U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

CINTESIS
Health. Research.

Accurate Glottal Source Estimation and Modelling

Bruno Miguel Silva Santos

Supervisor: Prof. Doutor Aníbal Ferreira, FEUP

Co-Supervisor: Prof. Doutor Jorge Spratley, FMUP

9th of July 2020

Cofinanciado por:

COMPETE 2020

PORTUGAL 2020

UNIÃO EUROPEIA
Fundo Europeu
de Desenvolvimento Regional

FCT
Fundação para a Ciência e a Tecnologia

OVERVIEW

CONTEXT

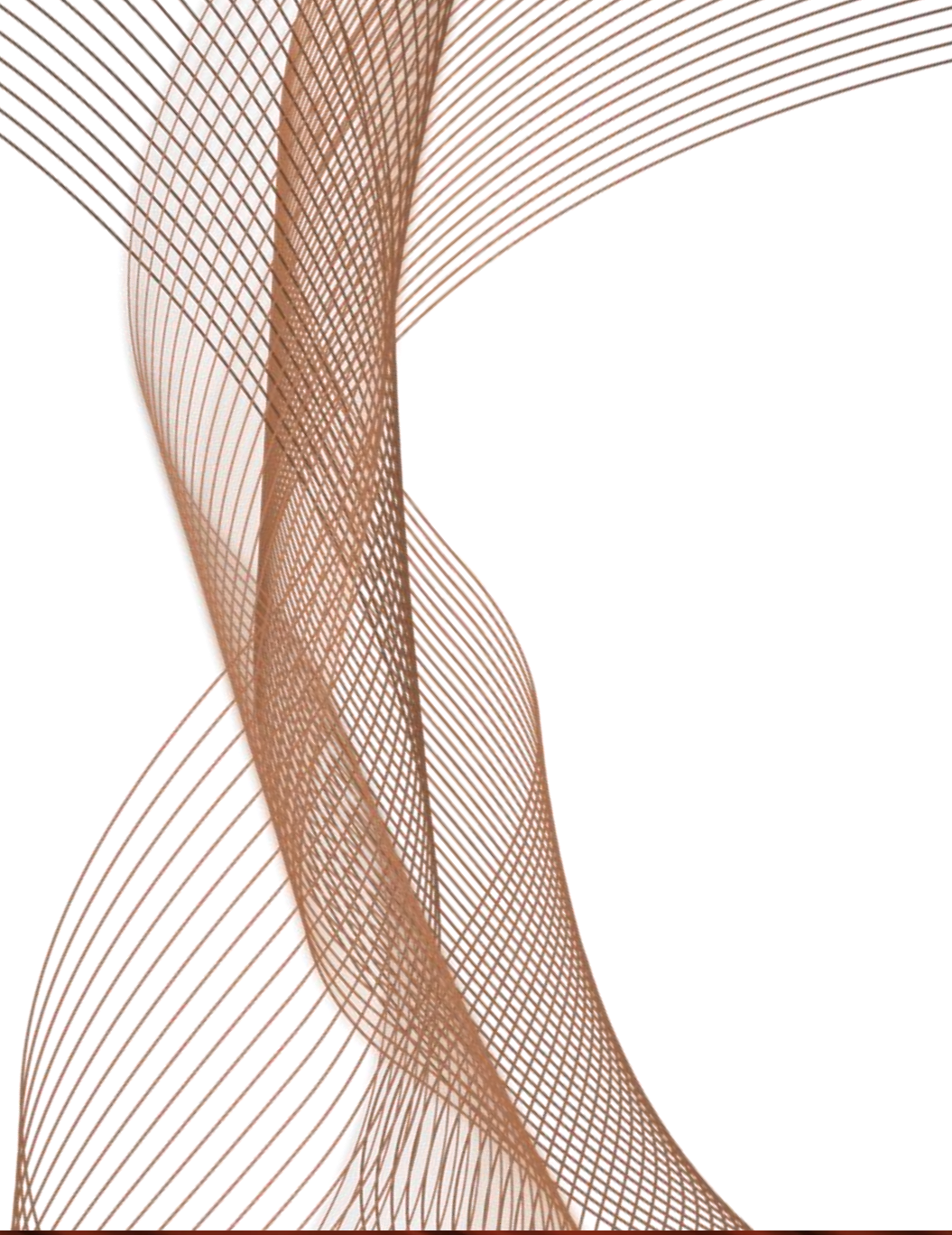
BACKGROUND

DATASET CREATION

GLOTTAL SOURCE CHARACTERIZATION

VOCAL TRACT CHARACTERIZATION

CONCLUSIONS



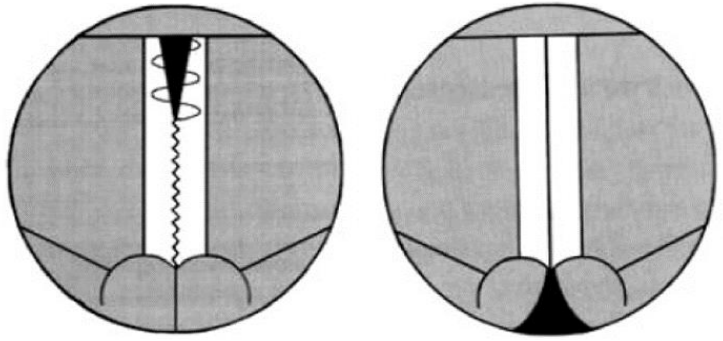
CONTEXT

- Normal vs whispered speech

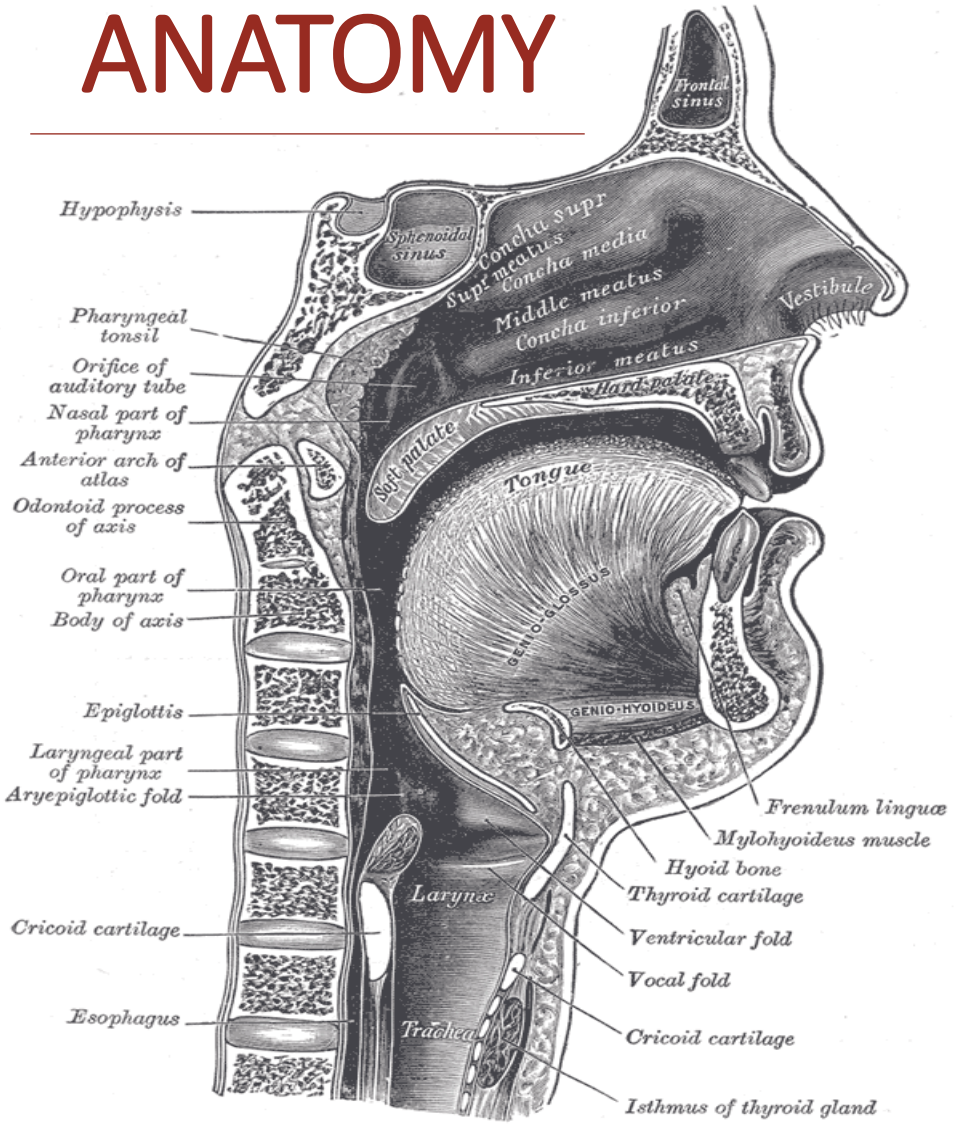
- DyNaVoiceR project

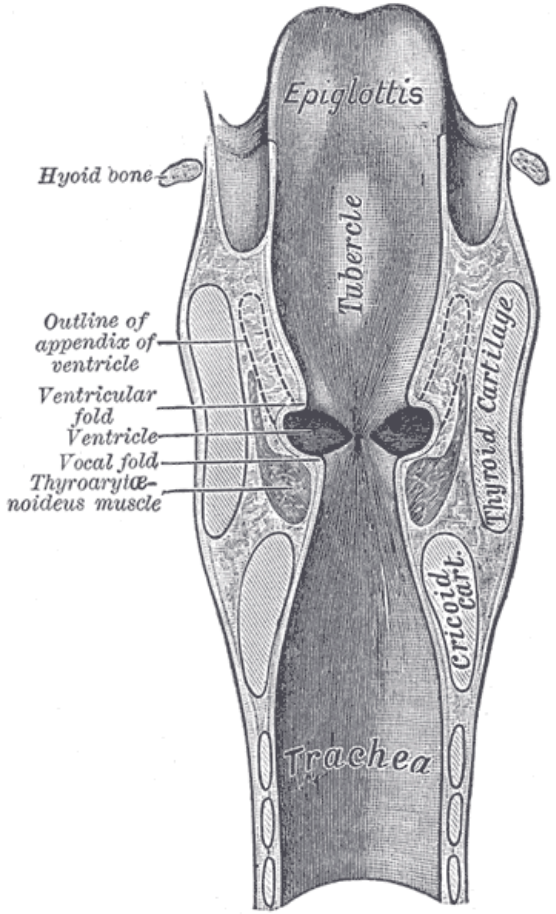
- Limitations of the idealized theoretical models

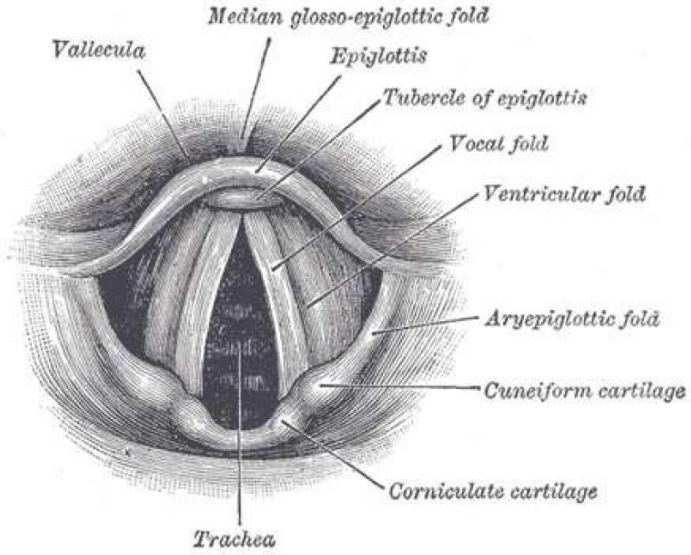
- Glottal source characterization



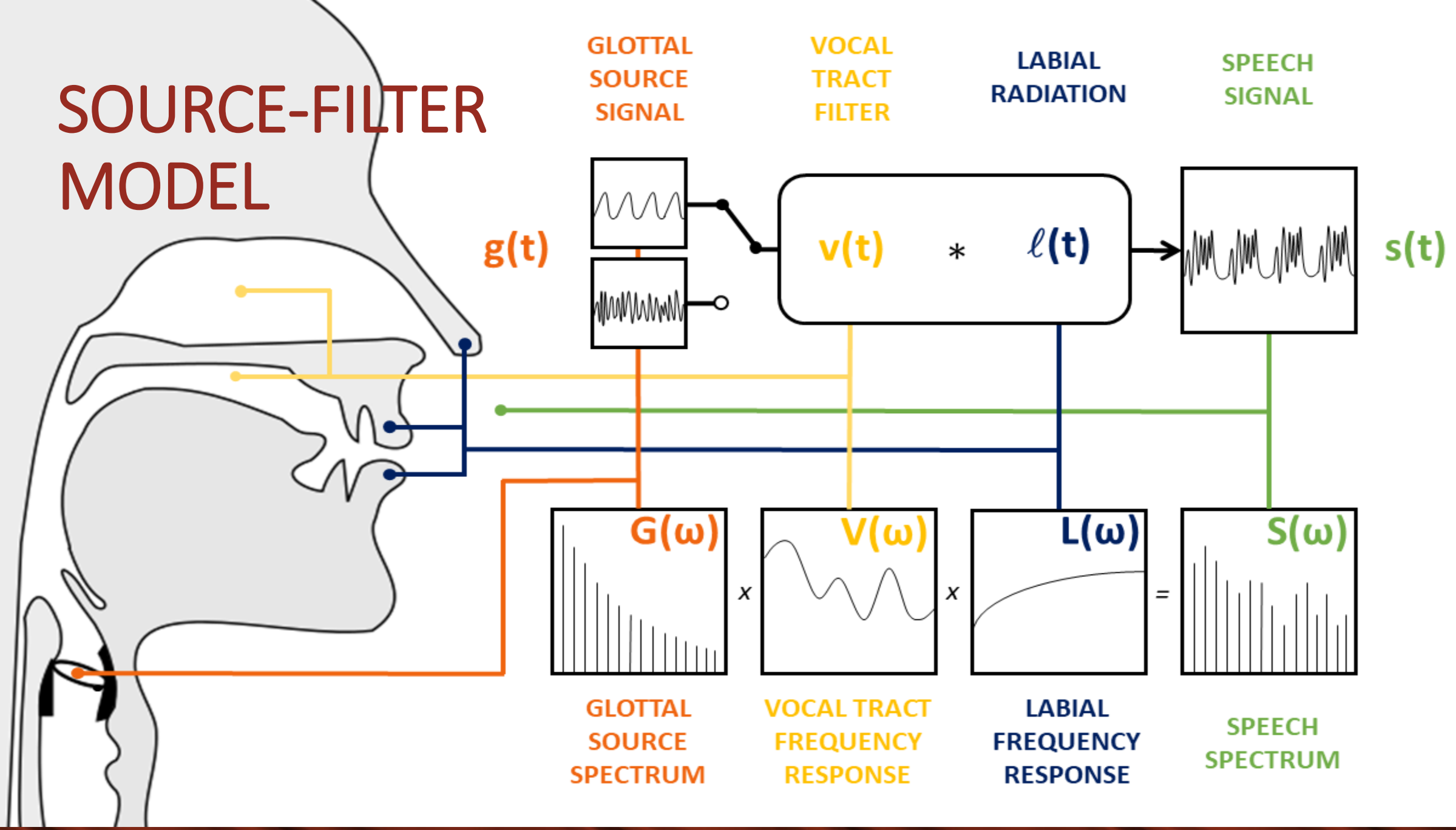
PHYSIOLOGY AND ANATOMY



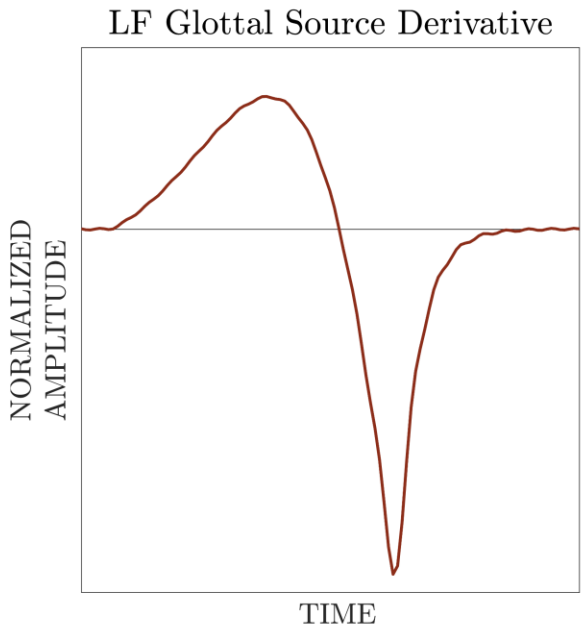


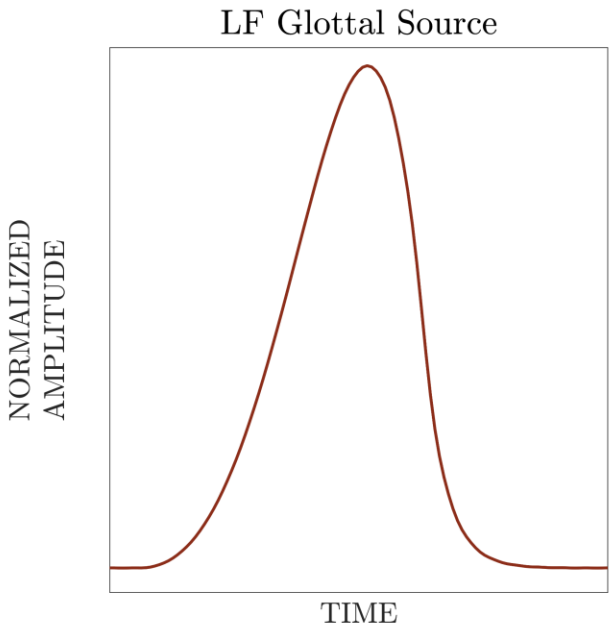


SOURCE-FILTER MODEL

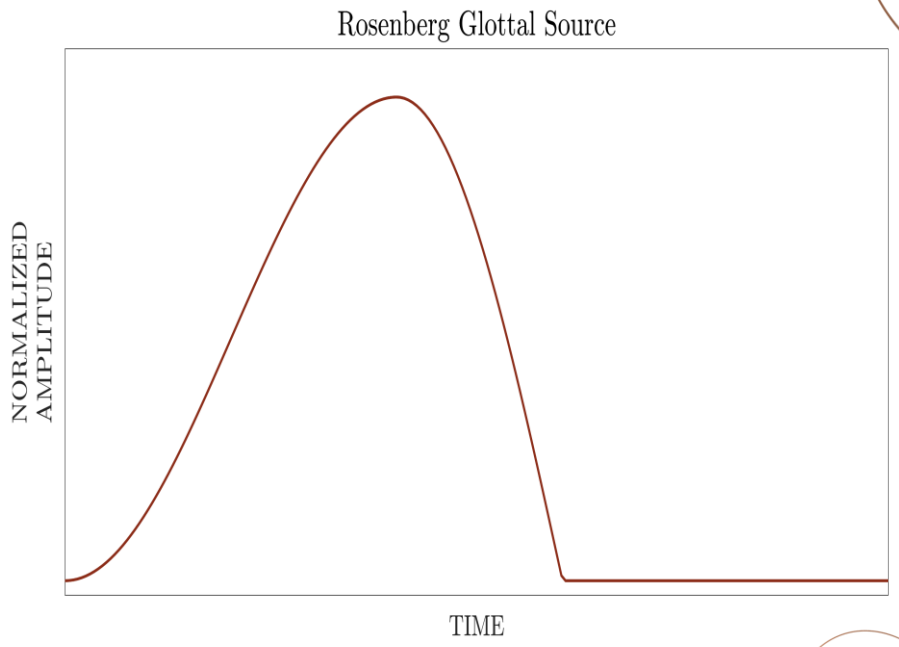


GLOTTAL SOURCE MODELS

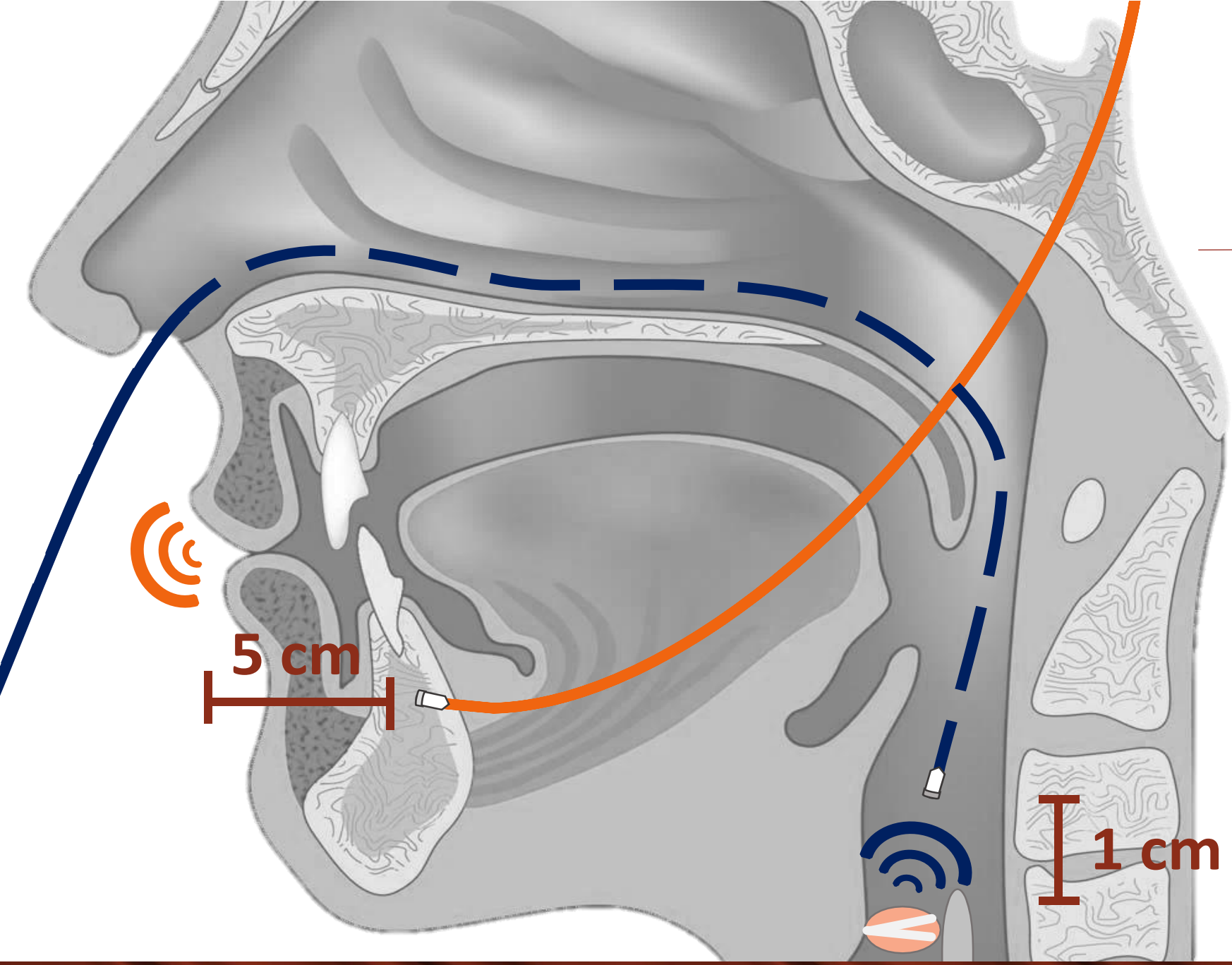




Liljencrants-Fant Model



Rosenberg Model



DATASET ACQUISITION

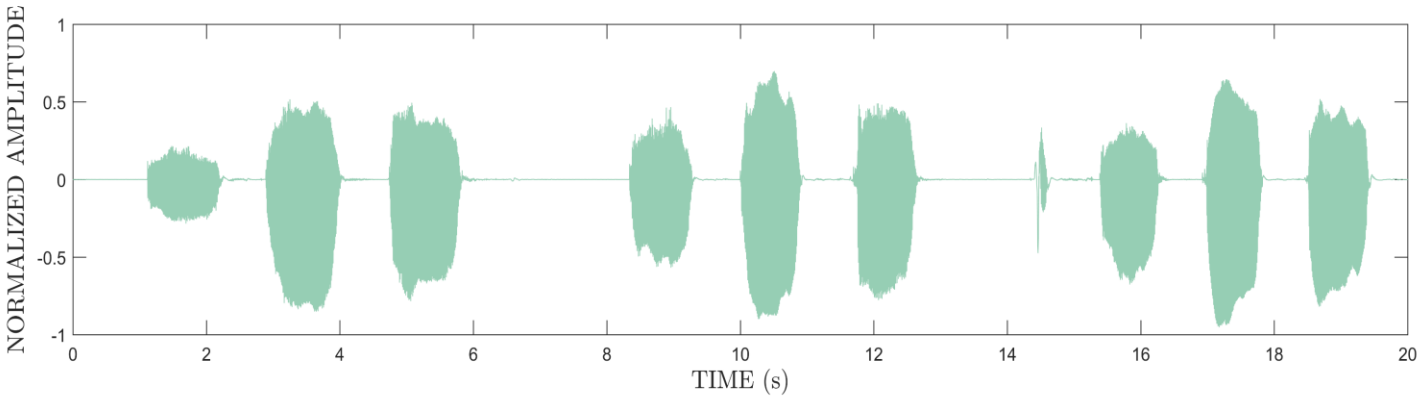
Material

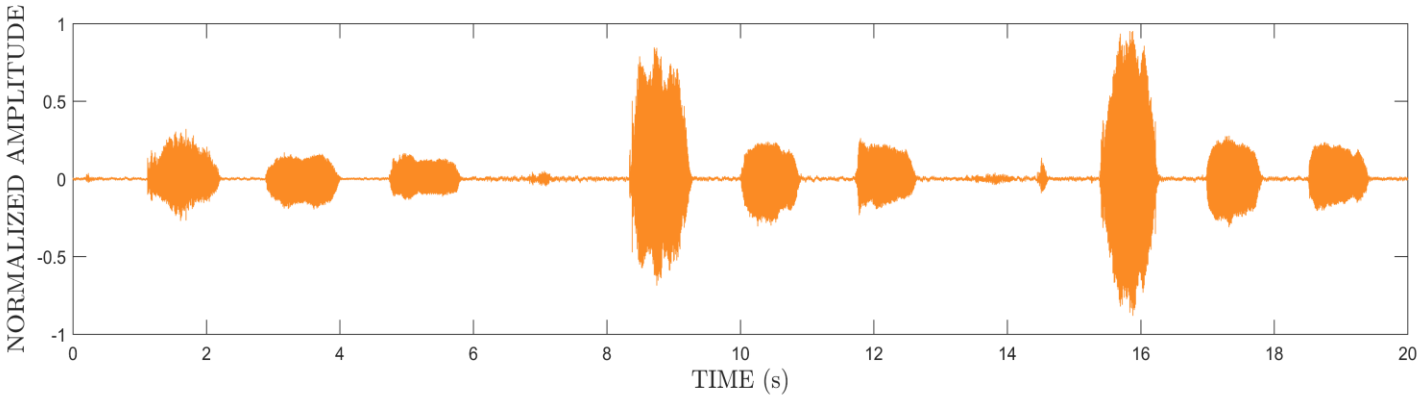
- two high quality **microphones** with extremely reduced dimensions (B6 Omnidirectional Lavalier);
- a 128 kHz USB **audio/MIDI interface** (Focusrite) with 2 stereo channels;
- two phantom **power adaptors** (MZA 900 P);
- a **flexible rhyno-laryngo fiberscope** (ENF-XP OLYMPUS);
- a **nasogastric tube** (6mm diameter).

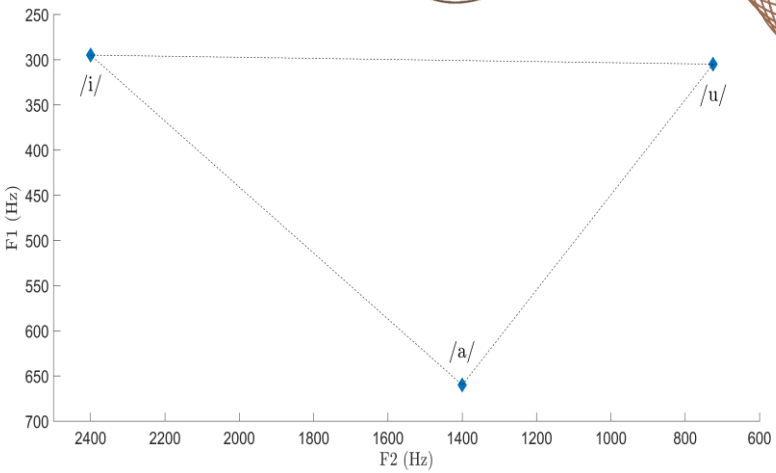
Criteria

- having at least 18 years of age;
- leading a healthy lifestyle, e.g. non-smoker;
- absence of voice disorders history;
- showing viability for the procedure after anterior rhinoscopy inspection.

DATASET CHARACTERIZATION



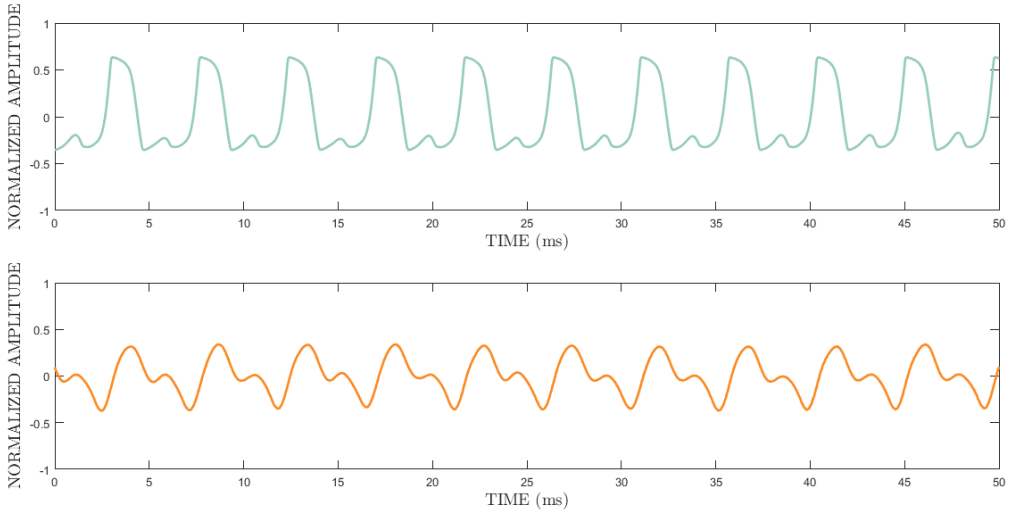




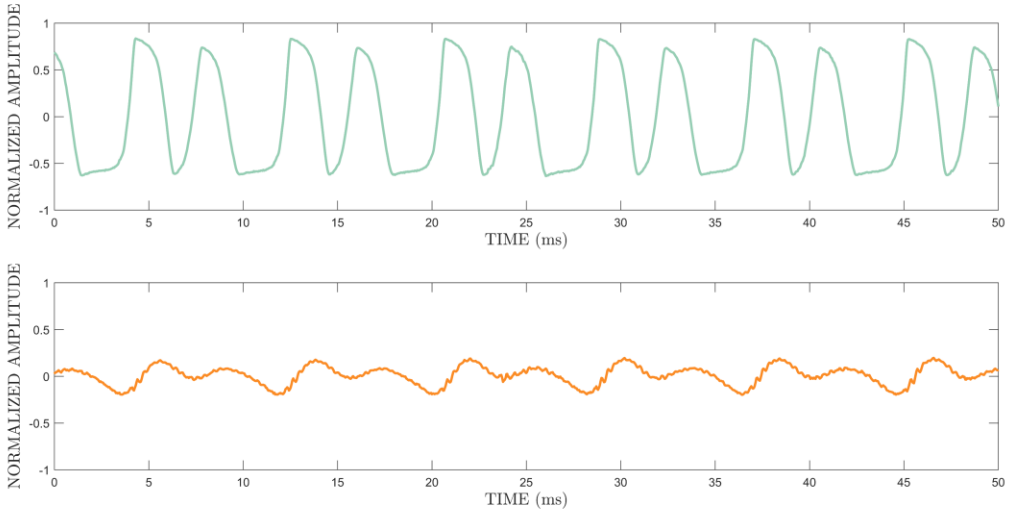
Speaker	Gender	Age
1	Male	25
2	Female	23
3	Female	19
4	Female	22
5	Male	27
6	Male	22

DATASET PRELIMINARY ANALYSIS

FEMALE SPEAKER



MALE SPEAKER

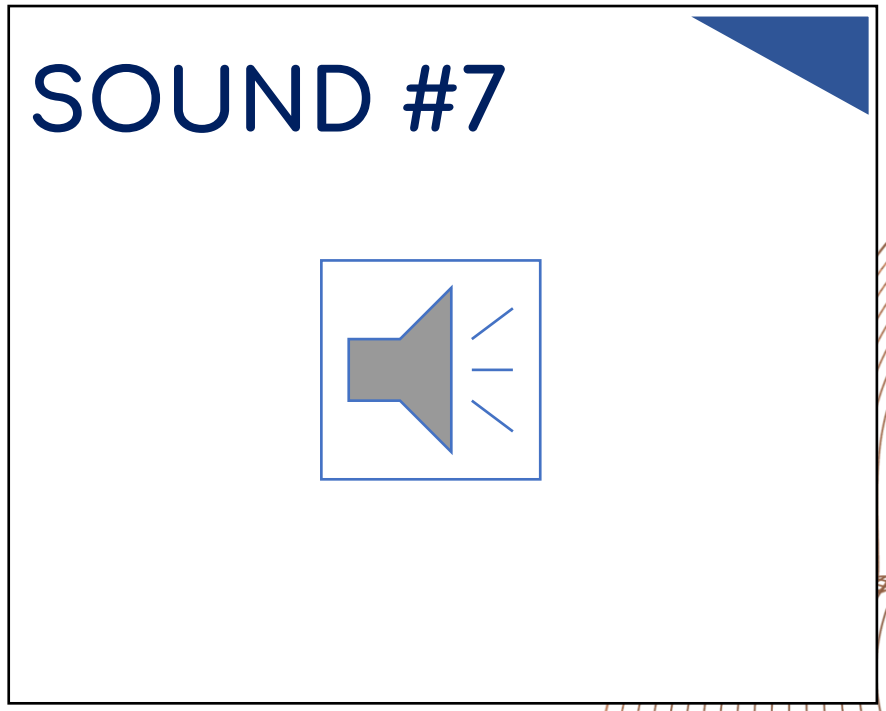


PERCEPTUAL TESTS

- **Test material:**
  - Natural glottal source signal of several sustained vowel utterances including */a/*, */i/* and */u/* by 6 different speakers

- **Task:**
  - Part #1 – identify which of the 9 oral vowels in the standard European Portuguese (/à/, /â/, /e/, /é/, /i/, /ê/, /ó/, /ô/, /u/) the given reference approximates most
  - Part #2 – identify which of the 3 vowels recorded (/à/, /i/, /u/) the given reference approximates most

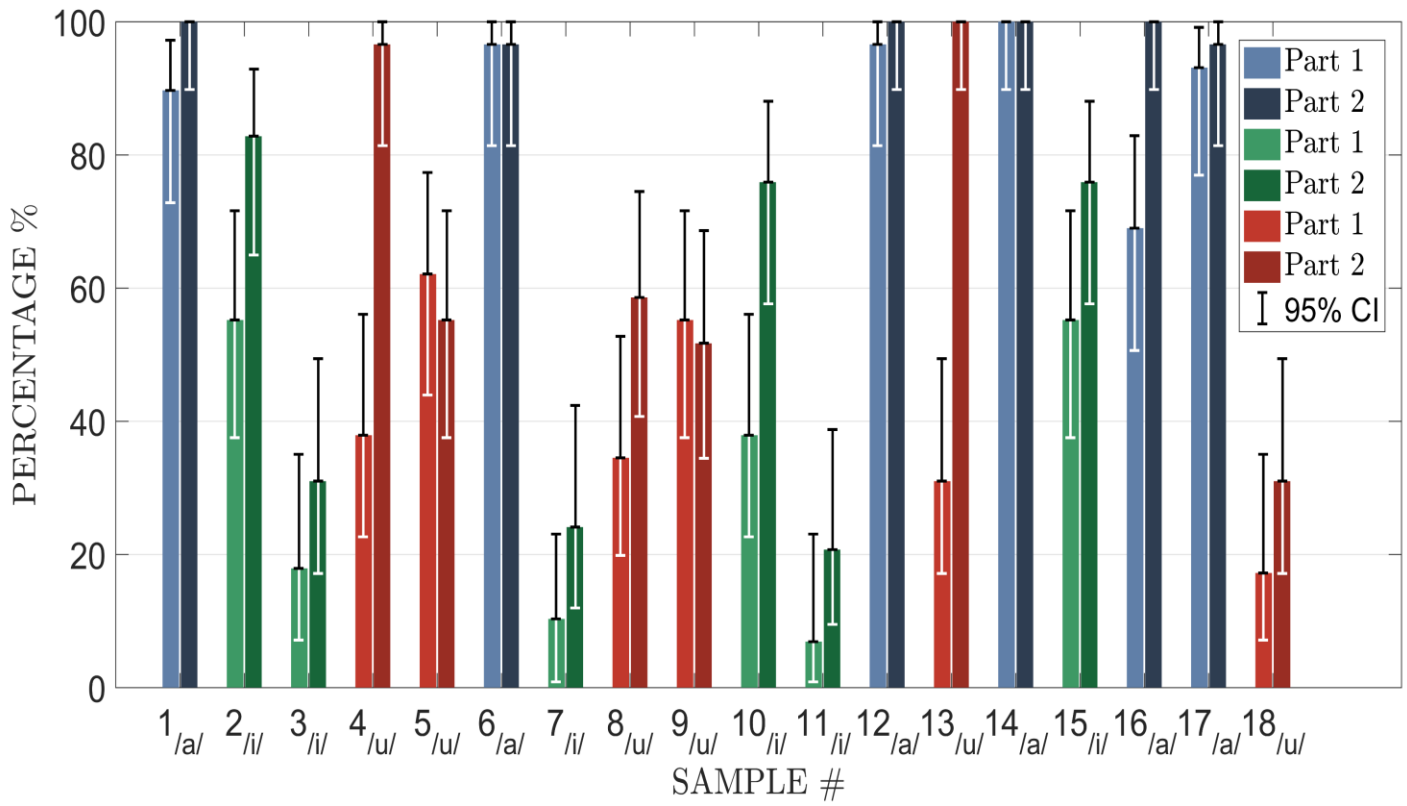
- **Evaluation:**
  - Correct vowel identification success rate



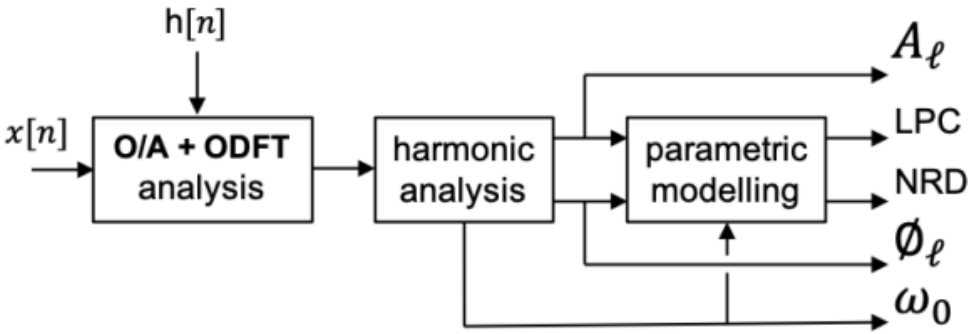
PERCEPTUAL TESTS RESULTS

- Participants are able to **effectively recognize** the vowel /a/, while having **difficulty** in recognizing the vowels /i/ and /u/
- Correct identification of the vowel /i/ shows a **statistically significant difference** between both parts (from **30.57%** to **51.73%**) with the **lowest** success rates
- Internal signals recorded for vowels /i/ and /u/ suffer less influence from the vocal tract, when compared to the vowel /a/

	/a/	/i/	/u/
PART I	90.83%	30.57%	39.65%
PART II	98.87%	51.73%	65.52%

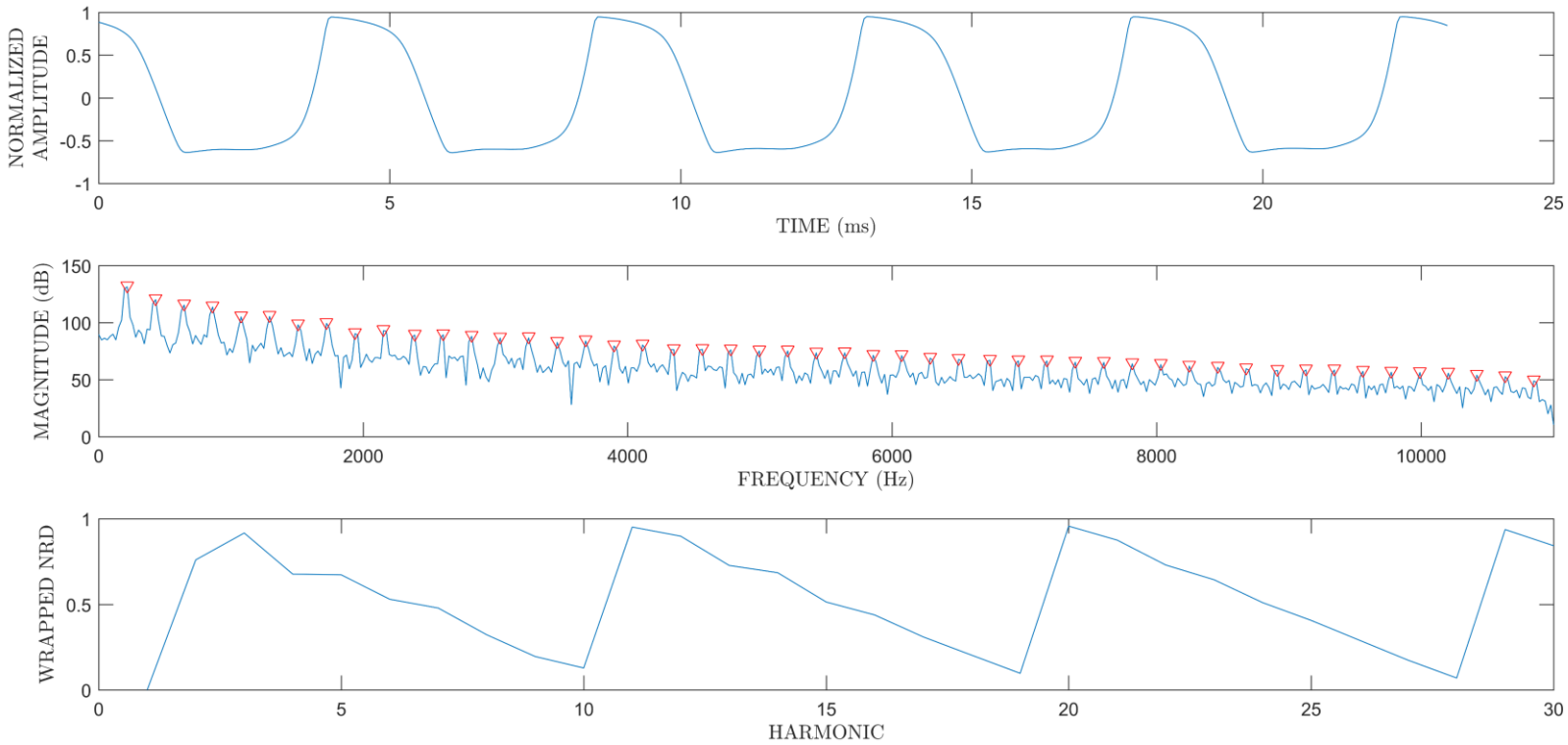


HARMONIC ANALYSIS

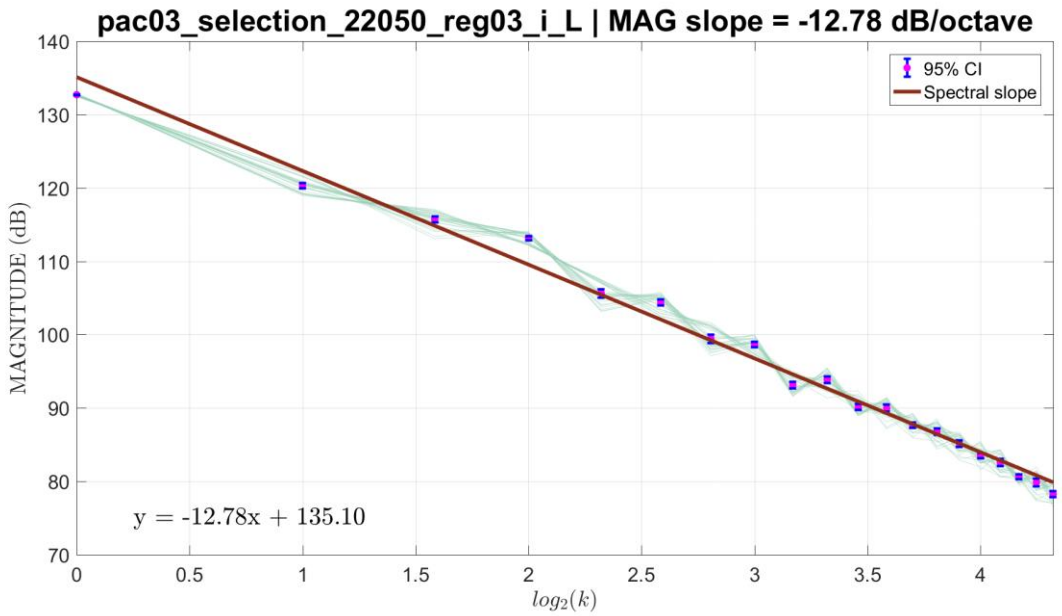


- **Signal segmentation**
  - sinusoidal window
  - 50% overlap-add analysis
- **Spectral magnitude analysis**
  - accurate harmonic magnitudes
  - all-pole (LPC) model
    1. dB interpolation between accurate harmonic magnitude
    2. Autocorrelation using Wiener-Khintchine theorem
    3. Parameters of the 22nd order all pole model through Levinson-Durbin recursion
- **Spectral phase analysis**
  - harmonic starting phases
  - Normalized Relative Delay (NRD) phase-related feature
    i. Time-shift invariant
    ii. Independent of the fundamental frequency

HARMONIC ANALYSIS

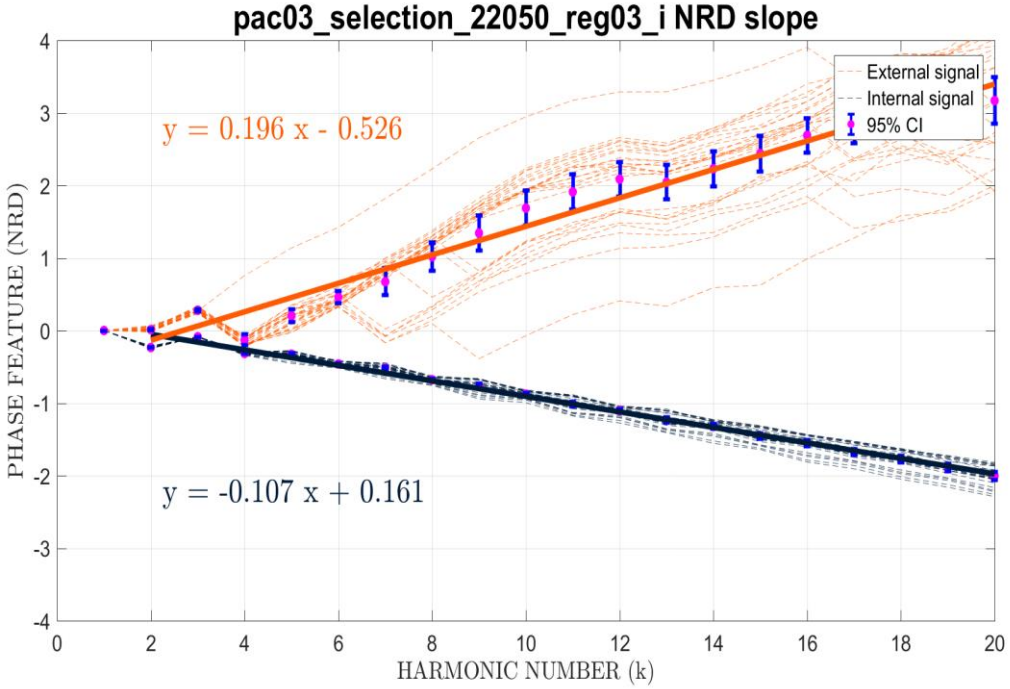


MAGNITUDE ANALYSIS



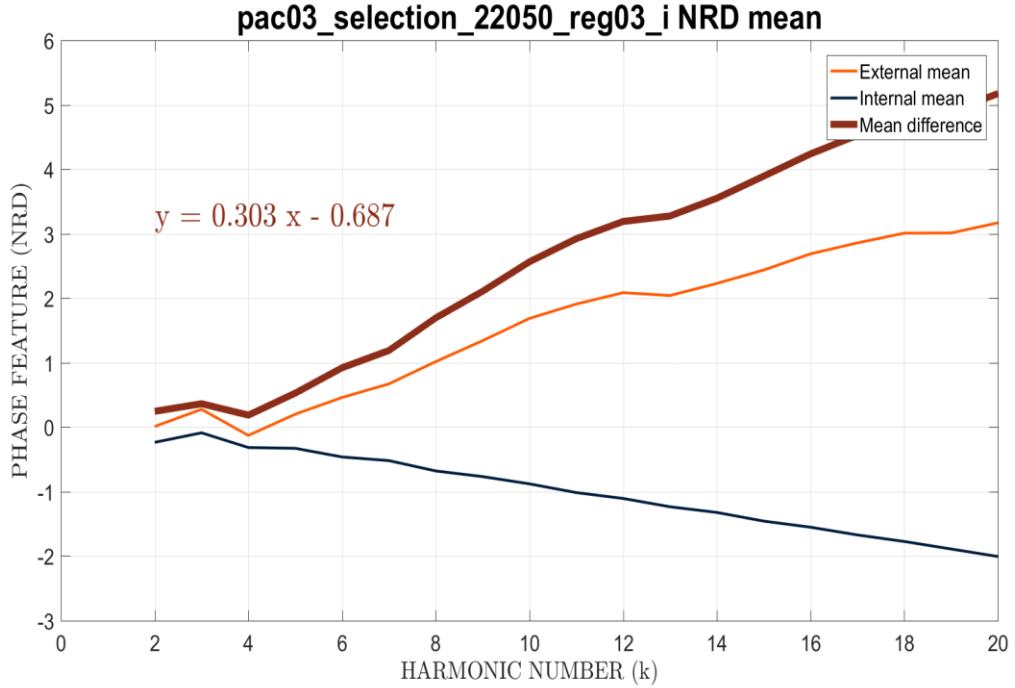
Speaker	Repetition #	/i/	/u/
2	1	-14.52	-13.61
	2	-13.06	-12.66
	3	-13.90	-12.78
3	1	-	-
	2	-12.72	-13.06
	3	-12.78	-12.30
4	1	-	-13.94
	2	-11.99	-13.36
	3	-11.09	-13.40
	$\overline{x}$	**-12.85**	**-13.15**
	σ	**± 0.83**	**± 0.45**
	$\overline{x}$/i/,/u/	**-13.01**	

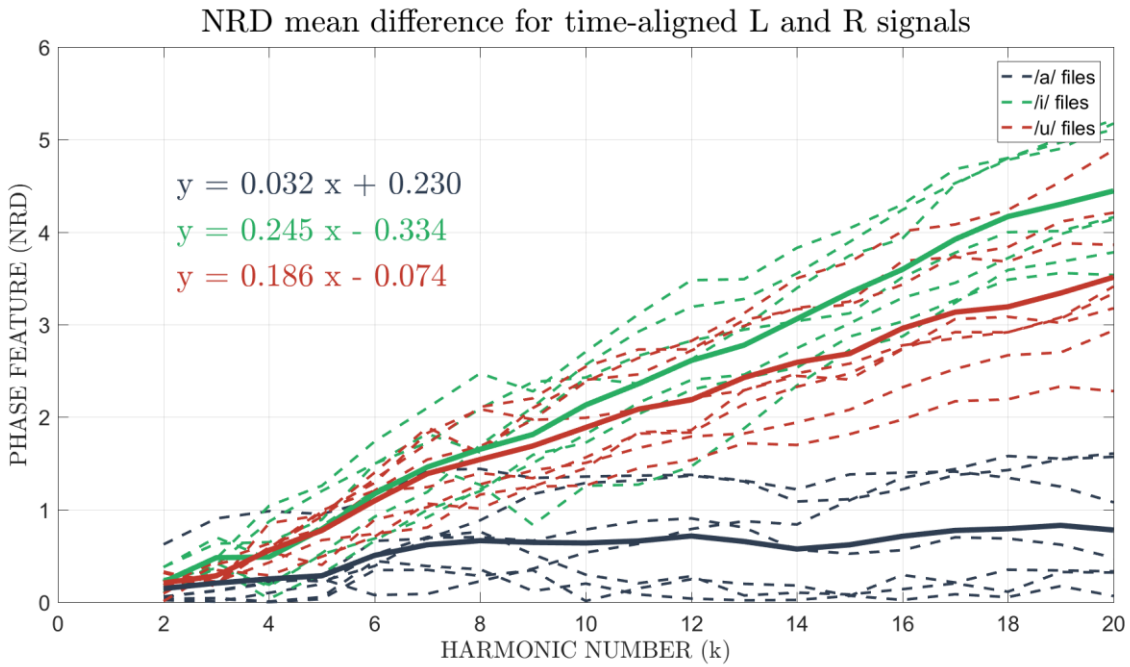
PHASE ANALYSIS



Speaker	Repetition #	/a/		/i/		/u/	
		L	R	L	R	L	R
2	1	0.084	0.077	-0.084	0.131	-0.119	0.031
	2	0.092	0.067	-0.029	0.132	-0.081	0.046
	3	0.101	0.098	-0.089	0.185	-0.123	0.138
3	1	-	-	-	-	-	-
	2	0.133	0.231	-0.077	0.161	-0.117	0.046
	3	0.106	0.186	-0.107	0.196	-0.147	0.078
4	1	-	-	-	-	-0.029	0.140
	2	0.148	0.169	-0.074	0.199	-0.058	0.124
	3	0.160	0.183	-0.037	0.156	-0.068	0.146
	$\overline{x}$	**0.118**	**0.144**	**-0.071**	**0.166**	**-0.093**	**0.094**
	σ	**±0.024**	**±0.055**	**±0.022**	**±0.023**	**±0.034**	**±0.043**

PHASE ANALYSIS





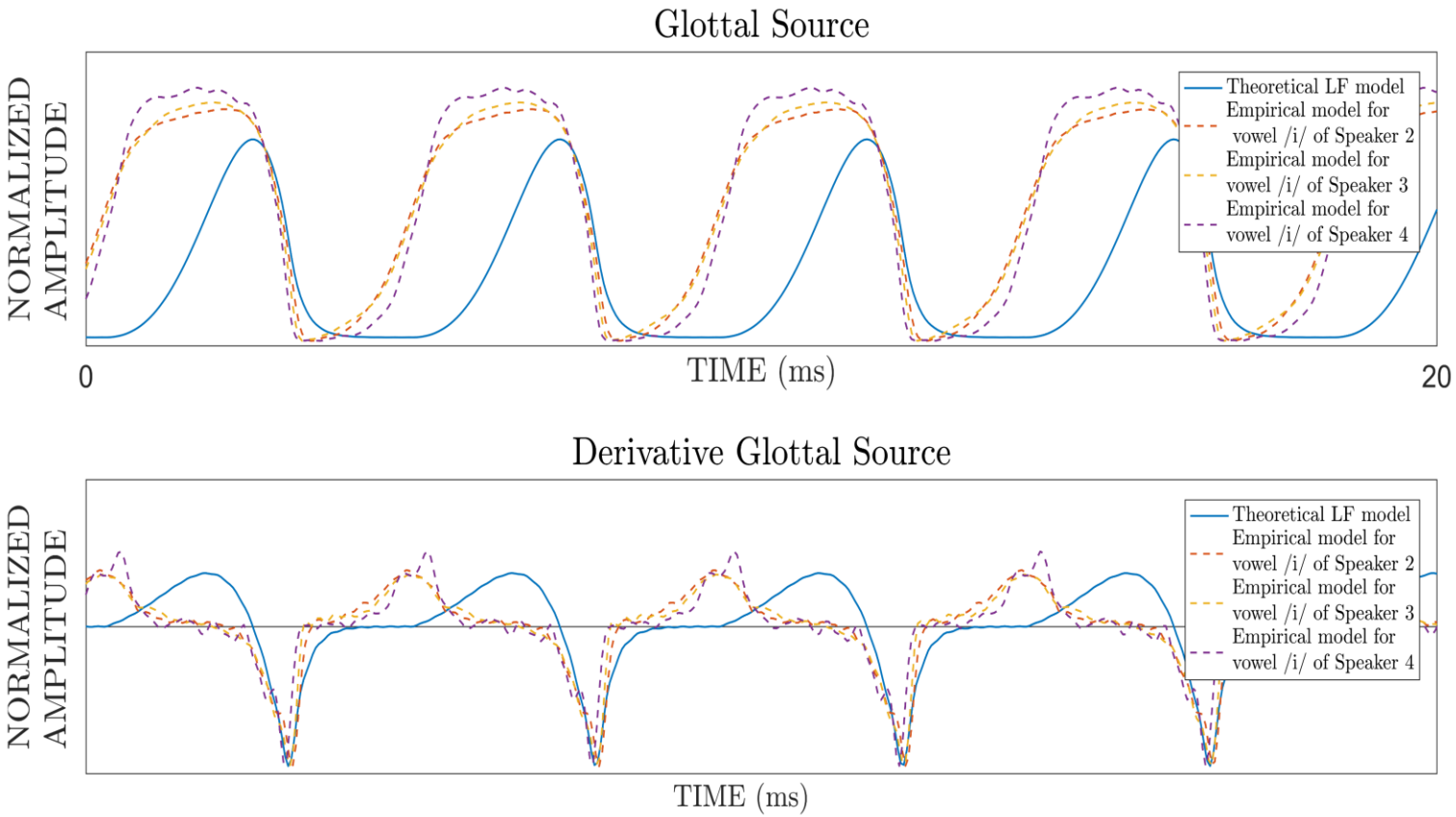
GLOTTAL SOURCE EMPIRICAL MODEL

$$gs(t) = \sum_{\ell=1}^{L} A_\ell \sin(\frac{2\pi}{T_0}\ell t + 2\pi NRD_\ell)$$

(glottal source synthesis)

$$dgs(t) = \sum_{\ell=1}^{L} \ell A_\ell \sin(\frac{2\pi}{T_0}\ell t + 2\pi NRD_\ell + \pi/2)$$

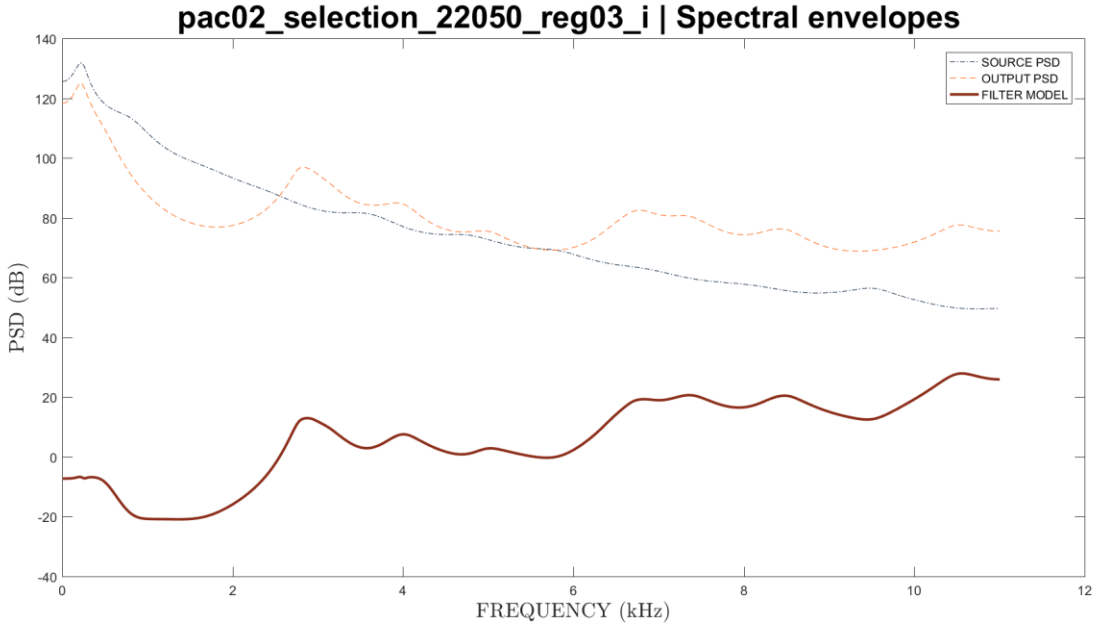
(glottal source derivative synthesis)

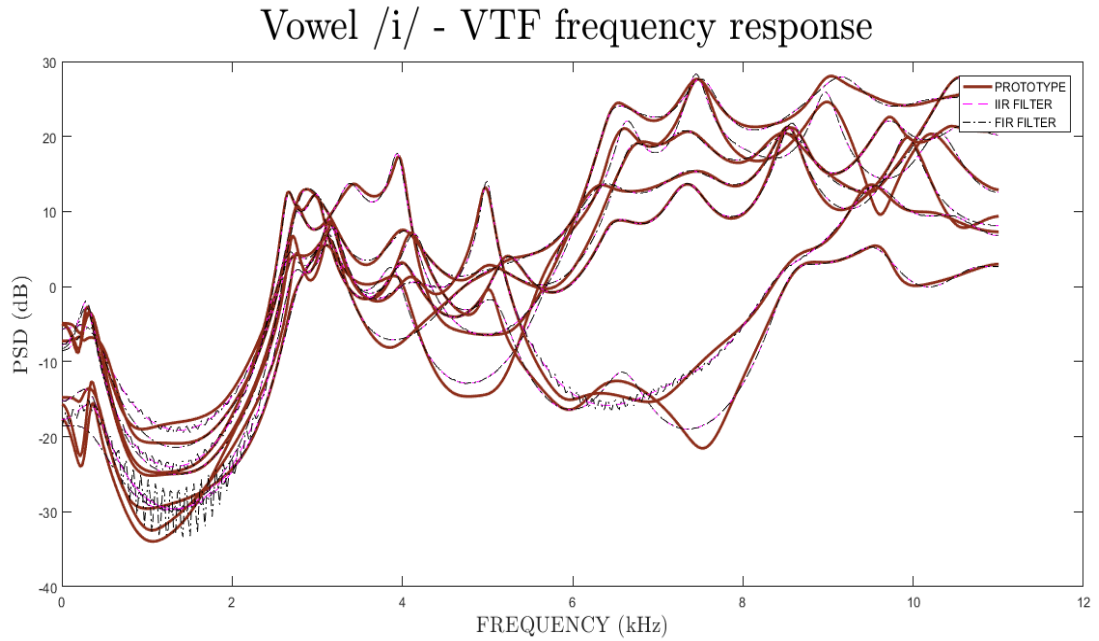


VOCAL TRACT FILTER ESTIMATION

1. Deconvolution approach
2. Adaptive filtering approach
3. Holistic filter design approach

- PSD of the prototype obtained from the difference between the spectral envelopes of the internal and external signals
- IIR and FIR filters design
  1. Autocorrelation using Wiener-Khintchine theorem
  2. Parameters of the 22nd order all pole model through Levinson-Durbin recursion
  3. Linear-phase property ensured by the use of a single band Parks-McClennan optimal equiripple design of order 500





SPEECH SYNTHETIC DATA

ORIGINAL SIGNAL /a/

ORIGINAL SIGNAL /i/

ORIGINAL SIGNAL /u/

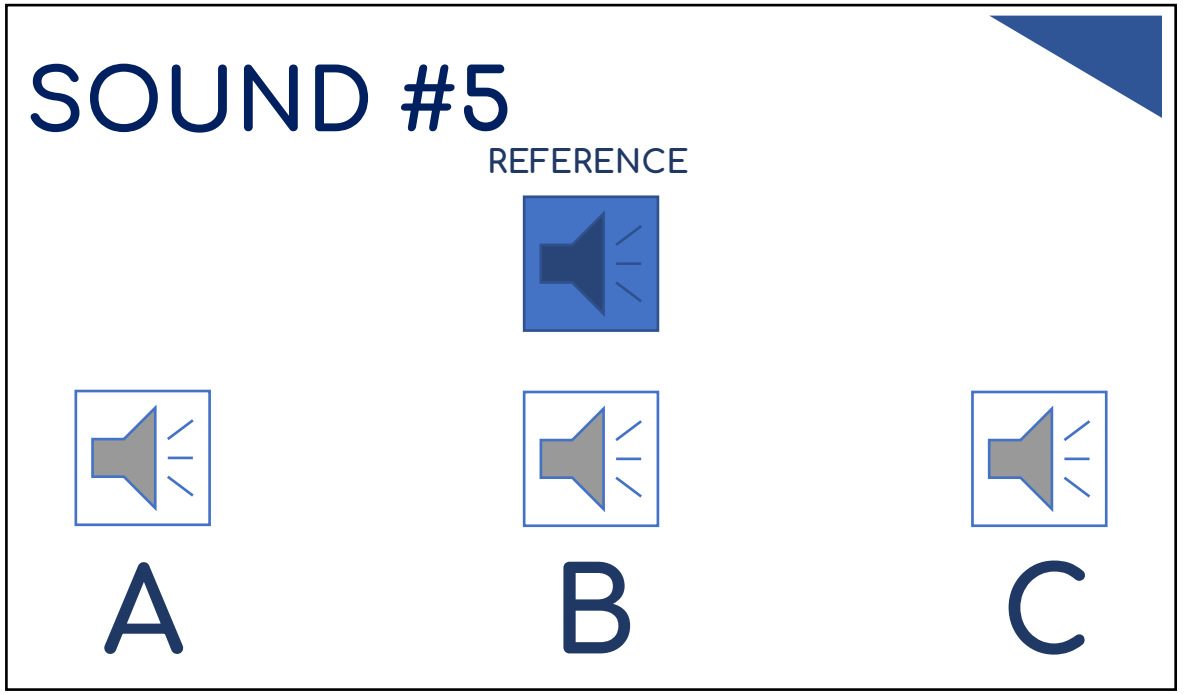
SYNTHETIC SIGNAL /a/

SYNTHETIC SIGNAL /i/

SYNTHETIC SIGNAL /u/

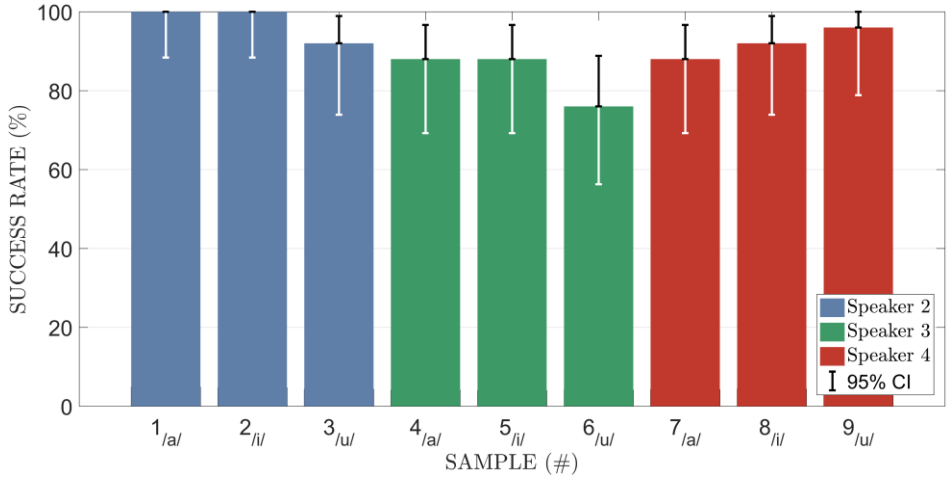
PERCEPTUAL TESTS

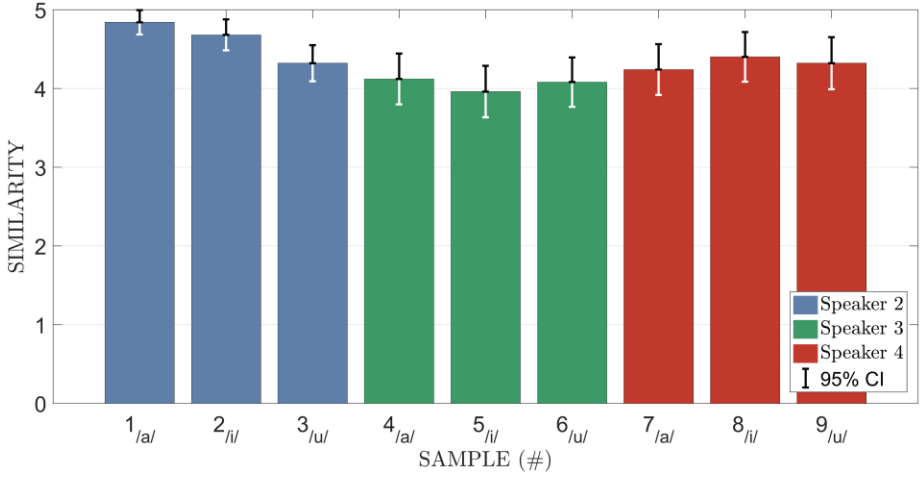
- **Test material:**
  - Synthetic signals generated using the real glottal source signals and the corresponding FIR filter obtained for each repetition according to speaker and vowel
- **Task:**
  - Question #1 – identify for the same vowel which of the 3 synthetic samples corresponds to the real speech sample given as reference
  - Question #2 – grade from 1 (low) to 5 (high) the similarity degree between the sample chosen and the sample given as reference
- **Evaluation:**
  - Correct speaker identification success rate
  - Similarity degree given by the participants



PERCEPTUAL TESTS RESULTS

- **Correct identification** of the speakers given as reference by the **majority** of the participants
- **Speaker 2** (blue) was the **most** and **Speaker 3** (green) the **least** accurately identified speaker
- Synthetic signals **resemble** the natural signals, since every sample achieved a mean score **above 4 out of 5** (very similar)
- Mean **success rate** of 91.11% and mean degree of **similarity** of **4.33**





TAKE HOME MESSAGES

- The waveshape of the internal signals shows **discrepancies** when compared to the **glottal source waveshape** of theoretical models.
- The internal recordings of the sustained vowel /a/ show a **significant influence of the vocal tract filter** when compared to vowels /i/ and /u/, which supports the idea that source and filter have **non-linear interactions**.
- The spectral slope of the **empirical spectral magnitude model** of both sustained vowels /i/ and /u/ of -13 dB/oct approximates more to the Rosenberg reference value of **-12 dB/oct**, rather than to the LF reference value of **-16 dB/oct**.
- The slope obtained for the **empirical spectral NRD models** of the internal and external signals resulted respectively in values of 0.118 and 0.144 for the vowel /a/, -0.071 and 0.166 for the vowel /i/ and -0.093 and 0.094 for the vowel /u/. The slope value obtained for the mean differences between the external and internal signals for the vowel /a/ was 0.032, for the vowel /i/ was 0.246 and, lastly, for the vowel /u/ was 0.186.
- The estimation of VTF was possible using the **holistic design filter approach**. However, **no difference** was noticed between the synthetic vowels generated with the **IIR and FIR filters**. Perceptual tests confirmed the **correct estimation of the VTF** with a **mean success rate** of 91,11% and a **mean degree of similarity** of 4,33.

FUTURE WORK

- Improve the dataset and record for **a larger variety of speakers** and **different types of phonation** (e.g. whispering)
- Estimate the glottal source using different state-of-the-art techniques developed in more recent studies and compare it with the glottal source empirical models obtained, in order to **validate** that the **recorded signal** obtained corresponds to the **real glottal source signal**
- Compare the empiric glottal source model obtained and the theoretical glottal source models described in the literature, regarding the **relation between the glottal source derivative behaviour and the physiological events**
- Study the cause of the difference in **NRD slope polarity** between the signals captured internally and externally, in the case of /i/ and /u/ vowels, which is probably related the **acoustic radiation effects**.

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

CINTESIS
Health. Research.

BRUNO MIGUEL SILVA SANTOS
up201504192@fe.up.pt

The glottal source signal still remains an undiscovered topic , however we hope to have opened the door to more research and developments in this field

Cofinanciado por:
COMPETE 2020
PORTUGAL 2020
UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional
FCT
Fundação para a Ciência e a Tecnologia