

FCT PROJECT “ASSISTIVE REAL-TIME TECHNOLOGY IN SINGING”

ESTIMATION OF THE GLOTTAL PULSE FROM SPEECH OR SINGING VOICE

Sandra Dias and Aníbal Ferreira

19.jan.2013

FEUP

TOPICS

- **Objectives**
- **Voice Production Models**
 - Glottal Pulse Models
 - Source-Filter Model
- **Glottal Pulse Estimation: Inverse Filtering**
- **Estimation of the Glottal Source in the Frequency Domain**
 - Principles
 - Physiological Signal Acquisition for Source and Filter Modelling
 - Hybrid LF-Rosenberg Glottal Source Model
 - New approach to glottal source estimation
 - Tests of the new approach in synthetic and real speech signals
- **Conclusions**

OBJETIVES

- **Robust glottal source estimation**

- Implemented in the frequency domain
- Based on the Normalized Relative Delay (NRDs) of the harmonics
- Included in Task 5 of the Project:

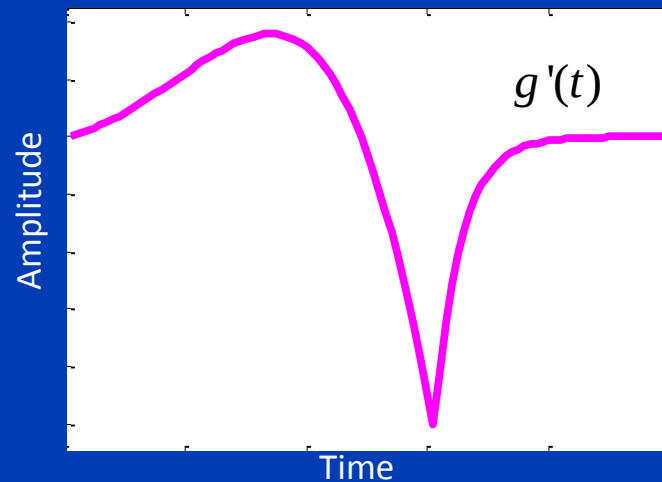
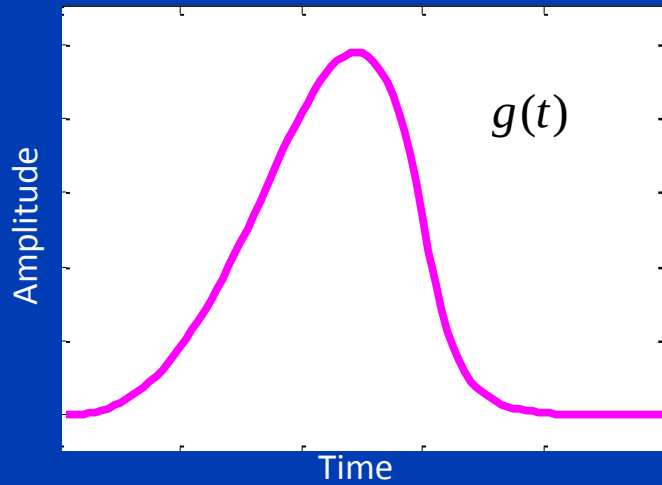
Robust real-time glottal pulse estimation from running singing

- **Contribute to:**

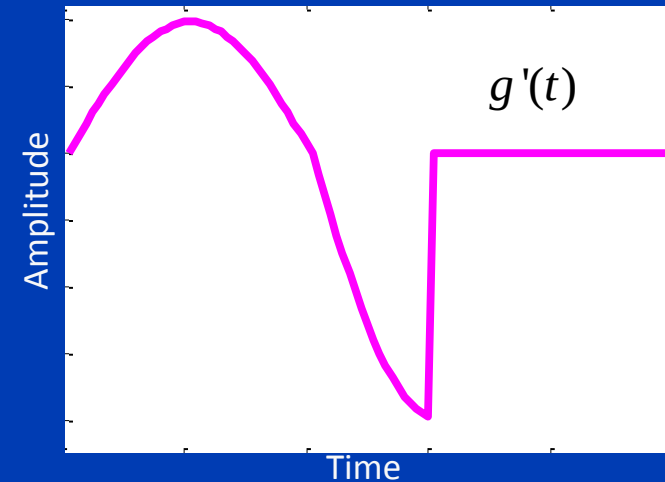
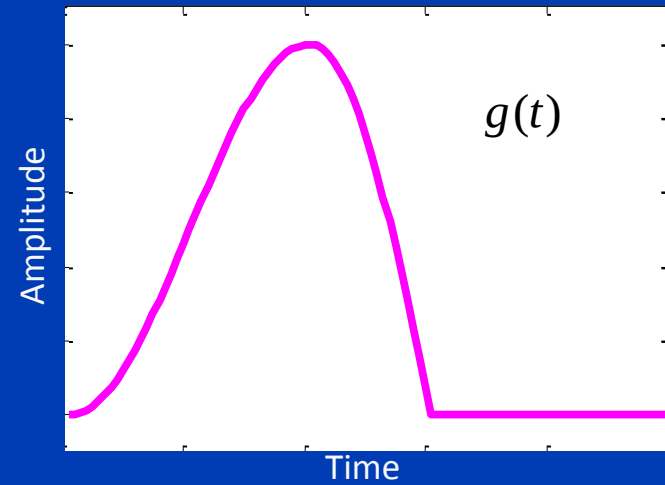
- the development of non-invasive procedures of estimation of the glottal pulse
- enhance scientific knowledge about the glottal pulse and speech analysis

GLOTTAL PULSE MODELS

- Liljencrants-Fant Model (LF)



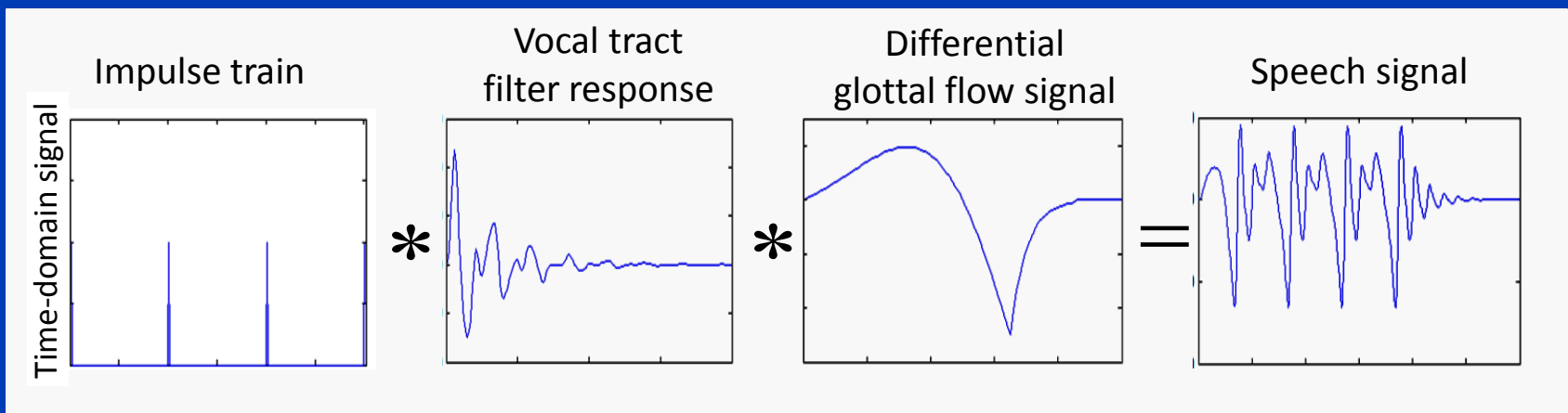
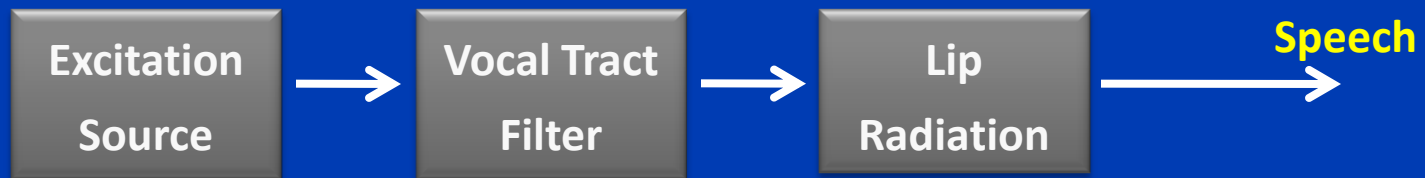
- Rosenberg Model (Ros)



SOURCE-FILTER MODEL

Based on Fant's source-filter theory:

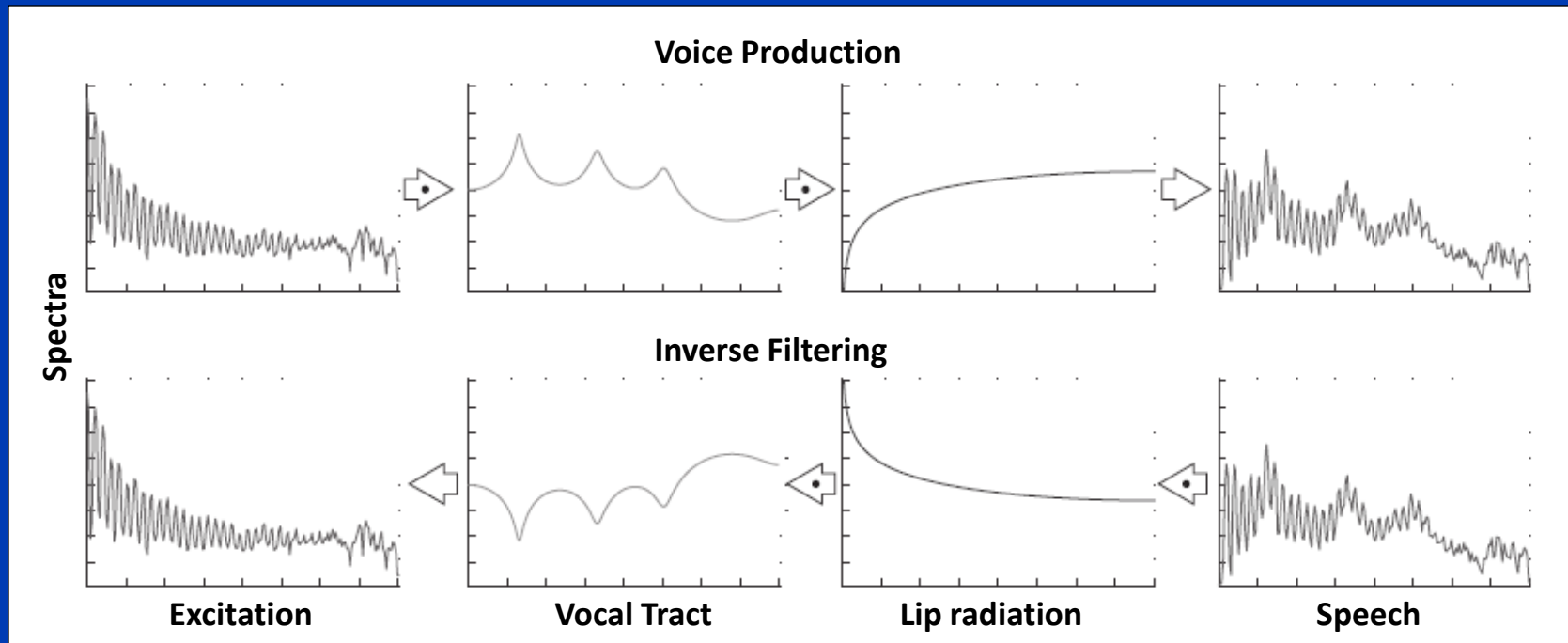
**The voice is the result of the convolution between
the excitation source and the filter system.**



GLOTTAL PULSE ESTIMATION: INVERSE FILTERING

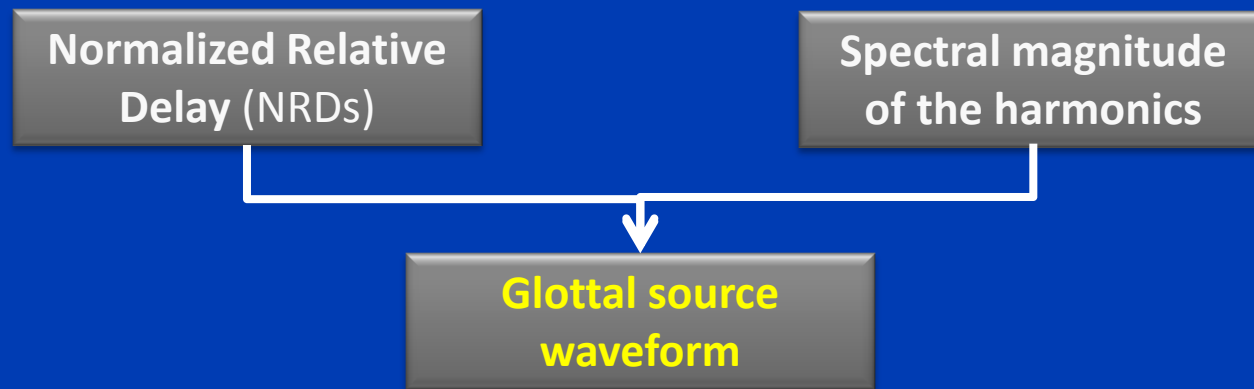
GOAL:

Estimate the glottal pulse by cancelling the spectral effects of the vocal tract and lip radiation on a speech signal.



ESTIMATION OF THE GLOTTAL SOURCE IN THE FREQUENCY DOMAIN

PRINCIPLES:

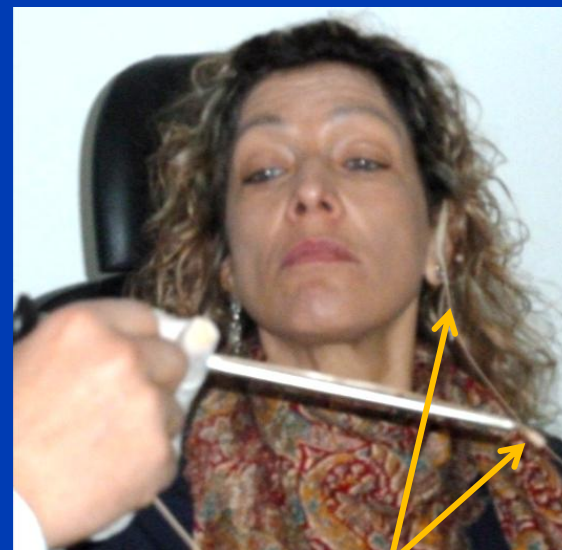
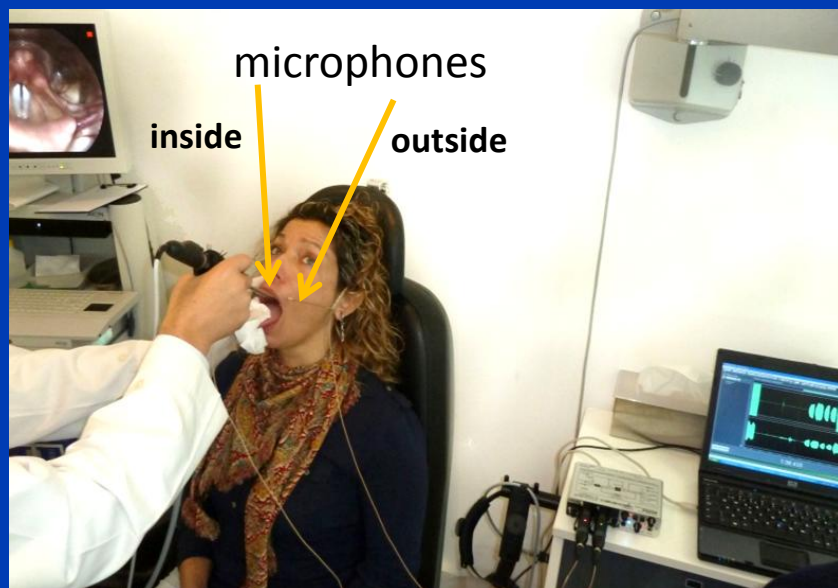


NORMALIZED RELATIVE DELAY (NRDs):

Relative delay difference between a harmonic and the fundamental frequency sinusoid.

ESTIMATION OF THE GLOTTAL SOURCE IN THE FREQUENCY DOMAIN

PSYCHOLOGICAL SIGNAL ACQUISITION FOR SOURCE AND FILTER MODELLING



Capture, at the same time:

- the acoustic signal with a microphone as close as possible to the glottis
- the speech signal outside.

(3 male and 3 female subjects)

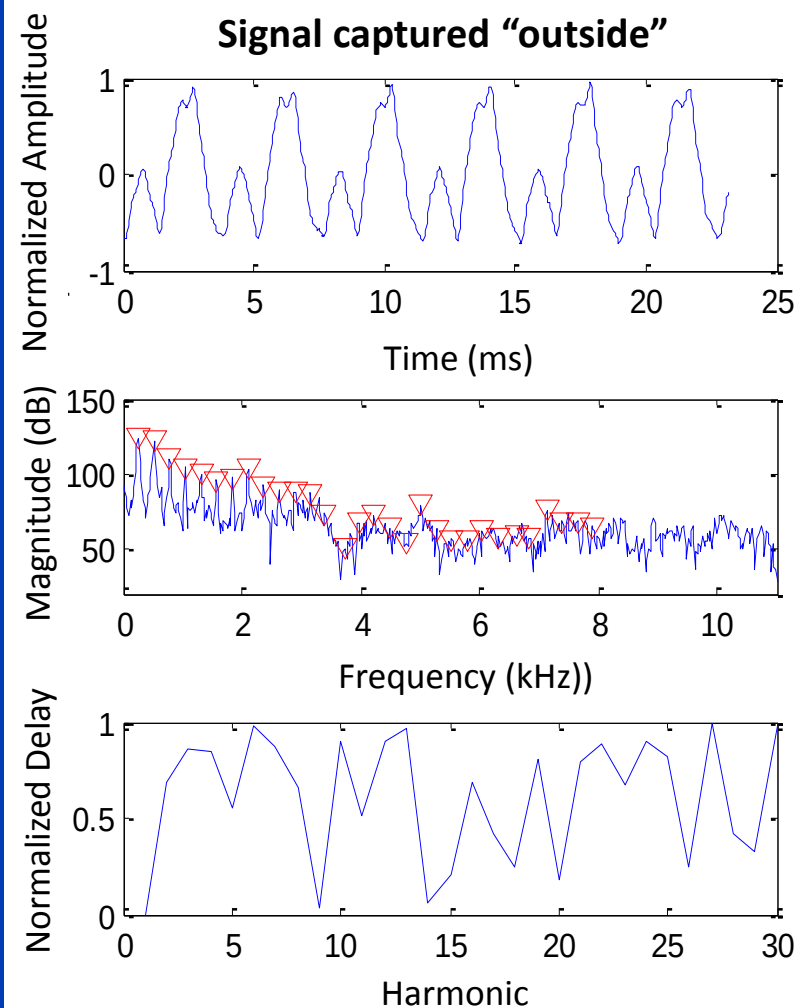
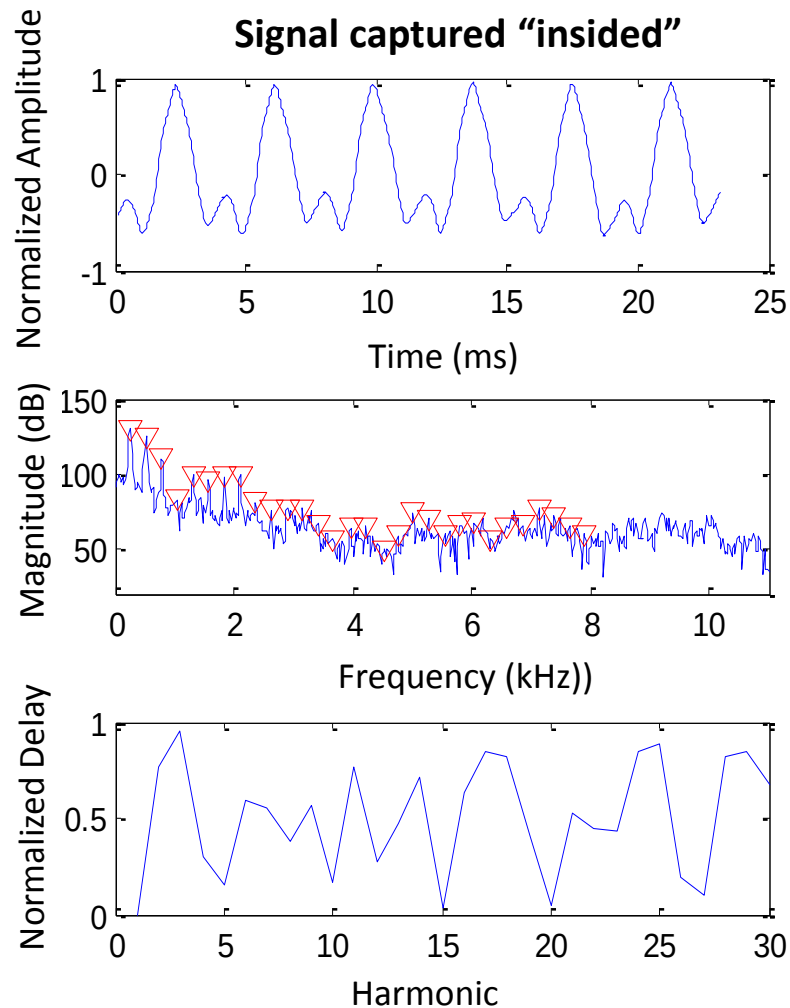


Objectives

Characterize:

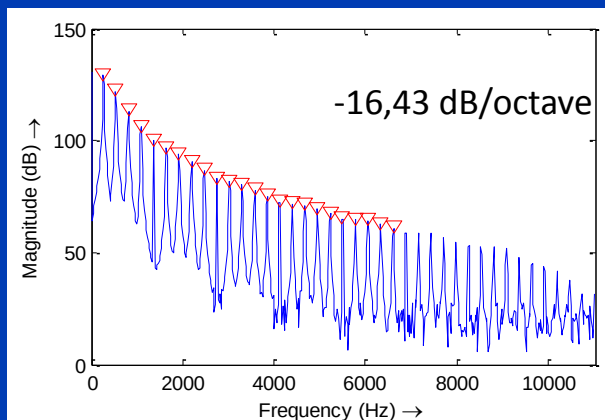
- the source signal
- the frequency response of the vocal tract filter, namely regarding the group delay.

ESTIMATION OF THE GLOTTAL SOURCE IN THE FREQUENCY DOMAIN

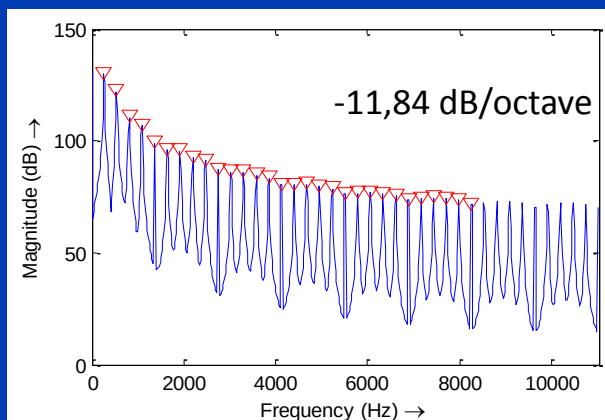


ESTIMATION OF THE GLOTTAL SOURCE IN THE FREQUENCY DOMAIN

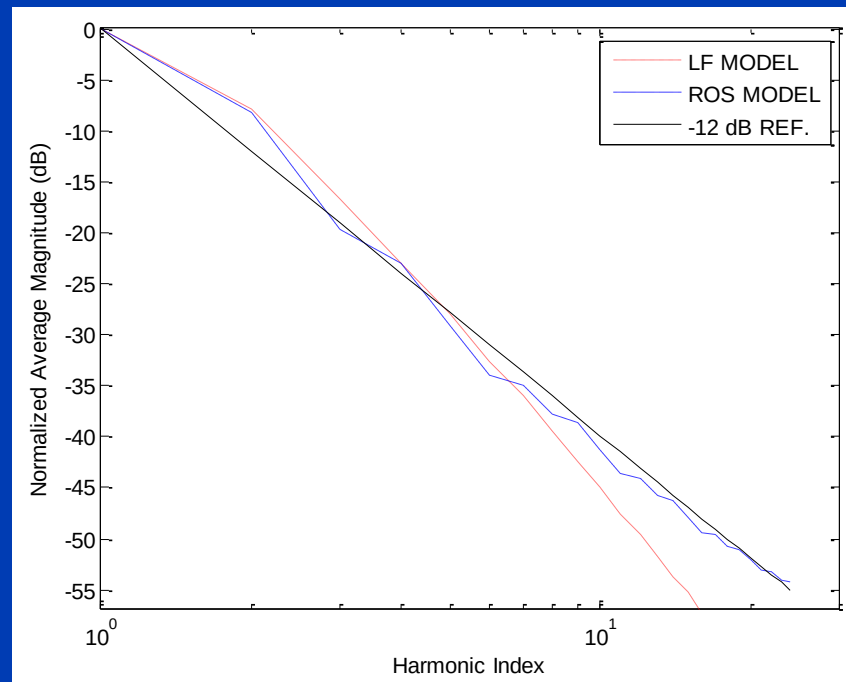
LEAST SQUARES REGRESSION MODEL OF THE MAGNITUDE OF THE HARMONICS



Magnitude Spectrum of the LF model



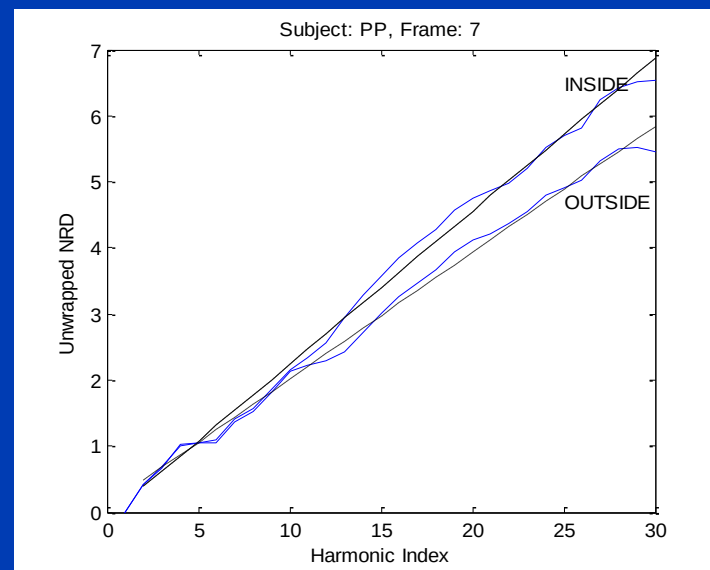
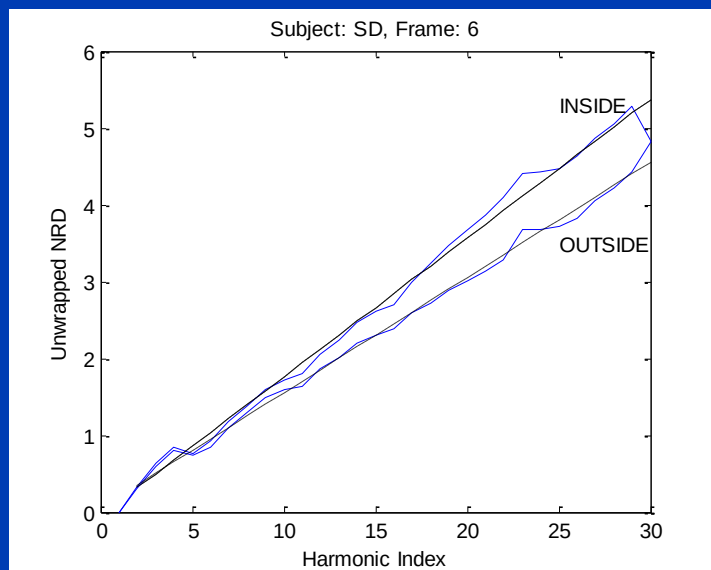
Magnitude Spectrum of the ROS model



SLOPE	Average /a/	DP /a/	Average /i/	DP /i/
Male	-9,6443	2,3266	-13,7305	2,8674
Female	-9,9334	3,3521	-13,6694	2,2575
AVERAGE	-9,7889	3,3949	-13,7000	2,2693

ESTIMATION OF THE GLOTTAL SOURCE IN THE FREQUENCY DOMAIN

LINEAR REGRESSION MODEL OF THE UNWRAPPED NRDs

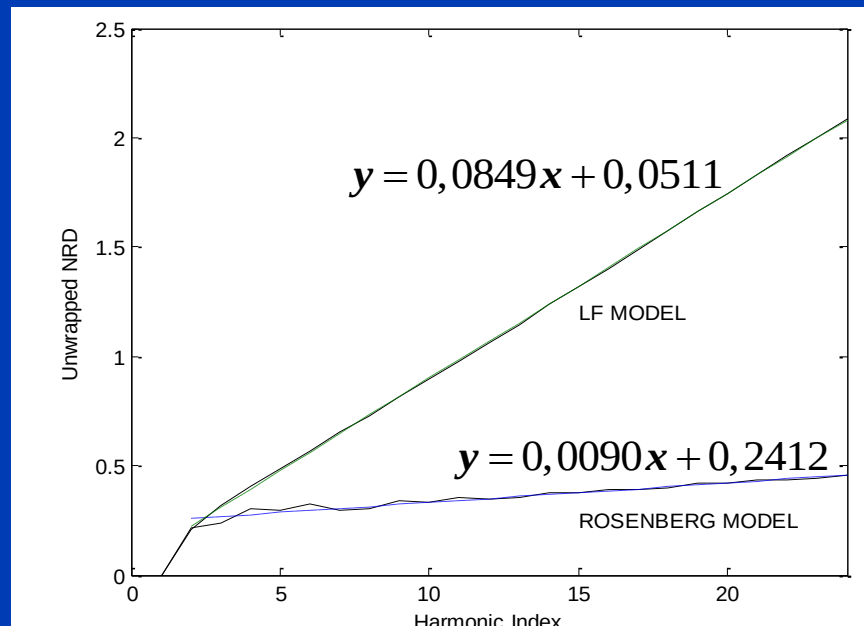


Difference, in modulus, between the slopes of the linear regression models corresponding to the unwrapped NRDs of the signals captured “inside” and “outside”

SLOPE	Average /a/	DP /a/	Average /i/	DP /i/
Male	0,0436	0,0063	0,0527	0,0463
Female	0,0422	0,0148	0,0868	0,0326
AVERAGE	0,0429	0,0108	0,0697	0,0421

ESTIMATION OF THE GLOTTAL SOURCE IN THE FREQUENCY DOMAIN

LINEAR REGRESSION MODEL OF THE UNWRAPPED NRDs



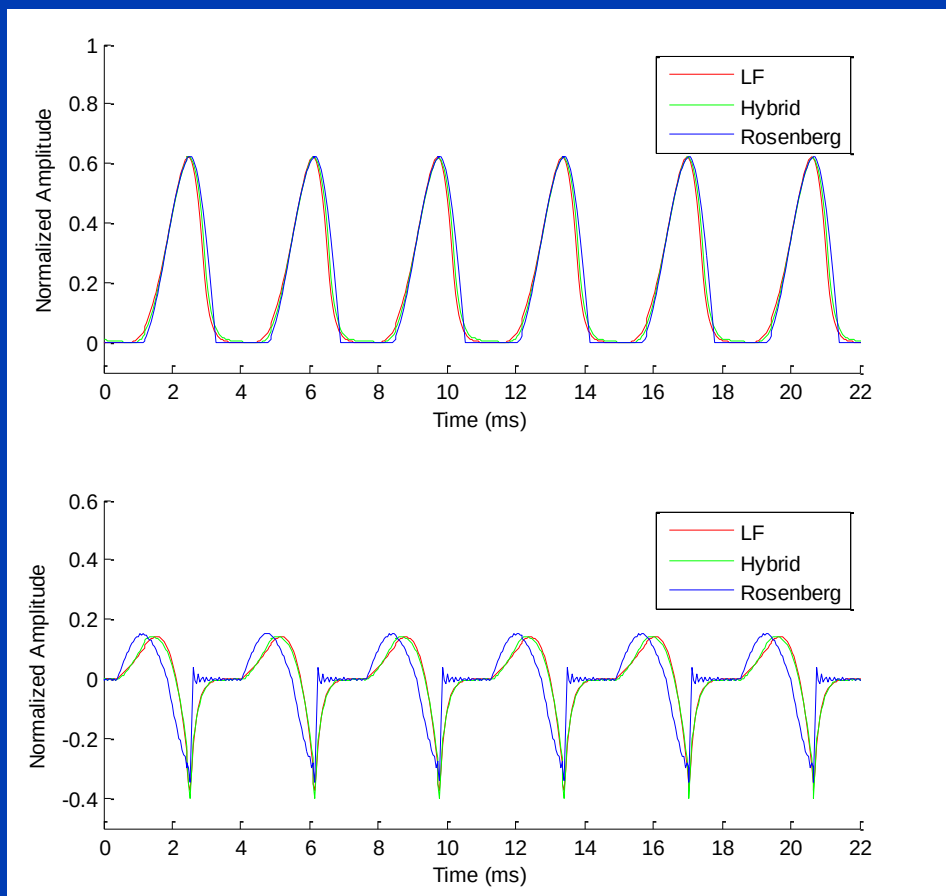
Unwrapped NRDs of the LF and Rosenberg glottal pulse models (solid lines) and the corresponding linear regression models (dashed lines).

Slope of the linear regression models corresponding to the unwrapped NRDs of the signals captured “inside”

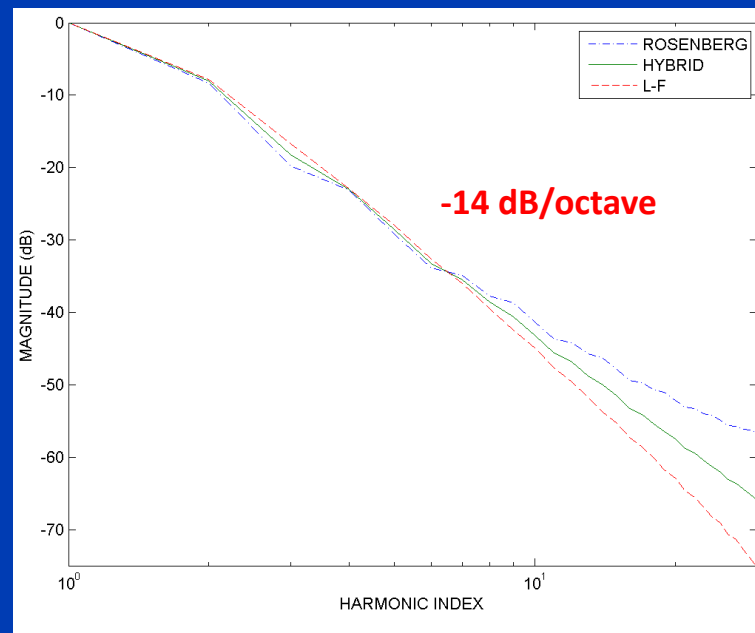
SLOPE	Average /a/	DP /a/	Average /i/	DP /i/
Male	0,1440	0,1055	0,0563	0,0887
Female	0,1311	0,0554	0,0981	0,0712
MÉDIA	0,1375	0,0806	0,0772	0,0797

ESTIMATION OF THE GLOTTAL SOURCE IN THE FREQUENCY DOMAIN

HYBRID LF-ROSENBERG GLOTTAL SOURCE MODEL



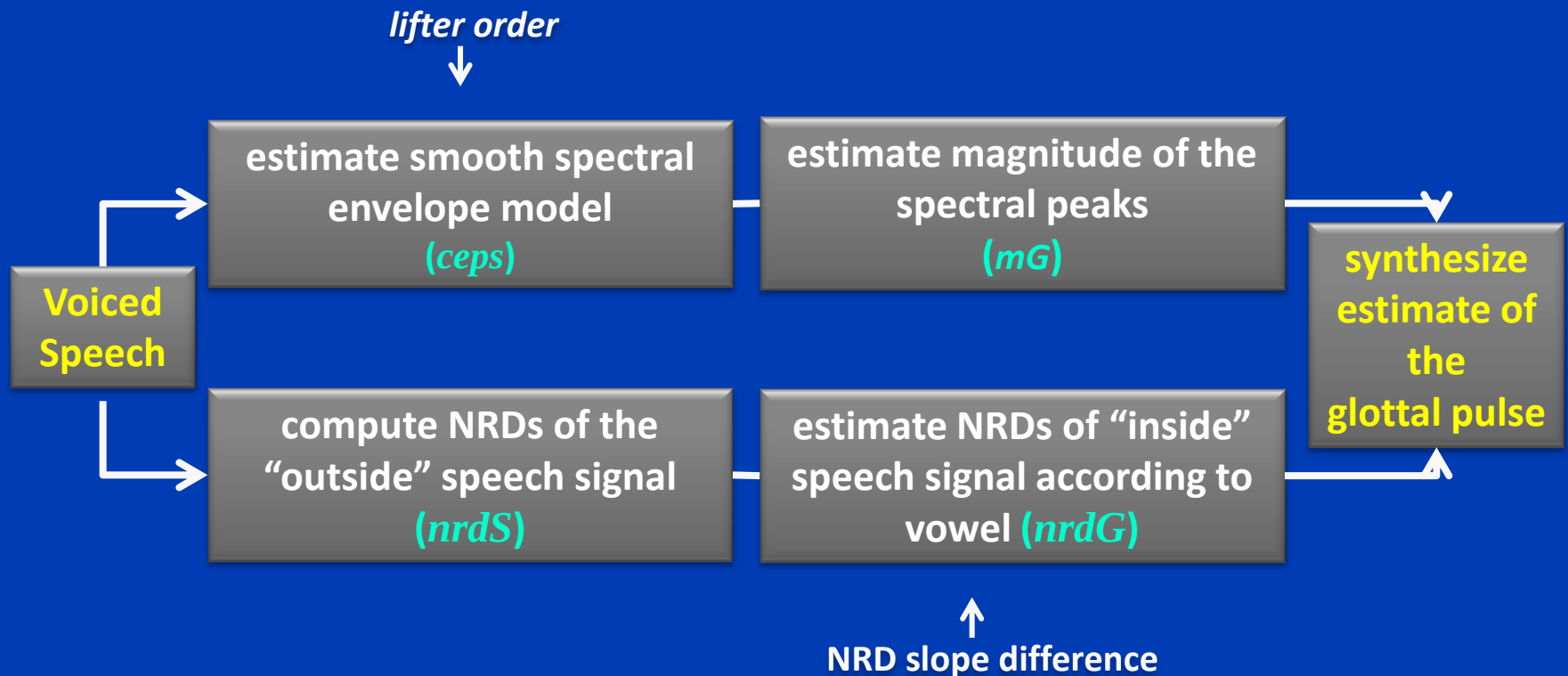
Time representation of the glottal pulse models and the corresponding derivatives.



Normalized magnitude of the first harmonics pertaining to the LF model, the Rosenberg model and the hybrid model.

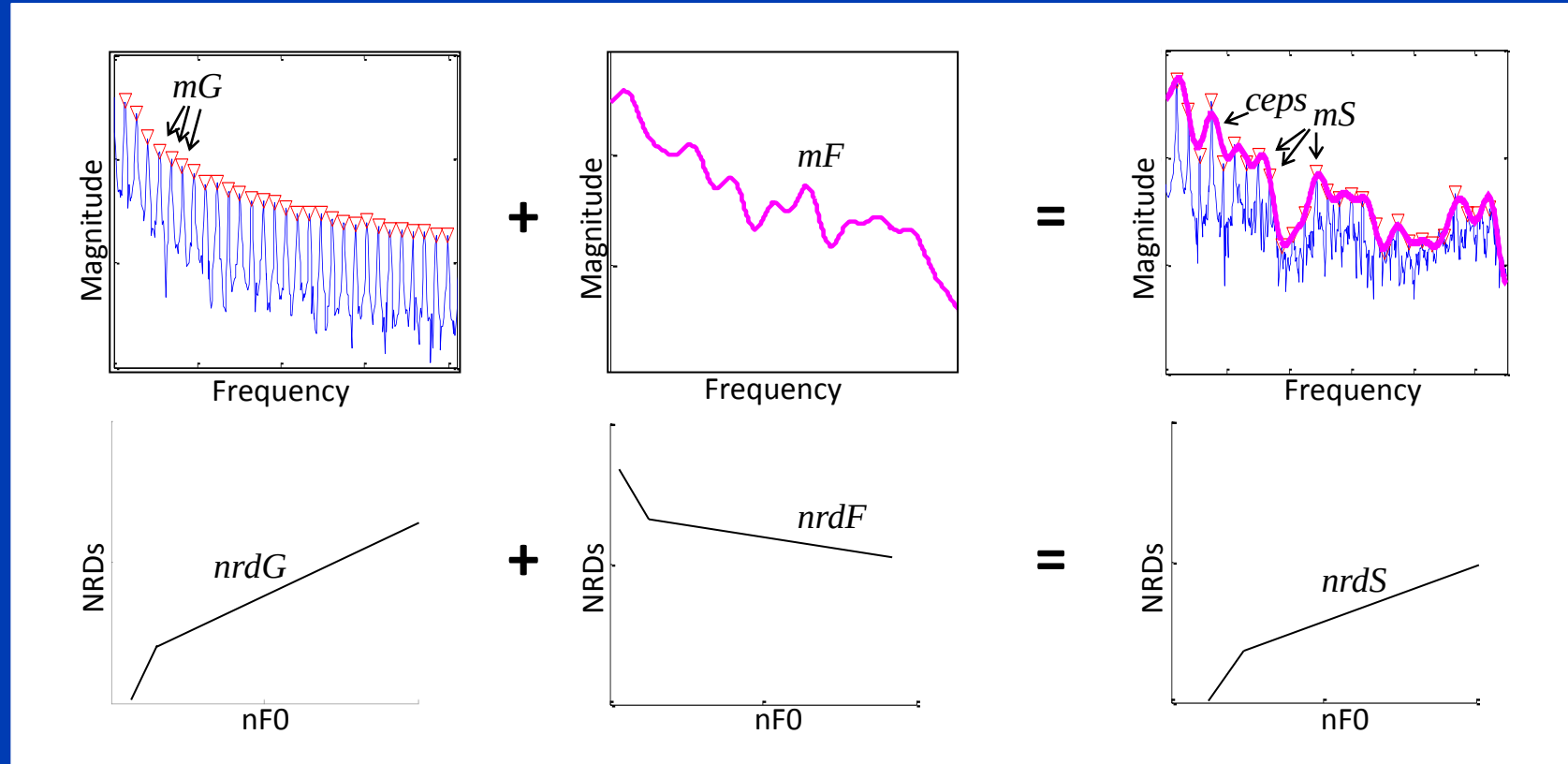
ESTIMATION OF THE GLOTTAL SOURCE IN THE FREQUENCY DOMAIN

NEW APPROACH TO GLOTTAL SOURCE ESTIMATION



ESTIMATION OF THE GLOTTAL SOURCE IN THE FREQUENCY DOMAIN

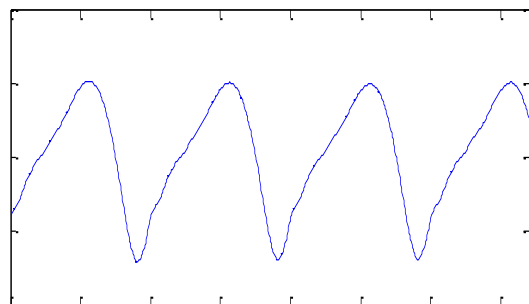
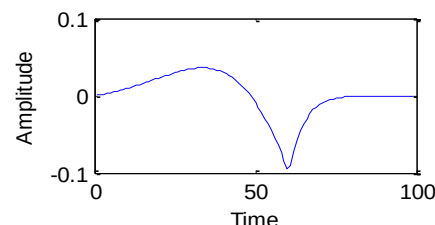
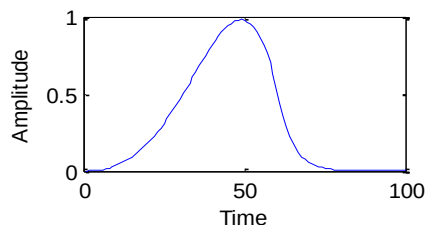
NEW GLOTTAL SOURCE ESTIMATION APPROACH



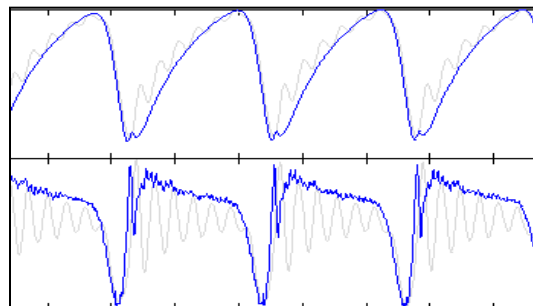
ESTIMATION OF THE GLOTTAL SOURCE IN THE FREQUENCY DOMAIN

TESTS WITH SYNTHETIC SPEECH SIGNALS OF VOWEL /A/

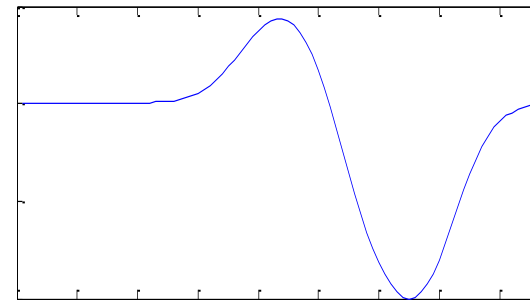
- LF glottal model



The glottal flow waveform estimated by the **new approach**



The glottal flow (top row) and derivative (bottom row) waveforms estimated by **IAIF**



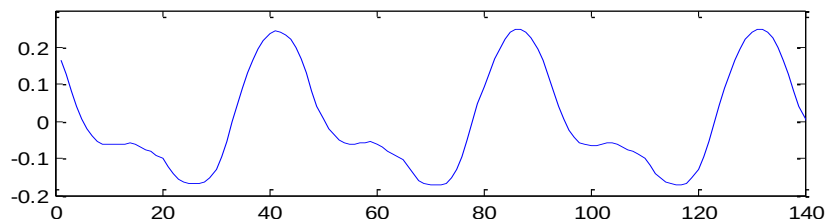
The glottal flow derivative estimation using **CC decomposition**

SNR	Male synthetic vowel /a/	Male synthetic vowel /i/	Female synthetic vowel /a/	Female synthetic vowel /i/	Average
IAIF	6,3832	6,0905	9,7968	8,7182	7,7472
New approach	7,3010	8,7168	9,7475	8,8672	8,6581

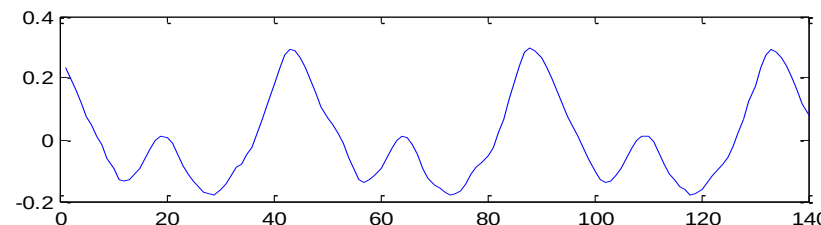
ESTIMATION OF THE GLOTTAL SOURCE IN THE FREQUENCY DOMAIN

TESTS WITH SPEECH SIGNALS FROM OUR DATA BASE

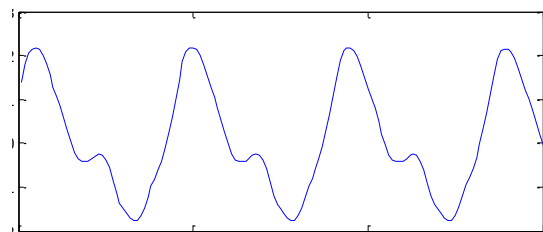
- Female vowel /i/



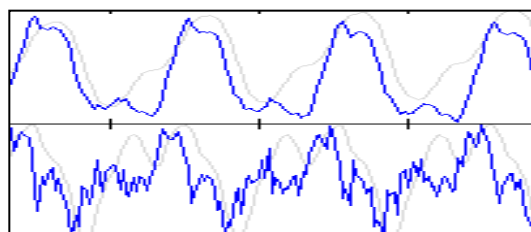
Signal captured "inside"



Signal captured "outside"

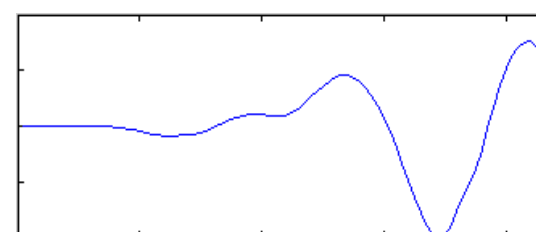


The glottal flow waveform estimated by the **new approach**



The glottal flow (top row) and derivative (bottom row) waveforms estimated by

IAIF



The glottal flow derivative estimation using **CC decomposition**

SNR	Female vowel/i/ (subject SD)	Female vowel /i/ (subject SF)	Female vowel /a/ (subject SF)	Average
IAIF	9,5403	10,3138	8,9982	9,6174
New approach	15,8266	11,2107	11,8187	12,952

CONCLUSIONS

WHATS NEW IN THIS STUDY?

- **Data base of speech signals time aligned:** the signal captured “inside” and “outside”
- **A new hybrid model** combining features of two reference glottal models - the Liljencrants-Fant and Rosenberg
- **A new frequency-domain approach to glottal inverse filtering**, using a phase-related feature based on the **Normalized Relative Delays (NRDs)** of the harmonics and an accurate determination of the **magnitude of the spectral peaks of the harmonics** of the speech signal

CONCLUSÕES

MAIN RESULTS:

- In the **estimation of the glottal source**, the **new approach** exhibited **better performance** when compared with representative state-of-the-art techniques (IAIF and CC decomposition).
- The results and the signals captured inside suggested that the **glottal pulse estimations do not have the bell-shape** and, therefore, that **idealistic glottal pulse models do not entirely correspond to real glottal pulse**, namely on the closing and closed phases.
- **Importance of phase features in speech analysis and synthesis**, namely the **NRDs**.