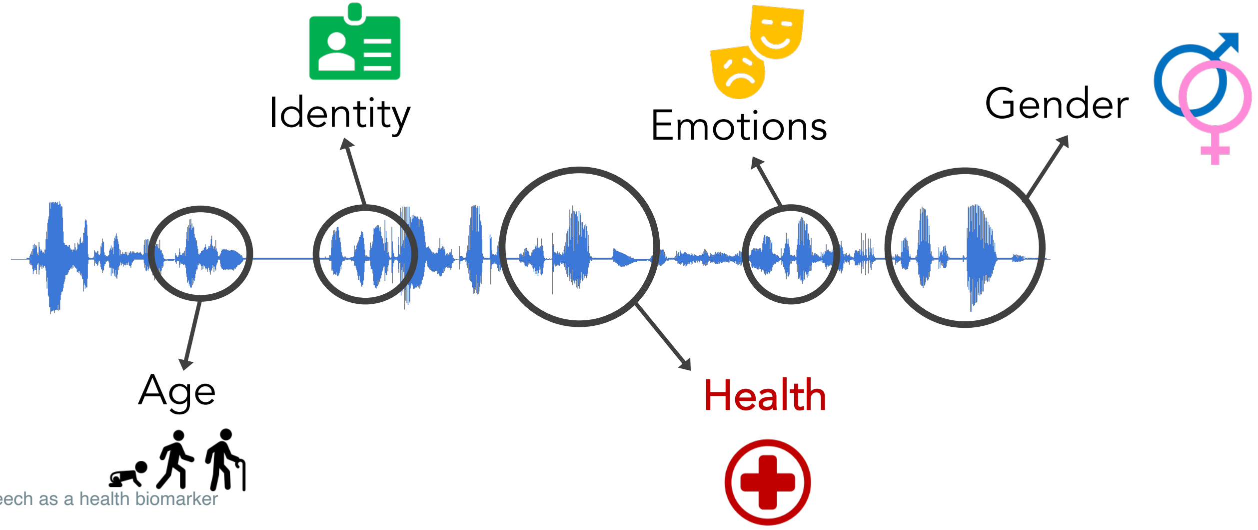
Speech signal: Physiology & Source/filter model





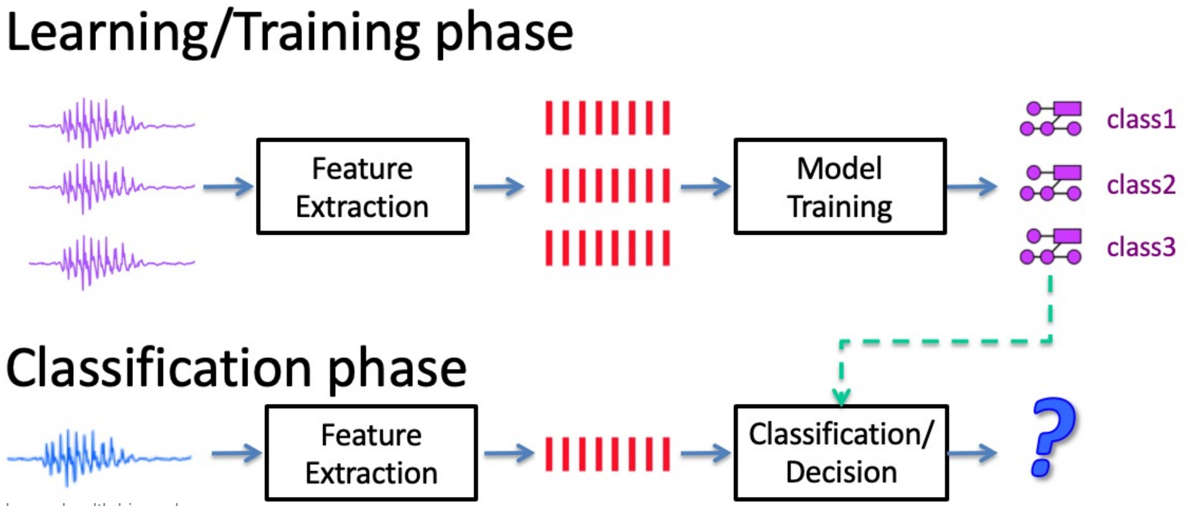
Speech as a carrier of information

- Speech carries a lot of information:
  - Of course information related to the message (LINGUISTIC?)...
  - ... but also, speaker traits (NON-LINGUISTIC/PARA-LINGUISTIC?):
    - Gender; Age; Language/accent; ID; Personality; Education; Intoxication; Sleepiness; Friendliness; Mood; Physical Stress; Cognitive Load; Emotion; Pathologies?



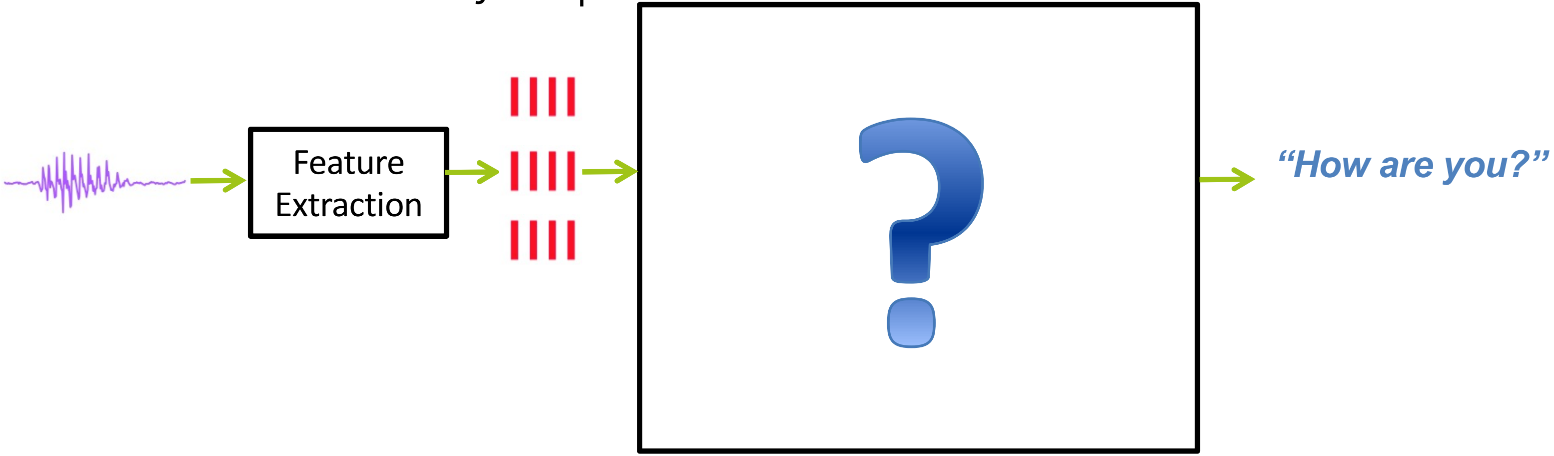
Speech classification (a.k.a. machine learning, IA, etc.)

- **Objective:** To convert a speech input sequence into a class label (or sequence of labels):
- The common blocks of any ML classification task are the front-end/feature extraction and the back-end/classification:
  - The classifier module is "learnt" using **data** during the training phase and used to classify new unseen data during test



A special case: Automatic Speech Recognition (ASR)

- **Goal** Given a sequence of observations determine which is the most likely sequence of words



  - Already decades of research on ASR (and other SLT related topics) → **Very challenging!!!**
  - **Related sub-tasks:** Isolated ASR, Continuous ASR, KWS, LVCSR, STD/Search on Speech, etc.

Outline

- Introduction to speech processing
- **Three generations of HLT applications in e-health @INESC-ID:**
  - 1G: Augmentative communication, assistive technologies and e-inclusion
  - 2G: Diagnosis and treatment of speech and language disorders
  - 3G: Speech as a health biomarker for speech-affecting diseases
- Challenges and open questions

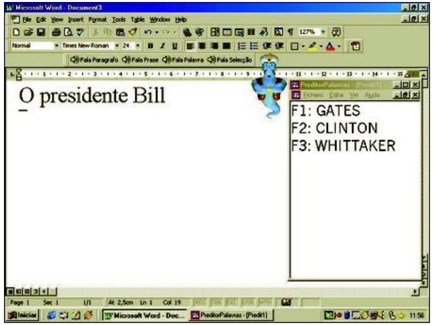
SLT applications in e-health @ INESC-ID
1G: Augmentative, assistive and e-inclusion applications

- Hearing impairment
  - Subtitling, meta-information, segmentation, events, etc.

- Visual impairment
  - Text-to-speech for content access, spoken books

SLT applications in e-health @ INESC-ID
1G: Augmentative, assistive and e-inclusion applications

- Cerebral palsy
  - Text-to-speech with virtual keyboards and word prediction (Eugenio)
- Physical impairments
  - Control, home-automation, etc.

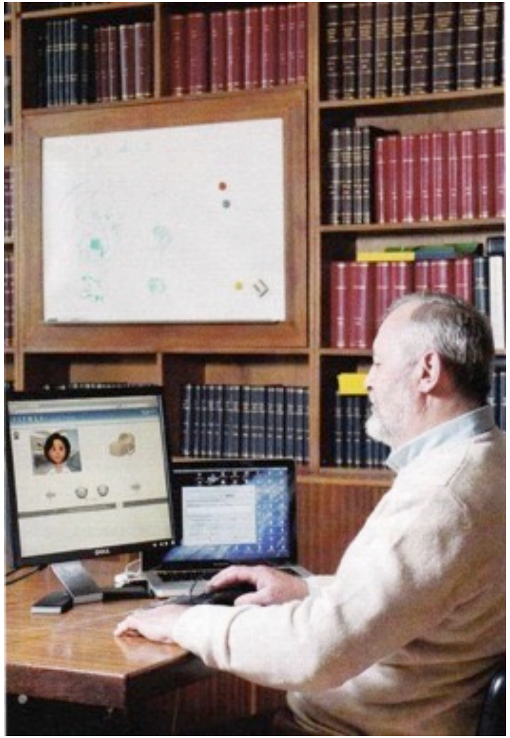


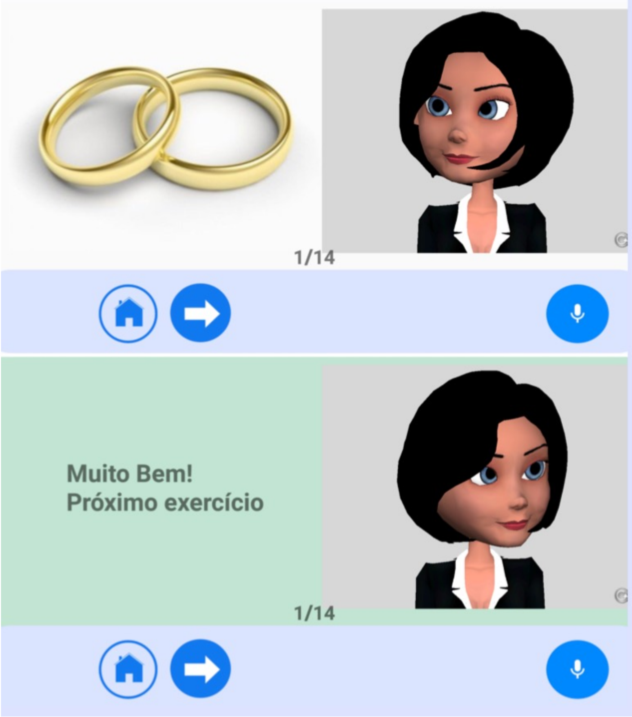
SLT applications in e-health @ INESC-ID
2G: Diagnosis and treatment of speech and language disorders

- Improved ASR and intelligibility objective measures useful in:
  - Diagnosis
    - Global and individual characteristics
    - Second opinion
    - Screening of populations
  - Therapy
    - Therapy control
    - Comparison of therapy methods
    - Computer-assisted and virtual/on-line therapy

SLT applications in e-health @ INESC-ID
2G: Diagnosis and treatment of speech and language disorders

Virtual Therapist for Aphasia Treatment

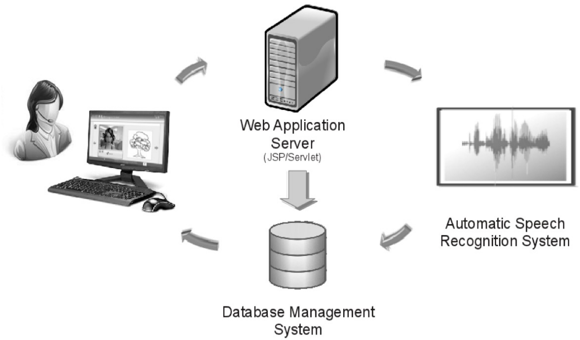




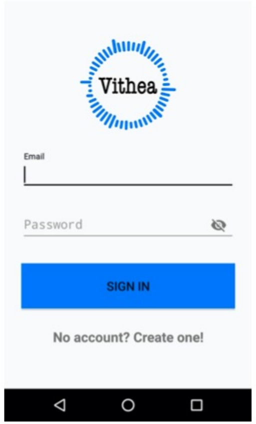
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

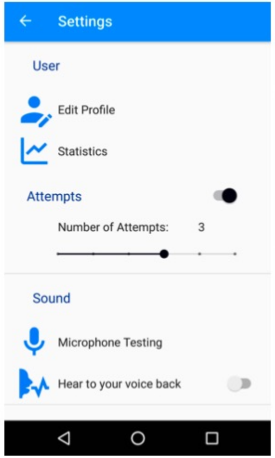
SLT applications in e-health @ INESC-ID
2G: Diagnosis and treatment of speech and language disorders

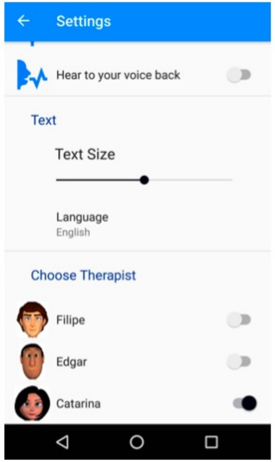
Virtual Therapist for Aphasia Treatment

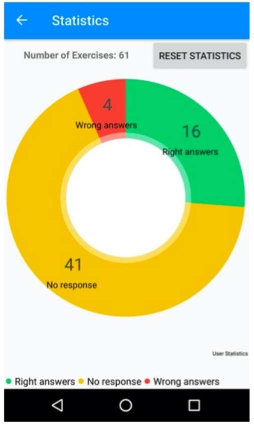


Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR



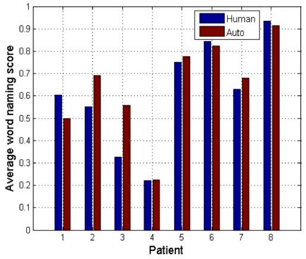


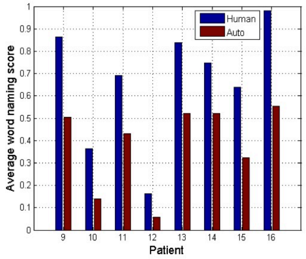


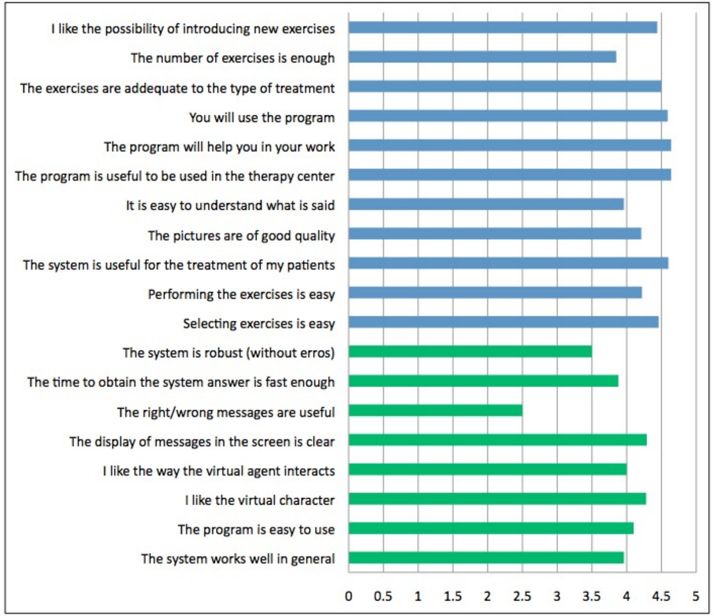


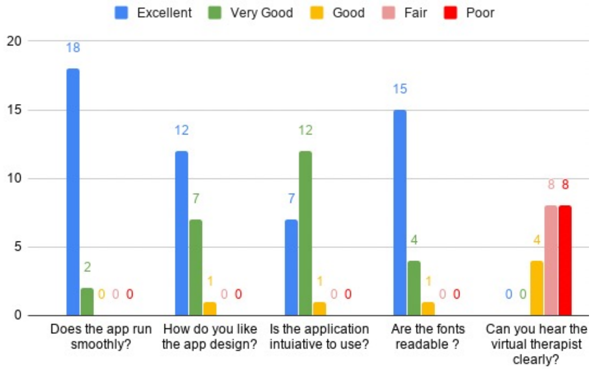
2G: Diagnosis and treatment of speech and language disorders

Virtual Therapist for Aphasia Treatment









Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

SLT applications in e-health @ INESC-ID
2G: Diagnosis and treatment of speech and language disorders

Virtual Therapist for Aphasia Treatment

Faculdade de Medicina Universidade de Lisboa

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

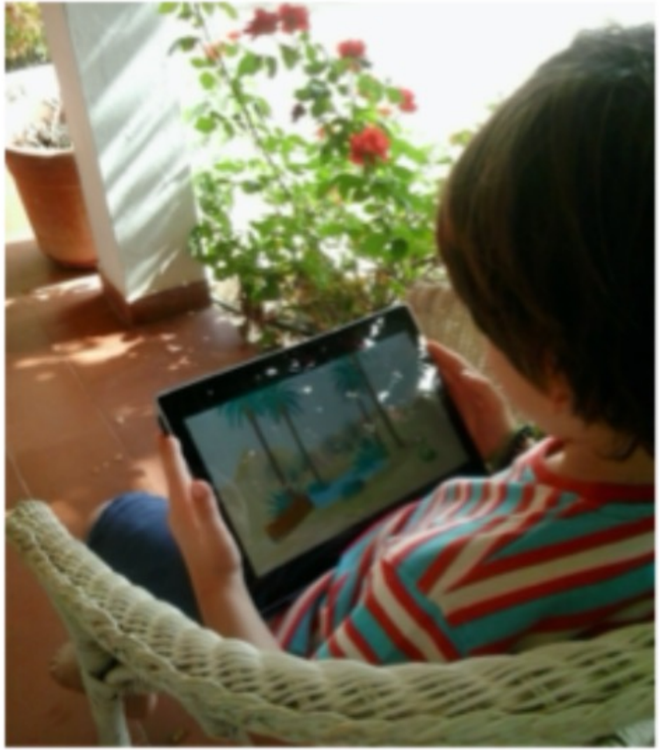
2G: Diagnosis and treatment of speech and language disorders

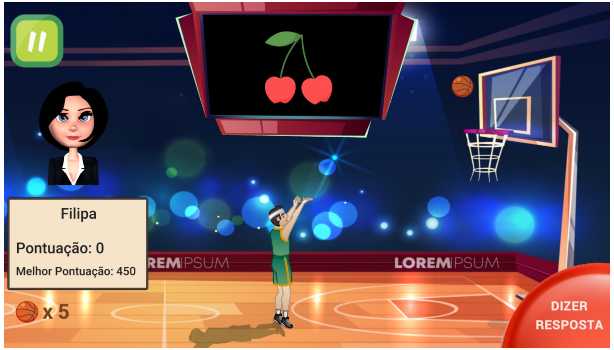
Serious digital games for speech and language therapy for children

BioVisualSpeech

Information and Communication Technologies Institute
Carnegie Mellon | PORTUGAL
AN INTERNATIONAL PARTNERSHIP

TAPAS





inesc id lisboa

SLT applications in e-health @ INESC-ID
2G: Diagnosis and treatment of speech and language disorders

Serious digital games for speech and language therapy for children

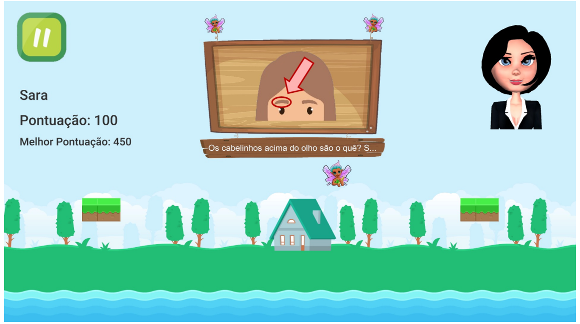
Figure 3. Stimulus used to suggest the word *mochila*

Table 3. Words with sibilants – number of children

Age	Girl	Boy	Total
5	20	19	39
6	35	35	70
7	51	33	84
8	39	50	89
9	37	37	74
Total	182	174	356

Table 4. Word samples

Phoneme	Incorrect phoneme occurrences	Total phoneme productions	Words with incorrect sibilants	Total number of words
ʃ	434	11202	422	11455
ʒ	240	2880	240	3333
s	354	6627	322	7024
z	154	2260	154	2616
Total	1182	22969	1138	22830



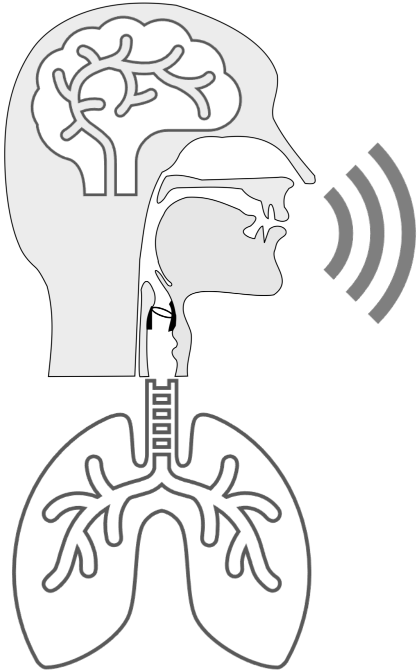
BioVisualSpeech

Information and Communication Technologies Institute
Carnegie Mellon | PORTUGAL
AN INTERNATIONAL PARTNERSHIP

TAPAS

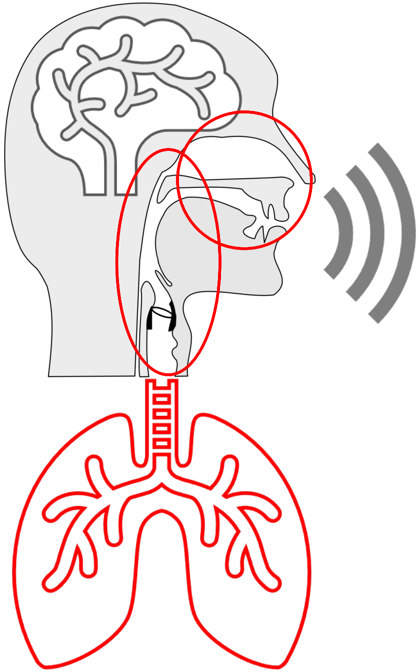
SLT applications in e-health @ INESC-ID
3G: Speech as a health biomarker

- Besides speech and language disorders, many other diseases affect speech in different ways:
    - Diseases that concern respiratory organs;
    - Neurodegenerative diseases;
    - Mood disorders;



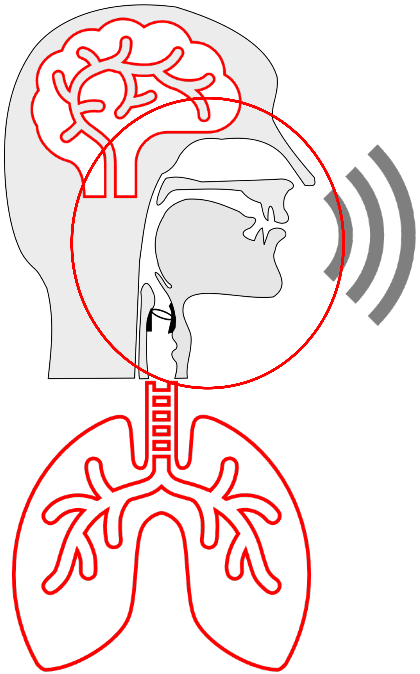
SLT applications in e-health @ INESC-ID
3G: Speech as a health biomarker

- **Diseases that concern respiratory organs, e.g.:**
  - Obstructive Sleep Apnea (OSA);
  - Common Cold;
  - COVID-19.



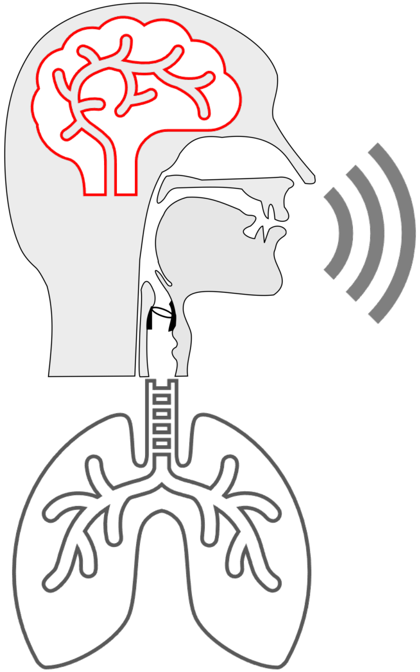
SLT applications in e-health @ INESC-ID
3G: Speech as a health biomarker

- Diseases that concern respiratory organs;
- **Neurodegenerative diseases, e.g.:**
  - Alzheimer's disease (AD);
  - Parkinson's disease (PD);
  - Huntington's disease (HD);
  - Amyotrophic lateral sclerosis (ALS).



SLT applications in e-health @ INESC-ID
3G: Speech as a health biomarker

- Diseases that concern respiratory organs;
- Neurodegenerative diseases;
- **Mood disorders, e.g.:**
  - Anxiety;
  - Depression;
  - Bipolar Disorder;
  - Post-traumatic stress disorder (PTSD).



SLT applications in e-health @ INESC-ID
3G: Speech as a health biomarker

- The potential of speech to act as a biomarker has bolstered several studies on the automatic detection of diseases, based on powerful **machine learning** classifiers, that either operate on features extracted from the speech signal, or directly on the signal itself.
- The fact that speech is **ubiquitous** and can be acquired **non-intrusively** makes it an inexpensive modality that may be used to identify high likelihoods of the presence of diseases, and the results of such early screening tests may act as alerts for users to seek medical assistance.
- Moreover, speech may be used in clinical facilities or at the patients' homes. It may also allow to remotely monitor the progress of a disease in order to adapt medication and support.

Speech as a health biomarker

Obstructive Speech Apnea (OSA)

By 2020, 230000 – 345000 people are expected to be **killed** in traffic accidents due to fatigue[1]

9% - 38% of the adult population suffers from **OSA** [3]

OSA speech anomalies: articulatory (slurred speech); phonation (larynx inflammation); resonance (vocal/nasal coupling)

1/3 adults suffers of **inadequate sleep** [2]

46% of OSA couples sleep in separate rooms [5]

Polysomnography (PSG) is the gold standard diagnosis: patients spend a night connected to electrodes

Speech as a health biomarker

Obstructive Speech Apnea (OSA)

Table 4. Best performing classifiers and feature sets for OSA detection, using PSD and WOSA corpus.

	RF features; SVM			OFS features; SVM			OFS features; SVM+LDA+kNN		
	TPR (%)	TNR (%)	WA (%)	TPR (%)	TNR (%)	WA (%)	TPR (%)	TNR (%)	WA (%)
PSD	92.0	65.0	80.0	88.0	75.0	82.2	88.0	80.0	84.0
PSD-b	85.0	68.2	76.2	70.0	77.3	73.8	80.0	72.7	76.2
WOSA	12.2	37.5	25.0	75.0	87.5	81.3	75.0	87.5	81.2
PSD+WOSA	50.0	25.0	37.5	75.0	62.5	68.8	75.0	62.5	68.8

Table 5. Comparison of the performance achieved per task and the relative frequency of nasal phonemes and diphthongs.

Task	Performance			Nasal Phonemes (%)	Diphthongs (%)
	TPR (%)	TNR (%)	WA (%)		
1	84.0	70.0	78.8	12.6	6.4
3.1	84.0	65.0	75.6	13.5	5.7
3.2	**92.0**	**75.0**	**84.4**	**25.0**	**10.0**
3.3	72.0	65.0	68.9	18.8	6.3
3.4	84.0	70.0	77.8	6.5	6.5
3.5	92.0	65.0	80.0	12.1	5.1
3.6	80.0	85.0	82.0	8.9	4.4
3.7	84.0	75.0	80.0	14.0	7.0
3.8	84.0	75.0	80.0	14.3	7.1
3.9	92.0	65.0	80.0	11.5	1.9
3.10	88.0	60.0	75.6	16.0	4.0
4	92.0	55.0	75.6	-	-

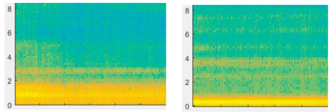
Portuguese Sleep Disorders (PSD) Corpus

Task 1: "The North Wind and the Sun"

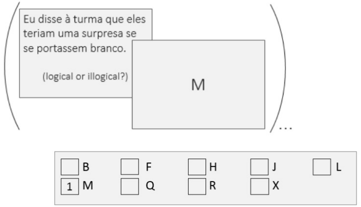
Source: https://en.wikipedia.org/wiki/The_North_Wind_and_the_Sun

Task 2: Elongated vowels

/a/ /i/

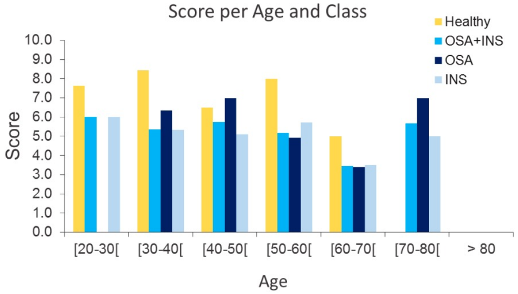


Task 3: Reading span task



Task 4: Describing the image

Source: http://time.com/4551131/Vincent-van-gogh-bedroom-bed-boxmeer/



Speech as a health biomarker

Alzheimer’s Disease

Alzheimer’s Disease (AD) is the most common form of Dementia

Between 2000 and 2050, the proportion of the world's population over 60 will double from 11% to 22%

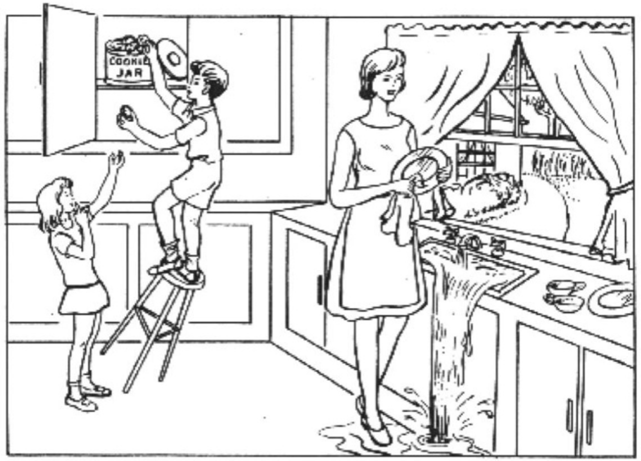
Besides alterations of memory and of spatial-temporal orientation, language impairments are also an important factor

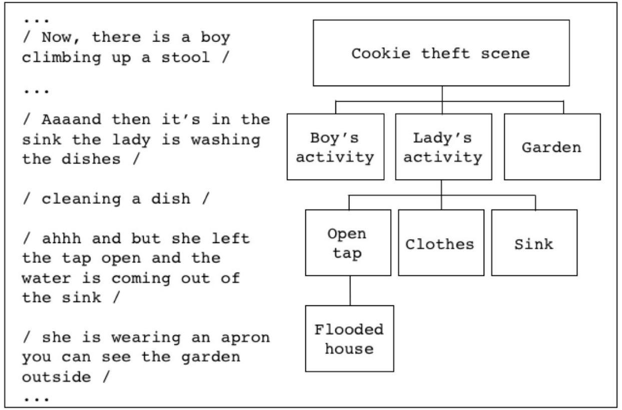
Prevalence increases with age: U.S. census reported that 3% of people aged 65-74, 17% of people aged 75-84, and 32% of people aged 85 and older have AD

Pharmacological treatments may temporarily improve the symptoms of the disease, but they can not stop or reverse its progression

Language impairments in AD speech: naming, word-finding difficulties, repetitions, overuse of indefinite and vague terms, inappropriate use of pronouns

Alzheimer's Disease: Topic coherence





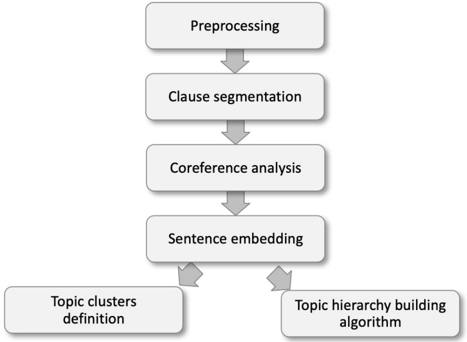
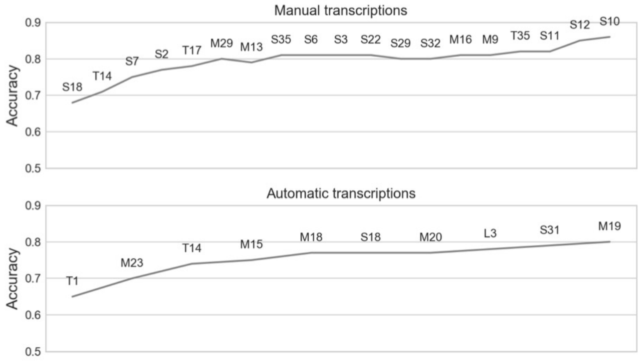


TABLE III
SUMMARY OF AD CLASSIFICATION RESULTS (AVG. AND RANGE ACCURACY %)

	Manual transcriptions			Automatic transcriptions
	Topic coherence	Linguistic	Fusion	Fusion
Accuracy	79.0±4.8	82.6±5.1	85.5±2.9	79.7±3.5



Speech as a health biomarker

Alzheimer's Disease: Challenge

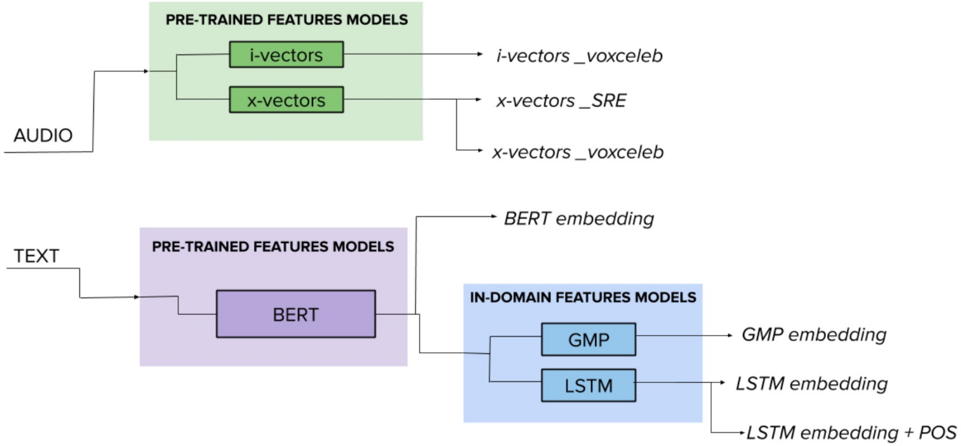


Table 2: *Results of different acoustic approaches on the development set*

	Accuracy	Precision	Recall	F1 Score
x-vectors_Vox	0.6818	0.6834	0.6919	0.6812
x-vectors_SRE	**0.7273**	**0.7273**	**0.7273**	**0.7273**
i-vectors_Vox	0.6818	0.7292	0.6818	0.6645
i-vectors_Vox_x-vectors_Vox	0.7273	0.7273	0.7273	0.7273
i-vectors_Vox_x-vectors_SRE	0.7273	0.7351	0.7273	0.7250

Table 3: *Results of different linguistic approaches on the development set*

	Accuracy	Precision	Recall	F1 Score
Global Max Pool.	0.7727	0.7947	0.7728	0.7684
LSTM-RNNs	0.8182	0.8182	0.8182	0.8182
LSTM-RNNs Pos	0.8636	0.8667	0.8637	0.8634
GMax/LSTM-RNNs/LSTM-RNNs-Pos	**0.9091**	**0.9091**	**0.9091**	**0.9091**
Sentence emb. - maj. vote	0.7727	0.7947	0.7728	0.7684

Table 4: *Results of different acoustic and linguistic approaches on the test set*

	Class	Accuracy	Precision	Recall	F1 Score
Fusion of system	AD	**0.8125**	0.9412	0.6667	0.7805
	non-AD		0.7419	0.9583	0.8364
Sentence embedding	AD	**0.7292**	0.8235	0.5833	0.6829
	non-AD		0.6774	0.8750	0.7636
x-vectors_SRE	AD	**0.5417**	0.5417	0.5417	0.5417
	non-AD		0.5417	0.5417	0.5417

Speech as a health biomarker

Parkinson's Disease

Parkinson's disease (PD) is a progressive degenerative disorder characterized by motor and non-motor symptoms.

Motor signs of PD include resting tremor, rigidity, bradykinesia, while non-motor symptoms include cognitive disorders, and sleep and sensory abnormalities.

Motor symptoms of PD influence also the speech production of language.

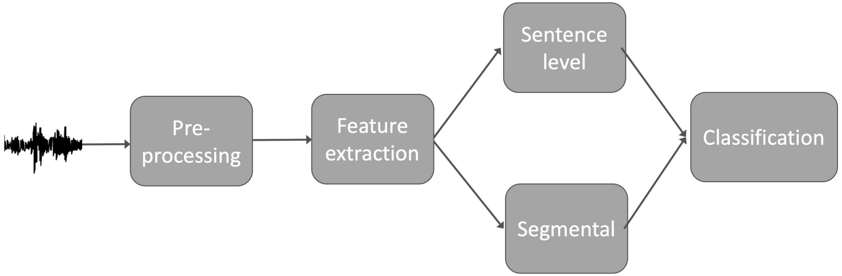
Dysarthria, which is characterized by a weakness, paralysis, or lack of coordination in the motor-speech system, is typically observed in PD patients and affects respiration, phonation, articulation and prosody.

As the disease progresses, patient alternate periods in which motor symptoms are mitigated due to medication intake (ON state), and periods with motor complications (OFF state).

The time that patients spend in the OFF condition is currently the main parameter used to assess pharmacological interventions.

Speech as a health biomarker

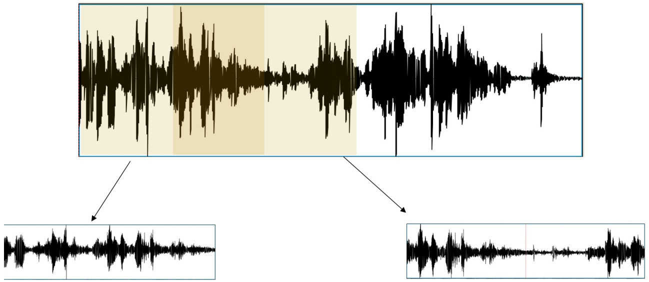
Parkinson's Disease: Study of production tasks



	PD patients	Healthy Control
N. of subjects:	75	65
Duration:	~7h 30m	~6h 31m

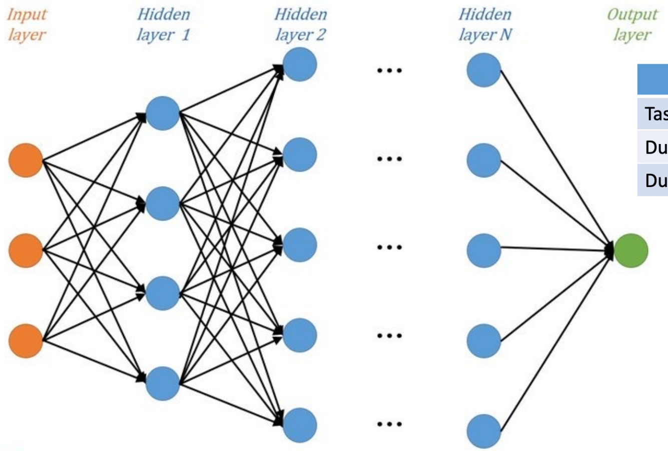
Task-dependent results on the classification task (PD vs. control)

	Accuracy (%)	
Task	**Sentence Level**	**Segmental**
Sustained vowel phonation (/a/)	55.00	58.14
Maximum phonation time (/a/)	60.00	75.65
Oral diadochokinesia	60.71	73.28
Reading of word	54.29	76.35
Reading of sentences	**62.14**	**81.74**
Reading of text	65.00	79.86
Storytelling guided by visual stimuli	**66.43**	**82.32**
Reading of prosodic sentences	**70.71**	**85.10**



Speech as a health biomarker

Parkinson's Disease: Monitoring medication sate



	Training set (avg. duration x speaker)	Test set (avg. duration x speaker)
Task:	Ddk, mpt, word, sent, pros, conv	/a/, text, frog
Duration ON:	12h 55m (10m)	2h 54m (2m)
Duration OFF:	13h 24m (10m)	3h 06m (2m)

Table 3: *Test results (utterance-level Acc - %) for speaker-dependent medication state assessment by task.*

Feature set	/a/	Reading text	Story telling
MFCC	86.49	87.16	92.57
MFCC+Δs	85.14	91.22	93.24
eGeMAPS	89.19	87.16	**95.27**
MFCC+PCA	86.49	89.19	94.59
MFCC+Δs+PCA	85.14	89.19	93.92
eGeMAPS+PCA	83.78	86.49	93.24

Table 2: *Optimal model configurations and results (utterance-level Acc - %) for speaker-dependent medication state assessment.*

Feature set	#Coefficients	Context	Input dim.	Architecture	α	Acc devel.	Acc test
MFCC	13	15	195	256, 128, 32, 1	0.01	93.92	88.74
MFCC+Δs	26	11	286	256, 128, 1	0.01	93.24	89.86
eGeMAPS	23	15	345	512, 128, 1	0.001	95.95	**90.54**
MFCC+PCA	13	15	95	512, 128, 1	0.03	91.89	90.09
MFCC+Δs+PCA	26	11	85	128, 64, 1	0.03	91.89	89.41
eGeMAPS+PCA	23	15	70	128, 64, 1	0.03	96.62	87.84

Instituto de Medicina Molecular

Outline

- Introduction to speech processing
- Three generations of HLT applications in e-health @INESC-ID:
  - 1G: Augmentative communication, assistive technologies and e-inclusion
  - 2G: Diagnosis and treatment of speech and language disorders
  - 3G: Speech as a health biomarker for speech-affecting diseases
- **Challenges and open questions**

Challenges and open questions

- There is a significant amount of research showing that speech has the potential to be a useful bio-marker for diagnosis and monitoring of several diseases.
- However, there are still significant limitations that prevent from broad application in real clinical settings:
  1. Data limitations:
     - Small data sets; unbalanced data; In-the-wild data; cross-domain, cross-lingual; privacy concerns
  2. Not realistic (or really valuable) tasks:
     - Binary classification of HC vs disease; labels not related with speech/language issues; cold vs COVID-19?
- Possible directions for future work:
  1. Mining on-line repositories; development of privacy-preserving protocols
  2. Towards "speech tests" equivalent to blood tests to help medical diagnosis

inesc id lisboa

DEFINING TECHNOLOGY

Thank you!

alberto.abad@inesc-id.pt