Séparation de la source glottique des influences du conduit vocal

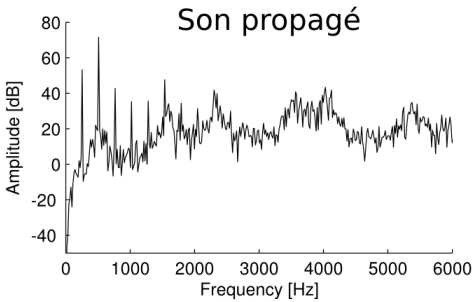
Estimation de paramètres glottiques et transformation de la voix utilisant un modèle glottique

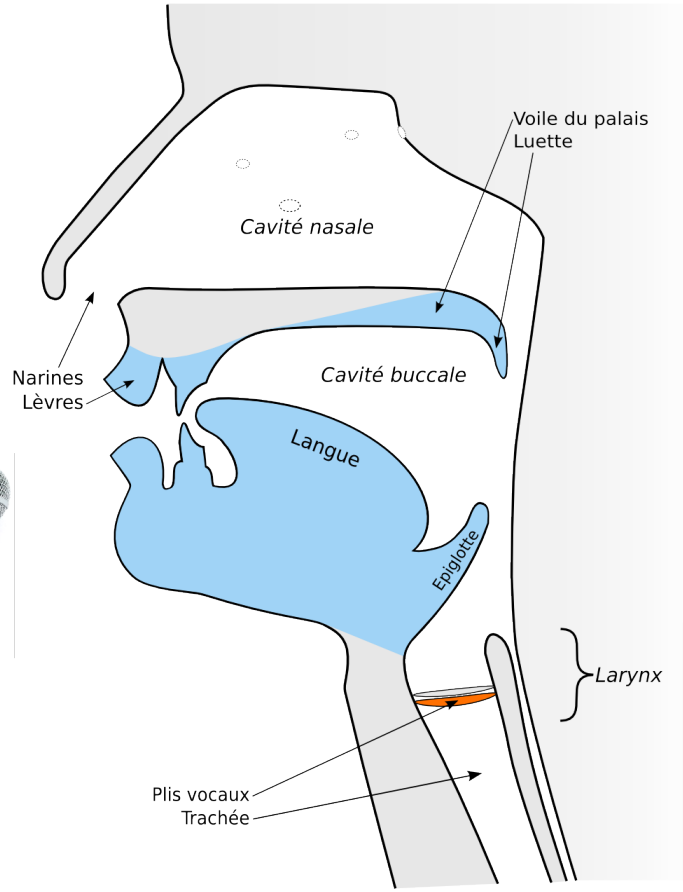
Gilles Degottex

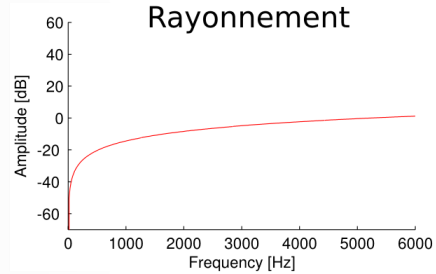
Encadrant	Axel Röbel
Directeur de thèse	Xavier Rodet
Rapporteurs	Yannis Stylianou et Christophe d'Alessandro
Examinateurs	Thierry Dutoit, Nathalie Henrich, Jean-Luc Zarader, Olivier Rosec, Olivier Boëffard

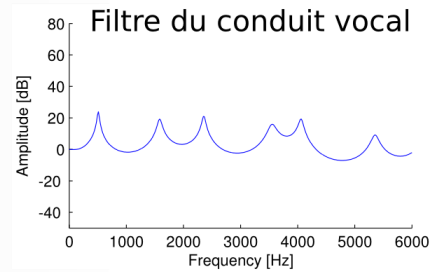
Ircam - CNRS-UMR9912-STMS - Equipe Analyse/Synthèse

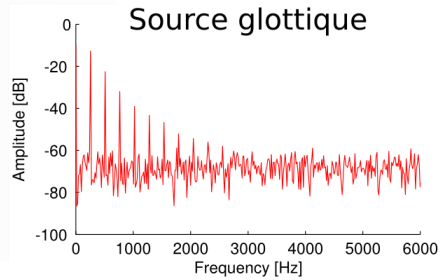
Objectif et problématique











Motivations et applications

Applications:

- La transformation de la voix
- La synthèse de la parole
- La conversion d'identité
- La synthèse d'expressivités

Motivations et applications

Applications:

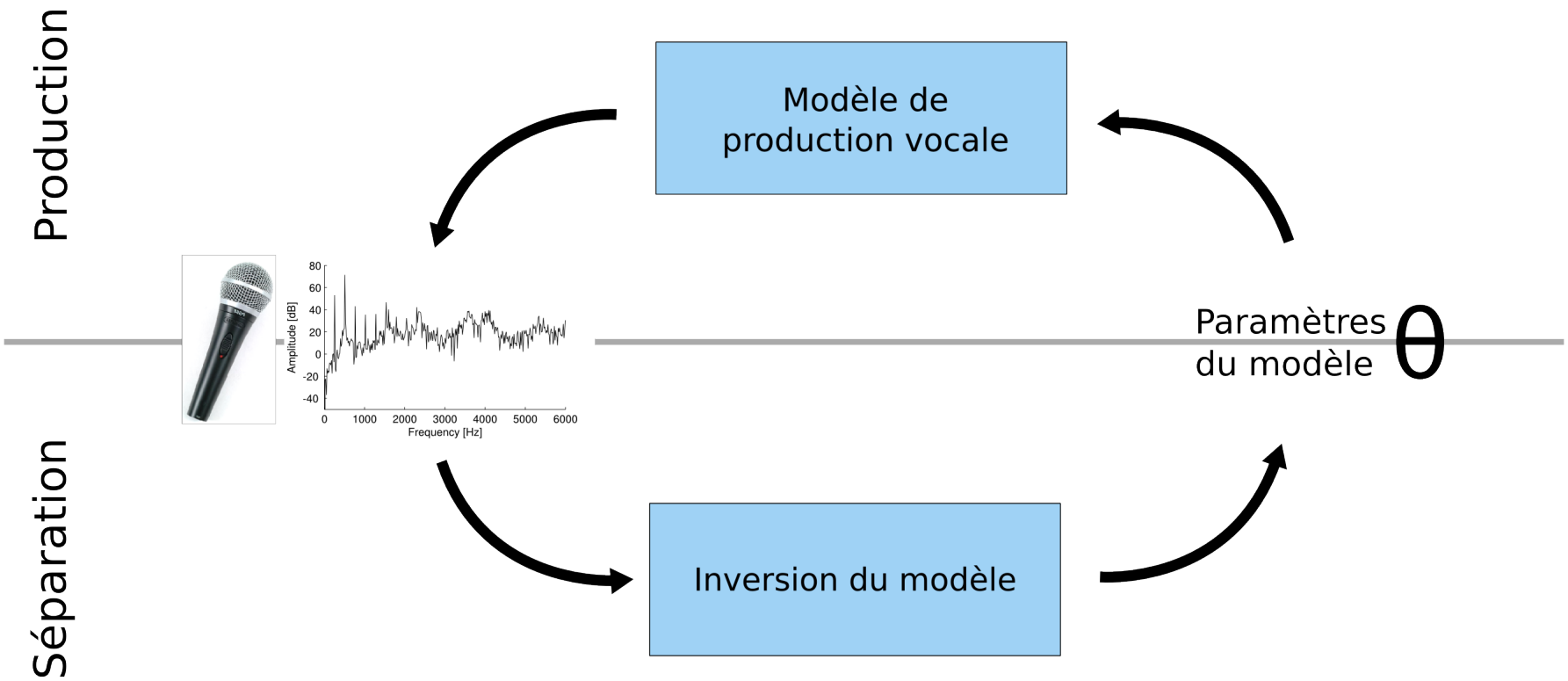
- La transformation de la voix
- La synthèse de la parole
- La conversion d'identité
- La synthèse d'expressivités

Pour ...

- La musique contemporaine, les installations sonores
- Musique et cinéma
- Les jeux vidéos
- Les télécommunications

Introduction
Approche

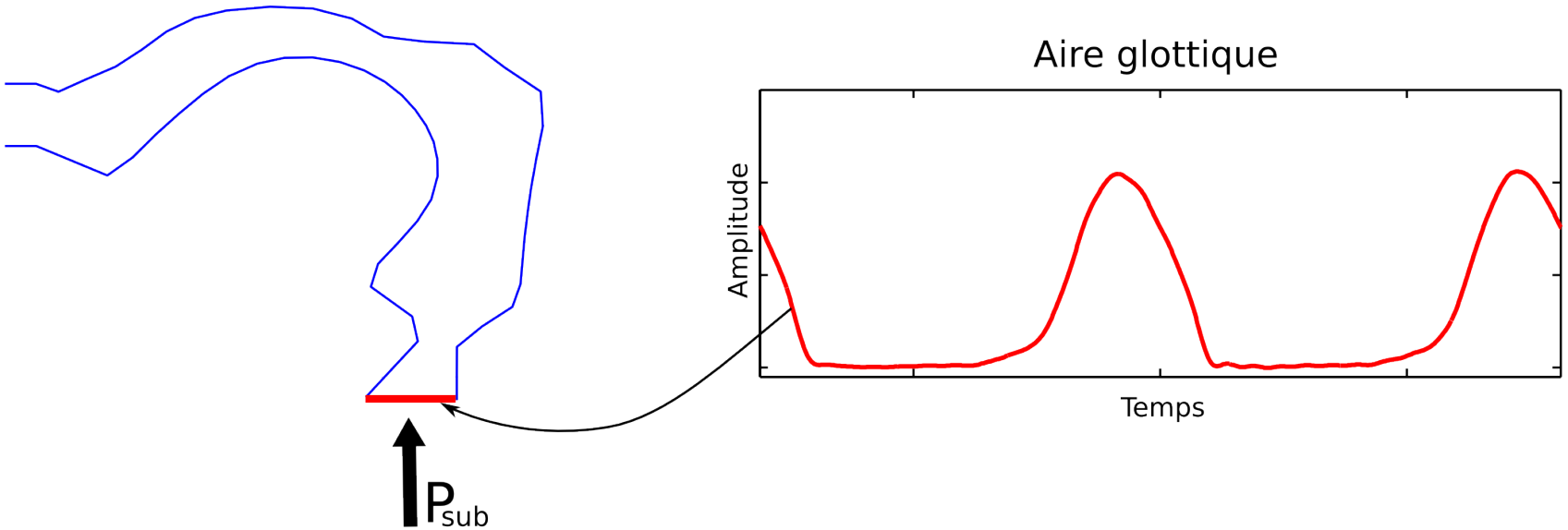
Inversion d'un modèle



Approche acoustique

Approche du modèle de Maeda

Géométrie du conduit vocal



Approche par modélisation des signaux

Modèle **source-filtre**:

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

$$\text{Son} = \text{Source glottique} \cdot \text{Conduit vocal} \cdot \text{Rayonnement}$$

Approche par modélisation des signaux

Modèle **source-filtre**:

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

$$\text{Son} = \text{Source glottique} \cdot \text{Conduit vocal} \cdot \text{Rayonnement}$$

Inversion:
Ex. expression générale de la source glottique:

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

Approche par modélisation des signaux

Modèle **source-filtre**:

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

$$\text{Son} = \text{Source glottique} \cdot \text{Conduit vocal} \cdot \text{Rayonnement}$$

Inversion:
Ex. expression générale de la source glottique:

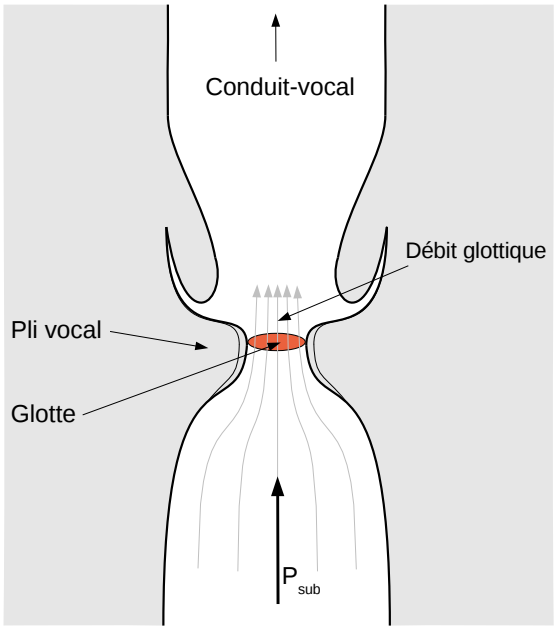
$$G(\omega) = \frac{S(\omega)}{C^{\theta_c}(\omega) \cdot L^{\theta_l}(\omega)}$$

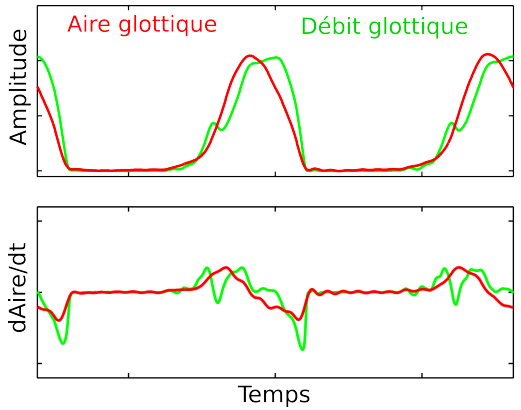
+ Simplicité d'inversion
− Forte approximation
hyp: suffisant pour la manipulation de la voix au niveau de la **perception**

Modélisation
La production vocale

La source glottique $G(\omega)$ - Plis vocaux, aire et débit glottique

©Erkki Bianco & Ircam property [1]





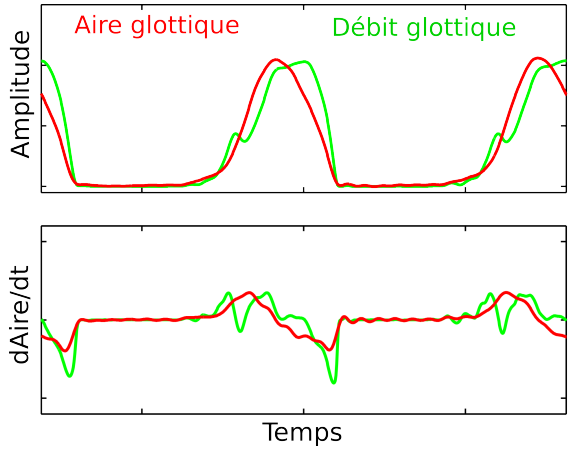
Couplage conduit-vocal → débit glottique:

Débit glottique Débit d'air passant à travers la glotte.

Source glottique Approx. suffisante pour manipuler la qualité vocale.

1 http://recherche.ircam.fr/anasyn/degottex/index.php?n=Main.IrcamUSC

La source glottique $G(\omega)$ - Modèle Transformed Liljencrants-Fant (LF)



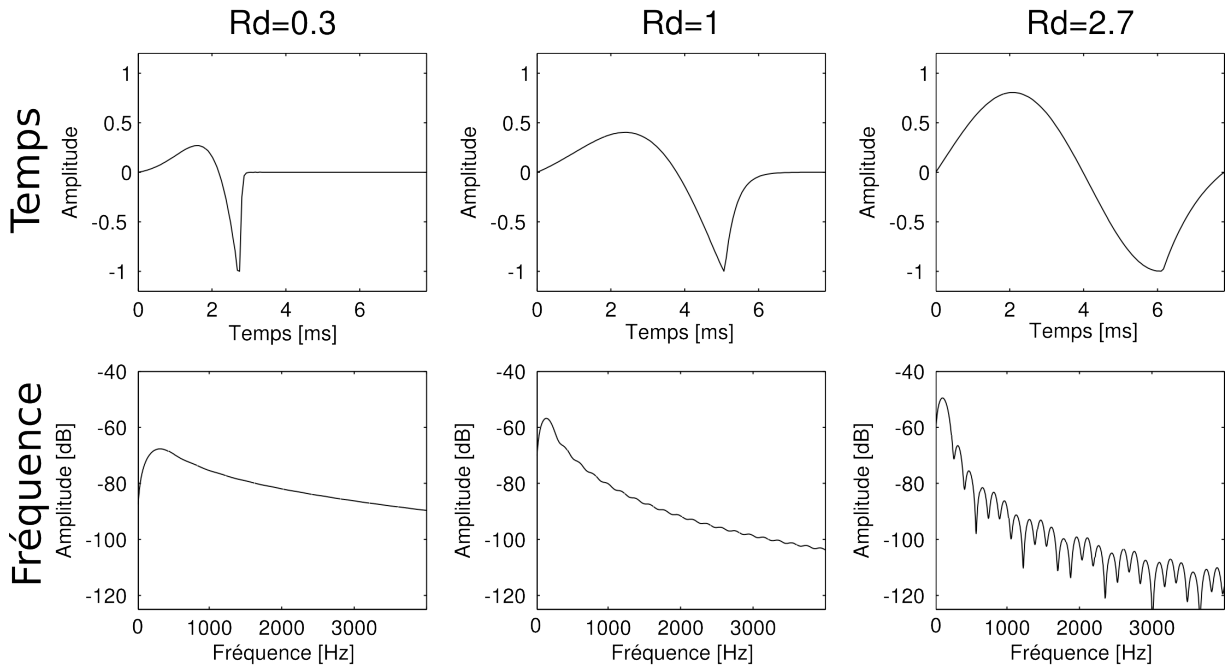
Modèle glottique = description analytique de la composante déterministe

$1/f_0$ Durée

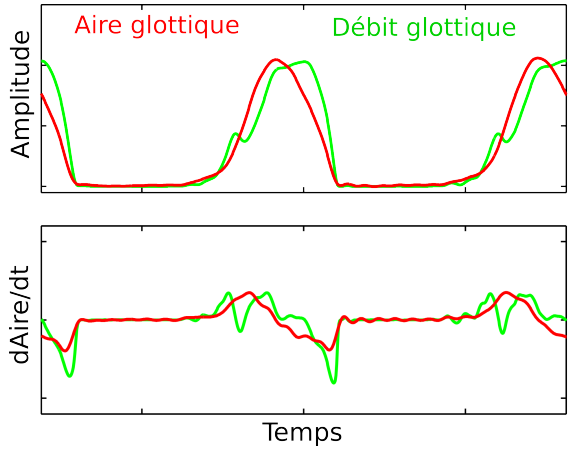
E Amplitude

ϕ Position

Rd Forme



La source glottique $G(\omega)$ - Modèle Transformed Liljencrants-Fant (LF)



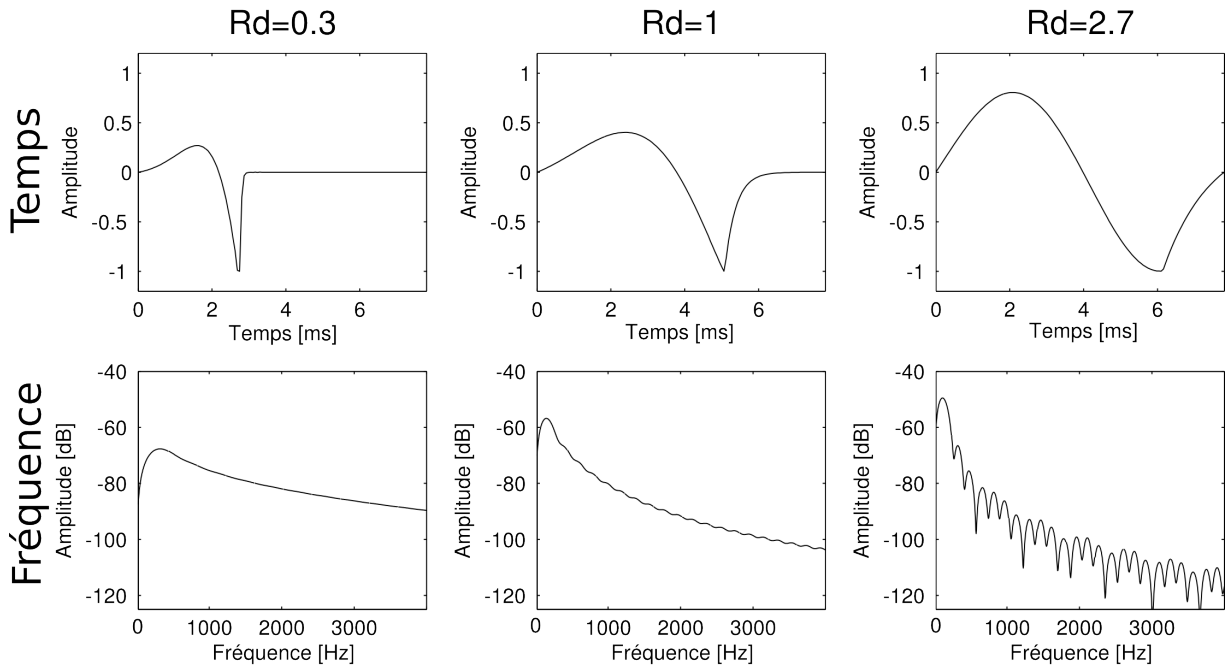
Modèle glottique = description analytique de la composante déterministe

$1/f_0$ Durée

E Amplitude

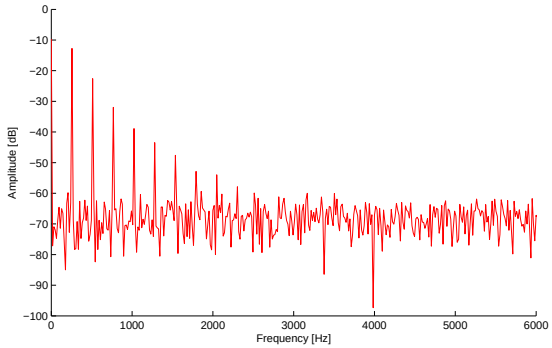
ϕ Position

Rd Forme



La source glottique $G(\omega)$ - Utilisation d'un modèle glottique

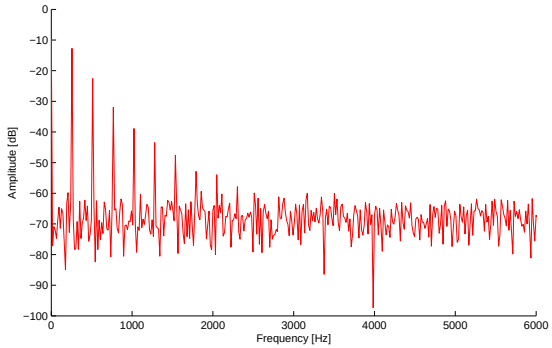
Source glottique



Déterministe + aléatoire

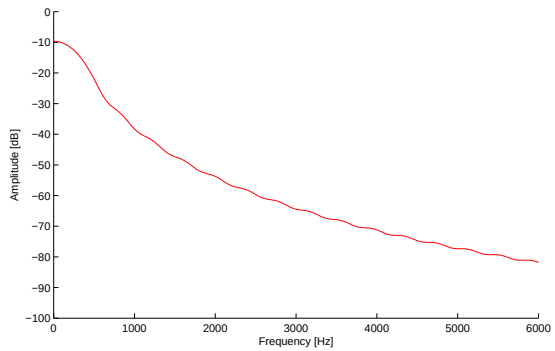
La source glottique $G(\omega)$ - Utilisation d'un modèle glottique

Source glottique



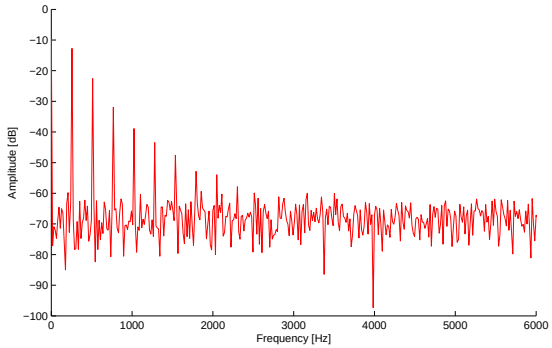
Déterministe + aléatoire

Modèle glottique LF



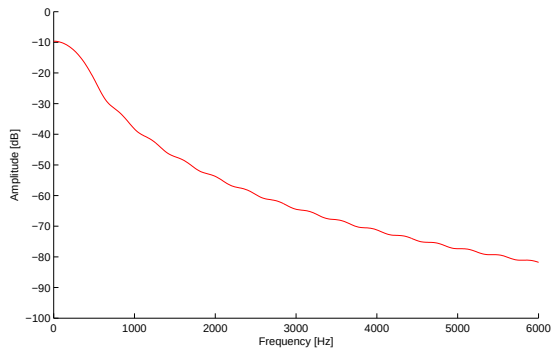
La source glottique $G(\omega)$ - Utilisation d'un modèle glottique

Source glottique



Déterministe + aléatoire

Modèle glottique LF

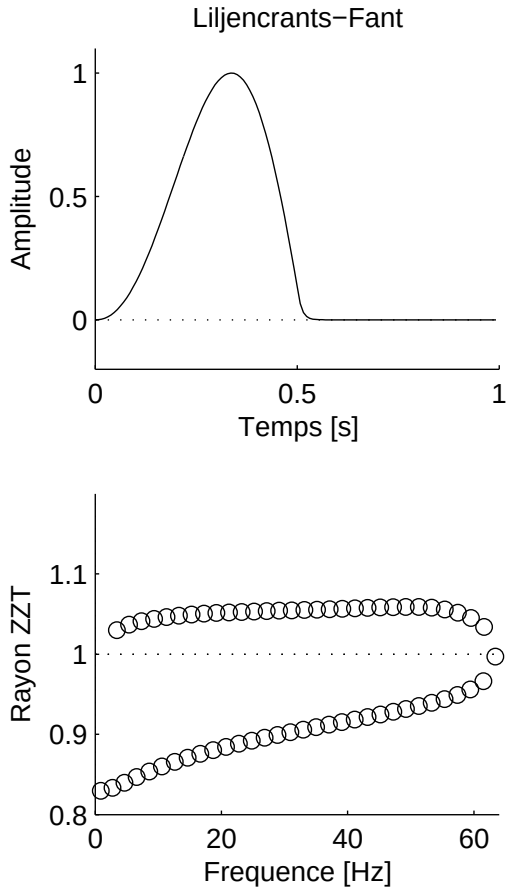


Étant donné un modèle glottique:

1. Comment estimer ses paramètres ?
2. Comment estimer le filtre du conduit-vocal ?
3. Comment transformer et synthétiser un signal vocal ?

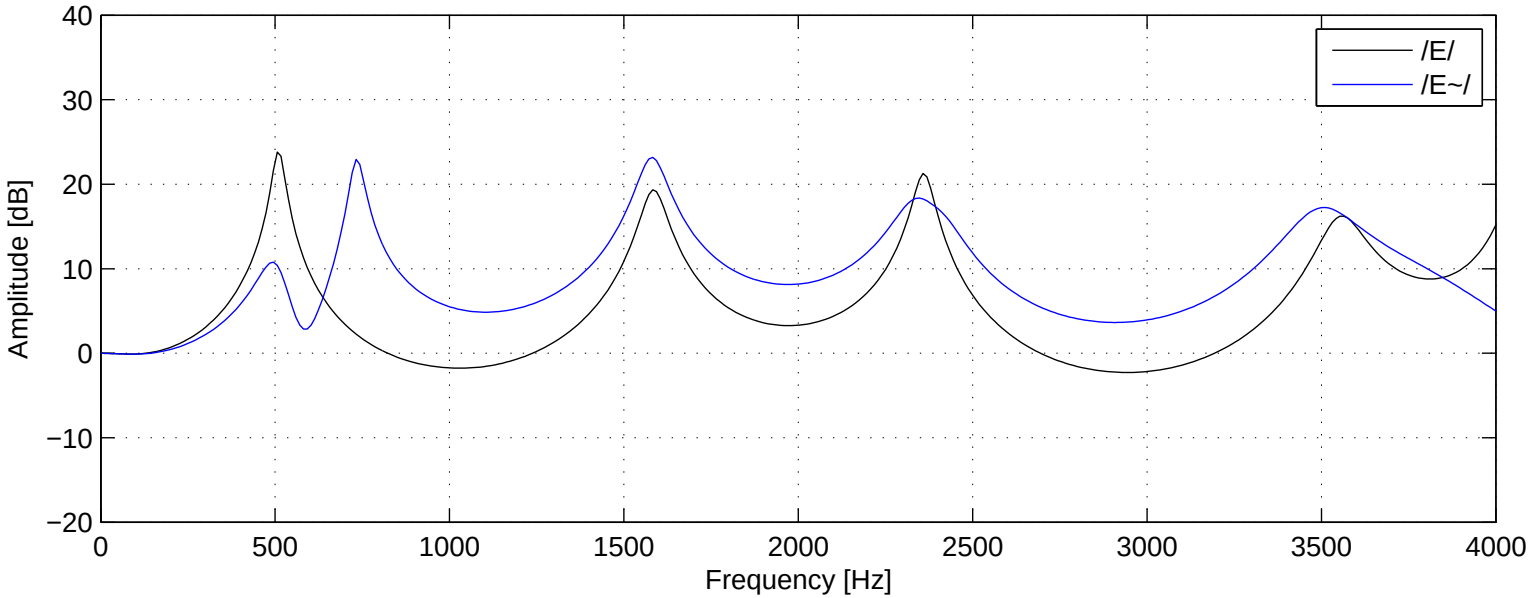
La source glottique $G(\omega)$ - Propriété phase-mixte du pulse glottique

Le pulse glottique est un **signal à phase mixte**.



Le filtre du conduit vocal $C(\omega)$

Il représente les résonances et anti-résonances du conduit-vocal.



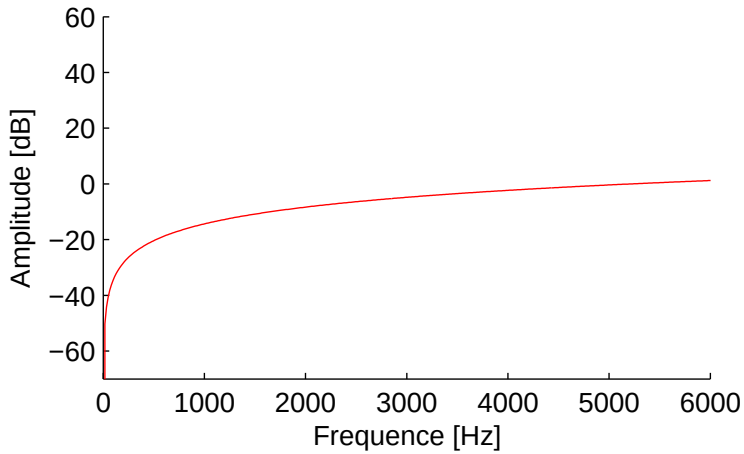
Passivité: Les pôles sont à l'intérieur du CU.
Postulat: Les zéros sont à l'intérieur du CU.
$\Rightarrow C(\omega)$ est à minimum de phase.

Aucune autre contrainte

Le rayonnement $L(\omega)$

Modèle constant, sans paramètres [1]:

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

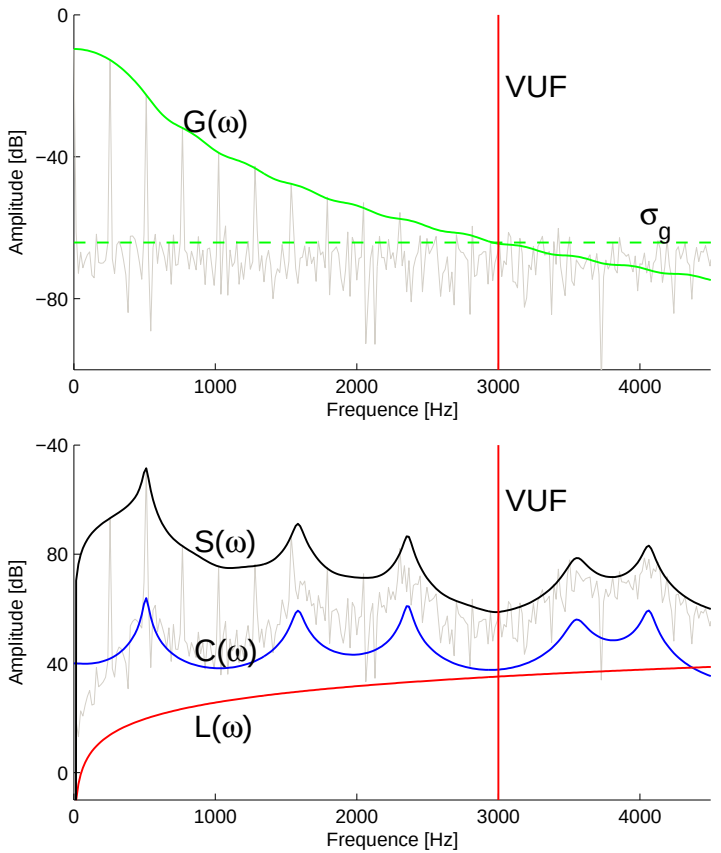


1 J.L. Flanagan, *Speech Analysis Synthesis and Perception*, Springer Verlag, 1972.

Modèle complet de la production vocale

$$S(\omega) = \left[e^{j\omega\phi} \cdot H^{f_0}(\omega) \cdot G^{Rd}(\omega) + N^{\sigma_g}(\omega) \right] \cdot C_-(\omega) \cdot j\omega$$

- $G^{Rd}(\omega)$ Forme du modèle glottique
- $e^{j\omega\phi}$ Position du modèle glottique
- $H^{f_0}(\omega)$ Périodicité
- $N^{\sigma_g}(\omega)$ Bruit de turbulence
- $C_-(\omega)$ Filtre du conduit-vocal
- $j\omega$ Rayonnement



Modèle complet de la production vocale

hyp: Séparable par une fréquence de voisement (VUF):

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

$G^{Rd}(\omega)$ Forme du modèle glottique

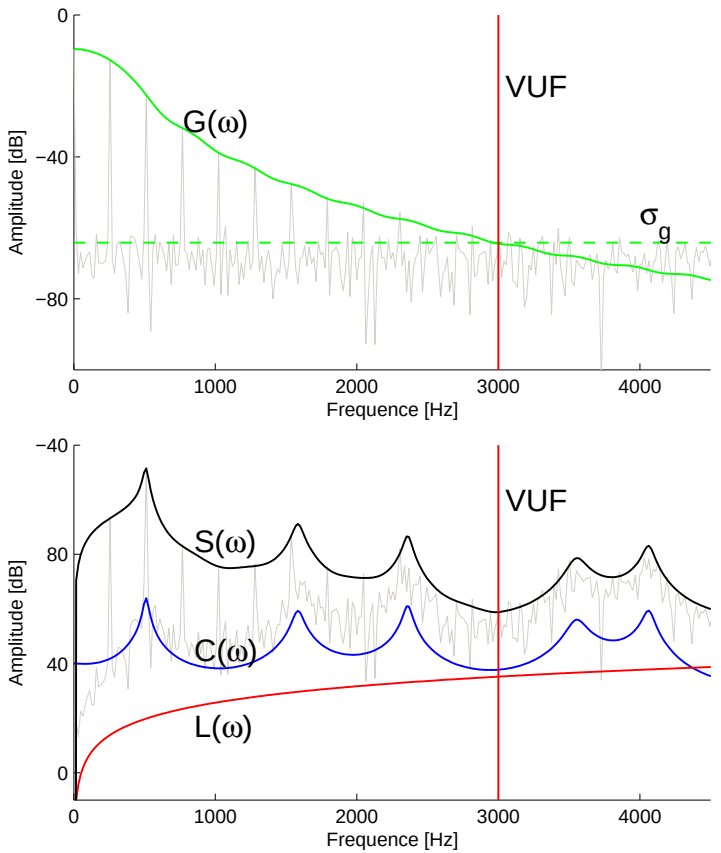
$e^{j\omega\phi}$ Position du modèle glottique

$H^{f_0}(\omega)$ Périodicité

$N^{\sigma_g}(\omega)$ Bruit de turbulence

$C_-(\omega)$ Filtre du conduit-vocal

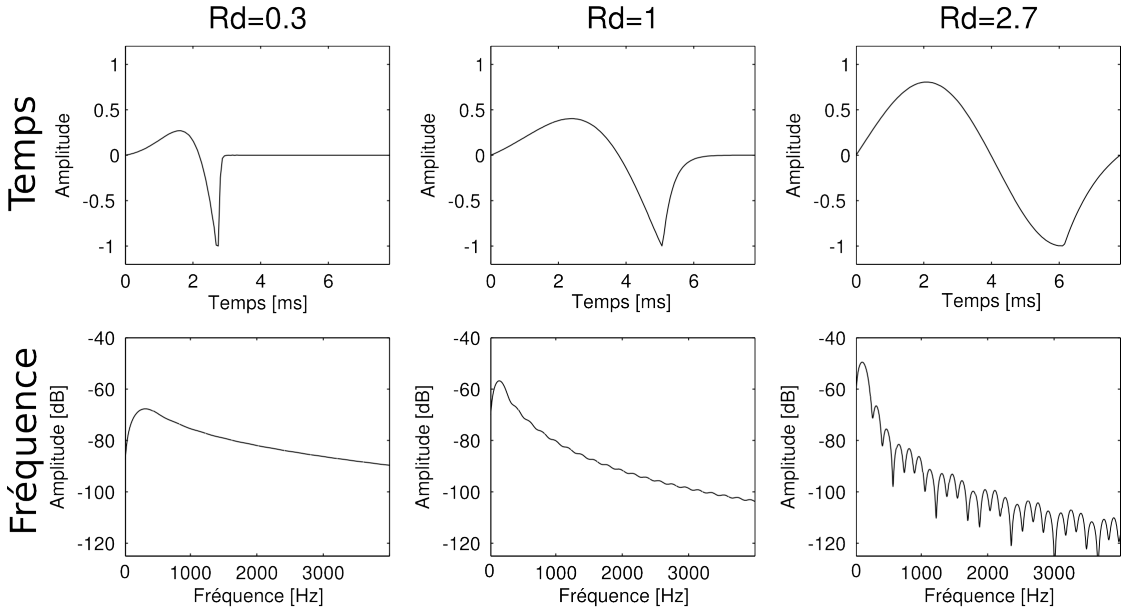
$j\omega$ Rayonnement



Analyse
Estimation d'un paramètre glottique

Modèle harmonique pour l'estimation de Rd

Estimation du paramètre de forme *Rd* du modèle LF



hyp: f_0 est connue $\Rightarrow$ **modèle harmonique:**

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

Notation indicée !

Expression du filtre du conduit-vocal

Le problème de séparation est représenté par $e^{jh\phi} \cdot G_h^{Rd}$ et C_{h-}

Leurs expressions générales:

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

Expression du filtre du conduit-vocal

Le problème de séparation est représenté par $e^{jh\phi} \cdot G_h^{Rd}$ et C_{h-}

Leurs expressions générales:

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

On contraint l'expression du conduit-vocal par $\mathcal{E}_-(.)$

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

Critère de minimisation de phase [1]

Le résiduel convolutif:

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

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

$\Rightarrow$

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

Idée

- Garantir un spectre d'amplitude unitaire
- Minimiser le spectre de phase

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) Phase quadratique moyenne

Pour le modèle de production vocale

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

Mean Squared Phase (MSP) Phase quadratique moyenne

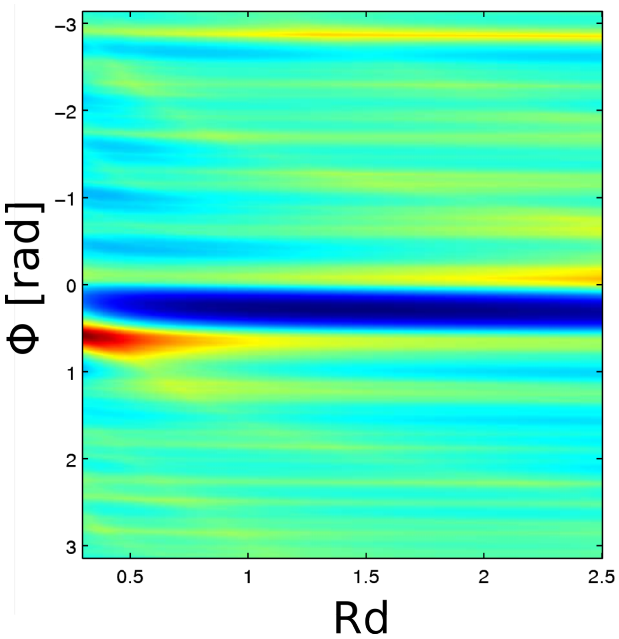
Pour le modèle de production vocale

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

On minimise la moyenne quadratique de la phase

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

Méthode MSP



Fonctions of Phase Distorsion (FPD) Fonctions de distorsion de phase

Nous avons proposé les **fonctions de distorsion de phase** de X_h:

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

Fonctions of Phase Distorsion (FPD) Fonctions de distorsion de phase

Nous avons proposé les **fonctions de distorsion de phase** de X_h:

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

$\mathcal{E}_-(.)$ Réalisation à minimum de phase
$\Rightarrow$ enlève le filtre du conduit-vocal.

Fonctions of Phase Distorsion (FPD) Fonctions de distorsion de phase

Nous avons proposé les **fonctions de distorsion de phase** de X_h:

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

$\mathcal{E}_-(.)$ Réalisation à minimum de phase
$\Rightarrow$ enlève le filtre du conduit-vocal.

Δ^2 Opérateur de différence de second ordre
$\Rightarrow$ supprime la composante linéaire.

Fonctions of Phase Distorsion (FPD) Fonctions de distorsion de phase

Nous avons proposé les **fonctions de distorsion de phase** de X_h:

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

$\mathcal{E}_-(.)$ Réalisation à minimum de phase
$\Rightarrow$ enlève le filtre du conduit-vocal.

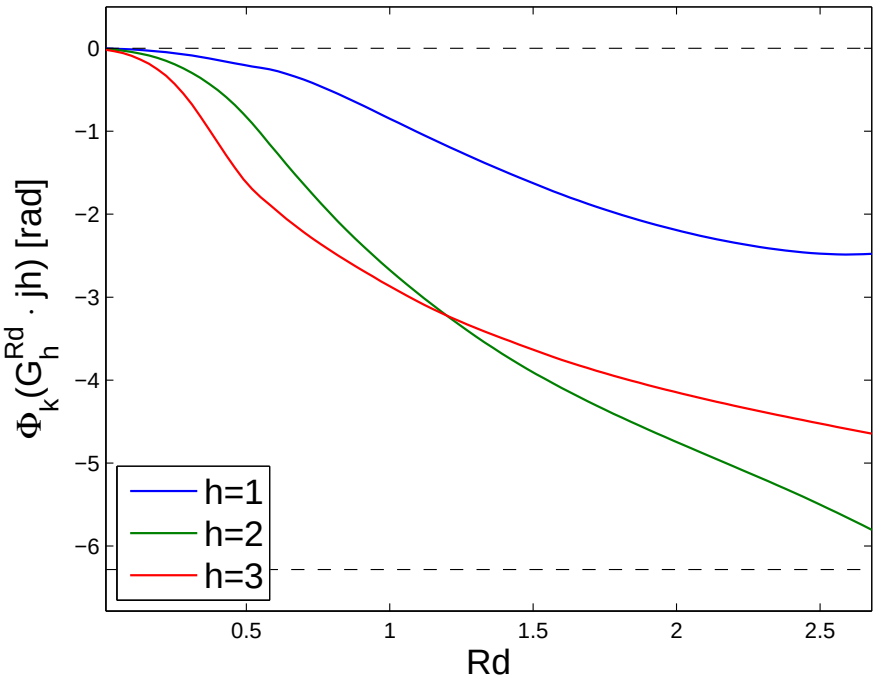
Δ^2 Opérateur de différence de second ordre
$\Rightarrow$ supprime la composante linéaire.

Δ^{-1} Opérateur d'anti-différence
$\Rightarrow$ représentation similaire au retard de groupe.

FPD - Exemple

Pour le modèle de Liljencrants-Fant:

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



FPD - Propriétés

FPD

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

Propriétés de $\Phi_k(G_h^{Rd})$ pour un modèle glottique:

1. Indépendantes de la position du pulse glottique.
2. Indépendantes de son amplitude E.
3. Indépendantes de la période du pulse glottique.
4. Indépendantes d'une composante à minimum de phase.

$\Rightarrow$ Uniquement reliées à la forme du pulse glottique.

FPD et minimisation de phase

FPD

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

Applicable au résiduel convolutif

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

FPD et minimisation de phase

FPD

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

Applicable au résiduel convolutif

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

On propose de minimiser l'erreur

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

Méthode MSPD^2

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

Méthode basée sur MSPD^2

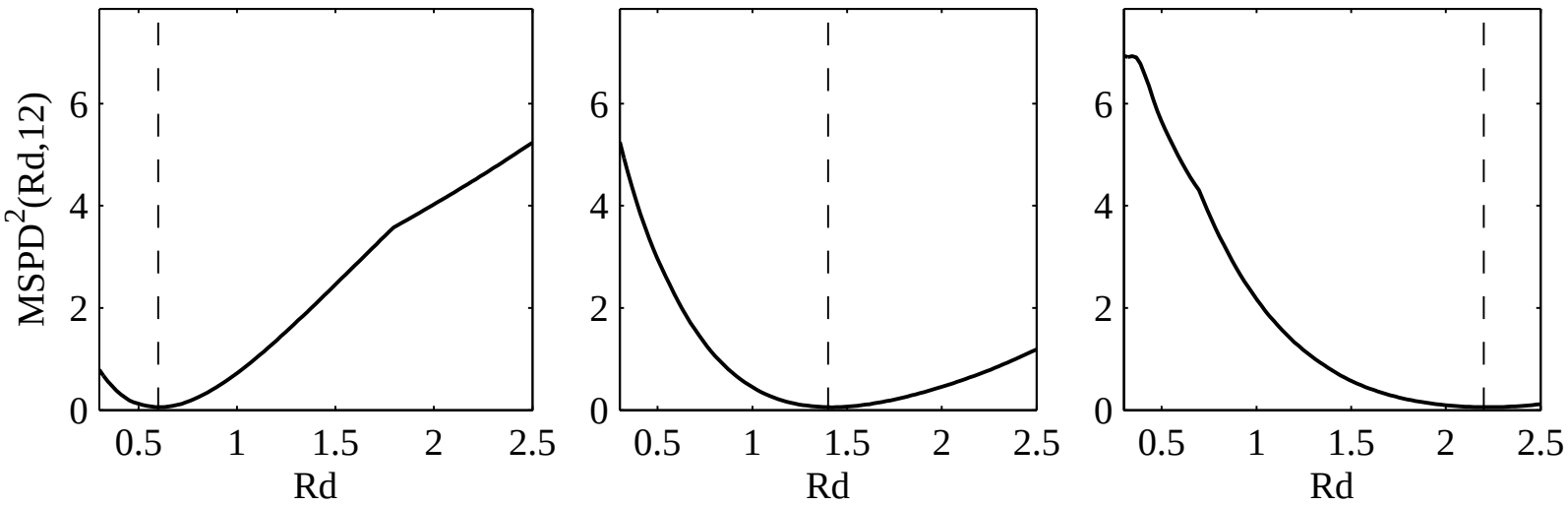


Figure: MSPD^2(Rd, 12) sur 3 signaux différents avec différentes valeurs de Rd.

FPD - Expression quasi explicite de *Rd*

But: Trouver une expression explicite de *Rd* à partir de 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 - Expression quasi explicite de *Rd*

But: Trouver une expression explicite de Rd à partir de S_h

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

On sépare les observations du modèle:

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

FPD - Expression quasi explicite de *Rd*

But: Trouver une expression explicite de Rd à partir de S_h

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

On sépare les observations du modèle:

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

$\Rightarrow$

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

FPD - Expression quasi explicite de Rd - Méthode

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

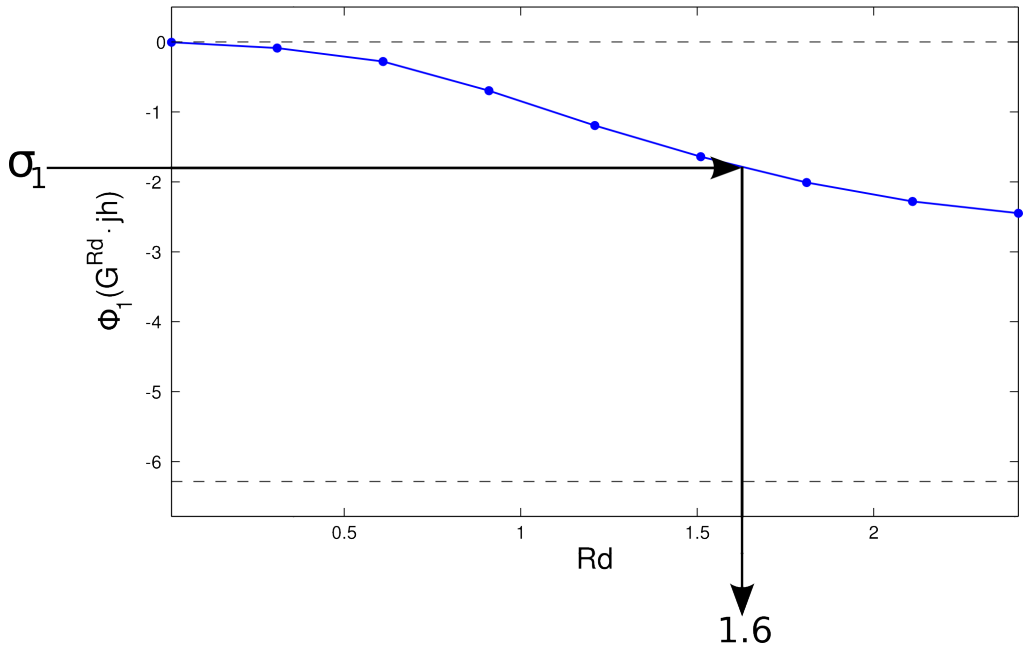
pour $\sigma_k = \Phi_k(S_h)$ trouver Rd : $\Phi_k(G_h^{Rd} \cdot jh) = \sigma_k$ Méthode FPD^{-1}

FPD - Expression quasi explicite de *Rd* - Méthode

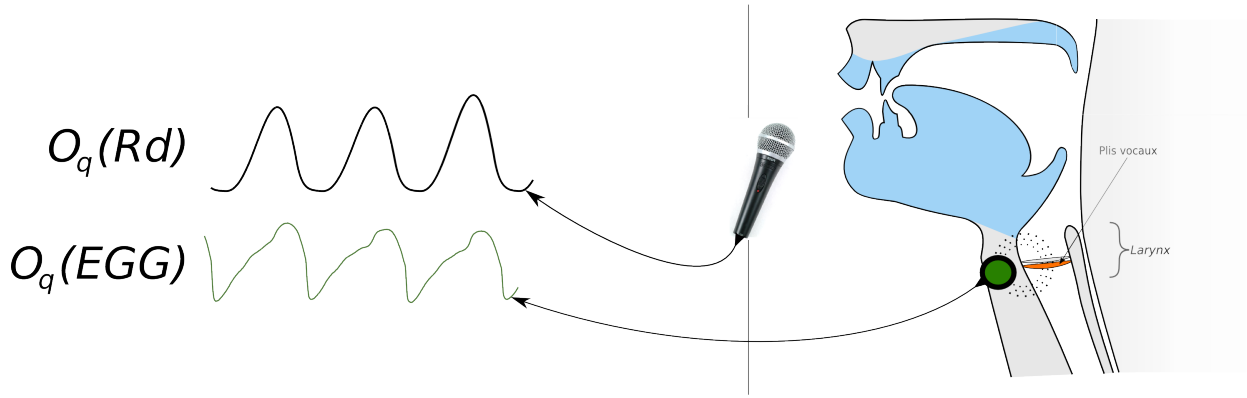
$$\Phi_k(S_h) = \Phi_k(G_h^{Rd} \cdot jh)$$

pour $\sigma_k = \Phi_k(S_h)$ trouver Rd : $\Phi_k(G_h^{Rd} \cdot jh) = \sigma_k$ Méthode FPD^{-1}

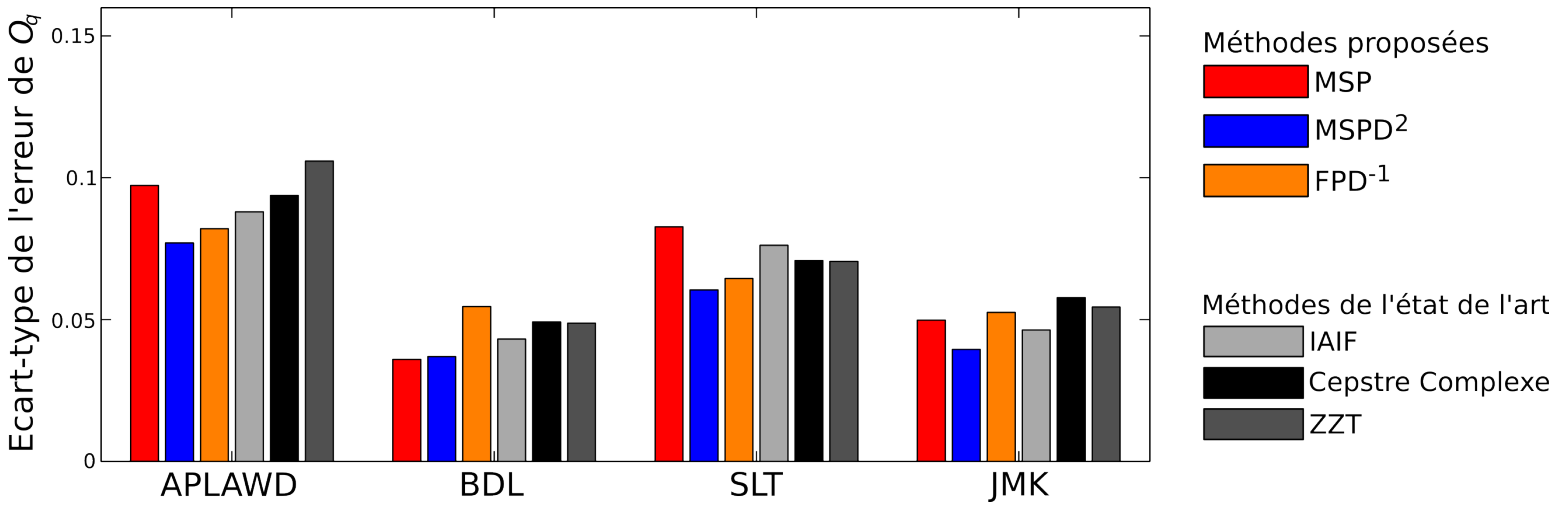
Approximation numérique par table de correspondance:



Évaluation



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



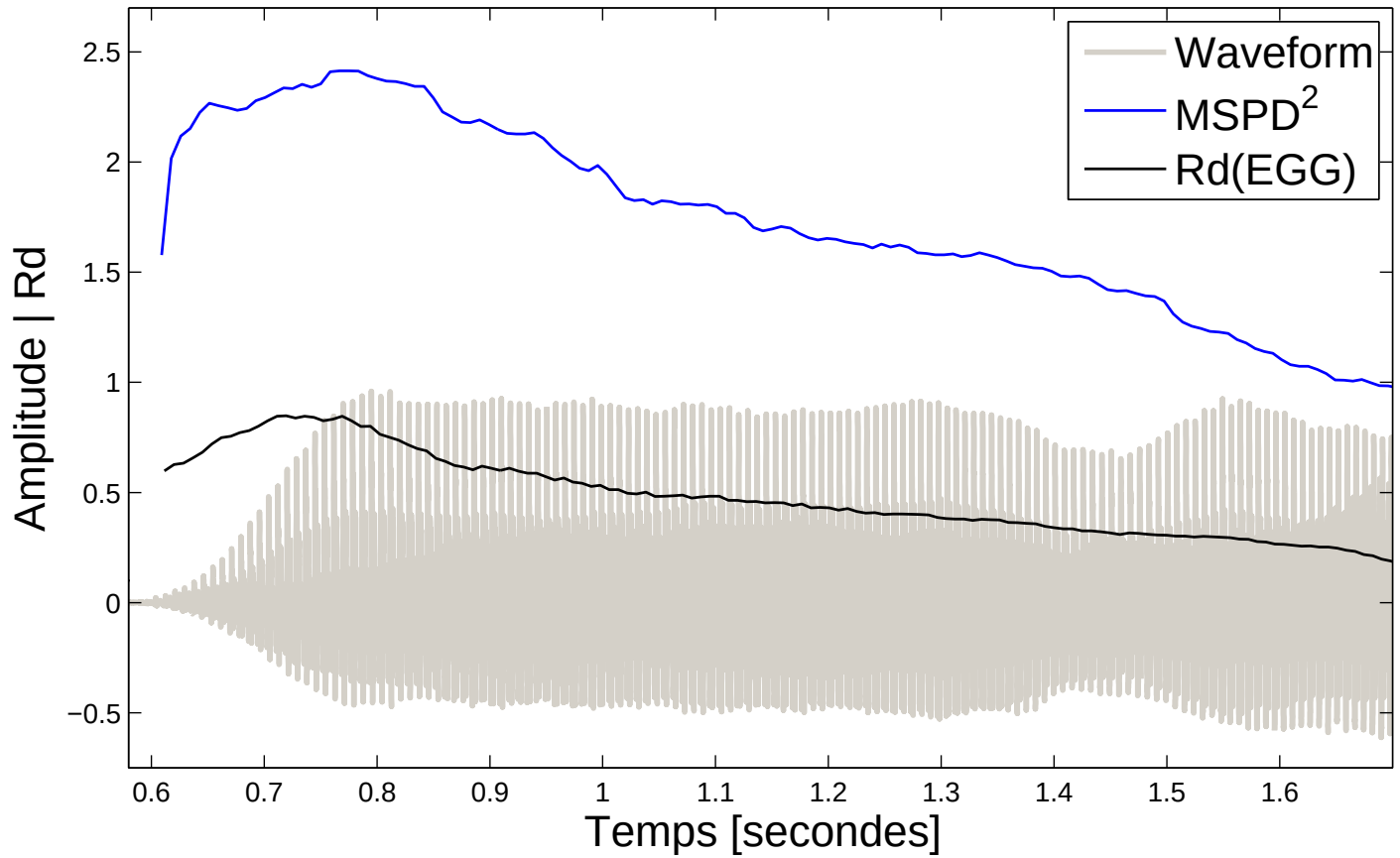
IAIF P. Alku, and H. Tiitinen and R. Naatanen, *A method for generating natural-sounding speech stimuli for cognitive brain research.*

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.*

Exemples d'estimation

Son



Application

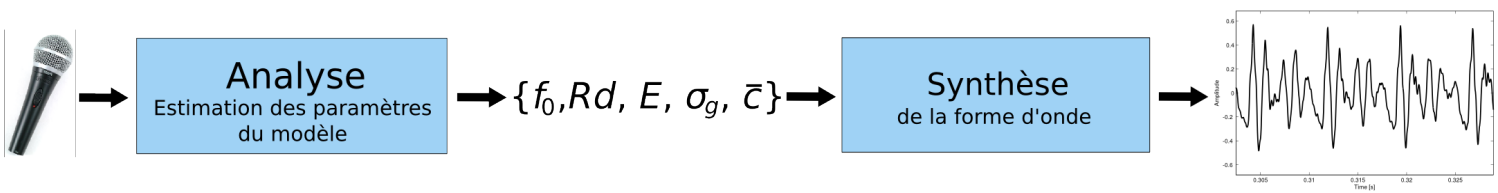
Méthode d'analyse/synthèse pour la transformation de la voix

Méthode d'analyse/synthèse - SVLN

Modèle de la production vocale pour SVLN

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

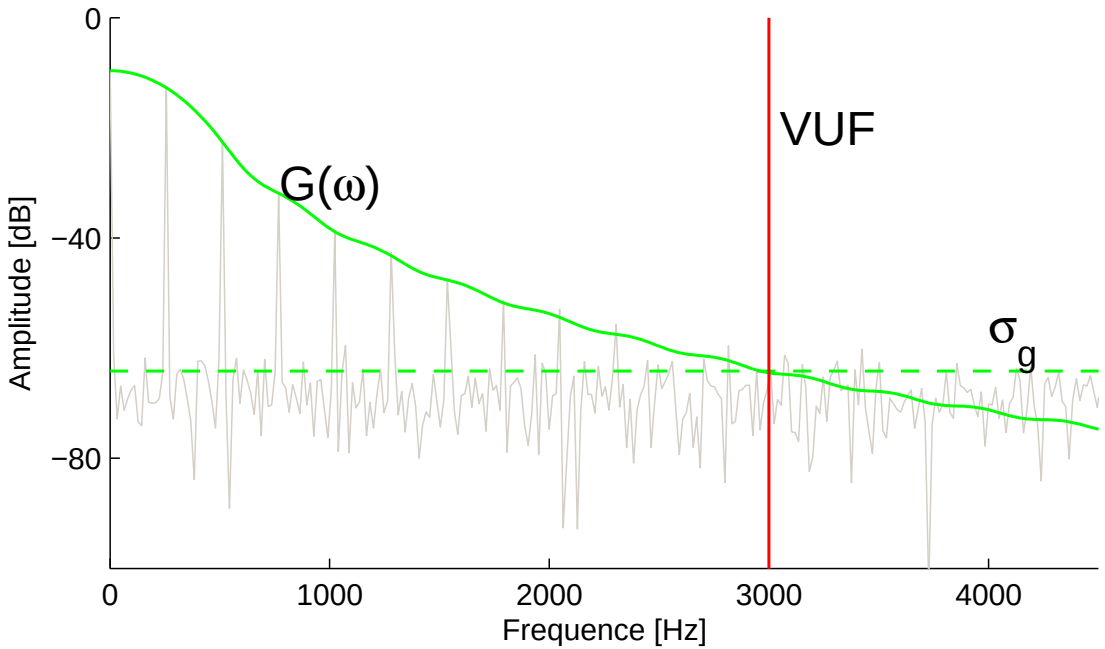
Procédure d'analyse/synthèse:



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

Analyse - Estimation des paramètres de la source glottique

- f_0 Connu *a priori*
- Rd Méthode basée sur MSPD[2]
- E Log énergie de la fenêtre
- σ_g Point de croisement entre $G(\omega)$ et VUF



VUF connu *a priori* par classification des pics spectraux.
Décision de voisement temporelle annotée manuellement.

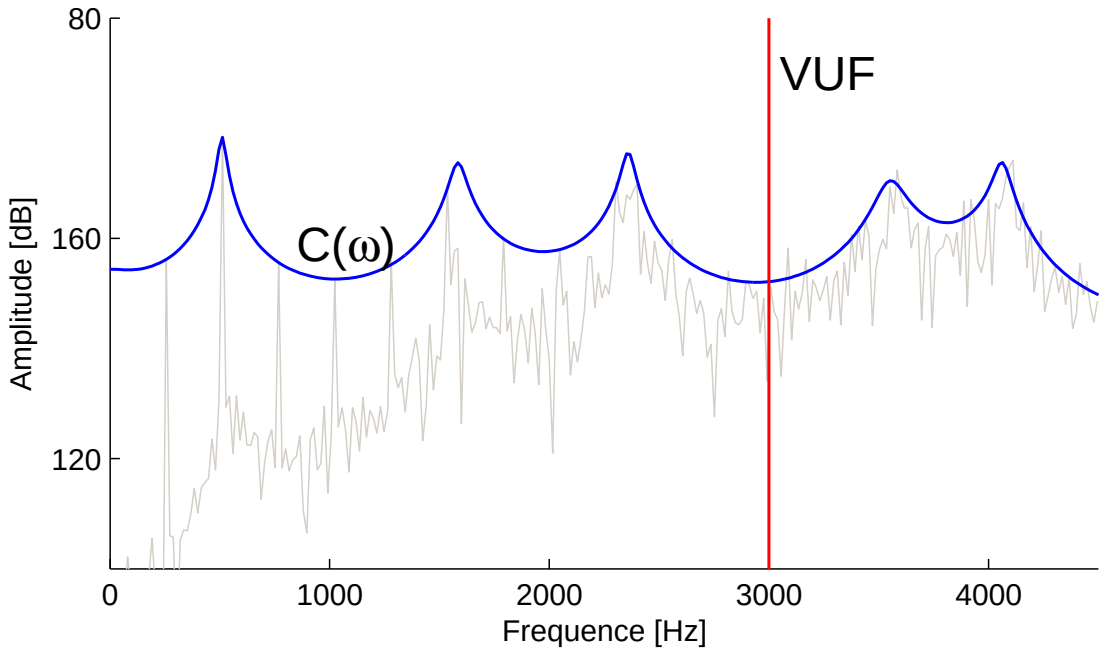
Analyse - Estimation du filtre du conduit vocal

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

$\mathcal{T}(.)$ La *True-envelope*

$\mathcal{P}(.)$ Le cepstre réel

$\gamma = \sum_t \text{win}[t]/(f_s/f_0)$ nombre de périodes dans la fenêtre.



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

Synthèse

- Synthétise période après période
- *Overlap-add* les périodes

Synthèse

- Synthétise période après période
- *Overlap-add* les périodes

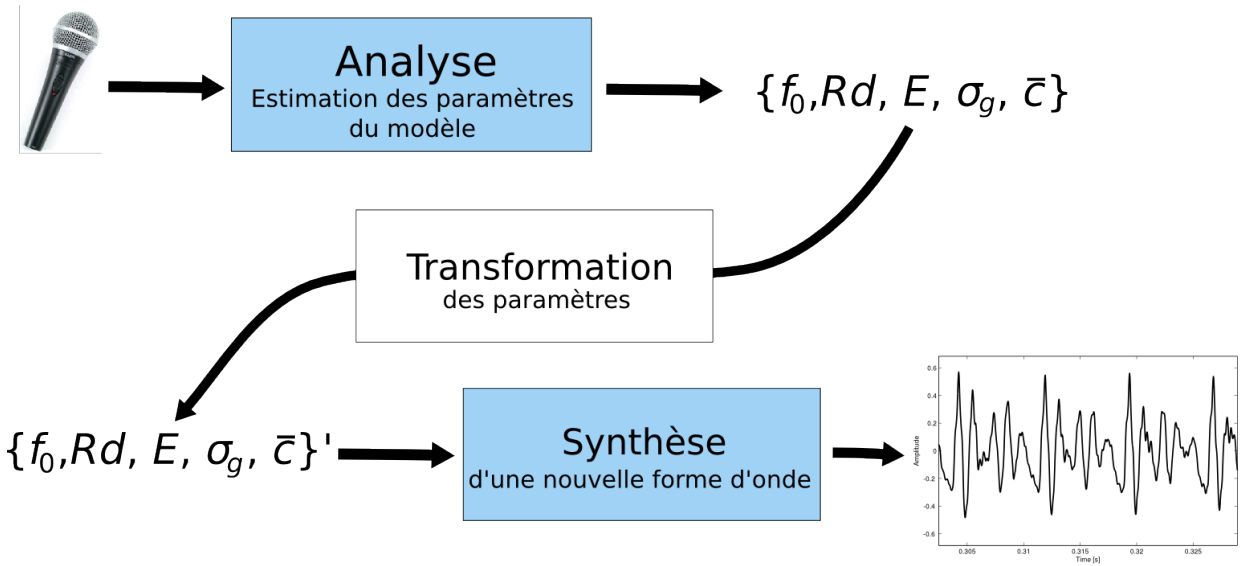
Pour un spectre observé $S(\omega)$:

$|S(\omega)|$ toujours reproduit

$\angle S(\omega)$ imposé par le modèle LF, le bruit Gaussien et $\angle C_{-}(\omega)$

Évaluation - Transformation de la voix

Procédure de transformation



Évaluation - Transposition

SVLN La méthode proposée -1oct Original +1oct

SHIP *SHape Invariant Phase vocoder*

STRAIGHT *Adaptive Interpolation of weiGHTed spectrum*

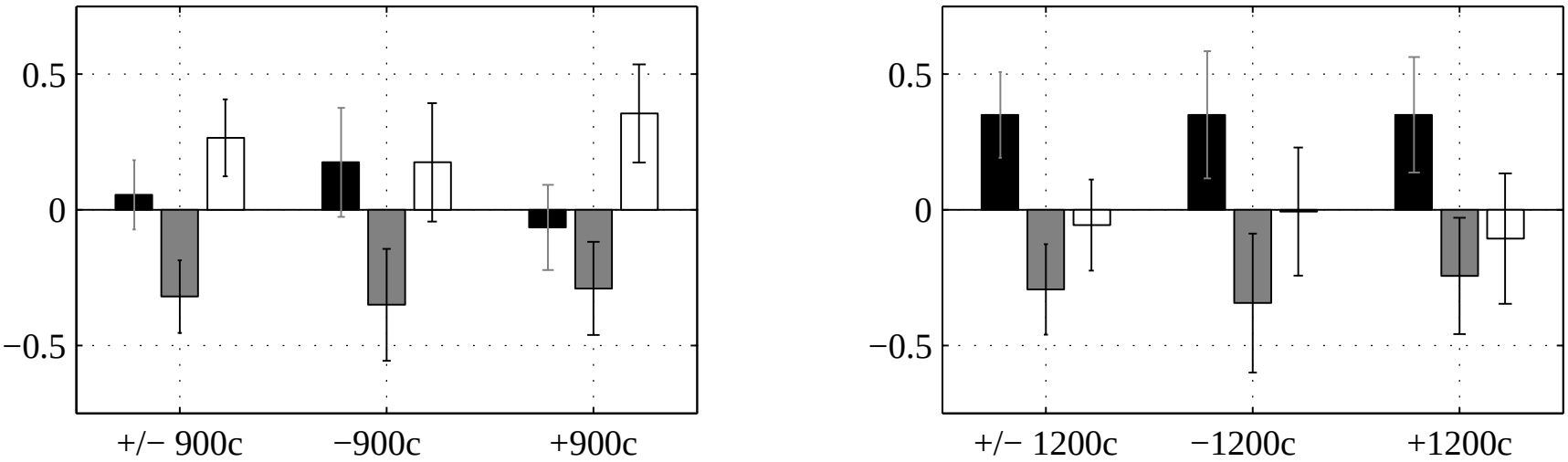


Figure: Préférences des méthodes pour différents facteurs de transposition (+/−T : sans distinction de direction).

SVLN est significativement préférée pour des transpositions fortes.

Évaluation - *Breathiness*

Rd est lié à la qualité vocale: tendu $\leftrightarrow$ relâché.

$$Rd' = c \cdot Rd$$

c=0.5 Original c=2

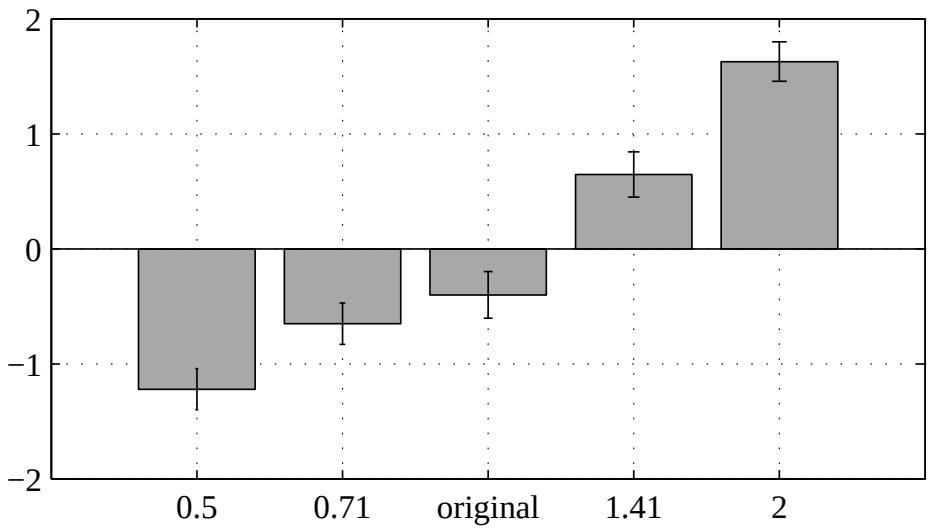


Figure: Scores de *breathiness*

Démonstrations

- Original
- Enfant
- Grave

- Original
- Enfant
- Grave

Conclusions

Contributions et directions futures

Contributions

- Estimation de paramètres d'un modèle glottique
  1. Minimisation de phase pour l'estimation de la forme.
     (Propriétés d'un modèle nécessaires pour une estimation fiable)
  2. Fonctions de distorsion de phase caractérisant le pulse glottique.
  3. Expression quasi explicite d'un paramètre de forme.

Contributions

- Estimation de paramètres d'un modèle glottique
  1. Minimisation de phase pour l'estimation de la forme.
     (Propriétés d'un modèle nécessaires pour une estimation fiable)
  2. Fonctions de distorsion de phase caractérisant le pulse glottique.
  3. Expression quasi explicite d'un paramètre de forme.

- Méthode d'analyse/synthèse SVLN
  1. Significativement préférée pour des transpositions fortes.
  2. Permet un contrôle de la *breathiness*

Directions futures

Modèles glottiques et leurs paramètres

1. Étudier les autres modèles avec de multiples paramètres (ex. LF et (O_q, α_m, Q_a))
2. Développer un modèle considérant la méthode d'estimation choisie. Utiliser les récentes avancées en vidéo-endoscopie à haute-vitesse.
3. Considérer le couplage débit glottique / conduit vocal

Directions futures

Modèles glottiques et leurs paramètres

1. Étudier les autres modèles avec de multiples paramètres (ex. LF et (O_q, α_m, Q_a))
2. Développer un modèle considérant la méthode d'estimation choisie. Utiliser les récentes avancées en vidéo-endoscopie à haute-vitesse.
3. Considérer le couplage débit glottique / conduit vocal

Méthode d'analyse/synthèse

- Modéliser les phases entre source déterministe et source aléatoire.

Merci pour votre attention