ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

Ψηφιακή Επεξεργασία Φωνής

Διάλεξη: Μοντελοποίηση Σήματος Πηγής με Χρήση Γλωττιδικών Μοντέλων

Παρουσίαση: Gilles Degottex

Στυλιανού Ιωάννης
Τμήμα Επιστήμης Υπολογιστών

Voice source modeling using a glottal model

Gilles Degottex

UoC-CSD/FORTH-ICS - Ircam/CNRS-UMR9912-STMS

Introduction

Motivations and Applications

Applications:

- Voice tranformation
- Speech synthesis
- Identity conversion
- Expressive synthesis

Motivations and Applications

Applications:

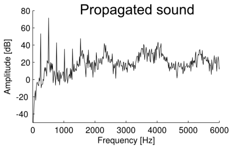
- Voice tranformation
- Speech synthesis
- Identity conversion
- Expressive synthesis

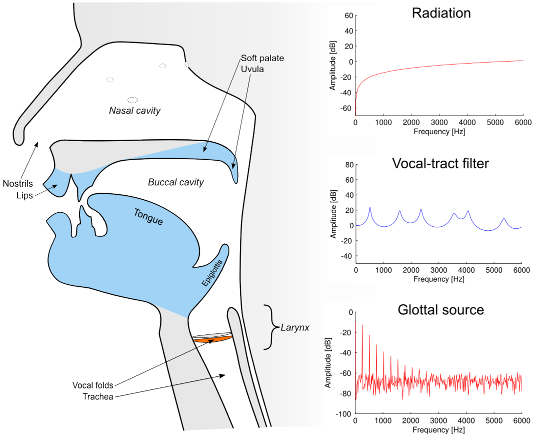
For ...

- Contemporary musique, sound installations
- Music and Cinema
- Video games
- Comunication technologies

Approach based on signal processing

Voice production





Approach based on signal processing

Source-filter model:

$$S(\omega) = G^{\theta_g}(\omega) \cdot C^{\theta_c}(\omega) \cdot L^{\theta_l}(\omega)$$

Waveform = Glottal-Source · Vocal-Tract-Filter · Radiation
with the parameters $\theta_g, \theta_c, \theta_l$ of each element

Approach based on signal processing

Source-filter model:

$$S(\omega) = G^{\theta_g}(\omega) \cdot C^{\theta_c}(\omega) \cdot L^{\theta_l}(\omega)$$

Waveform = Glottal-Source · Vocal-Tract-Filter · Radiation
with the parameters $\theta_g, \theta_c, \theta_l$ of each element

Model inversion:
E.G. General expression of the glottal source:

$$G(\omega) = \frac{S(\omega)}{C^{\theta_c}(\omega) \cdot L^{\theta_l}(\omega)}$$

Approach based on signal processing

Source-filter model:

$$S(\omega) = G^{\theta_g}(\omega) \cdot C^{\theta_c}(\omega) \cdot L^{\theta_l}(\omega)$$

Waveform = Glottal-Source · Vocal-Tract-Filter · Radiation
with the parameters $\theta_g, \theta_c, \theta_l$ of each element

Model inversion:
E.G. General expression of the glottal source:

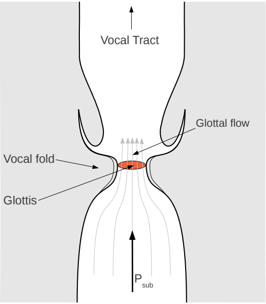
$$G(\omega) = \frac{S(\omega)}{C^{\theta_c}(\omega) \cdot L^{\theta_l}(\omega)}$$

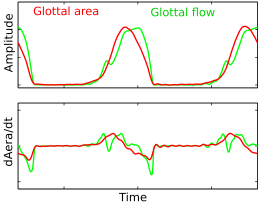
\+ Simplicity of inversion
− Strong approximation
hyp: Sufficient for voice manipulation with respect to **perception**

Glottal source modeling & Voice production model

The glottal source $G(\omega)$ - Vocal folds, Glottal area and Flow

©Erkki Bianco & Ircam property [1]



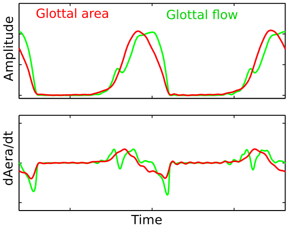


Glottal flow Air flow going through the glottis.

Glottal source in the source-filter model is an approximation of the glottal flow which should be suffisant for perceptual manipulation of the voice.

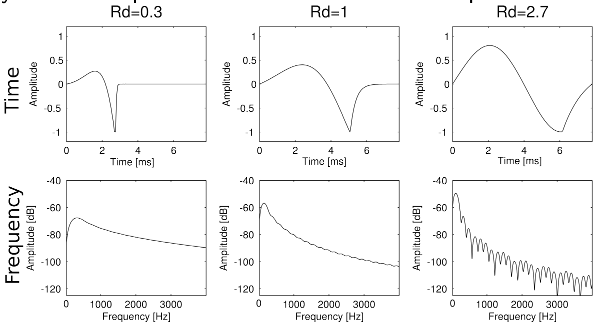
1 http://gillesdegottex.eu/IrcamUSC

The glottal source $G(\omega)$ - The Transformed Liljencrants-Fant model (LF)

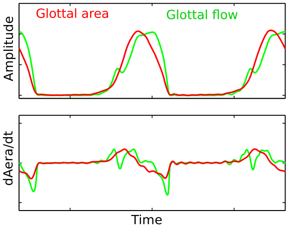


Glottal model = analytical description of the deterministic component

- $1/f_0$ Time duration
- E Amplitude
- ϕ Time position
- Rd Shape



The glottal source $G(\omega)$ - The Transformed Liljencrants-Fant model (LF)



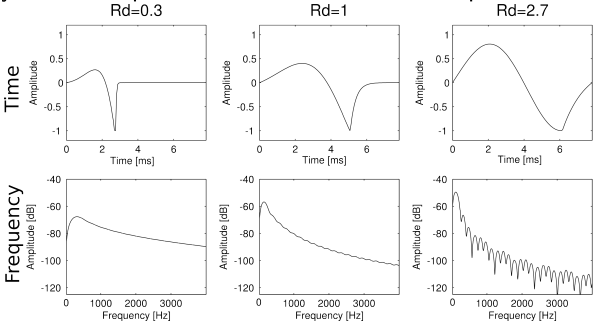
Glottal model = analytical description of the deterministic component

$1/f_0$ Time duration

E Amplitude

ϕ Time position

Rd Shape



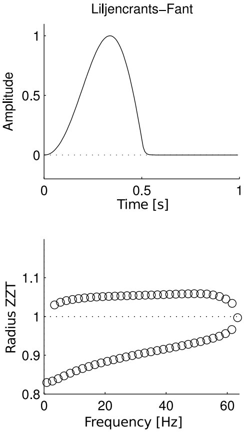
The glottal source $G(\omega)$ - Its use

Given a glottal model:

* How to estimate its parameters ?
* Comment estimer le filtre du conduit-vocal ?
* Comment transformer et synthétiser un signal vocal ?

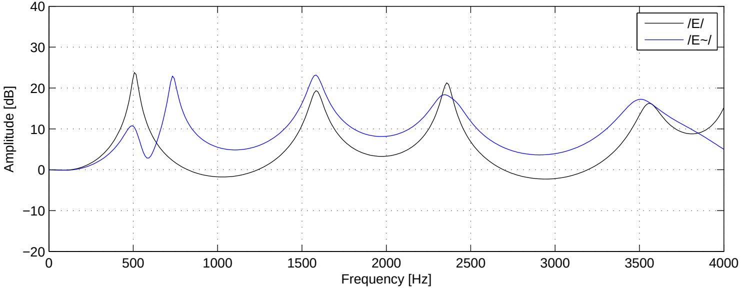
The glottal source $G(\omega)$ - Mixed-phase property of the glottal pulse

The glottal pulse is a **mixed-phase signal**.



The Vocal Tract Filter $C(\omega)$

It represents the resonnances and anti-resonnances of the vocal tract.



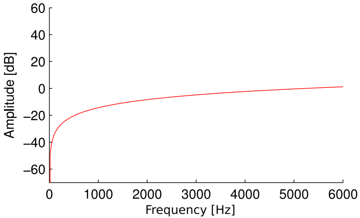
Passivity: The poles are inside the UC.
Postulate: The zeros are also inside the UC.
$\Rightarrow C(\omega)$ is minimum-phase.

This difference of phase property is the basis for parameter estimation.

The radiation $L(\omega)$

Constant model, without parameters [1]:

$$L(\omega) = j \cdot \omega$$



1 J.L. Flanagan, *Speech Analysis Synthesis and Perception*, Springer Verlag, 1972.

Complete model of the voice production

hyp: Split by a Voiced/Unvoiced Frequency (VUF):

$$S(\omega) = \begin{cases} e^{j\omega\phi} \cdot H^{f_0}(\omega) \cdot G^{(Rd,f_0)}(\omega) \cdot C_{-}(\omega) \cdot j\omega & \text{pour } \omega < \text{VUF} \\ N^{\sigma_g}(\omega) \cdot C_{-}(\omega) \cdot j\omega & \text{pour } \omega > \text{VUF} \end{cases}$$

$G^{(Rd,f_0)}(\omega)$ Shape of the glottal model

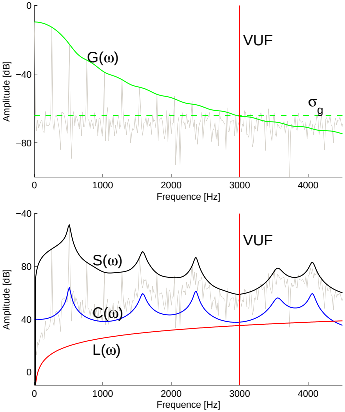
$e^{j\omega\phi}$ Time position of the shape

$H^{f_0}(\omega)$ Harmonicity

$N^{\sigma_g}(\omega)$ Turbulance noise

$C_{-}(\omega)$ Vocal Tract Filter (VTF)

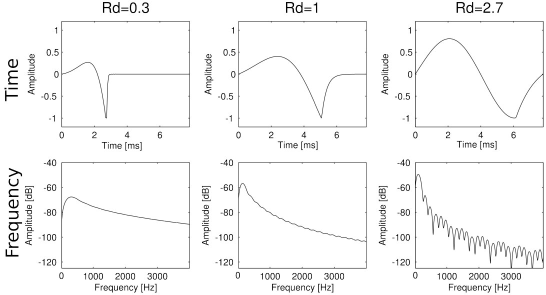
$j\omega$ Radiation



Estimation of the glottal parameters

Harmonic model for the Rd estimation

Estimation of the shape parameter *Rd* of the LF glottal model



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

$$\begin{array}{ccccccccc} S(\omega_h) & = & e^{j\omega_h\phi} & \cdot & G^{(Rd,f_0)}(\omega_h) & \cdot & C_{-}(\omega_h) & \cdot & j\omega_h \\ S_h & = & e^{jh\phi} & \cdot & G_h^{Rd} & \cdot & C_{h-} & \cdot & jh \end{array}$$

Indexed notation ! $X_h \equiv X(\omega_h)$

Vocal Tract Filter general expression

The general expression of the glottal source and the vocal tract filter are:

$$e^{jh\phi} \cdot G_h^{Rd} = \frac{S_h}{C_{h-} \cdot jh} \qquad C_{h-} = \frac{S_h}{e^{jh\phi} \cdot G_h^{Rd} \cdot jh}$$

Vocal Tract Filter general expression

The general expression of the glottal source and the vocal tract filter are:

$$e^{jh\phi} \cdot G_h^{Rd} = \frac{S_h}{C_{h-} \cdot jh} \qquad C_{h-} = \frac{S_h}{e^{jh\phi} \cdot G_h^{Rd} \cdot jh}$$

For the VTF, we force it to be minimum-phase using $\mathcal{E}_-(.)$

$$C_{h-} = \mathcal{E}_- \left(\frac{S_h}{G_h^{Rd} \cdot jh} \right)$$

Phase Minimization Criterion [1]

The convolutive residual:

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

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

$\Rightarrow$

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

Idea

- Ensure an unitary 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.

Mean Squared Phase (MSP)

In the context of the used voice production model

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

Mean Squared Phase (MSP)

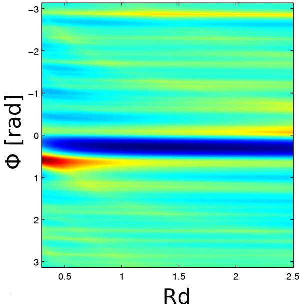
In the context of the used voice production model

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

Minimize the quadratic mean of the residual phase

$$\mathrm{MSP}(Rd, \phi, N) = \frac{1}{N} \sum_{h=1}^{N} \left(\angle R_h^{(Rd,\phi)} \right)^2$$

Method MSP



Functions of Phase Distorsion

Functions of Phase Distorsion (FPD)

Functions of Phase Distorsion of X_h:

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

Functions of Phase Distorsion (FPD)

Functions of Phase Distorsion 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 of X_h
$\Rightarrow$ remove the VTF C_{h-}

Functions of Phase Distorsion (FPD)

Functions of Phase Distorsion 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 of X_h
$\Rightarrow$ remove the VTF C_{h-}

Δ^2 2^{nd} order difference operator
$\Rightarrow$ remove the linear-phase component $e^{jh\phi}$

Functions of Phase Distorsion (FPD)

Functions of Phase Distorsion 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 of X_h
$\Rightarrow$ remove the VTF C_{h-}

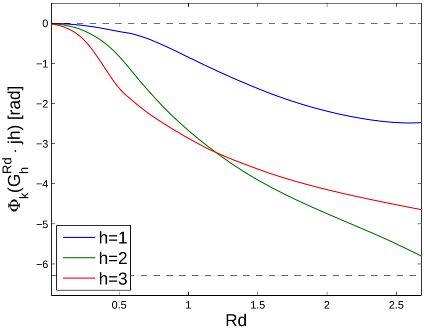
Δ^2 2^{nd} order difference operator
$\Rightarrow$ remove the linear-phase component $e^{jh\phi}$

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

FPD - Example

For the Liljencrants-Fant (LF) model:

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



FPD - Properties

FPD

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

Property of $\Phi_k(G_h^{Rd})$ for a glottal model:

1. Independent of the glottal pulse duration (period length)
2. Independent of the minimum-phase component
3. Independent of the time position of the glottal pulse
4. Independent of its amplitude E

$\Rightarrow$ Only related to the shape of the glottal pulse

FPD and phase minimization

The convolutive residual can be expressed as:

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

FPD and phase minimization

The convolutive residual can be expressed as:

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

To get ride fo the linear-phase term we can use the differance operators:

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

FPD and phase minimization

The convolutive residual can be expressed as:

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

To get ride fo the linear-phase term we can use the differance operators:

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

Which is equal to the FPD:

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

FPD and phase minimization

The convolutive residual can be expressed as:

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

To get ride fo the linear-phase term we can use the differance operators:

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

Which is equal to the FPD:

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

We therefore minimize the error:

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

Method MSPD^2

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

Method based on $MSPD^2$

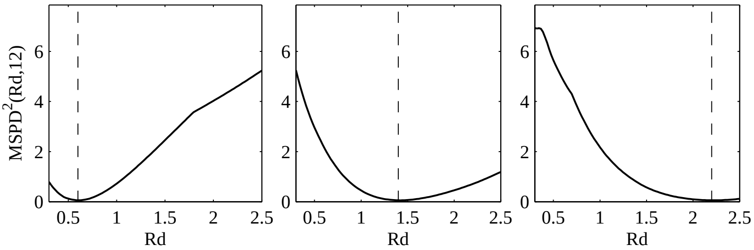


Figure: $MSPD^2(Rd, 12)$ with 3 different synthetic signals with different Rd values.

Quasi-closed form expression of the *Rd* parameter

FPD - Quasi-closed form expression of the *Rd* parameter

Goal: Find an explicit expression of Rd from S_h (e.g. $Rd = f(S_h)$)

$$S_h = e^{jh\phi} \cdot G_h^{Rd} \cdot \mathcal{E}_{-}\left(\frac{S_h}{G_h^{Rd} \cdot jh}\right) \cdot jh$$

FPD - Quasi-closed form expression of the *Rd* parameter

Goal: Find an explicit expression of Rd from S_h (e.g. $Rd = f(S_h)$)

$$S_h = e^{jh\phi} \cdot G_h^{Rd} \cdot \mathcal{E}_- \left(\frac{S_h}{G_h^{Rd} \cdot jh} \right) \cdot jh$$

We can distribute $\mathcal{E}_-(.)$ to its terms:

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

FPD - Quasi-closed form expression of the *Rd* parameter

Goal: Find an explicit expression of Rd from S_h (e.g. $Rd = f(S_h)$)

$$S_h = e^{jh\phi} \cdot G_h^{Rd} \cdot \mathcal{E}_- \left(\frac{S_h}{G_h^{Rd} \cdot jh} \right) \cdot jh$$

We can distribute $\mathcal{E}_-(.)$ to its terms:

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

And we put observations and models on each side of the equation:

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

FPD - Quasi-closed form expression of the *Rd* parameter

Goal: Find an explicit expression of Rd from S_h (e.g. $Rd = f(S_h)$)

$$S_h = e^{jh\phi} \cdot G_h^{Rd} \cdot \mathcal{E}_- \left(\frac{S_h}{G_h^{Rd} \cdot jh} \right) \cdot jh$$

We can distribute $\mathcal{E}_-(.)$ to its terms:

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

And we put observations and models on each side of the equation:

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

$\Rightarrow$

$$\Phi_k(S_h) = \Phi_k(G_h^{Rd} \cdot jh)$$

FPD - Expression quasi explicite de Rd - Méthode

$$\Phi_k(S_h) = \Phi_k(G_h^{Rd} \cdot jh)$$

$$\text{for } \sigma_k = \Phi_k(S_h) \text{ find } Rd : \Phi_k(G_h^{Rd} \cdot jh) = \sigma_k$$

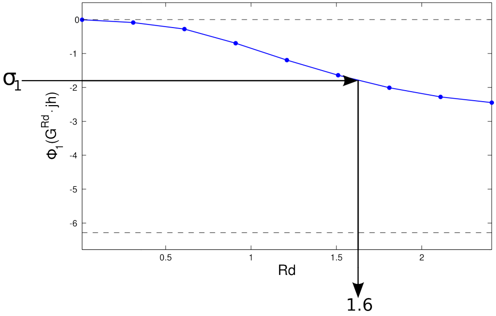
Method FPD^{-1}

FPD - Expression quasi explicite de *Rd* - Méthode

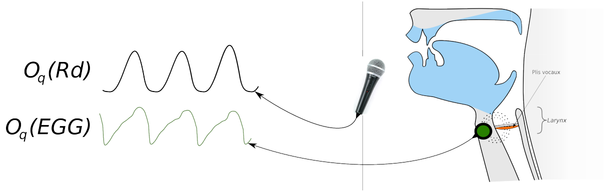
$$\Phi_k(S_h) = \Phi_k(G_h^{Rd} \cdot jh)$$

for $\sigma_k = \Phi_k(S_h)$ find Rd : $\Phi_k(G_h^{Rd} \cdot jh) = \sigma_k$ Method FPD^{-1}

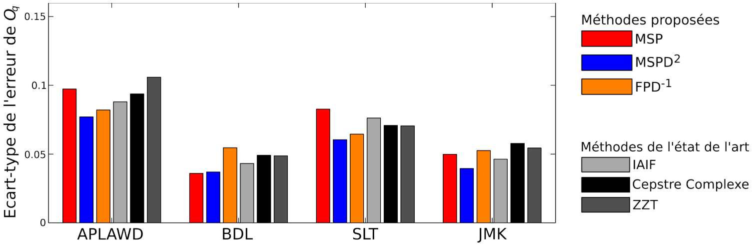
Numerical approximation using a lookup table:



Evaluation



We compare $O_q(Rd)$ vs. $O_q(EGG)$



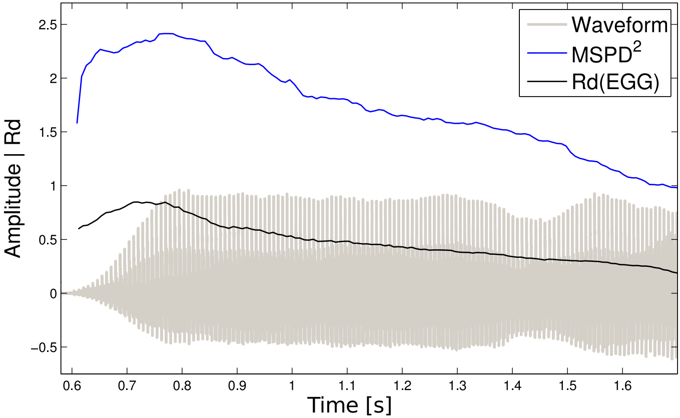
IAIF P. Alku, and H. Tiitinen and R. Naatanen, *A method for generating natural-sounding speech stimuli for cognitive brain research*.

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*.

Example of estimation

Son



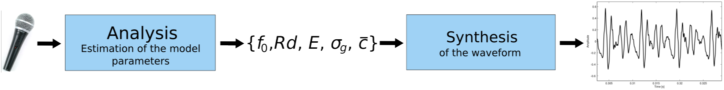
Application
Voice transformation

Analysis/Synthesis - SVLN

Model of the voice production used in SVLN

$$S(\omega) = \left[H^{f_0}(\omega) \cdot G^{Rd}(\omega) + N^{\sigma_g}(\omega)\right] \cdot C^{\bar{c}}_{-}(\omega) \cdot j\omega$$

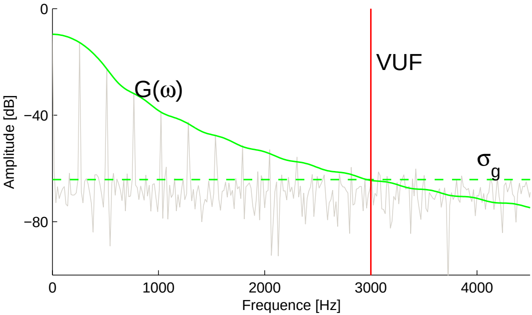
Analysis/Synthesis procedure



SVLN: *Separation of the Vocal-tract with the Liljencrants-Fant model plus Noise*

Analysis - Estimation of the glottal source parameters

- f_0 Known *a priori*
- Rd Method based on MSPD[2]
- E Log energy of the window
- σ_g Crossing point between $G(\omega)$ and VUF

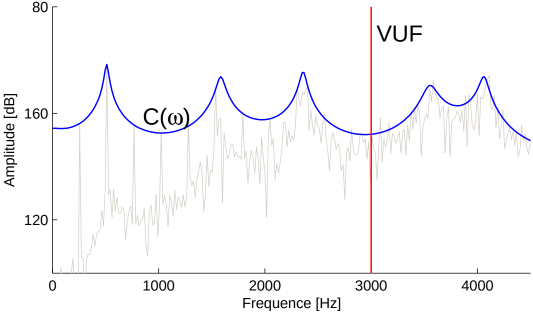


VUF known *a priori* by classification of the spectral peaks.

Analysis - Estimation of the VTF

$$C(\omega) = \begin{cases} \mathcal{T}\left(\frac{S(\omega)}{G^{Rd}(\omega)\cdot j\omega}\right)\cdot\gamma^{-1} & \text{if } \omega < \text{VUF} \\ \mathcal{P}\left(\frac{S(\omega)}{G^{Rd}(\text{VUF})\cdot j\omega}\right)\cdot\frac{\sqrt{\pi/2}}{\gamma\cdot e^{0.058}} & \text{if } \omega \geq \text{VUF} \end{cases}$$

$\mathcal{T}(.)$ The *True-envelope*
$\mathcal{P}(.)$ Real cepstrum
$\gamma = \sum_t \text{win}[t]/(f_s/f_0)$ number of periods in the window.



C. Yeh, *Multiple fundamental frequency estimation of polyphonic recordings*, Ph.D. thesis, UPMC, 2008.

Synthesis

- A sound chunk: Pulses $G(\omega)$ · VTF $C(\omega)$ · Radiation synthesis $L(\omega)$

Synthesis

- A sound chunk: Pulses $G(\omega) \cdot$ VTF $C(\omega) \cdot$ Radiation synthesis $L(\omega)$
- *Overlap-add* of the sound chunks

Synthesis

- A sound chunk: Pulses $G(\omega)$ · VTF $C(\omega)$ · Radiation synthesis $L(\omega)$
- *Overlap-add* of the sound chunks

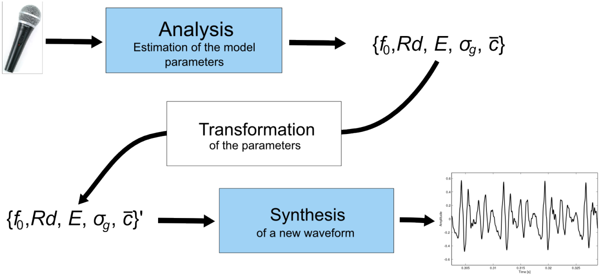
For any observed speech spectrum $S(\omega)$:

$|S(\omega)|$ always reproduced

$\angle S(\omega)$ imposed by the glottal model, the Gaussian noise and the phase of the VTF ($\angle C_{-}(\omega)$)

Voice transformation

Transformation process



Sound examples

http://gillesdegottex.eu/SVLN

Σας ευχαριστώ για την προσοχή σας

Τέλος Ενότητας

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Κρήτης**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημειώματα

Σημείωμα αδειοδότησης

.

Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Κρήτης, Στυλιανού Ιωάννης. «Ψηφιακή Επεξεργασία Φωνής. Μοντελοποίηση Σήματος Πηγής με Χρήση Γλωττιδικών Μοντέλων». Έκδοση: 1.0. Ηράκλειο/Ρέθυμνο 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: http://www.csd.uoc.gr/~hy578