Articulatory synthesis of continuous speech:
–
Global approach and copy synthesis

Benjamin Elie[1,2]

[1]LORIA, INRIA/CNRS, Nancy
[2]IADI, INSERM/CHU-Nancy

December 9, 2015

Résumé

Master 2

- Master ATIAM (*Acoustics, Signal Processing and Computer Science Applied to Music*), UPMC-Paris VI, 2009.
- Research internship at the Music Acoustics Lab of Sydney : *Characterization of Vocal Tract Acoustics in the Case of Oronasal Coupling*

Résumé

Master 2

- Master ATIAM (*Acoustics, Signal Processing and Computer Science Applied to Music*), UPMC-Paris VI, 2009.
- Research internship at the Music Acoustics Lab of Sydney : *Characterization of Vocal Tract Acoustics in the Case of Oronasal Coupling*

PhD thesis

PhD in **Acoustics** (LAUM and Télécom ParisTech, 2012): *Acoustic and vibratory characterization of stringed musical instruments – Application to lutherie assistance*. Supervised by François Gautier and Bertrand David

Résumé

Master 2

- Master ATIAM (*Acoustics, Signal Processing and Computer Science Applied to Music*), UPMC-Paris VI, 2009.
- Research internship at the Music Acoustics Lab of Sydney : *Characterization of Vocal Tract Acoustics in the Case of Oronasal Coupling*

PhD thesis

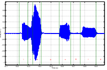
PhD in **Acoustics** (LAUM and Télécom ParisTech, 2012): *Acoustic and vibratory characterization of stringed musical instruments – Application to lutherie assistance*. Supervised by François Gautier and Bertrand David

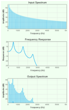
Postdoctoral experience

- ATER mechanics/acoustics (Université du Maine, 2012-2013)
- Postdoctoral fellowship Inria at LORIA. Articulatory synthesis: forward and inverse problem, (MULTISPEECH group), oct. 2013-2015.
- Postdoctoral fellowship CNRS at LORIA and IADI. Real-time acquisition of articulatory data by MRI techniques, 2015

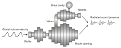
Classification of techniques for speech synthesis

1)

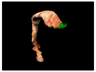




2)



3)



Signal-based approach ← → Physics-based approach

Concatenative | Source-filter | Simplified models | Finite elements

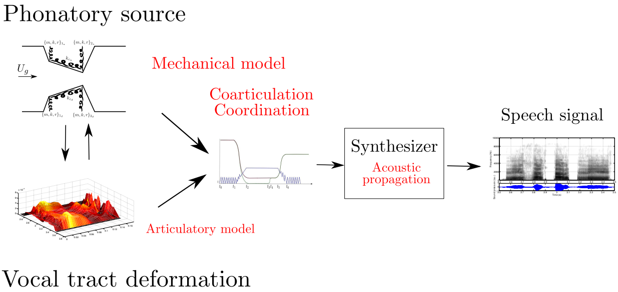
- Speech synthesis based on physical/acoustical models
- continuous time-domain, word/phrase level utterances
- simulation of acoustic and articulatory phenomena

1) http://www.phon.ucl.ac.uk/
2) http://www.vocaltractlab.de/
3) http://www.magic.ubc.ca/

Principle

Speech synthesis (utterances), **complete** and **realistic**, based on purely acoustical model

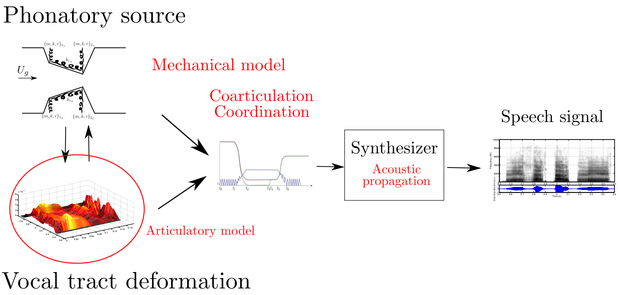
Example of an articulatory synthesizer



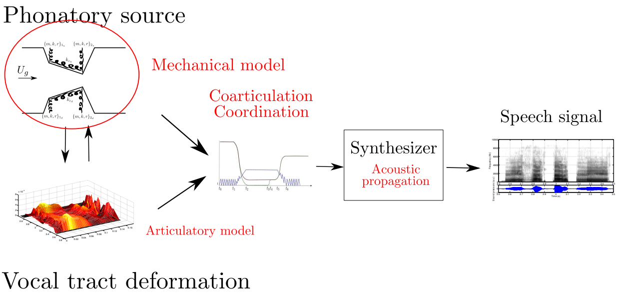
- Realistic acoustics
- Articulatory control

Applications: Medicine, audiovisual, language learning, text-to-speech...

Outline

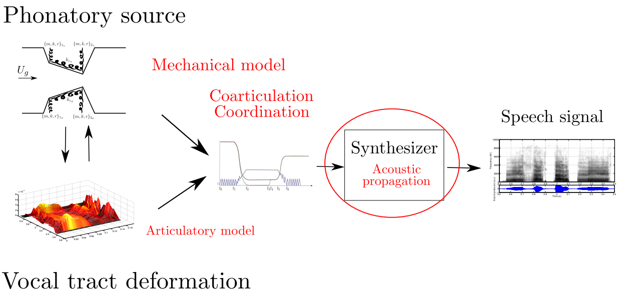


1. Articulatory model



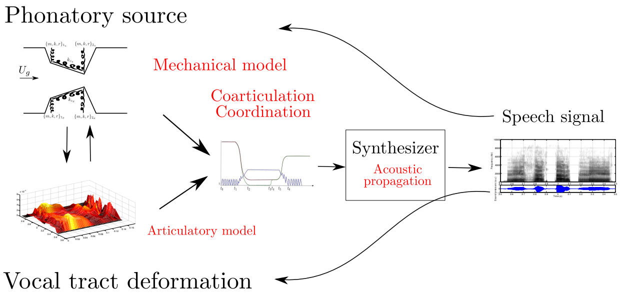
1. Articulatory model
2. Glottis model

Outline



1. Articulatory model
2. Glottis model
3. Acoustic propagation

Outline



1. Articulatory model
2. Glottis model
3. Acoustic propagation
4. Hybrid Synthesis

Plan

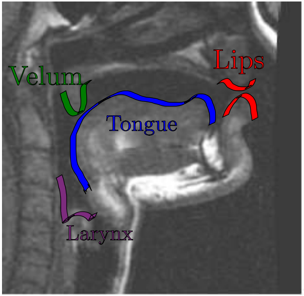
1. Articulatory model
2. Source model
3. Speech synthesis based on acoustical model
4. Copy and hybrid syntheses
5. Conclusions

Input parameters (Degrees of freedom)

2 categories :

- Vocal tract geometry (resonator)
- Glottal parameters (source)

Resonator :



Source :

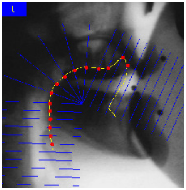
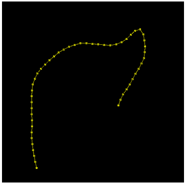
- Vocal folds partial abduction
- Fundamental frequency
- Laryngeal mechanisms
- Vocal folds asymmetry
- Sub-glottal pressure
- ...

Articulatory data

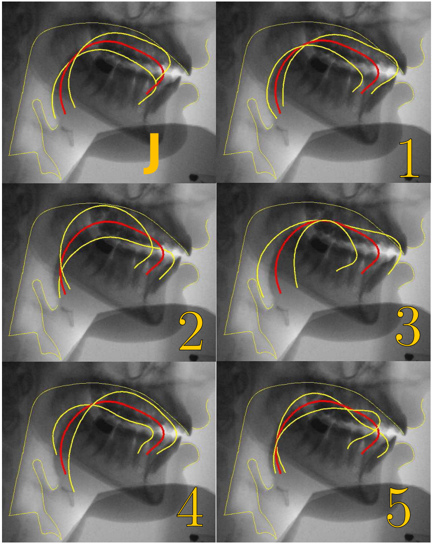
Making the articulatory model

- Large database
- Factorial analysis to reduce the number of components (PCA)
- Geometry of the vocal tract reduced to a few number of parameters

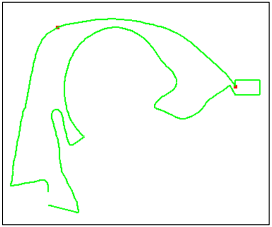
Which data ?


Tongue modes

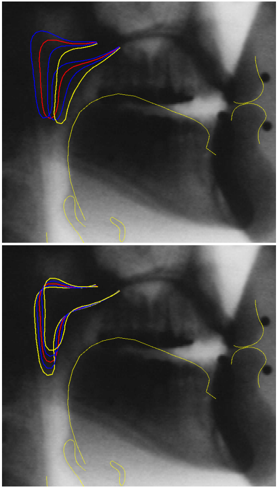


Fist mandible mode
and
5 first tongue modes

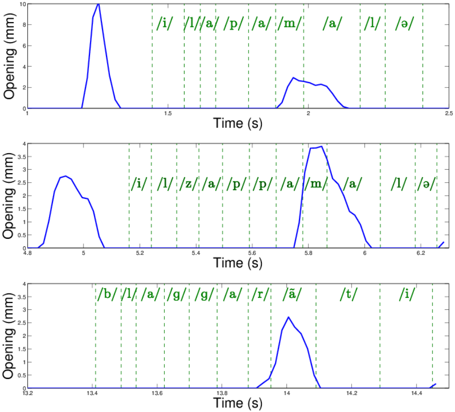


Complete model

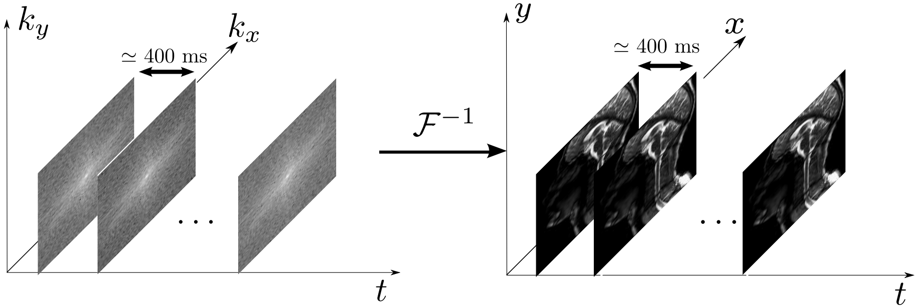
Velum modes



VPO for a few French utterances
(Laprie and Elie, ICPhS, 2015)



Acqusitions by MRI techniques: principles

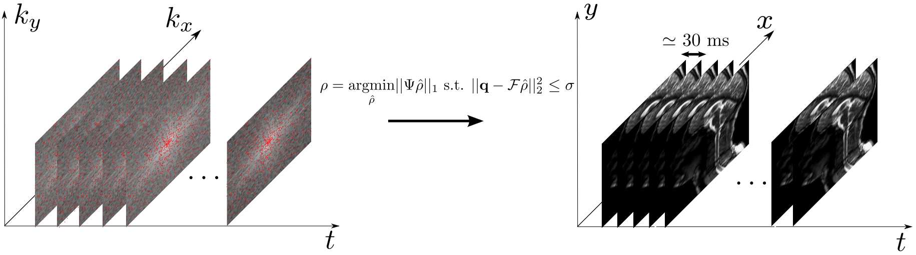


Reconstruction of midsagittal slices

- Full *k*-space sampling
  $\rightarrow$ bad temporal resolution

Sparse reconstruction (*Compressed Sensing*)

Using the sparsity for better temporal resolution

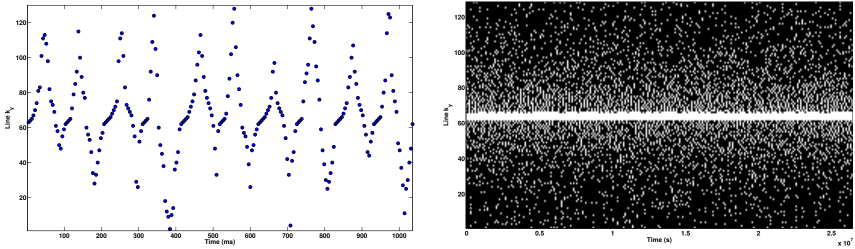


Sparse transform

- $x - f$ space : Temporal Fourier transform of the image space
- $w - f$ space : Temporal Fourier transform of the wavelet transform of the image space

Design of the subsampling pattern

Phase encoding (lines of the k-space)

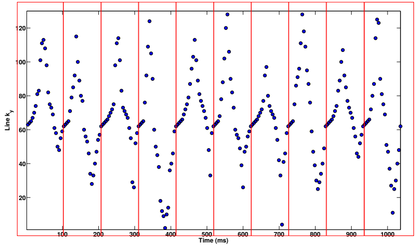


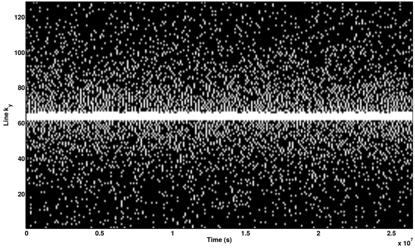
Suitable probability density

- full sampling of the central lines
- $pdf \propto 1/r^2$
- Partial phase line encoding for partial Fourier reconstruction

Design of the subsampling pattern

Phase encoding (lines of the *k*-space)





Suitable probability density

- full sampling of the central lines
- $pdf \propto 1/r^2$
- Partial phase line encoding for partial Fourier reconstruction

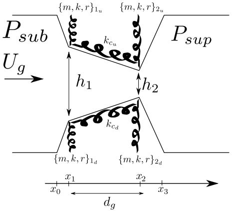
A few videos. . .

- "Des abat-jours" (/dezabaʒuʁ/), 36 Hz, 128×128 pixels
- /ara/, 45 Hz, 256×256 pixels

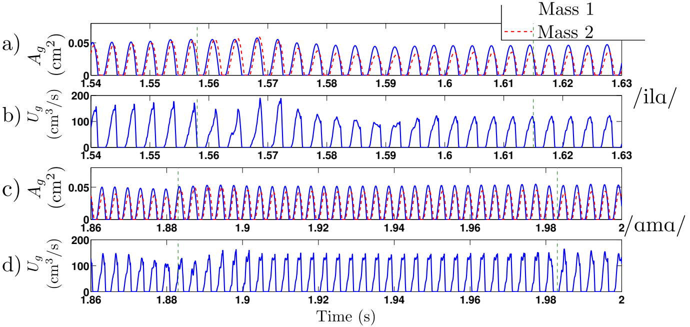
→ More data to be acquired

Plan

Self-oscillating model of the vocal folds



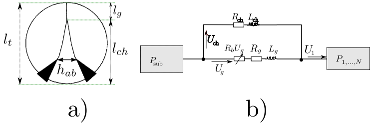
$$\mathbf{M\ddot{y}}+\mathbf{R\dot{y}}+\mathbf{Ky} = f(P_{sup}, P_{sub}, \theta_{geom})$$



Glottis partial closure

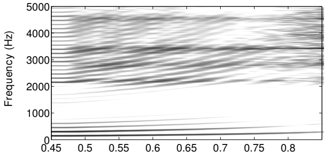
Glottal chink[a]

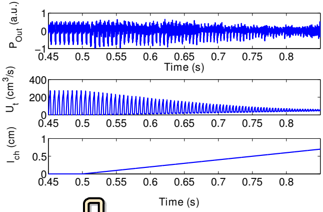
[a]Elie and Laprie, *Speech Commnication*, submitted



$$P_{sub} - P_1 = \Delta P_{close} + \frac{\partial}{\partial t}\left(L_1 U_{ch} + R_1 U_{ch}\right)$$

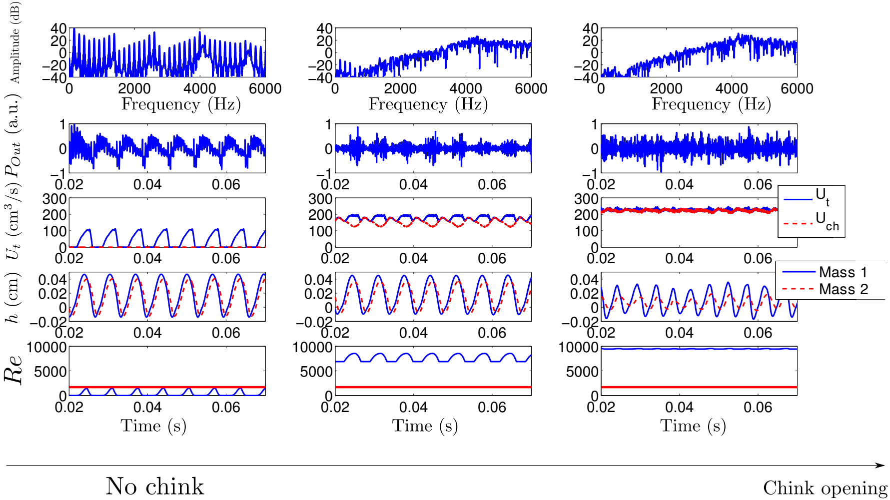
Exemple : /i/





Voiced sibilant fricative modeling : glottal partial closure

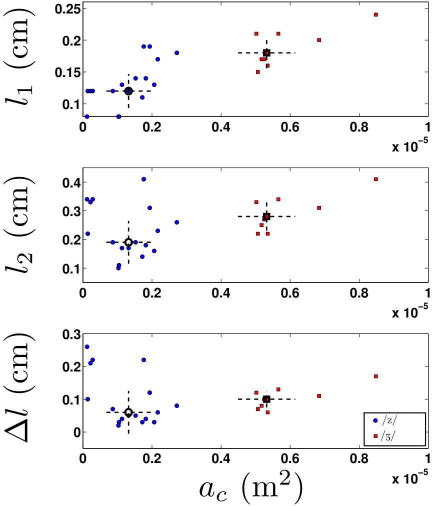
Voicing degree as a function of the glottal chink opening : /z/



→ **Glottal chink acts on voicing degree**

Example with sibilants

Conditions for producing voiced fricatives

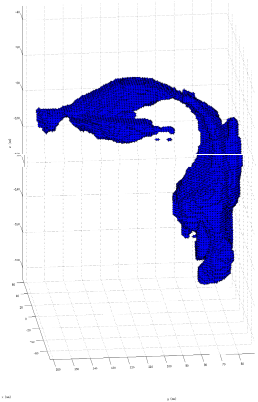


/z/ more likely devoiced than /ʒ/ ?

Plan

Simplification of the vocal tract geometry

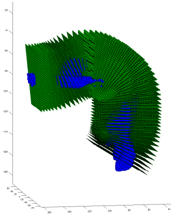
A complicated acoustic resonator



- Complex geometry
- Numerous side cavities

Simplification of the vocal tract geometry

A complicated acoustic resonator

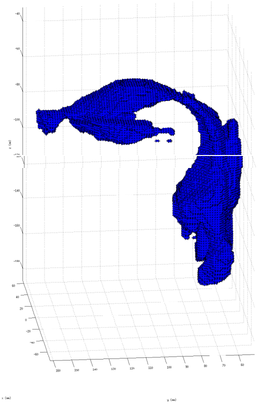


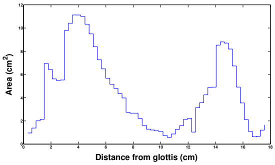
- Complex geometry
- Numerous side cavities

$\rightarrow$ segmentation of the vocal tract into a series of tubelets

Vocal tract sampling

Definition of the area function

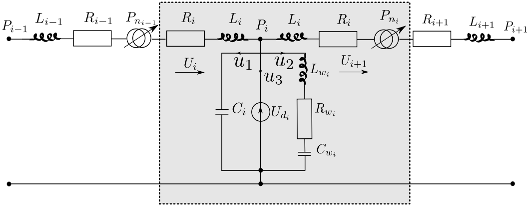




Vocal tract geometry defined by two vectors **a** and **l**.

From the area function to speech: electric analogy

Equivalent lumped circuit elements of a tubelet[1]:



Electric	Acoustic
Current	Volume velocity u
Voltage	Pressure p
R_i	Energy loss
C_i	Air compliance
L_i	Air inertance
R_{w_i}	Wall resistance
C_{w_i}	Wall compliance
L_{w_i}	Wall inertance
U_{d_i}	Flow source
P_{n_i}	Friction noise source

Propagation equations

$$\begin{aligned} P_{i-1} - P_i &= \frac{\partial}{\partial t}\left[(L_{i-1} + L_i)\, U_i\right] + \left(R_{i-1} + R_i + R_{n_{i-1}}\right) U_i + P_{n_{i-1}} \\ U_i - U_{i+1} &= u_1 + u_2 + u_3 \end{aligned}$$

[1] S. Maeda, *Speech Commnication*, 1982

Matrix form of the propagation equations

Solution for the sampled system

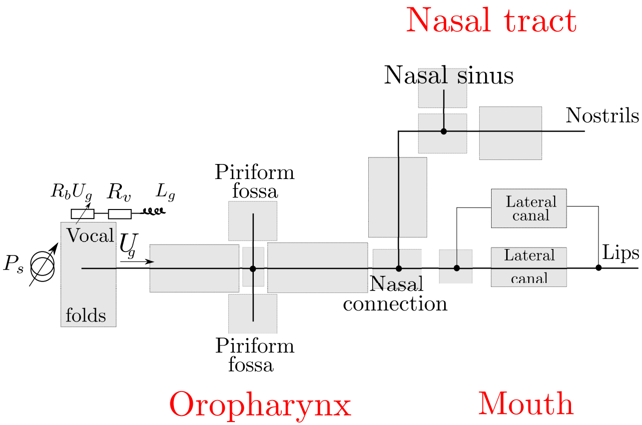
$$b_{i-1}(n)U_{i-1}(n) + Z_i(n)U_i(n) + b_i(n)U_{i+1}(n) = F_i(n)$$

$$\downarrow$$

$$\underbrace{\begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \\ \vdots \\ F_M \\ F_{M+1} \end{bmatrix}}_{\mathbf{f}} = \underbrace{\begin{bmatrix} Z_1 & b_1 & 0 & & \dots & & 0 \\ b_1 & Z_2 & b_2 & 0 & & & \\ 0 & \ddots & \ddots & \ddots & 0 & & \vdots \\ & 0 & b_{n-1} & Z_n & b_n & 0 & \\ \vdots & & 0 & \ddots & \ddots & \ddots & 0 \\ & & & 0 & b_{M-1} & Z_M & b_M \\ 0 & & \dots & & 0 & b_M & Z_{M+1} \end{bmatrix}}_{\mathbf{Z}} . \underbrace{\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \\ \vdots \\ U_M \\ U_{M+1} \end{bmatrix}}_{\mathbf{u}}$$

Generalization to a waveguide network

Simultaneous consideration of several side branches



Waveguide network: matrix form

$$\begin{bmatrix} \mathbf{f}^{(1)} \\ \mathbf{f}^{(2)} \\ \vdots \\ \mathbf{f}^{(\mathcal{N})} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}^{(1)} & \mathbf{C}_{(1,2)}^T & \cdots & \mathbf{C}_{(1,\mathcal{N})}^T \\ \mathbf{C}_{(1,2)} & \mathbf{Z}^{(2)} & & \\ \vdots & & \ddots & \\ \mathbf{C}_{(1,\mathcal{N})} & & & \mathbf{Z}^{(\mathcal{N})} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{u}^{(1)} \\ \mathbf{u}^{(2)} \\ \vdots \\ \mathbf{u}^{(\mathcal{N})} \end{bmatrix}$$

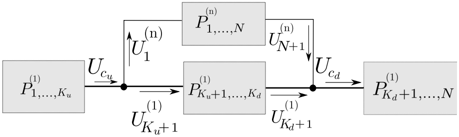
Sparse coupling matrices

4 cases for $\mathbf{C}_{(m,n)}$

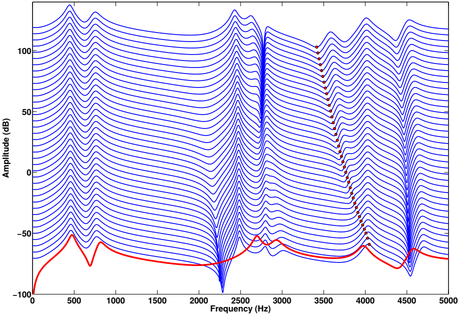
- m and n independent $\rightarrow \mathbf{C}_{(m,n)} = \mathbf{0}$
- m parent of $n \rightarrow \mathbf{C}_{(m,n)} = \mathbf{0}$, except $c_{1,K} = b_K^{(m)}$ and $c_{1,K+1} = Z_C^{(m,n)}$
- m and n twins, from parent $p \rightarrow \mathbf{C}_{(m,n)} = \mathbf{0}$, except $c_{1,1} = Z_C^{(p,n)}$
- n anabranch of $m \rightarrow \mathbf{C}_{(m,n)} = \mathbf{0}$, except $c_{1,K_u} = b_{K_u}^{(m)}$, $c_{1,K_u+1} = Z_{C_u}^{(m,n)}$, $c_{N+1,K_d+2} = b_{K_d+1}^{(m)}$ and $c_{N+1,K_d+1} = Z_{C_d+1}^{(m,n)}$

Bilateral consonants

Anastomoses :[2]



Validation : effect of the bilateralization on the VT transfer functions



[2] Elie and Laprie, *Speech Commnication*, submitted

Bernoulli term: introduction of non-linearity at the glottis

Small constriction → Bernoulli non negligible

$$\Delta P = \frac{1}{2}\rho v^2 = \frac{\rho}{2}\frac{U_1^2}{a_g^2}$$

$$P_{sub} - P_1 = R_b U_1^2 + R_v U_g + \frac{\partial}{\partial t}\left[L_g U_1\right]$$

$$F_1 = Z_1 U_1 + b_1 U_2 + R_b U_1^2$$

↓

Bernoulli term: introduction of non-linearity at the glottis

Small constriction → Bernoulli non negligible

$$\Delta P = \frac{1}{2}\rho v^2 = \frac{\rho}{2}\frac{U_1^2}{a_g^2}$$

$$P_{sub} - P_1 = R_b U_1^2 + R_v U_g + \frac{\partial}{\partial t}\left[L_g U_1\right]$$

$$F_1 = Z_1 U_1 + b_1 U_2 + R_b U_1^2$$

$$\downarrow$$

$$\mathbf{f} = \mathbf{Z}\mathbf{u}_Z + \mathbf{Q}\mathbf{u}_Q$$

$$\mathbf{u}_Z = \left[U_1, \ldots, U_{N+1}\right]^T, \quad \mathbf{u}_Q = \left[U_1^2, 0, \ldots, 0\right]^T$$

Non linear system, more delicate to solve

Generalized matrix form

$$\begin{bmatrix} \mathbf{f}^{(1)} \\ \mathbf{f}^{(2)} \\ \vdots \\ \mathbf{f}^{(\mathcal{N})} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}^{(1)} & \mathbf{C}_1^{(2)\,T} & \dots & \mathbf{C}_1^{(\mathcal{N})\,T} \\ \mathbf{C}_1^{(2)} & \mathbf{Z}^{(2)} & & \\ \vdots & & \ddots & \\ \mathbf{C}_1^{(\mathcal{N})} & & & \mathbf{Z}^{(\mathcal{N})} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{u}^{(1)} \\ \mathbf{u}^{(2)} \\ \vdots \\ \mathbf{u}^{(\mathcal{N})} \end{bmatrix}$$

With vocal folds:

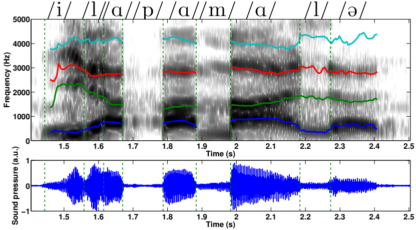
$$\mathbf{f} = \mathbf{Z}\mathbf{u}_Z + \mathbf{Q}\mathbf{u}_Q,$$

$\mathbf{u}_Q = [U_1^2, U_2^2, \dots, U_N^2]^T$, $\mathbf{Q}$ contains 0, except $Q_{1,1} = Be$
Rearrangement : $\mathbf{Z}^{-1}\mathbf{f} = \mathbf{u}_Z + \mathbf{Z}^{-1}\mathbf{Q}\mathbf{u}_Q$

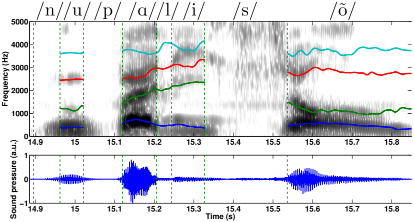
Plan

A few results from X-ray images...

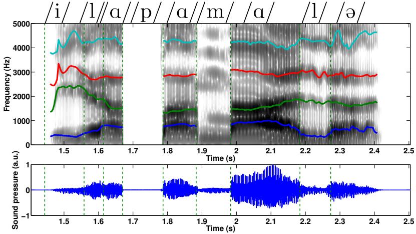
a) Original



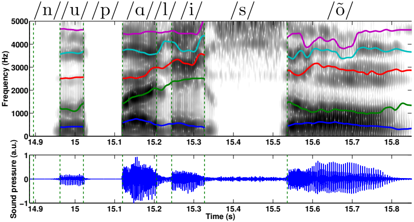
a) Original



b) copy



b) copy



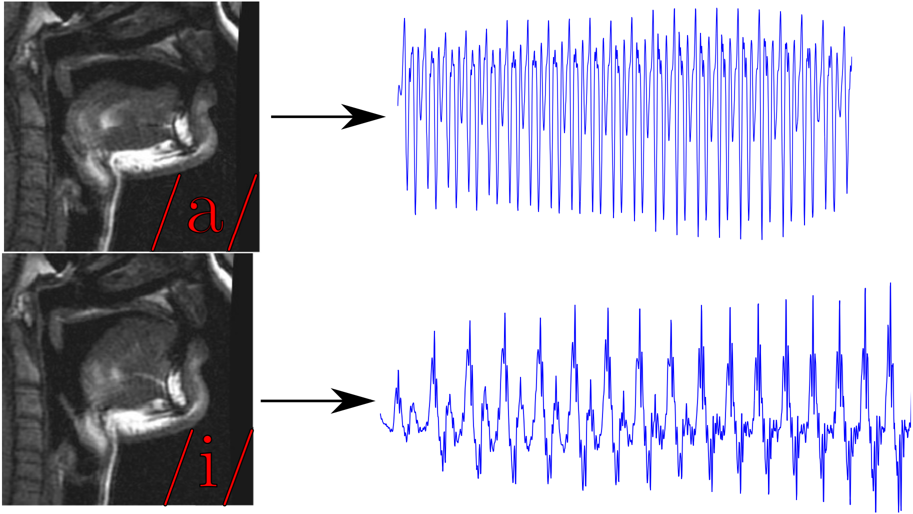
Details: /nupɑlisõ/

Observation of acoustics phenomena during speech production

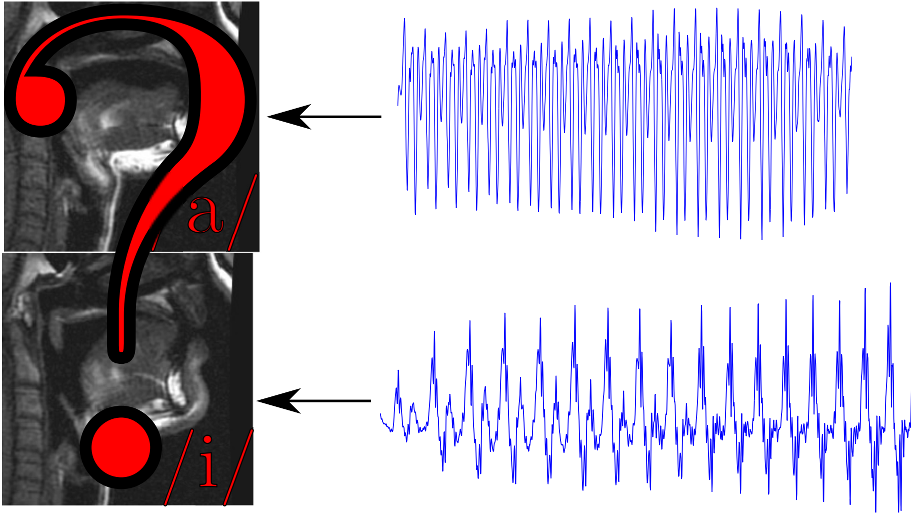
- Vocal folds oscillation
- Pressure wave propagation along the vocal tract and/or the nasal tract

Example for utterance /nupɑlisõ/ (run video)

Acoustic inversion



Acoustic inversion



Acoustic to area inversion : principle

Search input parameters $\mathbf{p}$ so that

$$\mathbf{s} = \mathcal{L}(\mathbf{p})$$

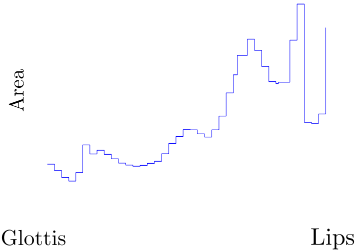
- $\mathbf{p}$: Vocal Tract (VT) configuration, *e.g.* area function, length function, articulatory parameters...
- $\mathbf{s}$: acoustic vector, *e.g.* formant frequency, cepstral coefficients...
- $\mathcal{L}(\mathbf{p})$: operator giving the acoustic vector for a certain VT configuration $\mathbf{p}$.

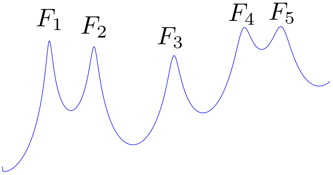
Chosen parameters

- Input parameter : Area function of the vocal tract

$$\mathbf{p} = [a_1, a_2, \ldots, a_N, l_1, l_2, \ldots, l_N]^T$$

- Output acoustic vector : M first formant frequencies ($\mathbf{s} = \mathbf{f} = [F_1, \ldots F_M]^T$)
- Model $\mathcal{L}$: Chain matrix paradigm[3] $\rightarrow$ VT transfer function (formant frequencies = resonance frequencies)





[3] Sondhi and Schroeter, 1987

Iterative method

Inverse Jacobian technique

1. Computation of $\mathbf{J} = \frac{\partial \mathbf{f}}{\partial \mathbf{p}}$, so that $\Delta \mathbf{f} = \mathbf{J} \Delta \mathbf{p}$
2. $\Delta \mathbf{p} = \mathbf{J}^T \Delta \mathbf{f}$

- Requires an initial function $\mathbf{p}_0$
- $\Delta \mathbf{p} = \mathbf{J}^T \Delta \mathbf{f}$ is true only for small variations. The process is performed iteratively until $\Delta \mathbf{f}$ vanishes

How to compute J

Fant and Pauli theory of perturbation

$$\left[\frac{\Delta F_m}{F_m}\right]_{\mathbf{a}} = \sum_{n=1}^{N} S_n^a(F_m)\frac{\Delta a_n}{a_n} \qquad \left[\frac{\Delta F_m}{F_m}\right]_{\mathbf{l}} = \sum_{n=1}^{N} S_n^l(F_m)\Delta\lambda_n$$

$$\Delta\lambda_n = -\frac{\Delta l_n}{l_n + \Delta l_n}$$

$$S_n^a(F_m) = \frac{\mathcal{T}_n(F_m) - \mathcal{V}_n(F_m)}{\mathcal{H}(F_m)} \qquad S_n^l(F_m) = \frac{\mathcal{T}_n(F_m) + \mathcal{V}_n(F_m)}{\mathcal{H}(F_m)}$$

Sensitivity matrix

$$\mathbf{J_a} = \begin{bmatrix} S_1^a(F_1) & \cdots & S_N^a(F_1) \\ \vdots & \ddots & \vdots \\ S_1^a(F_M) & \cdots & S_N^a(F_M) \end{bmatrix} \qquad \mathbf{J_l} = \begin{bmatrix} S_1^l(F_1) & \cdots & S_N^l(F_1) \\ \vdots & \ddots & \vdots \\ S_1^l(F_M) & \cdots & S_N^l(F_M) \end{bmatrix}$$

Area and length iterative deformations

$$\Delta \mathbf{f} = [\mathbf{J_a} | \mathbf{J_l}] \, \Delta \mathbf{p} \qquad \Delta \mathbf{f} = \left[\frac{F_1 - F_1'}{F_1'}, \ldots, \frac{F_M - F_M'}{F_M'} \right]^T$$

At each iteration $k+1$ (Elie and Laprie, EUSIPCO, 2014)

- $\mathbf{a}_{k+1} = \mathbf{a}_k + \psi_{\mathbf{a}} \mathbf{A}_k \mathbf{J_a}^T \Delta \mathbf{f}_k$
- $\mathbf{l}_{k+1} = \text{diag} \left(\frac{1}{1+\delta\lambda_1}, \frac{1}{1+\delta\lambda_2}, \ldots, \frac{1}{1+\delta\lambda_N} \right) \mathbf{l}_k, \; \delta\lambda = \psi_{\mathbf{l}} \mathbf{J_l}^T \Delta \mathbf{f}_k$
- Repeat process until $|\Delta \mathbf{f}|_1 < \epsilon$ ($\simeq 1\%$)

Biomechanical constraints

Minimizing the potential energy: avoid unrealistic configurations

$$\mathcal{C}_{\mathcal{V}} = 2\left[\mathbf{p} - \mathbf{p}_0\right] ||\mathbf{p} - \mathbf{p}_0||_2^2$$

- $\mathbf{p}_0$ = neutral position, according to the lip opening area (Maeda articulatory model[a])

[a] Maeda, 1979

Minimizing the kinetic energy: avoid unrealistic movements

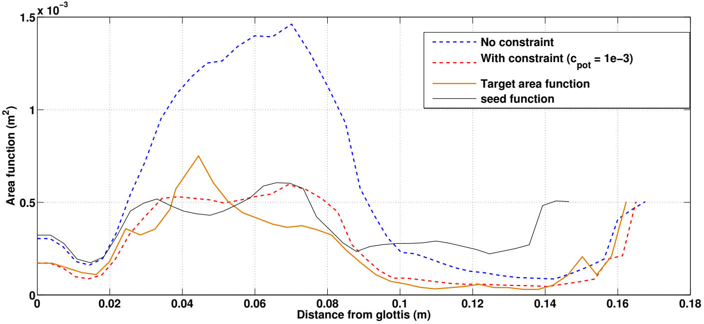
$$\mathcal{C}_{\mathcal{T}}(t) = \frac{\partial \mathcal{T}_{art}(t)}{\partial \mathbf{p}(t)} ||\Delta\mathbf{p}(t)||_2^2,$$

$$\frac{\partial \mathcal{T}_{art}}{\partial \mathbf{p}}(t) = \begin{cases} 2\Delta\mathbf{p}(t), & t = 1 \\ 2\left[\Delta\mathbf{p}(t) - \Delta\mathbf{p}(t-1)\right], & 2 \leq t \leq t_{max} - 1 \\ 2\Delta\mathbf{p}(t-1), & t = t_{max} \end{cases}$$

Results with the constrained algorithm

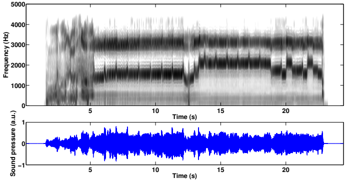
Addition of the constraints

$$\tilde{\mathbf{a}}_{k+1} = \tilde{\mathbf{a}}_k + \tilde{\mathbf{A}}_k[\underbrace{(1 - c_{kin} - c_{pot})}_{\text{Weighting coefficients}} \tilde{\mathbf{J}}_{\mathbf{a}}^T \delta\tilde{\mathbf{f}}_k + c_{kin}\tilde{\mathcal{C}}_{\mathcal{T}} + c_{pot}\tilde{\mathcal{C}}_{\mathcal{V}}]$$



Application to singing techniques

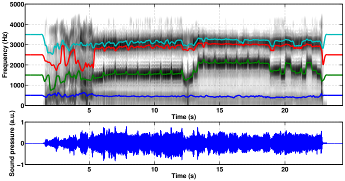
Example : Overtone singing



Source : http://www.crem-cnrs.fr/clefs-ecoute/animations/diphonique/hai1.html

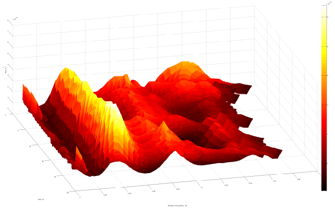
Application to singing techniques

Example : Overtone singing

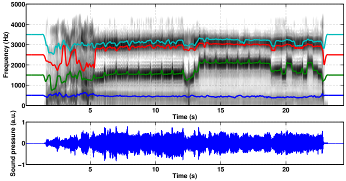


Source : `http://www.crem-cnrs.fr/clefs-ecoute/animations/diphonique/hai1.html`

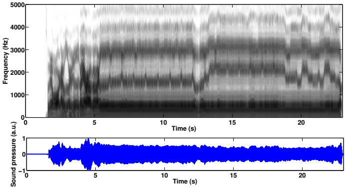
Application to singing techniques



Example : Overtone singing



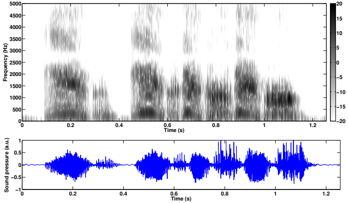
Source : `http://www.crem-cnrs.fr/clefs-ecoute/animations/diphonique/hai1.html`



Synthesis

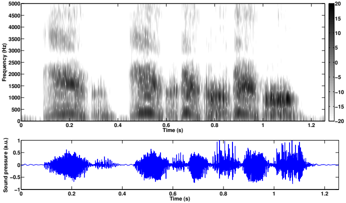
Application to bioacoustics

Example : Diane monkeys

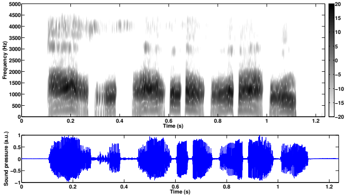


Application to bioacoustics

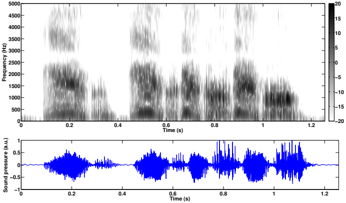
Example : Diane monkeys



Application to bioacoustics



Example : Diane monkeys



Synthesis

Plan

1. Articulatory model
2. Source model
3. Speech synthesis based on acoustical model
4. Copy and hybrid syntheses
5. Conclusions

General conclusion

Good points

- accurate copies of formant trajectories and phonetic contrasts
- access to aerodynamic quantities
- consideration of the coupling VF/VT
- integration of glottal chink
- simultaneous consideration of several side cavities
- hybrid synthesis

Limitations

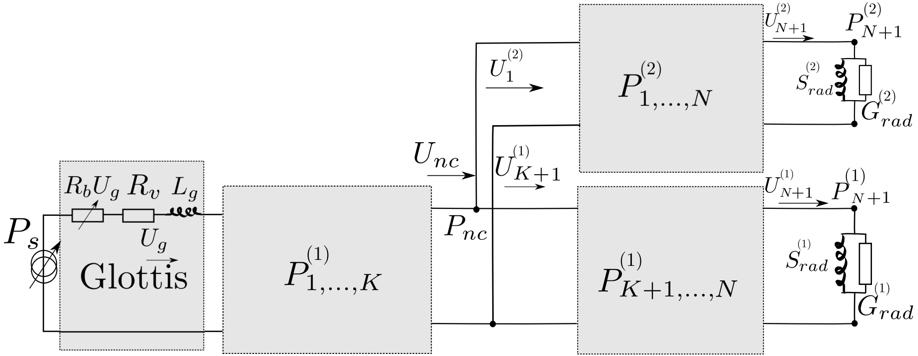
- 2D modeling of the VT
- numerous control parameters for the VF
- unrealistic spectral tilt
- stability problems
- frication noise generation

Future works

Further works

- 3D modeling of the VT (3D cineMRI)
- large articulatory database (vocal techniques, voice expressions. . .)
- glottal source parameters (in vivo acquisitions, fluid-structure interactions. . .)
- tongue, lips and velum oscillations (trill and click consonants)
- acoustic-articulatory inversion for all natural classes
- finer model for frication noise

Modélisation de la cavité nasale



$$
\begin{aligned}
F_K^{(1)} &= b_{K-1}^{(1)} U_{K-1}^{(1)} + H_K^{(1)} U_K^{(1)} + b_K^{(1)} U_{K+1}^{(1)} + b_K^{(1)} U_1^{(2)}, \\
F_{K+1}^{(1)} &= b_K^{(1)} U_K^{(1)} + H_{K+1}^{(1)} U_{K+1}^{(1)} + b_{K+1}^{(1)} U_{K+2}^{(1)} + H_C^{(1,2)} U_1^{(2)} \\
F_1^{(2)} &= H_1^{(2)} U_1^{(2)} + b_1^{(2)} U_2^{(2)} + b_K^{(1)} U_K^{(1)} + H_C^{(1,2)} U_{K+1}^{(1)},
\end{aligned}
$$

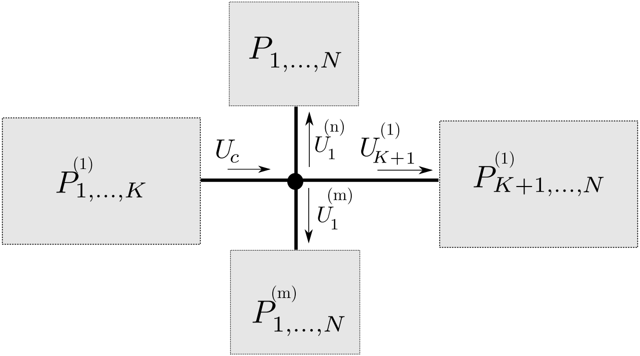
New matrix form

$$
\left[\begin{array}{c} F_1^{(1)} \\ F_2^{(1)} \\ \vdots \\ F_K^{(1)} \\ F_{K+1}^{(1)} \\ \vdots \\ F_{N+1}^{(1)} \\ \hline F_1^{(2)} \\ \vdots \\ F_{N+1}^{(2)} \end{array}\right]
=
\left[\begin{array}{ccccccc|ccc}
H_1^{(1)} & b_1^{(1)} & 0 & & & & & & & 0 \\
b_1^{(1)} & H_2^{(1)} & b_2^{(1)} & 0 & & & & & & \\
0 & \ddots & \ddots & \ddots & 0 & & & & & \\
 & 0 & b_{K-1}^{(1)} & H_K^{(1)} & b_K^{(1)} & 0 & & b_K^{(1)} & & \\
 & & 0 & b_K^{(1)} & H_{K+1}^{(1)} & b_{K+1}^{(1)} & 0 & H_C^{(1,2)} & & \\
 & & & & \ddots & \ddots & \ddots & & & \\
 & & & & & b_N^{(1)} & H_{N+1}^{(1)} & 0 & & \\
\hline
 & & & b_K^{(1)} & H_C^{(1,2)} & & 0 & H_1^{(2)} & b_1^{(2)} & \\
 & & & & & & & \ddots & \ddots & \ddots \\
0 & & & & & & & & b_N^{(2)} & H_{N+1}^{(2)}
\end{array}\right]
\cdot
\left[\begin{array}{c} U_1^{(1)} \\ U_2^{(1)} \\ \vdots \\ U_K^{(1)} \\ U_{K+1}^{(1)} \\ \vdots \\ U_{N+1}^{(1)} \\ \hline U_1^{(2)} \\ \vdots \\ U_{N+1}^{(2)} \end{array}\right]
$$

$$\downarrow$$

$$
\left[\begin{array}{c} \mathbf{f}^{(1)} \\ \hline \mathbf{f}^{(2)} \end{array}\right]
=
\left[\begin{array}{c|c} \mathbf{L}^{(1)} & \mathbf{C}_{(1,2)}^T \\ \hline \mathbf{C}_{(1,2)} & \mathbf{L}^{(2)} \end{array}\right]
\cdot
\left[\begin{array}{c} \mathbf{u}^{(1)} \\ \hline \mathbf{u}^{(2)} \end{array}\right]
$$

Particular case: 1. twin side branches (*e.g.*: piriform fossa)



$$
\begin{aligned}
F_K^{(1)} &= b_{K-1}^{(1)} U_{K-1}^{(1)} + H_K^{(1)} U_K^{(1)} + b_K^{(1)} U_{K+1}^{(1)} + b_K^{(1)} U_1^{(m)} + b_K^{(1)} U_1^{(n)} \\
F_{K+1}^{(1)} &= b_K^{(1)} U_K^{(1)} + H_{K+1}^{(1)} U_{K+1}^{(1)} + b_{K+1}^{(1)} U_{K+2}^{(1)} + H_C^{(1,m)} U_1^{(m)} + H_C^{(1,n)} U_1^{(n)} \\
F_1^{(m)} &= H_1^{(m)} U_1^{(m)} + b_1^{(m)} U_2^{(m)} + b_K^{(1)} U_K^{(1)} + H_C^{(1,m)} U_{K+1}^{(1)} + H_C^{(1,m)} U_1^{(n)} \\
F_1^{(n)} &= H_1^{(n)} U_1^{(n)} + b_1^{(n)} U_2^{(n)} + b_K^{(1)} U_K^{(1)} + H_C^{(1,n)} U_{K+1}^{(1)} + H_C^{(1,n)} U_1^{(m)}
\end{aligned}
$$

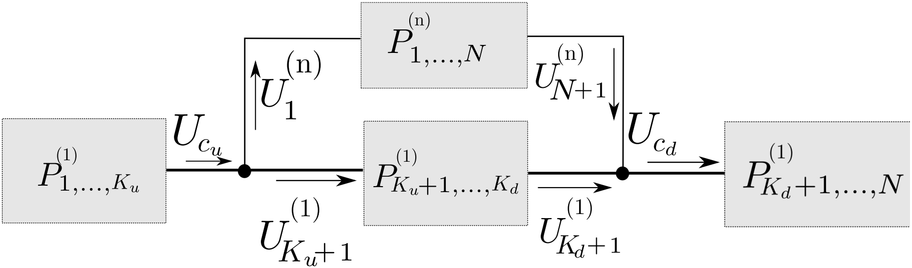
Piriform fossa : matrix form

$$
\left[\begin{array}{c}
F_1^{(1)} \\ F_2^{(1)} \\ \vdots \\ F_K^{(1)} \\ F_{K+1}^{(1)} \\ \vdots \\ F_{N+1}^{(1)} \\ \hline
F_1^{(m)} \\ \vdots \\ F_{N+1}^{(m)} \\ \hline
F_1^{(n)} \\ \vdots \\ F_{N+1}^{(n)}
\end{array}\right]
=
\left[\begin{array}{ccccccc|ccc|ccc}
H_1^{(1)} & b_1^{(1)} & 0 & & & & & & & 0 & & & 0 \\
b_1^{(1)} & H_2^{(1)} & b_2^{(1)} & 0 & & & & & & & & & \\
0 & \ddots & \ddots & \ddots & 0 & & & & & & & & \\
 & 0 & b_{K-1}^{(1)} & H_K^{(1)} & b_K^{(1)} & 0 & & b_K^{(1)} & & & b_K^{(1)} & & \\
 & & 0 & b_K^{(1)} & H_{K+1}^{(1)} & b_{K+1}^{(1)} & 0 & H_C^{(1,2)} & & & H_C^{(1,m)} & & \\
 & & & & \ddots & \ddots & \ddots & & & & & & \\
 & & & & & b_N^{(1)} & H_{N+1}^{(1)} & 0 & & & 0 & & \\ \hline
 & & & b_K^{(1)} & H_C^{(1,m)} & & 0 & H_1^{(m)} & b_1^{(m)} & & H_C^{(1,m)} & & \\
 & & & & & & & \ddots & \ddots & \ddots & & & \\
 & & & & & & & 0 & b_N^{(m)} & H_{N+1}^{(m)} & 0 & & \\ \hline
 & & & b_K^{(1)} & H_C^{(1,n)} & & 0 & H_C^{(1,n)} & & 0 & H_1^{(n)} & b_1^{(n)} & \\
 & & & & & & & & & & \ddots & \ddots & \ddots \\
0 & & & & & & & & & & 0 & b_N^{(m)} & H_{N+1}^{(m)}
\end{array}\right]
\cdot
\left[\begin{array}{c}
U_1^{(1)} \\ U_2^{(1)} \\ \vdots \\ U_K^{(1)} \\ U_{K+1}^{(1)} \\ \vdots \\ U_{N+1}^{(1)} \\ \hline
U_1^{(m)} \\ \vdots \\ U_{N+1}^{(m)} \\ \hline
U_1^{(n)} \\ \vdots \\ U_{N+1}^{(n)}
\end{array}\right]
$$

$$\downarrow$$

$$
\begin{bmatrix} \mathbf{f}^{(1)} \\ \mathbf{f}^{(m)} \\ \mathbf{f}^{(n)} \end{bmatrix}
=
\begin{bmatrix}
\mathbf{L}^{(1)} & \mathbf{C}_{(1,m)}^{T} & \mathbf{C}_{(1,n)}^{T} \\
\mathbf{C}_{(1,m)} & \mathbf{L}^{(m)} & \mathbf{C}_{(m,n)}^{T} \\
\mathbf{C}_{(1,n)} & \mathbf{C}_{(m,n)} & \mathbf{L}^{(n)}
\end{bmatrix}
\cdot
\begin{bmatrix} \mathbf{u}^{(1)} \\ \mathbf{u}^{(m)} \\ \mathbf{u}^{(n)} \end{bmatrix}
$$

Particular case: 2. Anabranch (*e.g.*: lateral consonants)



$$
\begin{aligned}
F^{(1)}_{K+1} &= b^{(1)}_K U^{(1)}_K + H^{(1)}_{K+1} U^{(1)}_{K+1} + b^{(1)}_{K+1} U^{(1)}_{K+2} + H^{(1,n)}_{C+1} U^{(n)}_{N+1} \\
F^{(1)}_{K+2} &= b^{(1)}_{K+1} U^{(1)}_{K+1} + H^{(1)}_{K+2} U^{(1)}_{K+2} + b^{(1)}_{K+2} U^{(1)}_{K+3} + b^{(1)}_{K+1} U^{(n)}_{N+1} \\
F^{(n)}_{N+1} &= b^{(n)}_N U_N(n) + H^{(n)}_{N+1} U^{(n)}_{N+1} + b^{(1)}_{K+1} U^{(1)}_{K+2} + H^{(1,n)}_{C+1} U^{(1)}_{K+1},
\end{aligned}
$$

Lateral consonants : matrix form

$$
\left[\begin{array}{c}
F_1^{(1)} \\ F_2^{(1)} \\ \vdots \\ F_{K_u}^{(1)} \\ F_{K_u+1}^{(1)} \\ \vdots \\ F_{K_d+1}^{(1)} \\ F_{K_d+2}^{(1)} \\ \vdots \\ F_{N+1}^{(1)} \\ \hline F_1^{(n)} \\ \vdots \\ F_{N+1}^{(n)}
\end{array}\right]
=
\left[\begin{array}{cccccccccc|ccc}
H_1^{(1)} & b_1^{(1)} & 0 & & & & & & & 0 & & & 0 \\
b_1^{(1)} & H_2^{(1)} & b_2^{(1)} & 0 & & & & & & & & & \\
0 & \ddots & \ddots & \ddots & 0 & & & & & & & & \\
 & 0 & b_{K_u-1}^{(1)} & H_{K_u}^{(1)} & b_{K_u}^{(1)} & 0 & & & & & b_{K_u}^{(1)} & & \\
 & & 0 & b_{K_u}^{(1)} & H_{K_u+1}^{(1)} & b_{K_u+1}^{(1)} & 0 & & & & H_{C_u}^{(1,m)} & & \\
 & & & & \ddots & \ddots & \ddots & & & & & & \\
 & & & & 0 & b_{K_d}^{(1)} & H_{K_d+1}^{(1)} & b_{K_d+1}^{(1)} & 0 & & & & \\
 & & & & & 0 & b_{K_d+1}^{(1)} & H_{K_d+2}^{(1)} & b_{K_d+2}^{(1)} & 0 & & & H_{C_d+1}^{(1,n)} \\
 & & & & & & & \ddots & \ddots & \ddots & & & b_{K_d+1}^{(1)} \\
0 & & & & & & & & b_N^{(1)} & H_{N+1}^{(1)} & 0 & & \\
\hline
 & & & b_{K_u}^{(1)} & H_{C_u}^{(1,n)} & & & & & 0 & H_1^{(n)} & b_1^{(n)} & \\
 & & & & & & & & & & \ddots & \ddots & \ddots \\
0 & & & & & & & H_{C_d+1}^{(1,n)} & b_{K_d+1}^{(1)} & & 0 & b_N^{(m)} & H_{N+1}^{(m)}
\end{array}\right]
\cdot
\left[\begin{array}{c}
U_1^{(1)} \\ U_2^{(1)} \\ \vdots \\ U_{K_u}^{(1)} \\ U_{K_u+1}^{(1)} \\ \vdots \\ U_{K_d+1}^{(1)} \\ U_{K_d+2}^{(1)} \\ \vdots \\ U_{N+1}^{(1)} \\ \hline U_1^{(n)} \\ \vdots \\ U_{N+1}^{(n)}
\end{array}\right]
$$

$$\downarrow$$

$$
\left[\begin{array}{c} \mathbf{f}^{(1)} \\ \mathbf{f}^{(n)} \end{array}\right]
=
\left[\begin{array}{cc} \mathbf{L}^{(1)} & \mathbf{C}_{(1,n)}^{T} \\ \mathbf{C}_{(1,n)} & \mathbf{L}^{(n)} \end{array}\right]
\cdot
\left[\begin{array}{c} \mathbf{u}^{(1)} \\ \mathbf{u}^{(n)} \end{array}\right]
$$

Existing methods

Machine learning

- Neural network
- Hidden Markov chains
- Gaussian mixture

Time-consuming, require large database, adapted for only one speaker

Analysis-by-synthesis methods

- Codebook search
- Iterative methods

Require appropriate model of vocal tract

Non-exhaustive review of iterative methods using sensitivity functions

Authors	Yu (1993)	Carré (2004)	Bunton and Story (2013)
Ac. vector	Formants	Formants	Formants
Input vector	Fourier coeff.	**a**	**a**
VT length	Arbitrary	Arbitrary	Arbitrary
Regul.	None	None	Lip aperture

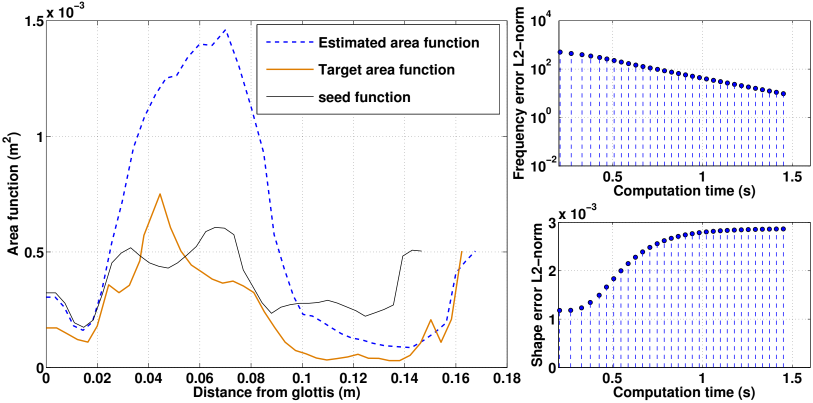
Non-exhaustive review of iterative methods using sensitivity functions

Authors	Yu (1993)	Carré (2004)	Bunton and Story (2013)	Elie and Laprie (2014)
Ac. vector	Formants	Formants	Formants	Formants
Input vector	Fourier coeff.	**a**	**a**	**a** and **l**
VT length	Arbitrary	Arbitrary	Arbitrary	Estimated
Regul.	None	None	Lip aperture	Lip aperture Shape, traj.

Propositions:

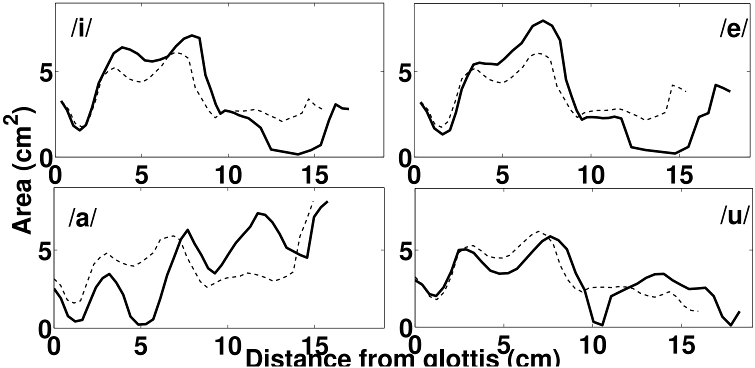
- estimate both area (**a**) and length (**l**) functions
- add biomechanical constraints for better regularization

Example of ill behavior of the unconstrained algorithm



- Front cavity is well-estimated, but back cavity is unrealistic
- The formant frequency difference is converging but the shape difference is diverging

Results for a french native speaker /i/, /e/, /a/, /u/

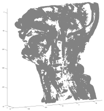


Plan

Des coupes IRM vers le volume

Exemple d'un /u/

→



Ligne médiane

Référence: coupe médio-sagittale

Ligne médiane

Référence: coupe médio-sagittale

→ On ajoute la ligne médiane

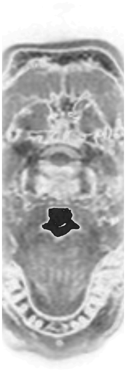
Plans tangents

On récupère les projections du volume sur les plans tangents

Exploration du conduit vocal:

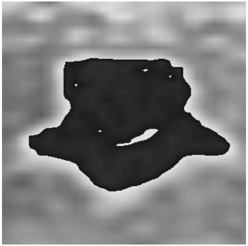
Récupération des contours des tranches:

Délimitation de la surface



Délimitation de la surface

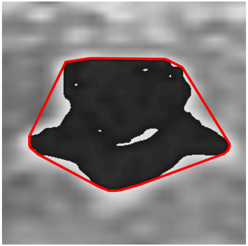
Zoom



Délimitation de la surface

Calcul de l'enveloppe convexe:

Plus petit ensemble convexe autour du conduit vocal



Les contours du CV pas forcément convexes

→ Besoin de déformer l'enveloppe

Délimitation de la surface

Déformation de l'enveloppe convexe:

Substitution de chaque point de l'enveloppe par le point du CV le plus proche

Contours bien définis par un polygone

On récupère alors l'aire du polygone

Finalement...