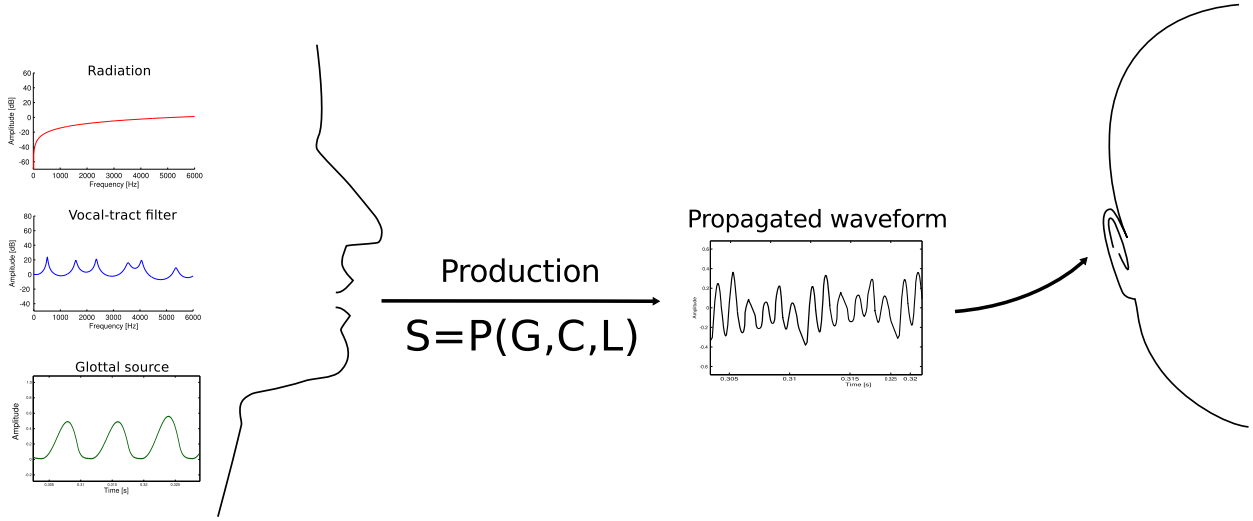
Function of phase-distortion for glottal model estimation

Gilles Degottex, Axel Roebel, Xavier Rodet

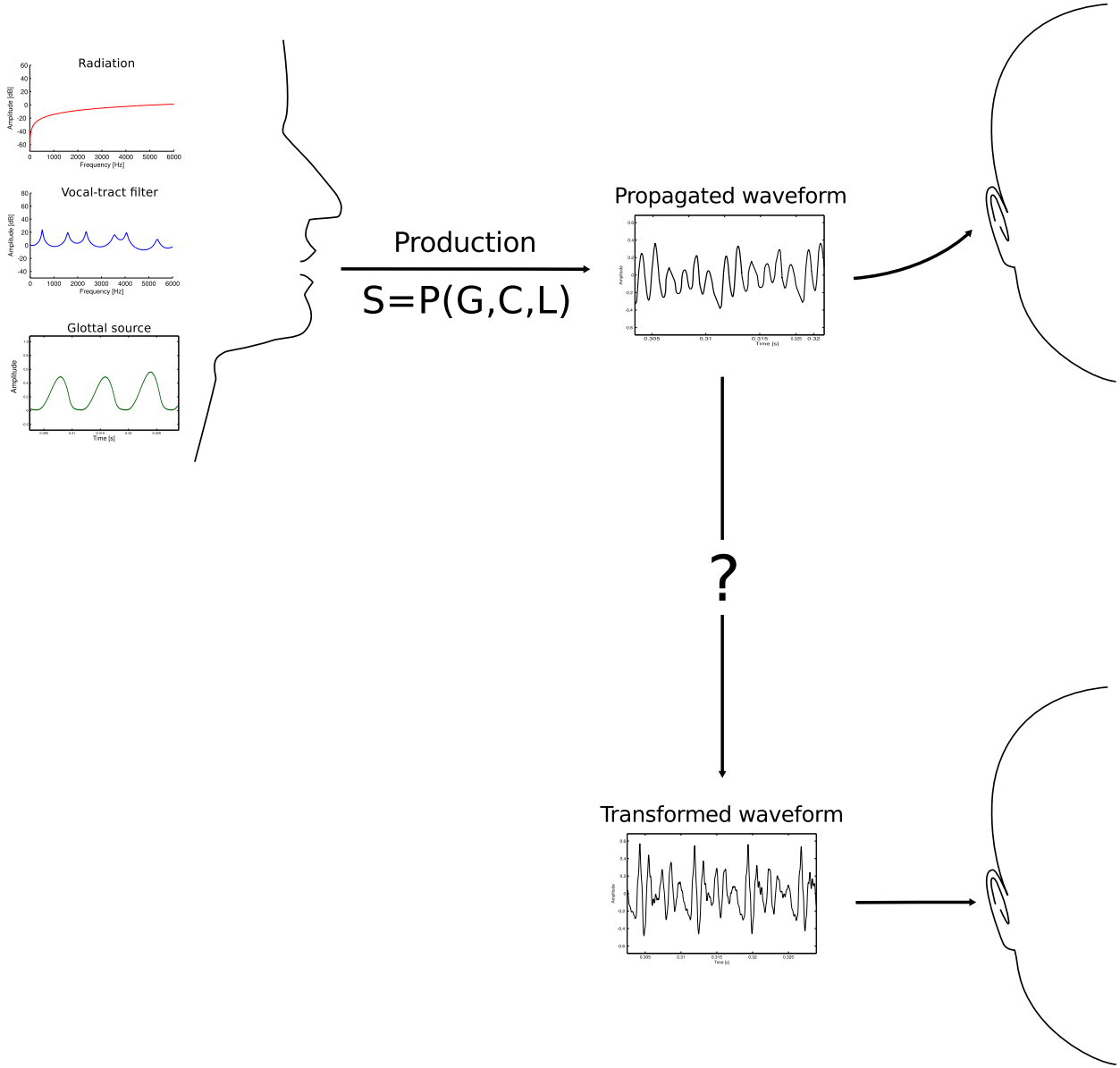
Institut de Recherche et Coordination Acoustique/Musique (Ircam), Analysis-Synthesis Team
Centre National de la Recherche Scientifique (CNRS) UMR9912-STMS
Swiss National Science Foundation (SNSF)

Introduction
Research context

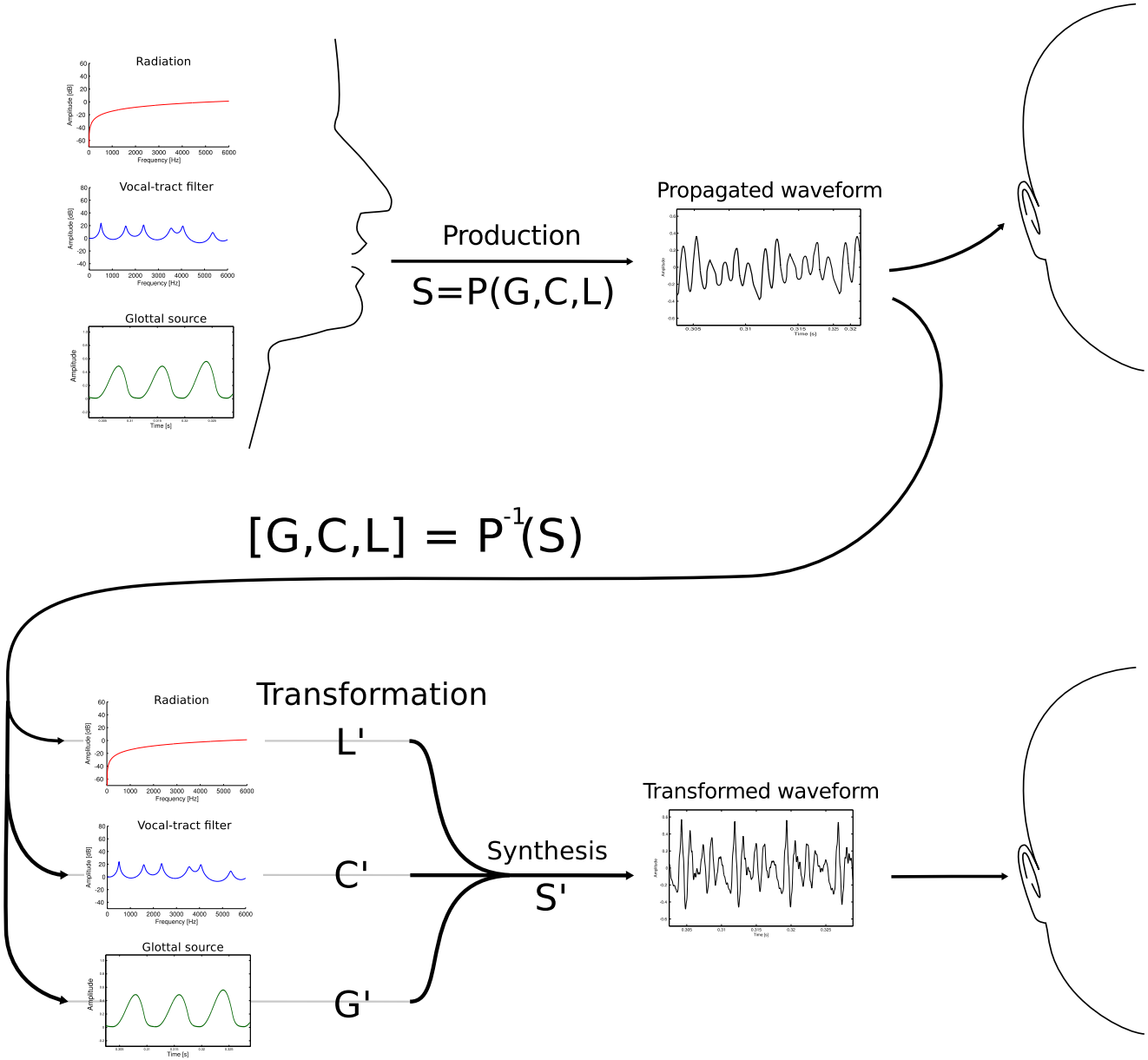
Voice production



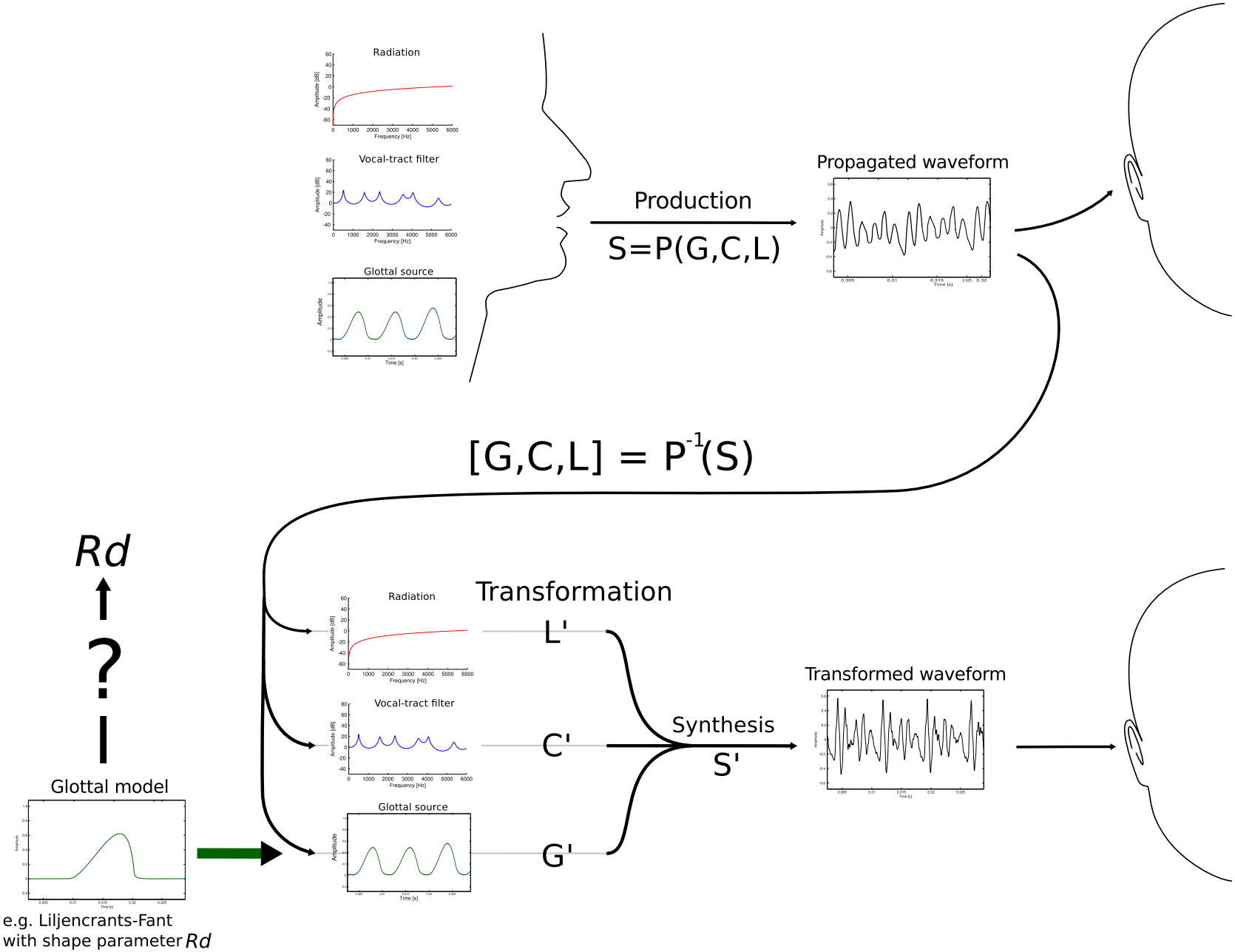
Voice transformation



Voice separation



Glottal model estimation

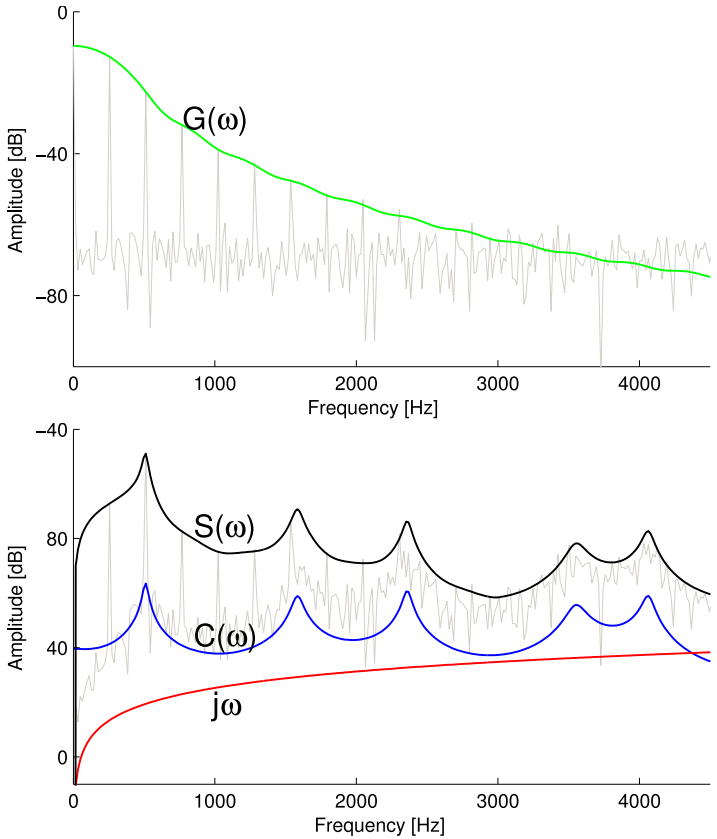


Model
of the voice production

Voice production model

$$S(\omega) = \overbrace{H^{f_0}(\omega) \cdot e^{j\omega\phi} \cdot G^{Rd}(\omega)}^{\text{Glottal source}} \cdot \overbrace{C_-(\omega)}^{\text{Vocal-tract}} \cdot \overbrace{j\omega}^{\text{Radiation}}$$

- $G^{Rd}(\omega)$ Shape of the glottal model
- $e^{j\omega\phi}$ Position of the glottal pulse
- $H^{f_0}(\omega)$ Periodicity
- $C_-(\omega)$ Vocal-tract filter (minimum-phase)
- $j\omega$ Radiation



Harmonic model

hyp: f_0 is known $\Rightarrow$ **use a harmonic model:**

$$\begin{array}{ccccccccc} S(\omega_h) & = & e^{j\omega_h\phi} & \cdot & G^{Rd}(\omega_h) & \cdot & C_-(\omega_h) & \cdot & j\omega_h \\ S_h & = & e^{jh\phi} & \cdot & G_h^{Rd} & \cdot & C_{h-} & \cdot & jh \end{array}$$

Notation using indices!

proposed idea

Function of Phase Distortion (FPD)

Function of Phase Distortion (FPD)

We propose the **Function of Phase Distortion (FPD)** of X_h:

$$\Phi_k(X_h) = \Delta^{-1}\Delta^2 \angle \left(\frac{X_h}{\mathcal{E}_-(X_h)} \right)$$

Function of Phase Distortion (FPD)

We propose the **Function of Phase Distortion (FPD)** of X_h:

$$\Phi_k(X_h) = \Delta^{-1}\Delta^2 \angle \left(\frac{X_h}{\mathcal{E}_-(X_h)} \right)$$

$\mathcal{E}_-(.)$ Minimum-phase realization
$\Rightarrow$ e.g. remove the vocal-tract filter

Function of Phase Distortion (FPD)

We propose the **Function of Phase Distortion (FPD)** of X_h:

$$\Phi_k(X_h) = \Delta^{-1}\Delta^2 \angle \left(\frac{X_h}{\mathcal{E}_-(X_h)} \right)$$

$\mathcal{E}_-(.)$ Minimum-phase realization
$\Rightarrow$ e.g. remove the vocal-tract filter

Δ^2 Second order difference operator
$\Rightarrow$ remove any linear-phase component

Function of Phase Distortion (FPD)

We propose the **Function of Phase Distortion (FPD)** of X_h:

$$\Phi_k(X_h) = \Delta^{-1}\Delta^2 \angle \left(\frac{X_h}{\mathcal{E}_-(X_h)} \right)$$

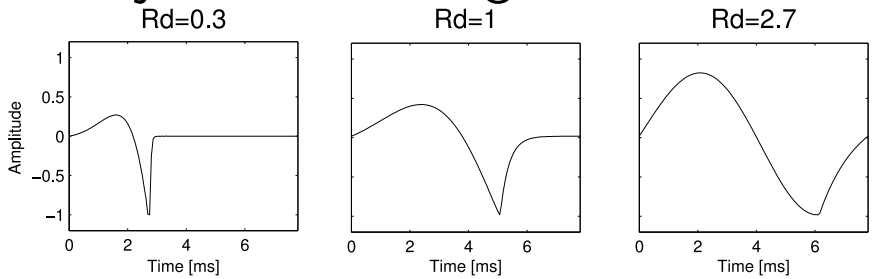
$\mathcal{E}_-(.)$ Minimum-phase realization
$\Rightarrow$ e.g. remove the vocal-tract filter

Δ^2 Second order difference operator
$\Rightarrow$ remove any linear-phase component

Δ^{-1} Anti-difference operator
$\Rightarrow$ obtain a similar representation to the group-delay

FPD - Exemple for the LF model parametrized by *Rd*

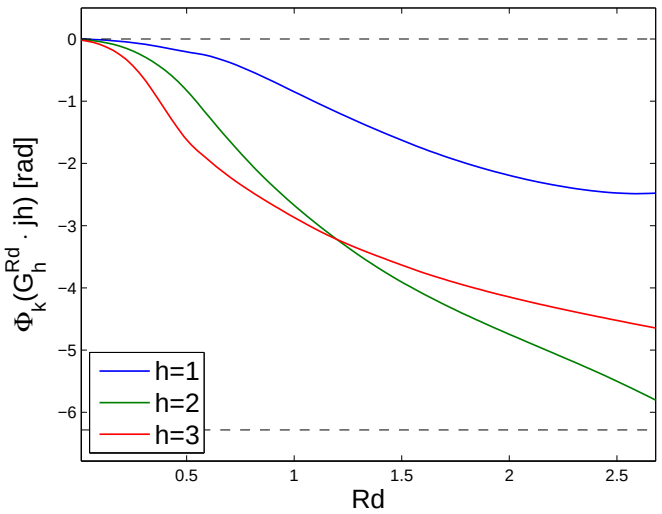
For the Transformed Liljencrants-Fant glottal model [1]



$$\Phi_k(G_h^{Rd} \cdot jh) = \Delta^{-1}\Delta^2 \angle \left(\frac{G_h^{Rd} \cdot jh}{\mathcal{E}_-(G_h^{Rd} \cdot jh)} \right)$$

1. Independent of the time position
2. Independent of the amplitude
3. Independent of the duration $(\frac{1}{f_0})$
4. Independent of a minimum-phase component

$\Rightarrow$ Related to the pulse shape only



1 G. Fant. *The LF-model revisited. transformations and frequency domain analysis*, STL-QPSR, 36(2-3):119-156, 1995.

Application
Estimation of the Rd parameter

Criterion of phase minimization [1]

Convolutive residual:

$$R_h = \frac{S_h}{M_h^{(Rd,\phi)}}$$

$$M_h^{(Rd,\phi)} = S_h \quad \Leftrightarrow \quad R_h^{(Rd,\phi)} = 1 \; \forall h$$

$\Rightarrow$

$$|R_h^{(Rd,\phi)}| = 1 \quad \text{et} \quad \angle R_h^{(Rd,\phi)} = 0 \quad \forall h$$

Idea

- Ensure a constant amplitude spectrum
- Minimize the phase spectrum

1 R. Smits and B. Yegnanarayana, *Determination of Instants of Significant Excitation in Speech Using Group Delay Function*, IEEE Trans. Speech and Audio Processing, vol. 3, pp. 325–333, 1995.

FPD and phase minimization

Given the voice production model

$$M_h = e^{jh\phi} \cdot G_h^{Rd} \cdot C_{h-} \cdot jh$$

the Function of Phase Distortion

$$\Phi_k(X_h) = \Delta^{-1}\Delta^2 \angle \left(\frac{X_h}{\mathcal{E}_-(X_h)} \right)$$

can be applied to the convolutive residual

$$R_h^{Rd} = \frac{S_h}{M_h^{(Rd,\phi)}} = \frac{S_h}{e^{jh\phi} \cdot G_h^{Rd} \cdot \mathcal{E}_-(S_h/G_h^{Rd} \cdot jh) \cdot jh} = e^{-jh\phi} \frac{S_h/G_h^{Rd} \cdot jh}{\mathcal{E}_-(S_h/G_h^{Rd} \cdot jh)}$$

FPD and phase minimization

Given the voice production model

$$M_h = e^{jh\phi} \cdot G_h^{Rd} \cdot C_{h-} \cdot jh$$

the Function of Phase Distortion

$$\Phi_k(X_h) = \Delta^{-1}\Delta^2 \angle \left(\frac{X_h}{\mathcal{E}_-(X_h)} \right)$$

can be applied to the convolutive residual

$$R_h^{Rd} = \frac{S_h}{M_h^{(Rd,\phi)}} = \frac{S_h}{e^{jh\phi} \cdot G_h^{Rd} \cdot \mathcal{E}_-(S_h/G_h^{Rd} \cdot jh) \cdot jh} = e^{-jh\phi} \frac{S_h/G_h^{Rd} \cdot jh}{\mathcal{E}_-(S_h/G_h^{Rd} \cdot jh)}$$

We propose to minimize the error

$$\text{MSPD}^2(Rd, N) = \frac{1}{N} \sum_{k=1}^{N} \left(\Phi_k(S_h/G_h^{Rd} \cdot jh) \right)^2$$

MSPD^2: *Mean Squared Phase using the 2nd order phase Difference*

FPD and phase minimization - Exemples of error functions

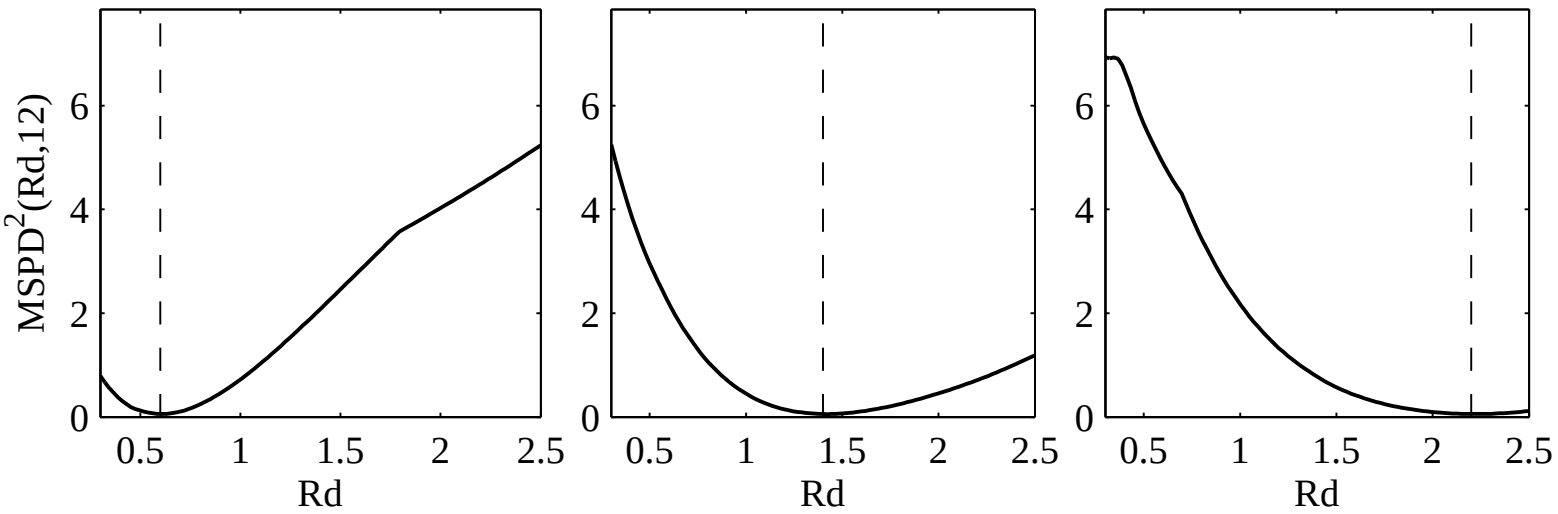
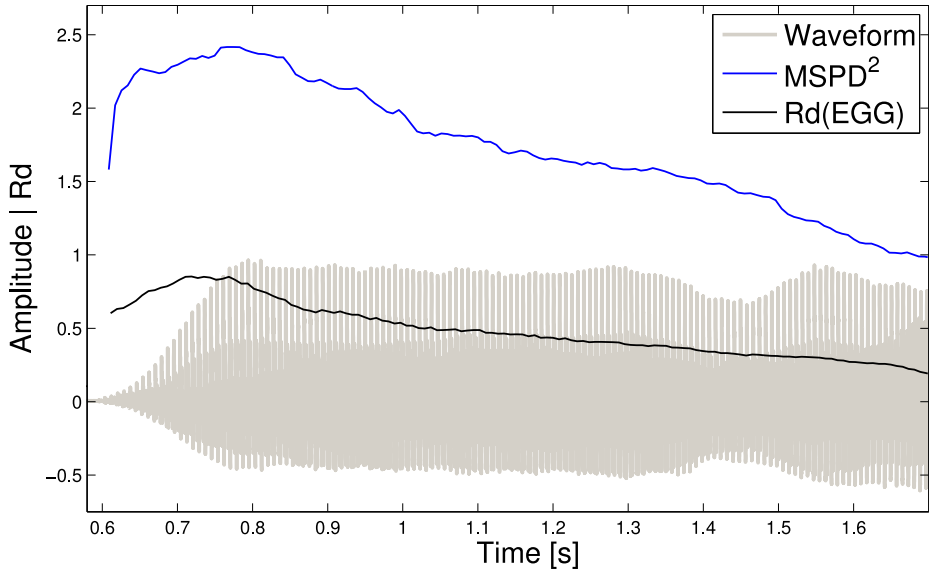


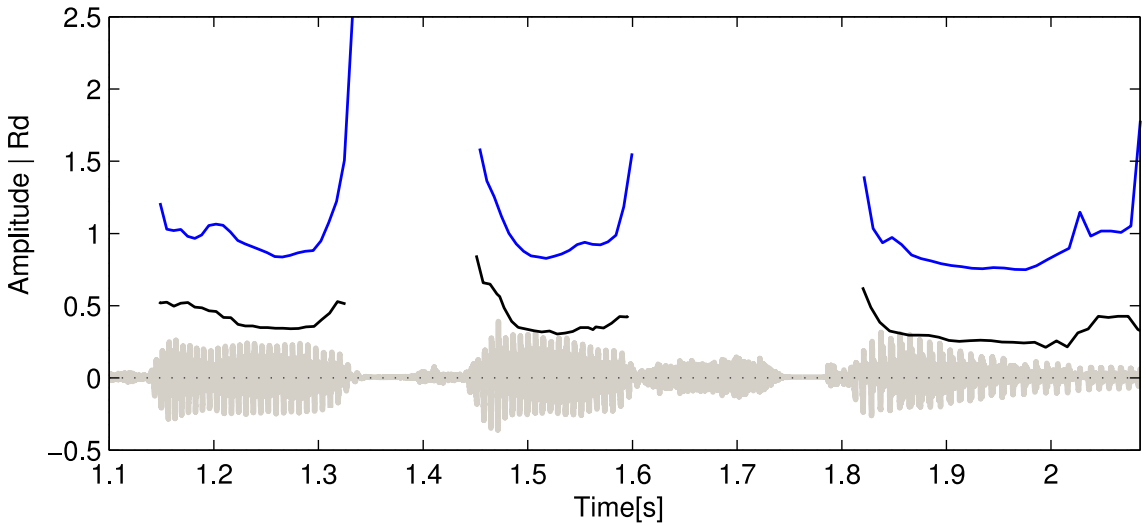
Figure: $\mathrm{MSPD}^2(Rd, 12)$ for 3 synthetic signals with different Rd values

The global minimum reached by a Brent's method.

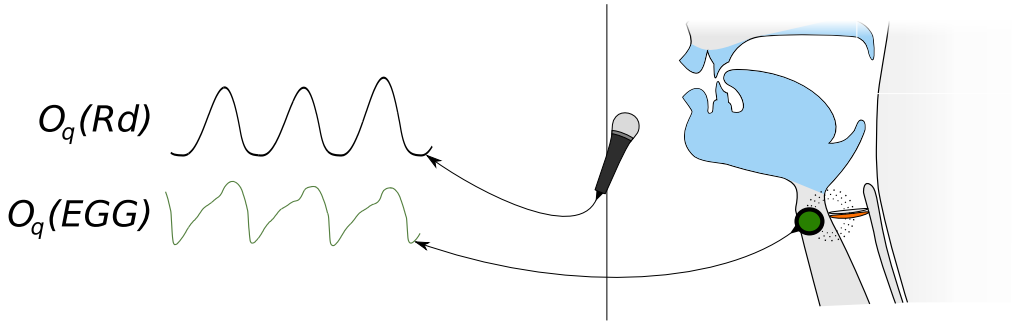
Evaluation
Comparison with Electroglottographic signals

Estimation exemples

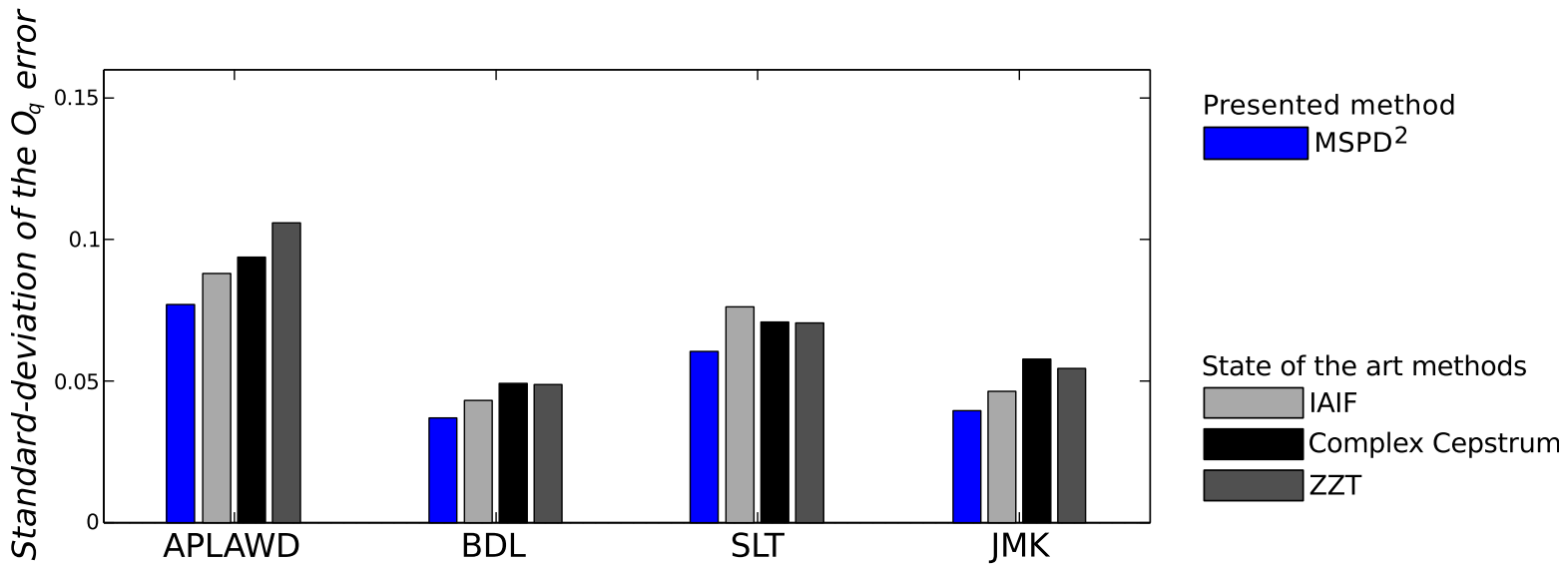




Evaluation



Comparison of $O_q(Rd)$ vs. $O_q(EGG)$



IAIF P. Alku, and H. Tiitinen and R. Naatanen, *A method for generating natural-sounding speech stimuli for cognitive brain research*.

CC T. Drugman, B. Bozkurt and T. Dutoit, *Complex Cepstrum-based Decomposition of Speech for Glottal Source Estimation*.

ZZT B. Bozkurt, B. Doval, C. d'Alessandro and T. Dutoit, *ZZT representation with application to source-filter separation in speech*.

Conclusions

- **Function of phase distortion** related to the pulse **shape** only.
- Can be used to estimate parameters of glottal models.

Σας ευχαριστώ για την προσοχή σας
Thank you for your attention
Merci pour votre attention