TRANSFORMATION DE LA VOIX À L'AIDE D'UN MODÈLE DE SOURCE GLOTTIQUE

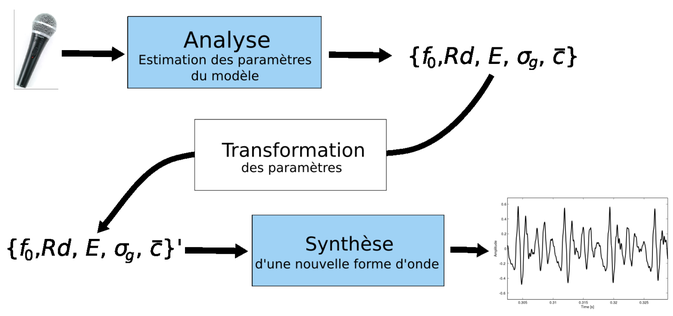
Gilles Degottex, Axel Röbel, Xavier Rodet

IRCAM - CNRS-UMR9912-STMS, Équipe Analyse/Synthèse, Paris

gilles.degottex@ircam.fr

Nous présentons ci-dessous, les résultats d'une méthode d'analyse/synthèse pour la transformation de la voix. Cette méthode est appelée *Separation of the Vocal-tract with the Liljencrants-Fant model plus Noise* (SVLN). Cette méthode utilise une description analytique de la composante déterministe de la source vocale, un modèle glottique.

Applications: la transformation de la voix, la synthèse de la parole, la synthèse d'expressivités, la conversion de la voix ... les arts contemporains, les industries de la musique et du cinéma, les jeux vidéos, etc.



MODÈLE 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 de Liljencrants-Fant [1]

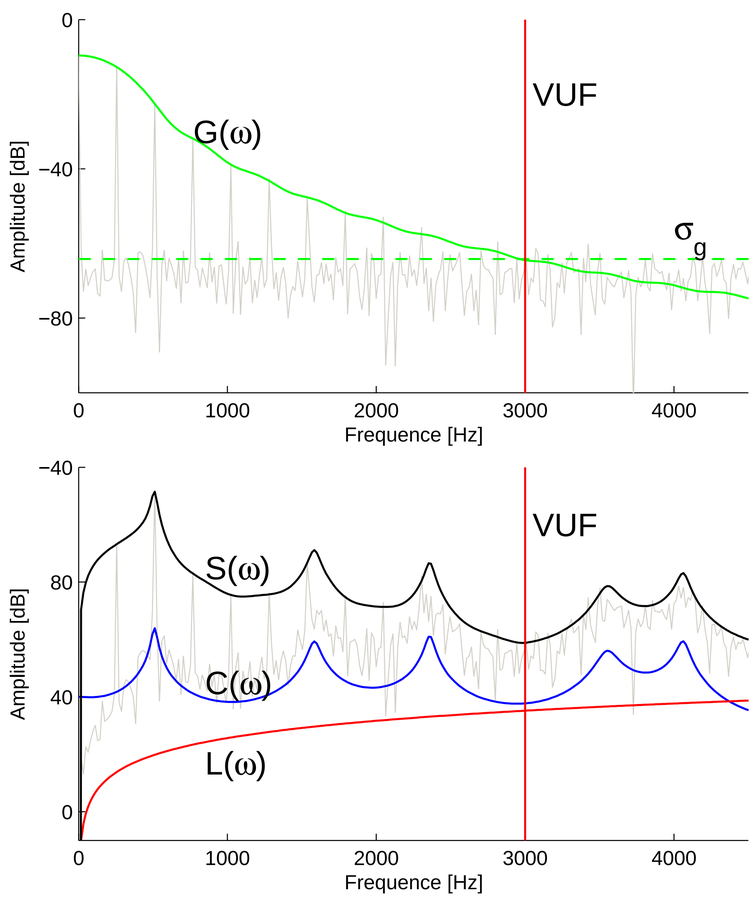
$e^{j\omega\phi}$ Position du modèle glottique

$H^{f_0}(\omega)$ Périodicité

$N^{\sigma_g}(\omega)$ Bruit de turbulence, *hyp:* Gaussien d'écart-type σ_g.

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

$j\omega$ Rayonnement



ANALYSE/SYNTHÈSE SVLN

ANALYSE - PARAMÈTRES DE LA SOURCE GLOTTIQUE

f_0 Connu *a priori*

Rd Méthode basée sur MSPD^2 [2]

E Log énergie de la fenêtre

σ_g Point de croisement entre $G(\omega)$ et VUF

VUF connu *a priori*

Décision de voisement temporelle annotée manuellement.

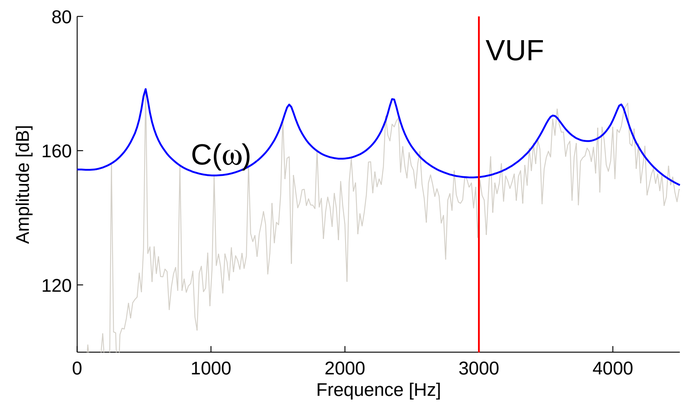
ANALYSE - FILTRE DU CONDUIT-VOCAL

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

$\mathcal{T}(.)$ La *True-envelope* [3]

$\mathcal{P}(.)$ Le cepstre réel tronqué (+ correction du gain [4])

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



SYNTHÈSE

- Période après période
- *Overlap-add* des périodes

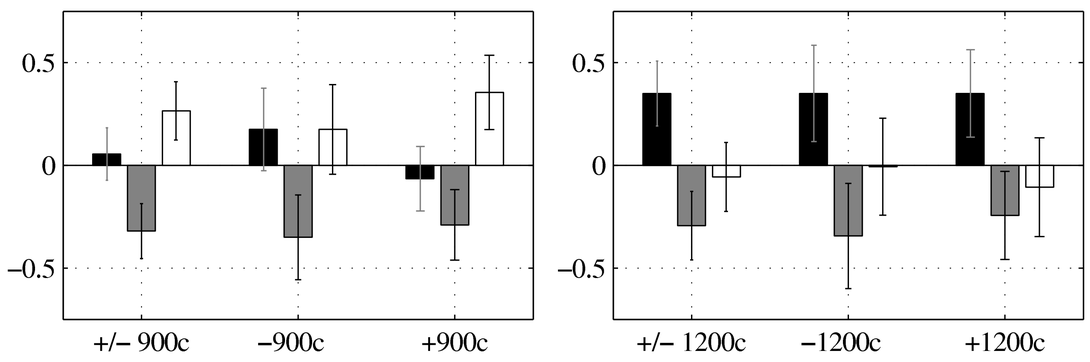
ÉVALUATION

TRANSPOSITION $f_0' = 2^{c/1200} \cdot f_0$

SVLN La méthode proposée

SHIP *SHape Invariant Phase vocoder* [5]

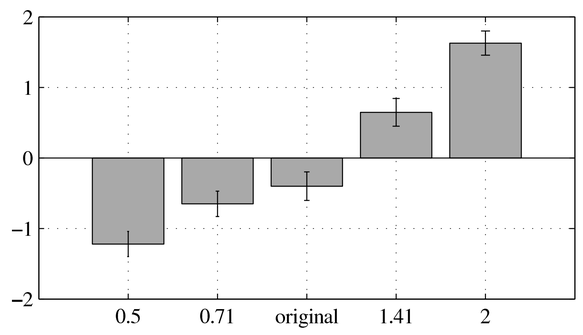
STRAIGHT *Adaptive Interpolation of weiGHTed spectrum* [6]



Préférences des méthodes pour différents facteurs de transposition

(+/–T : sans distinction de direction).

Breathiness $Rd' = c \cdot Rd$



CONCLUSIONS

Pour un spectre observé $S(\omega)$:

$|S(\omega)|$ toujours reproduit

$\angle S(\omega)$ imposé par le modèle LF, le bruit Gaussien et $\angle C(\omega)$

Les test de préférences permettent de dire que SVLN est:

- Préférée pour des transpositions fortes.
- Permet un contrôle de la *breathiness*.

[1] G. Fant, "The LF-model revisited. transformations and frequency domain analysis.," *STL-QPSR*, vol. 36, no. 2-3, pp. 119–156, 1995.

[2] G. Degottex, A. Roebel, and X. Rodet, "Phase minimization for glottal model estimation," *IEEE ASLP*, vol. PP, no. 99, pp. 1 –1, 2010.

[3] A. Roebel and X. Rodet, "Efficient spectral envelope estimation and its application to pitch shifting and envelope preservation," in *DAFx*, 2005.

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

[5] Axel Roebel, "A shape-invariant phase vocoder for speech transformation," in *DAFx*, 2010.

[6] H. Kawahara, I Masuda-Katsuse, and A. Cheveigne, "Restructuring speech representations using a pitch-adaptative time-frequency smoothing and an instantaneous-frequency-based f0 extraction: Possible role of a repetitive structure in sounds," in *Speech Communication*, 1999, vol. 27.