

Friday 8 December 2006

Three-dimensional modelling of the speech organs from MRI images for nasals production

Articulatory-acoustic characterization of the velum movements

Antoine Serrurier



M. Gang Feng,
M. Philip Hoole,
M. Yves Laprie,
M. Pierre Badin,
M. Shinji Maeda,
M. Gérard Bailly,

Président
Rapporteur
Rapporteur
Directeur de thèse
Examinateur
Examinateur

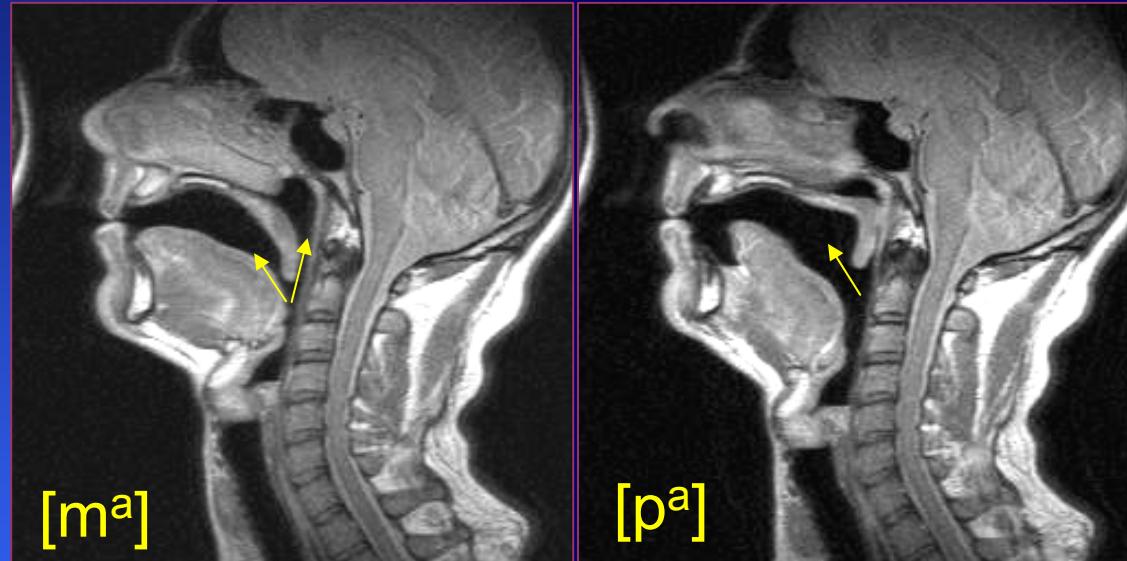


Plan

- Nasality: definition, objectives, literature
- Articulatory data
- Three-dimensional articulatory model
- Articulatory-acoustic modelling
- Conclusion and perspectives

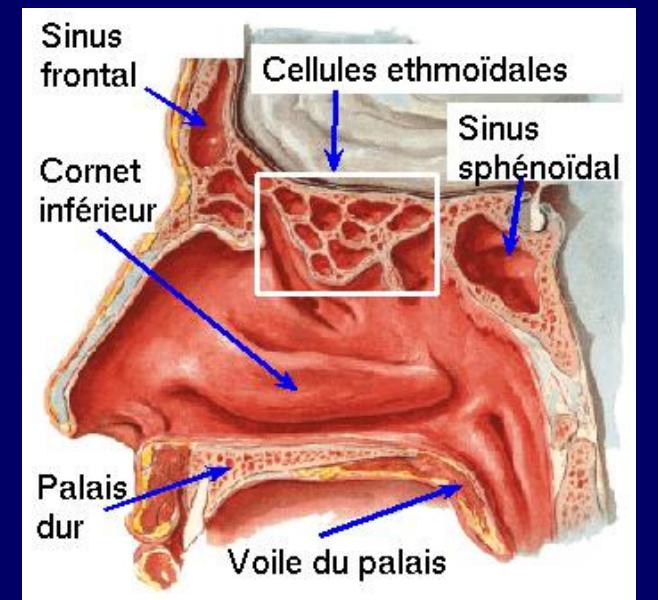
Nasality : definition

- « The term of nasality refers to sounds produced with an enough low velum which allows an audible air flow through the nasal tract. » (Crystal, 1997)



Complex structure of the nasal cavities and sinuses

Opening / Closing of velopharyngeal port



Objective : understand nasality

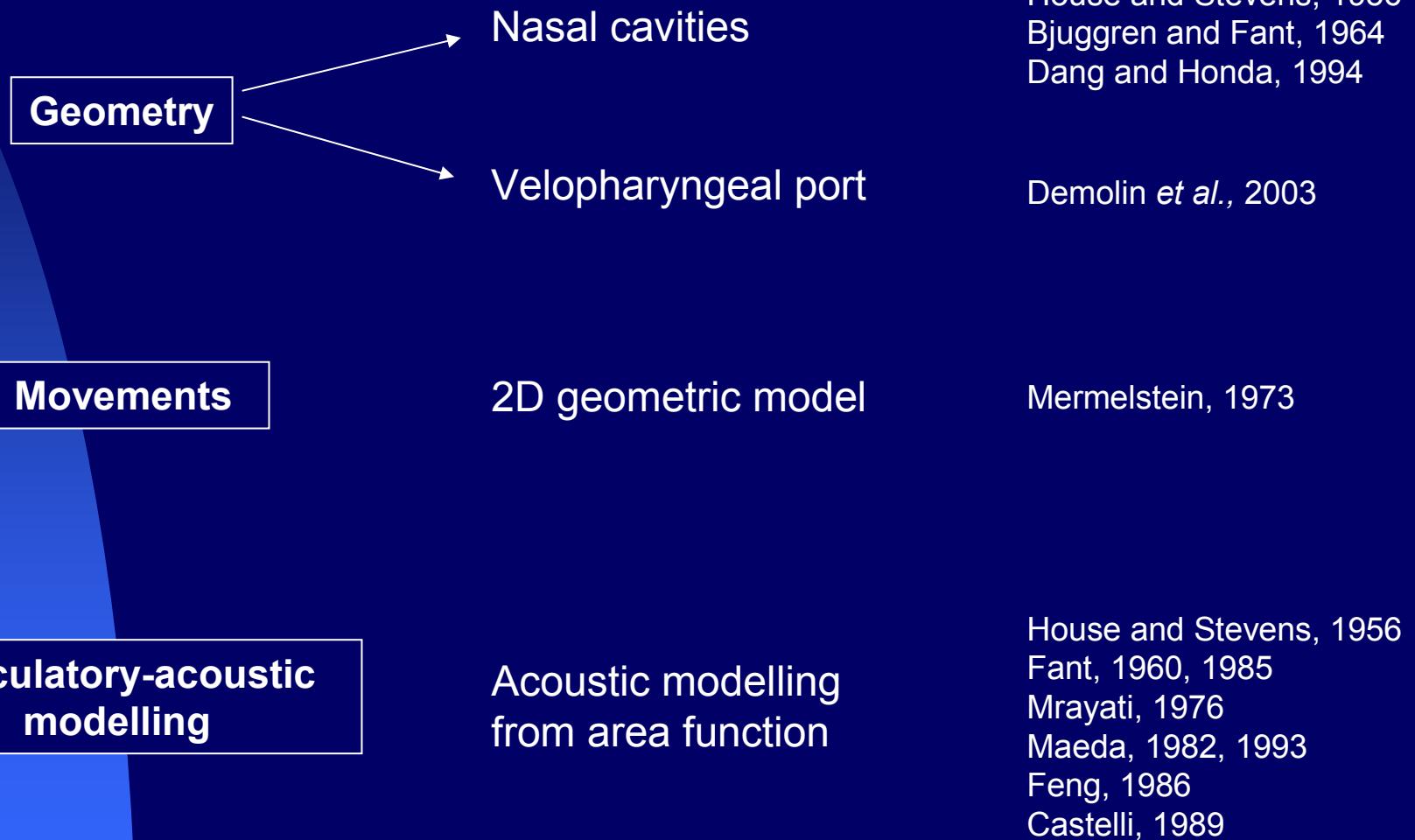
Characterization and modelling of the articulatory gesture corresponding to lowering of the velum

- Global complex Geometry
- Nature and dimensionality of the movements

Characterization and modelling of the acoustic consequences

Articulatory Synthesis

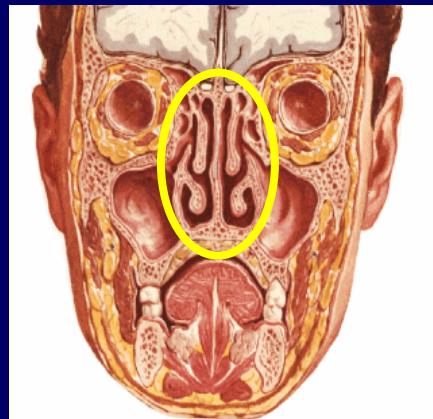
Articulatory-acoustic studies : literature



Articulatory model

- 3D model : accurate description of the complex geometry

Nasal cavities



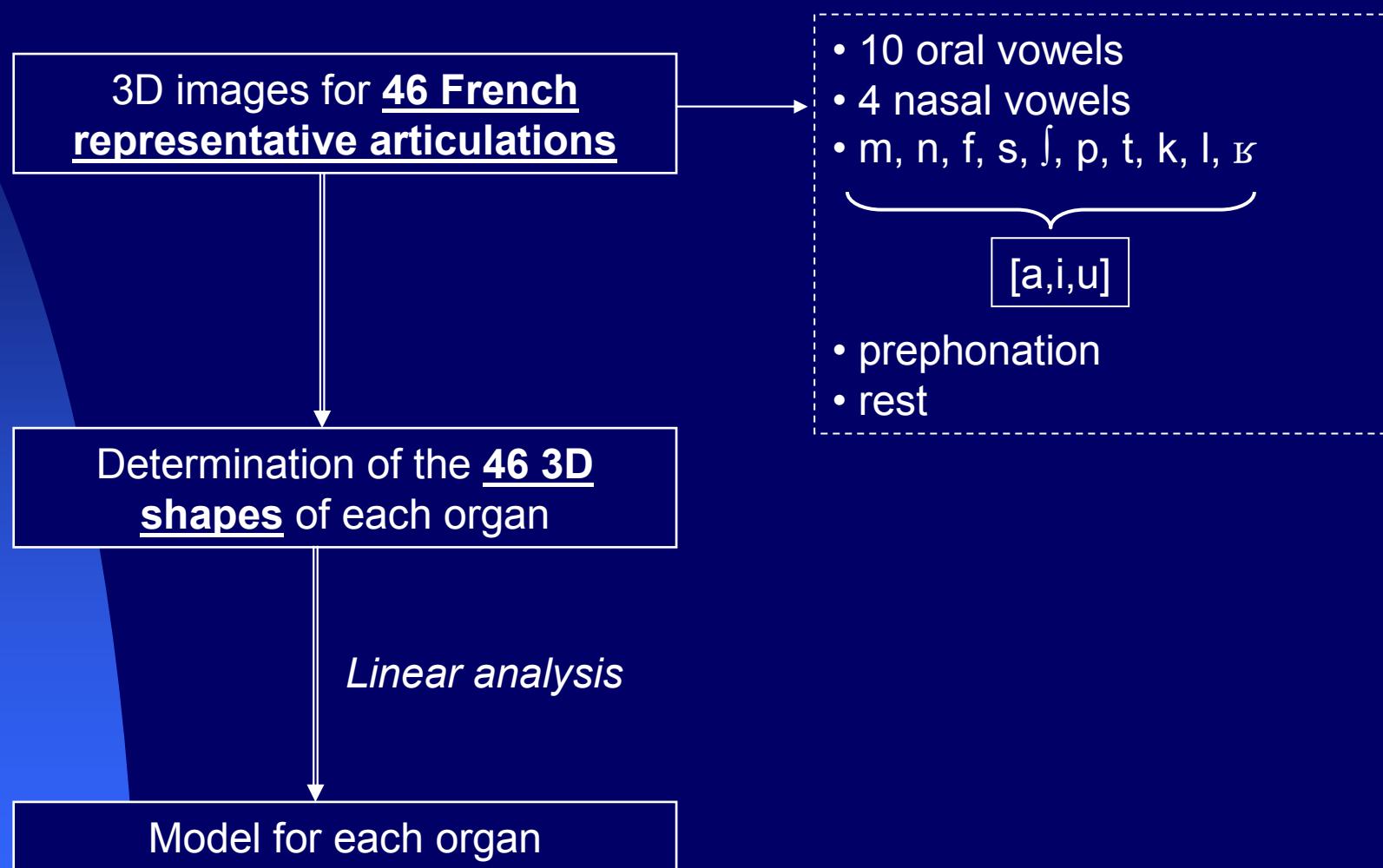
Velum



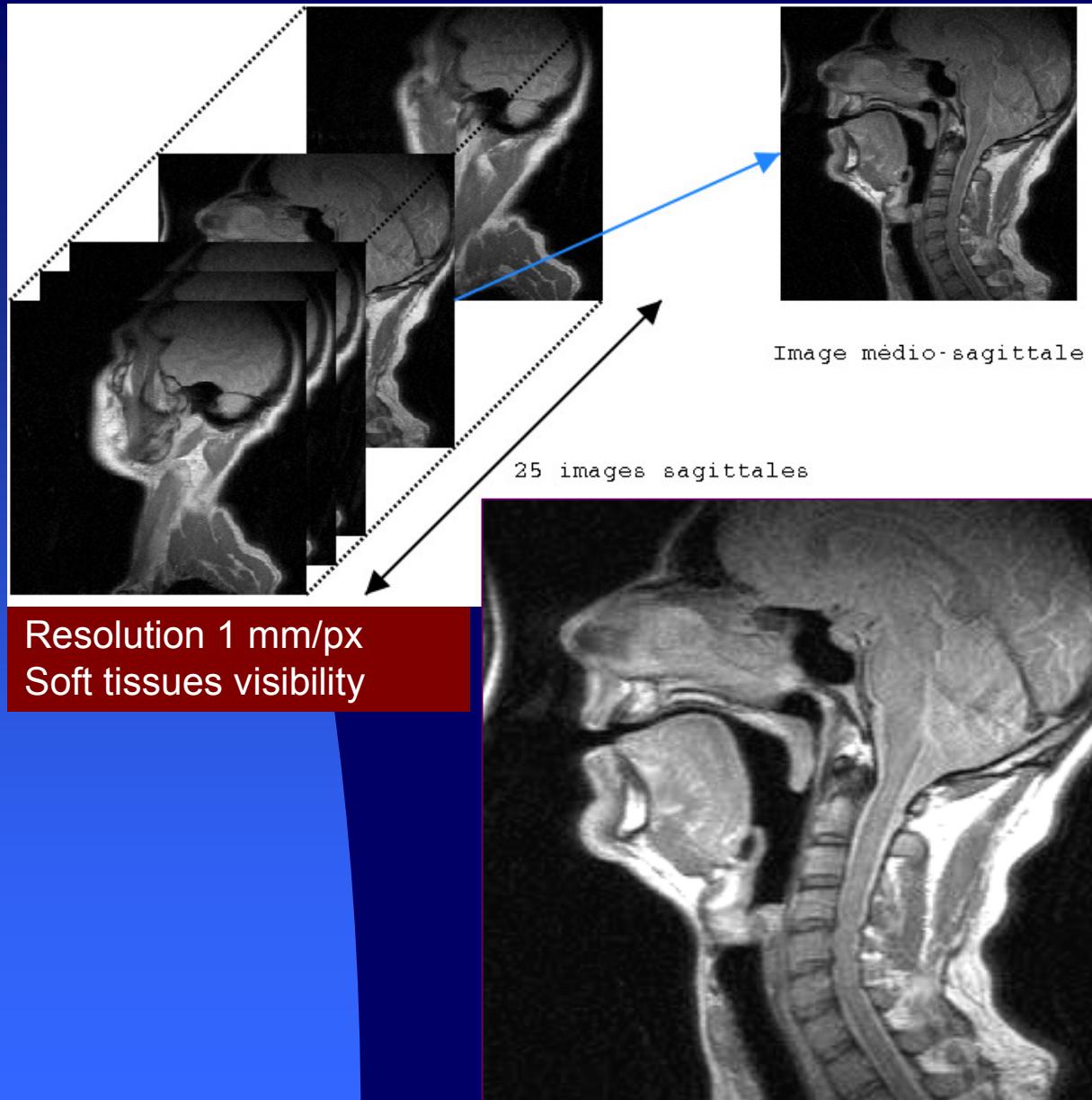
- Data based model
- Organ shape model
 - **Velum**
 - **Nasopharyngeal walls**
 - **Nasal cavities**
- Single subject



Modelling process



MRI images

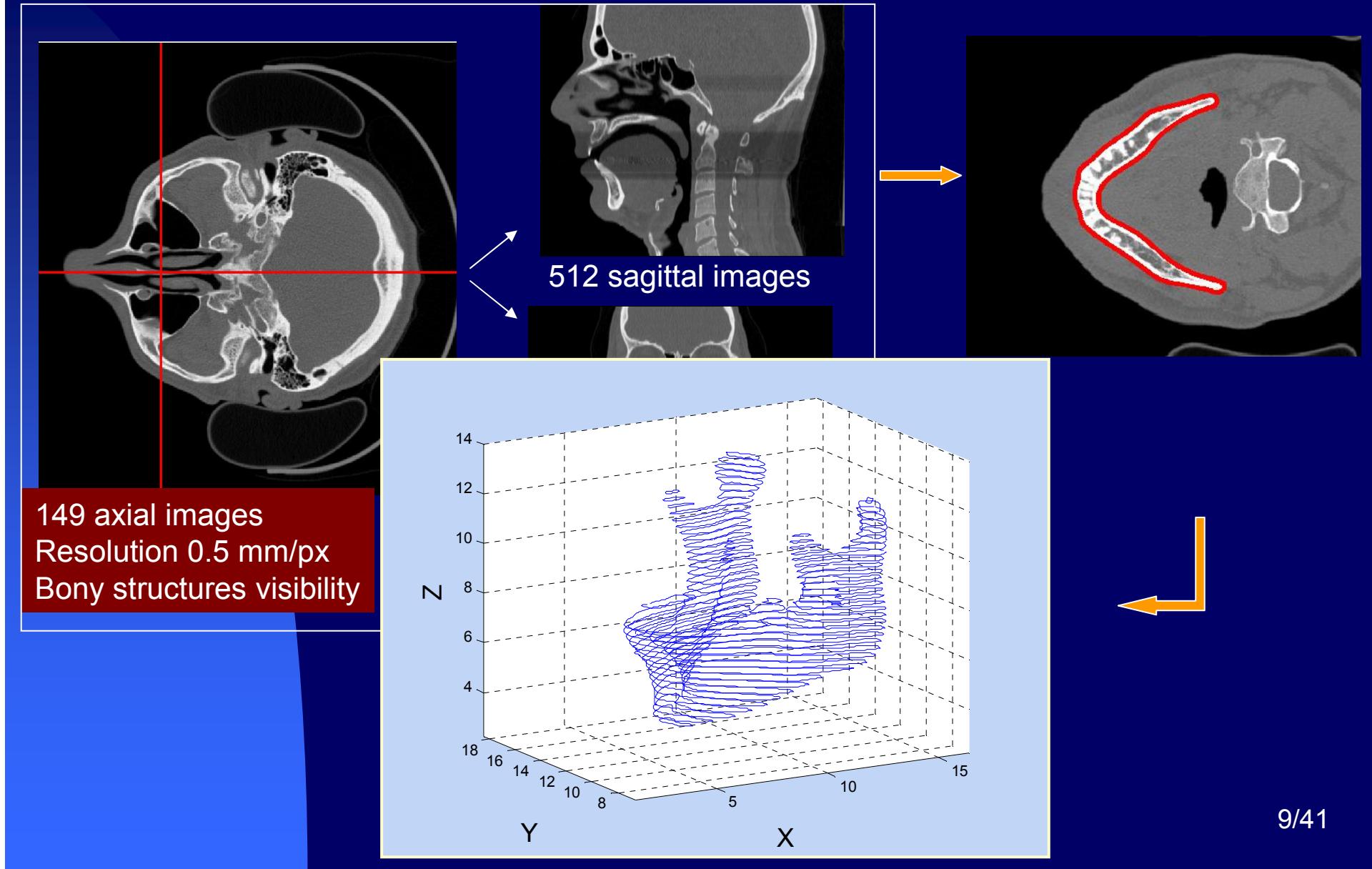


46 sets of sagittal 3D images to align

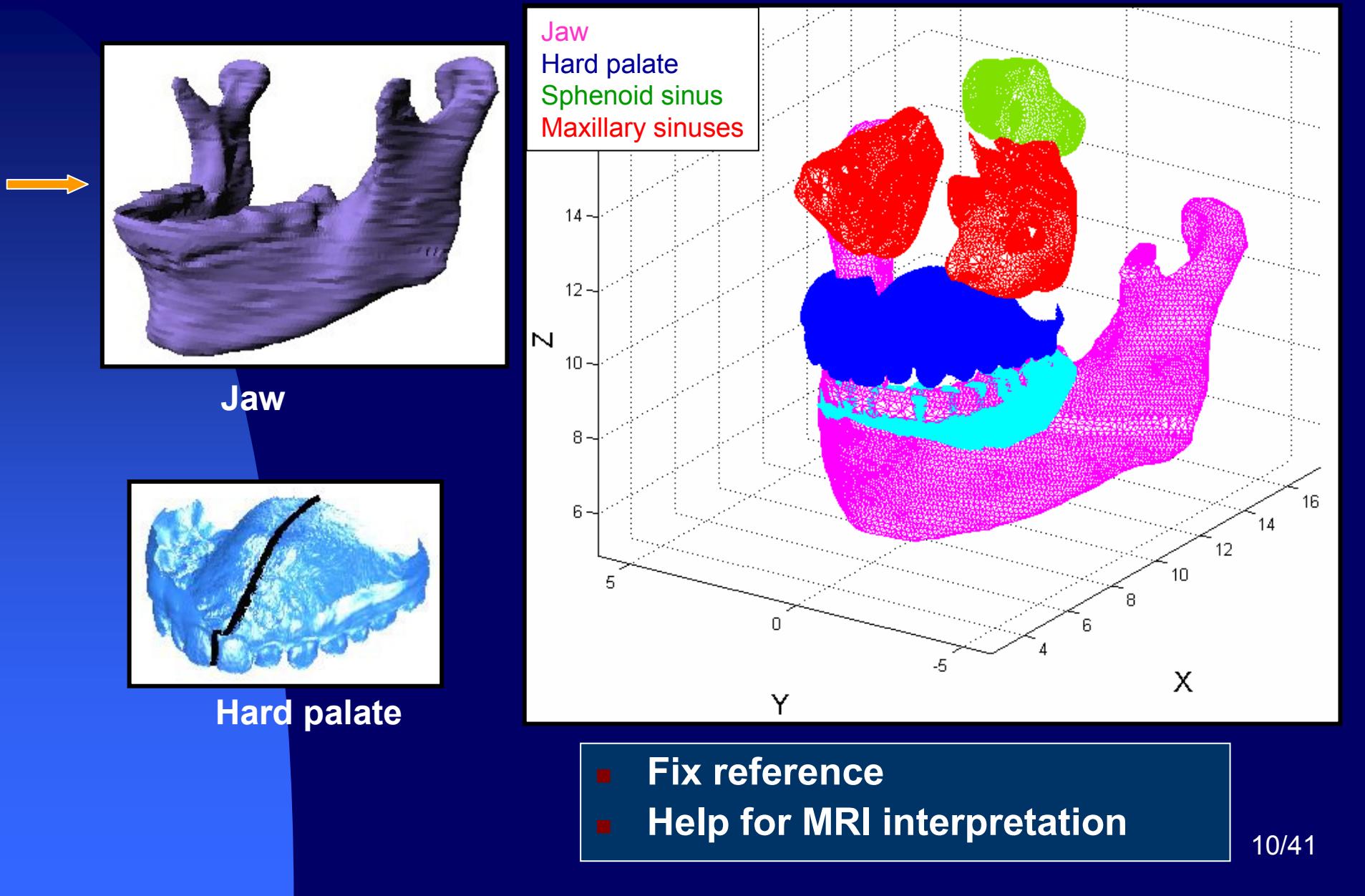
Need of 3D reconstruction of the rigid structures

No bony structure visibility

Tomodensitometric images

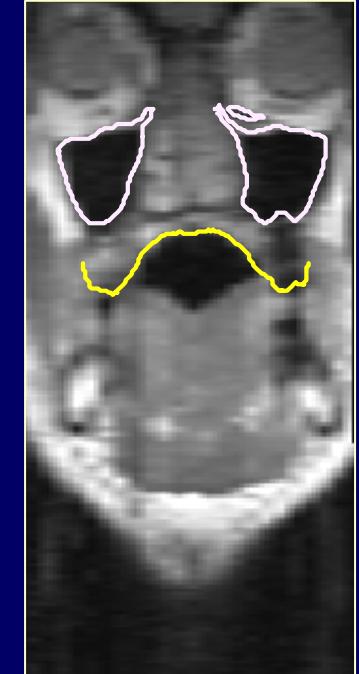
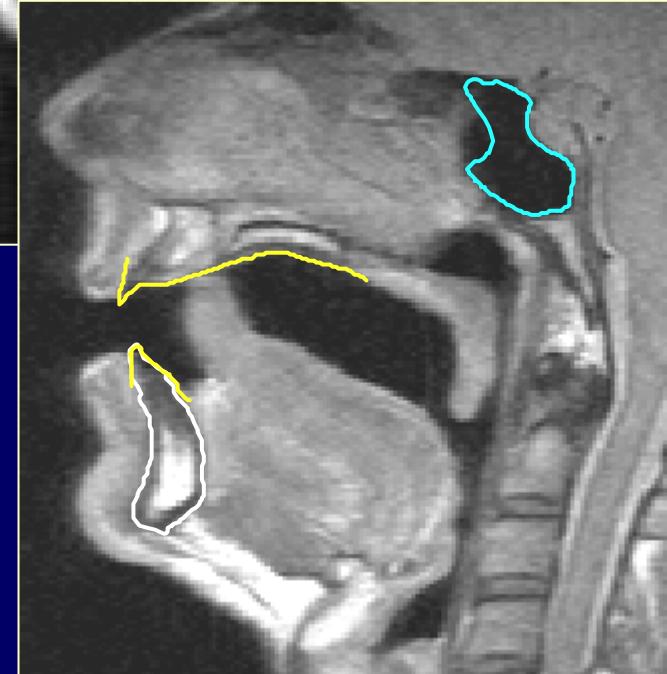
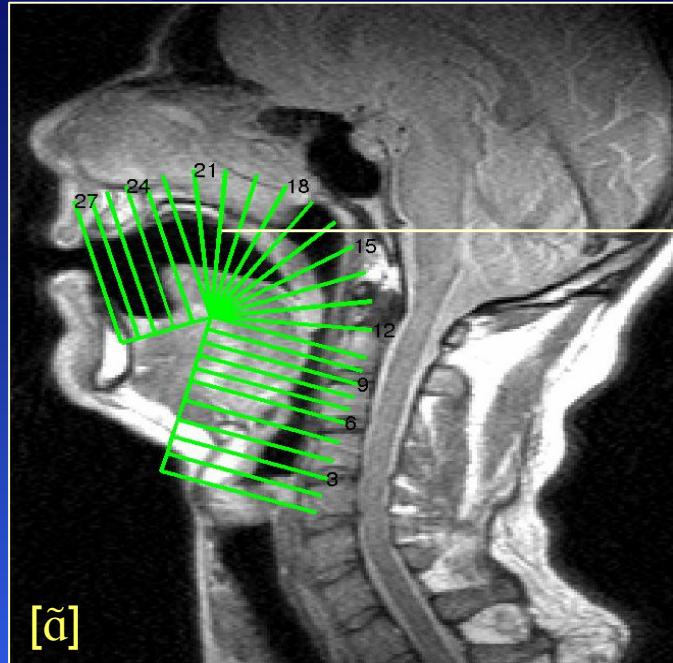


Rigid structures



Pre-processing of MRI images

1. Perpendicular re-slicing of images



2. Superimposition of the rigid structures on the MRI

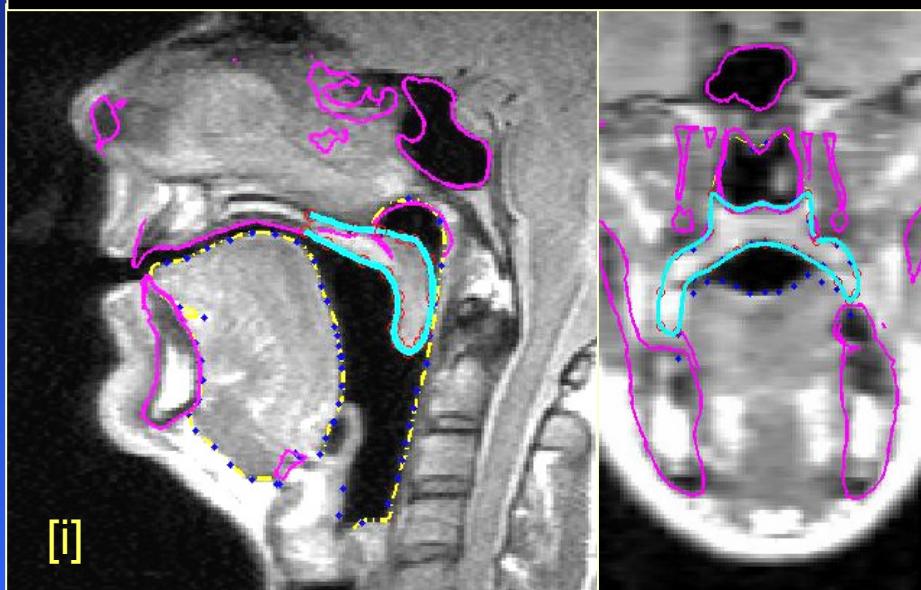
- Alignment in a common reference coordinate system
- More accurate MRI interpretation

Soft structures (1)

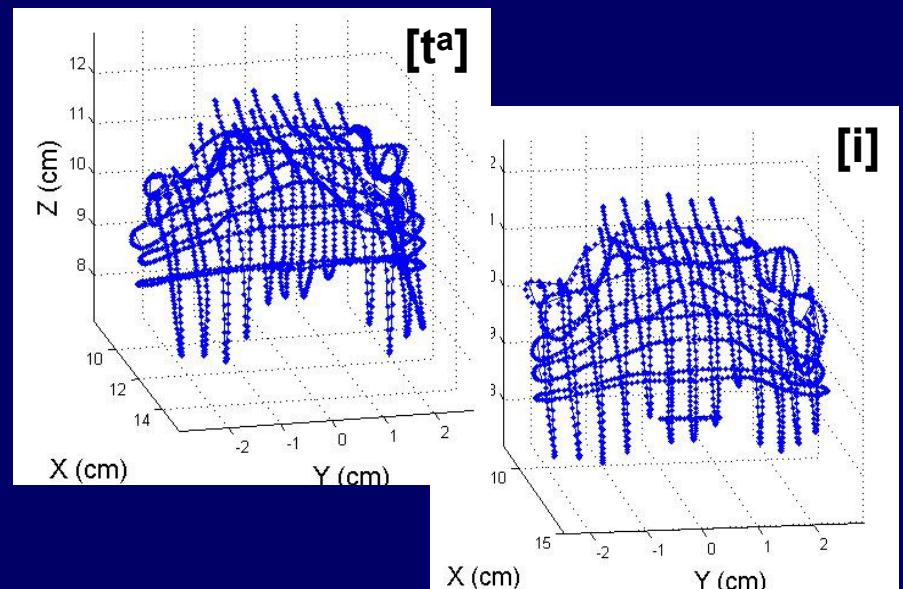
3. Manual outlines :

- velum
- pharyngeal wall

Superimposed rigid structures
Manual edited soft structures

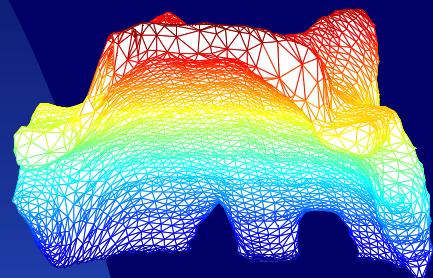


4. 46 3D shapes of velum and nasopharyngeal wall

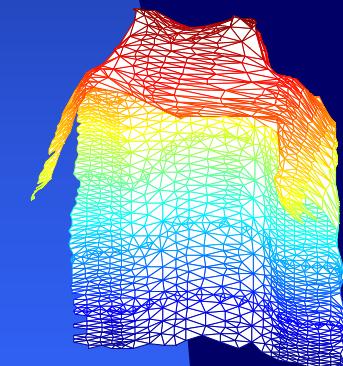


Soft structures (2)

4. Generic mesh definition

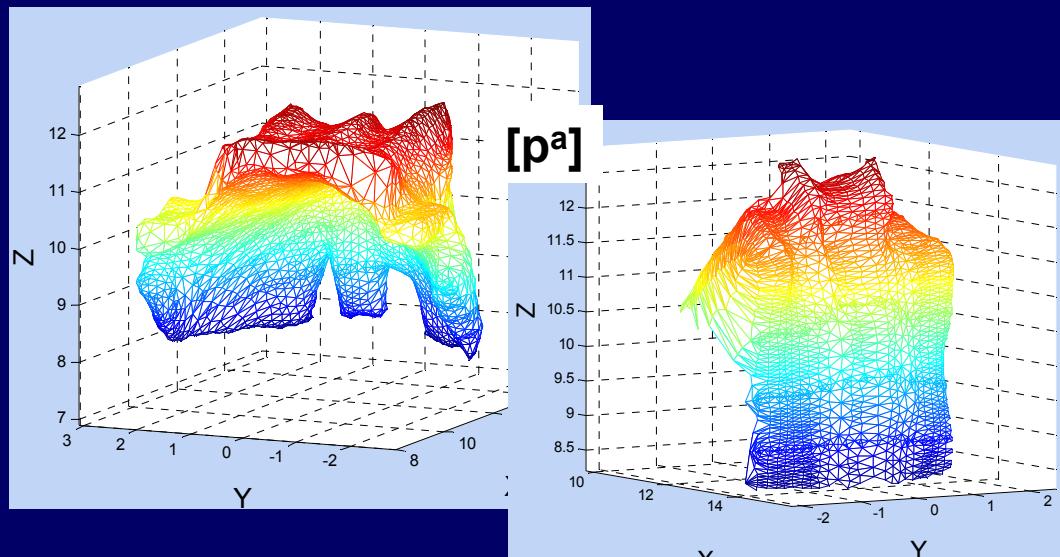


Velum: 5239 vertices



Nasopharyngeal wall: 2110 vertices

5. Fitting of the generic mesh towards the 46 articulatory targets

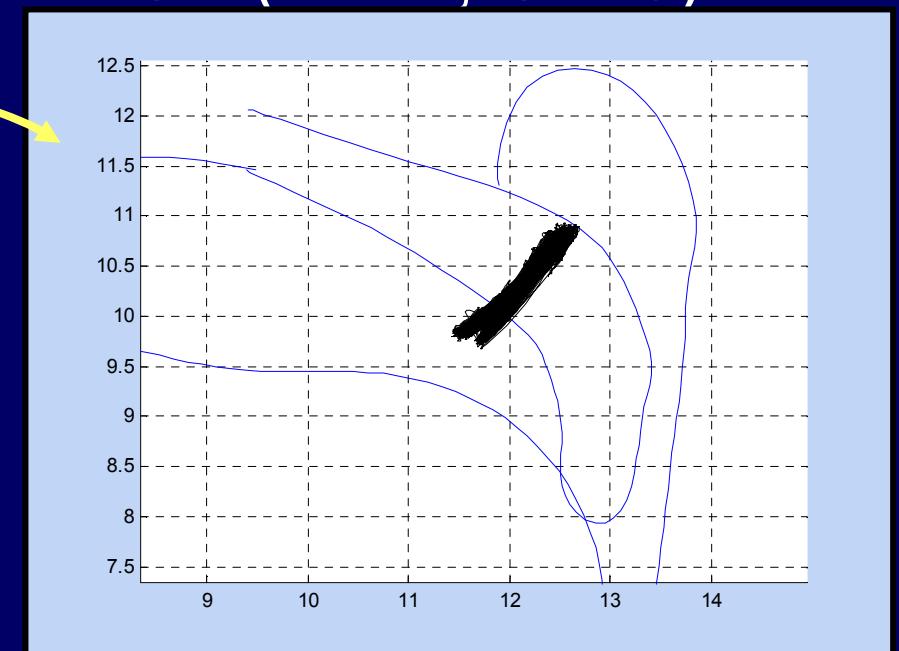
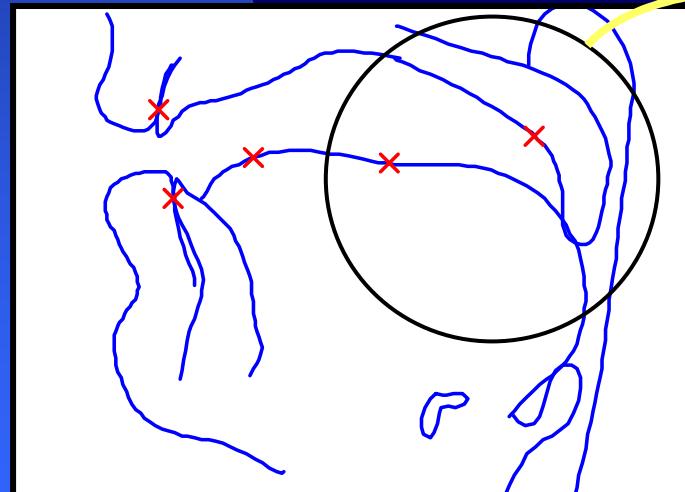


× 46 articulations of the corpus

Coherence between this process and the real deformations?

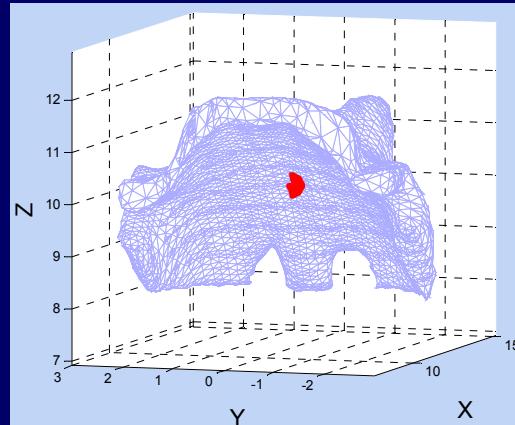
EMA recording

- *Electromagnetic Articulography* : recording of a flesh point position in the midsagittal plan
- **Dynamic speech corpus** : VCV ($V=14$, $C=16$)

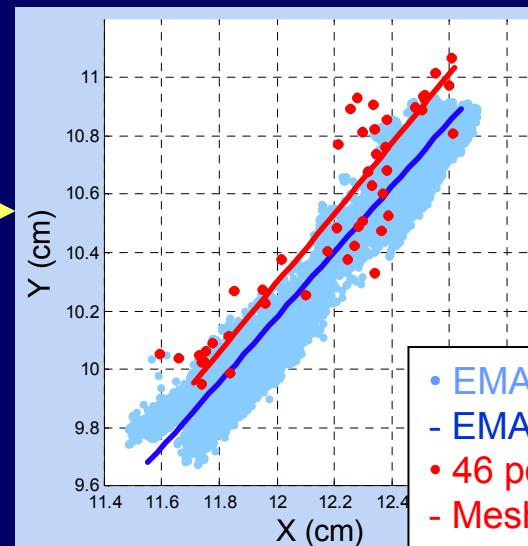
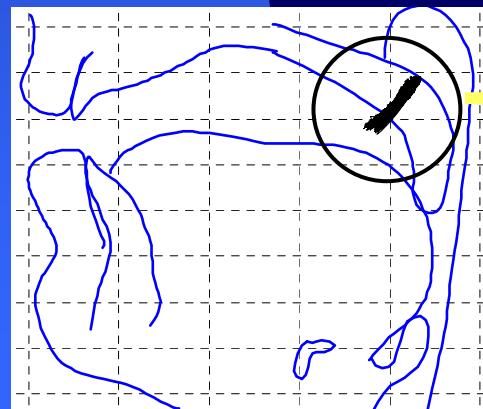


Coherence meshes - EMA

- Determination of a pseudo-EMA vertex on the generic mesh



- Position of this point through the mesh deformation process



⇒ Validation of the approach

- EMA recording
- EMA PCA
- 46 points of the fitted mesh
- Mesh point PCA

Nasal cavities (1)

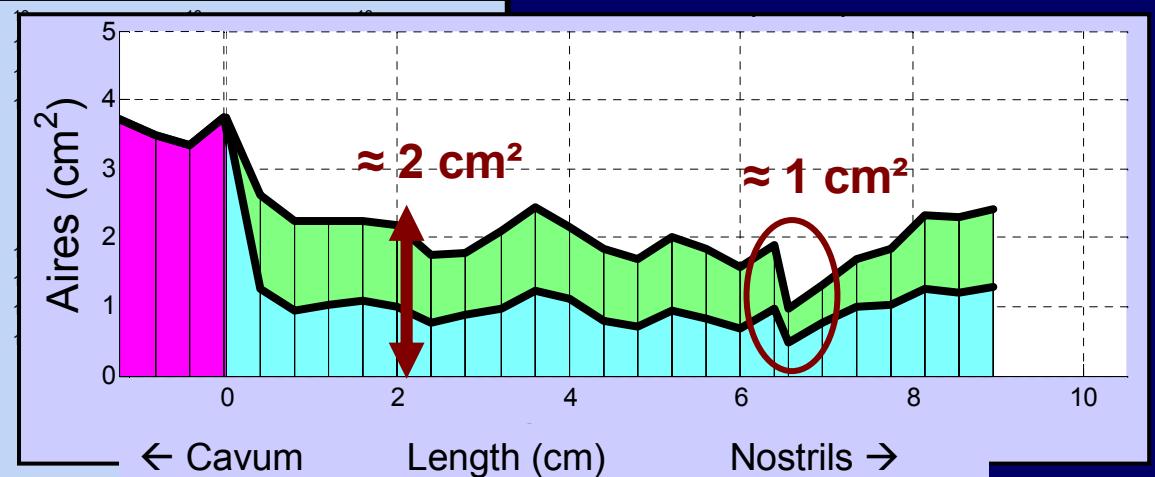
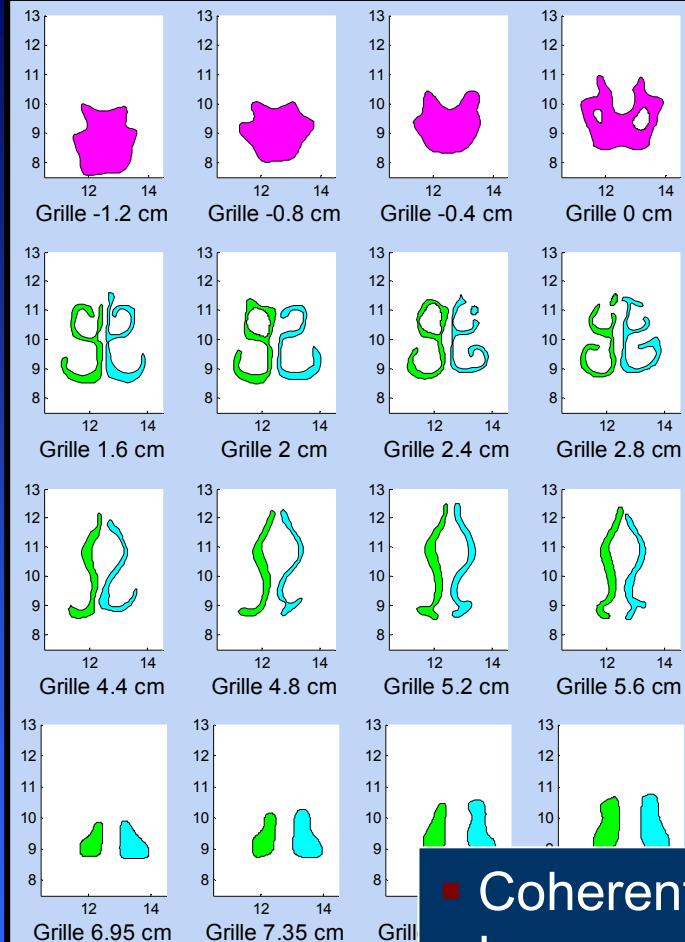


Coronal MRI
Resolution 1 mm/px
Mucosa visibility



Nasal cavities (2)

Middle
Right
Left



- Coherent with Dang et al., 1994 values
- Lower areas than Bjuggren and Fant, 1964
- Higher nostril constriction than House and Stevens, 1956

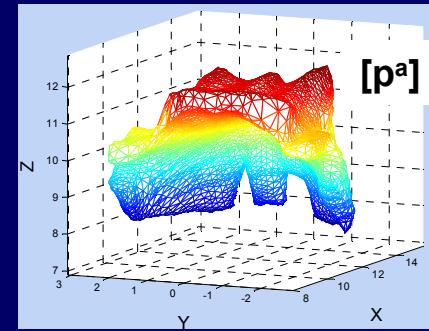
Articulatory data : conclusion

- Alignment of different nature data
- 3D meshes of the rigid structures
- 3D meshes of the soft tissues fitted to the 46 articulatory targets
 - Velum and nasopharyngeal wall
 - Accuracy < 1 mm
 - Coherent for a statistical analysis
- Coherence between 3D static and dynamic EMA data
- Nostril constriction $\approx 1 \text{ cm}^2$

Linear modelling

46 speech task representative observations:

- Velum: 5239×3 variables
- Nasopharyngeal wall : 2110×3 variables



Principal Component Analysis

Principal Components \Leftrightarrow Control parameters of the model

Evaluation

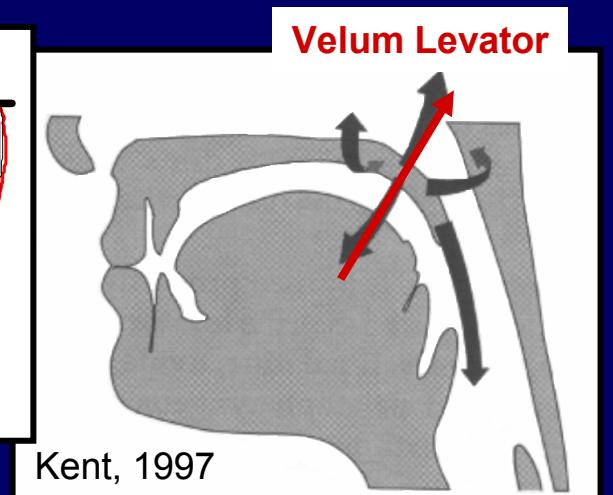
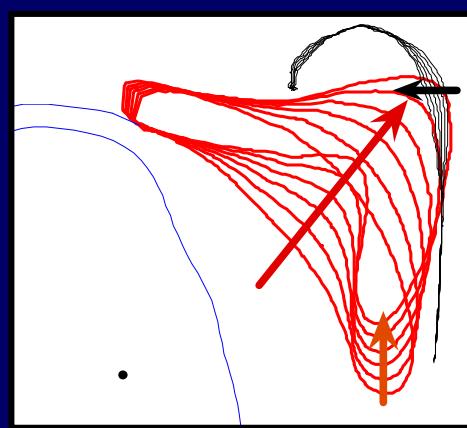
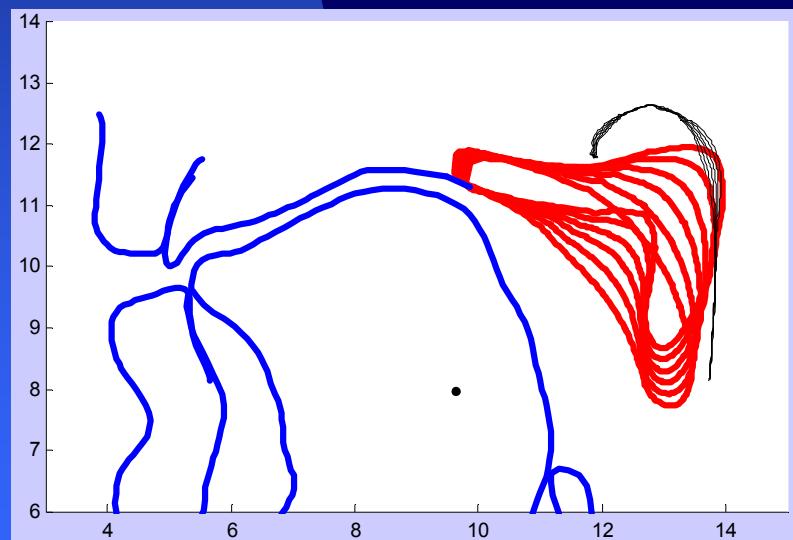
- Data variance explanation (%) for each component
- RMS reconstruction error (cm) by using the model

1st parameter : VL

PCA on the 3D velum

VL: 1 st PCA parameter	Var. Ex. (%) (on ± 1.5 cm)	Cum. RMS recons. error (cm)
Velum	83 %	0.08 cm
Pharyngeal wall	47 %	0.07 cm

→ 1 dominant parameter



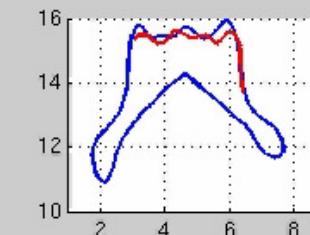
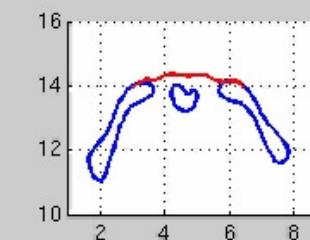
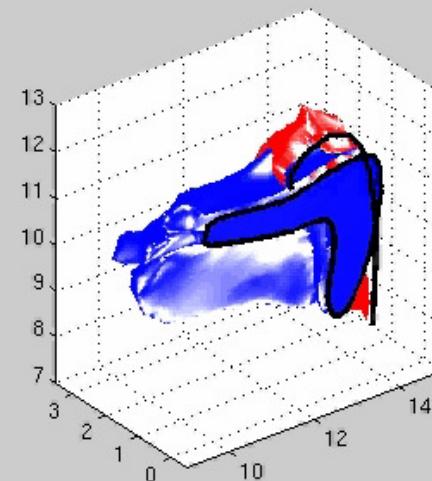
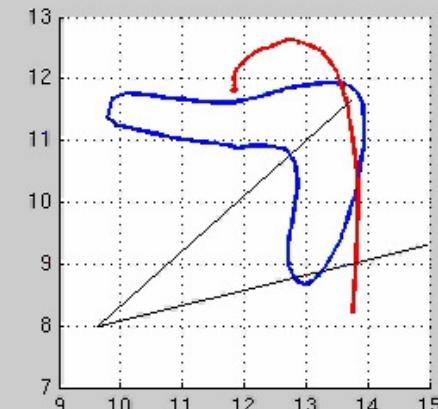
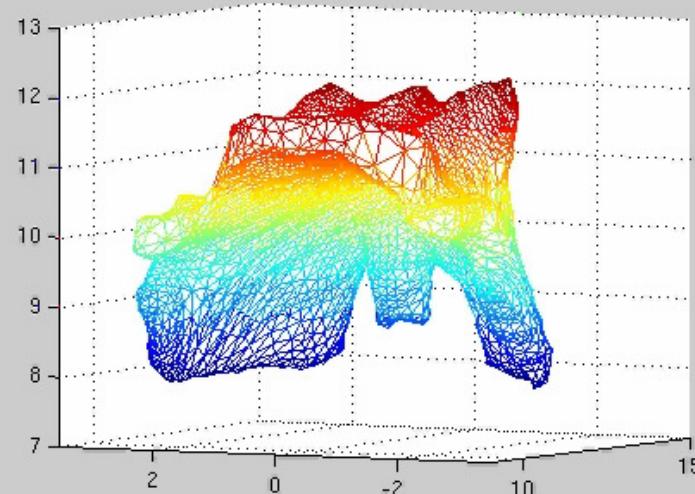
VL = *Velum Levator* and *Passavant's Pad*

Velum

Nasopharyngeal wall

Midsagittal plan

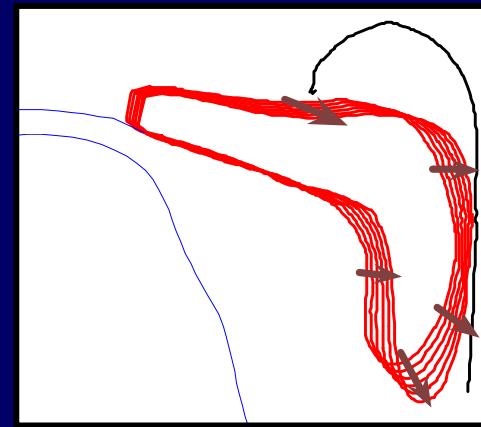
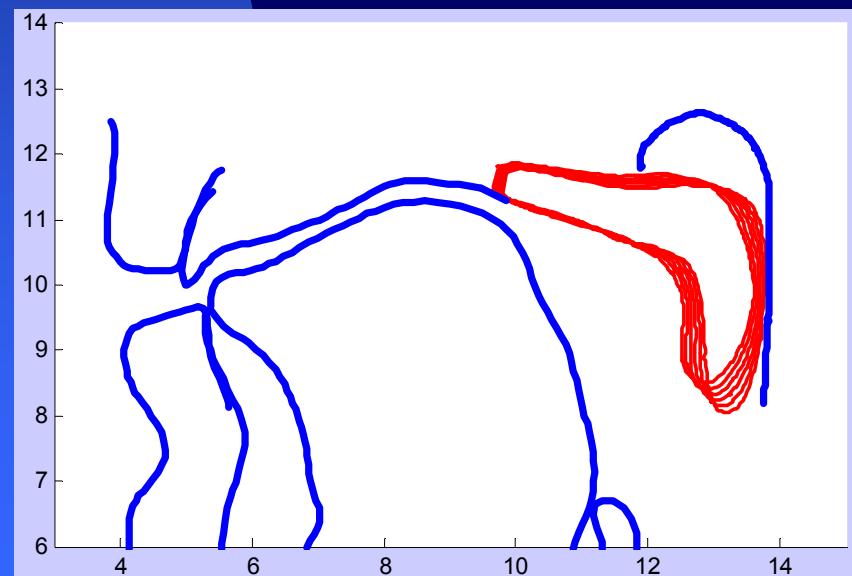
1st parameter : VL



2nd parameter : VS

PCA on the 3D velum

VS: 2 nd PCA parameter	Var. Ex. (%) (on ± 1.5 cm)	Cum. RMS recons. error (cm)
Velum	6 % (cum. 89 %)	0.06 cm
Pharyngeal wall	5 % (cum. 52 %)	0.06 cm



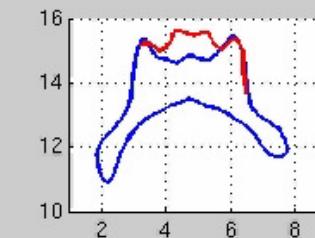
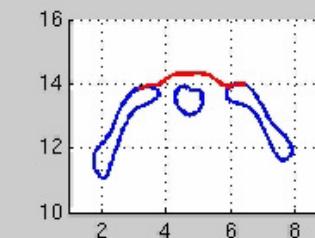
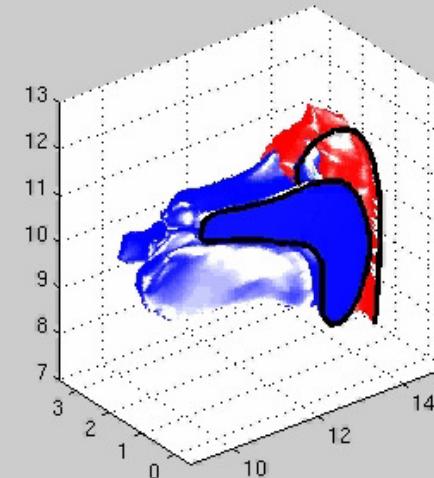
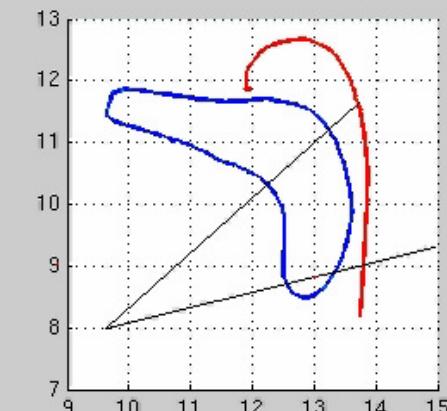
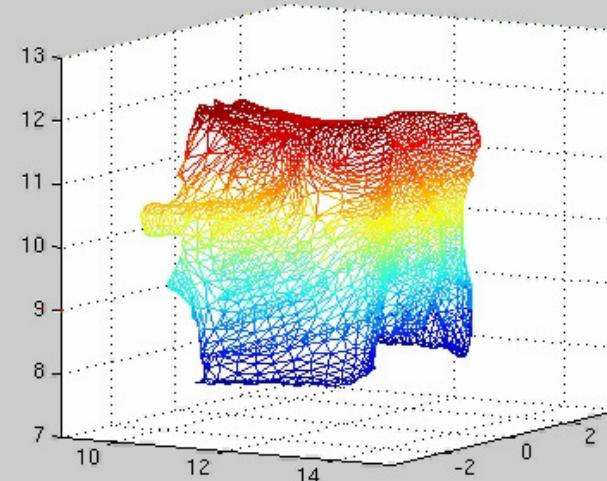
- Small amplitude
- Influential on nasal coupling area

Velum

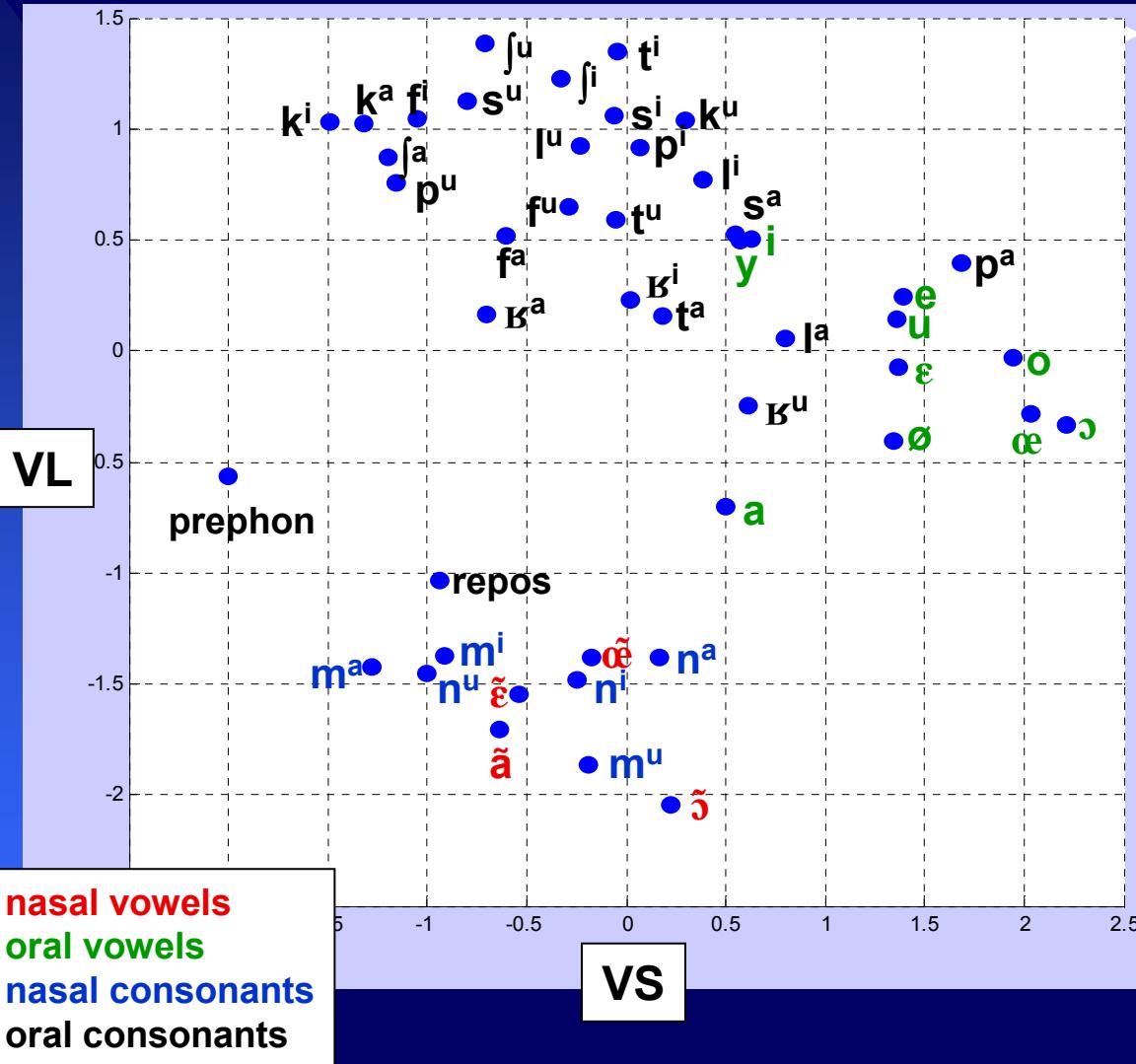
Nasopharyngeal wall

Midsagittal plan

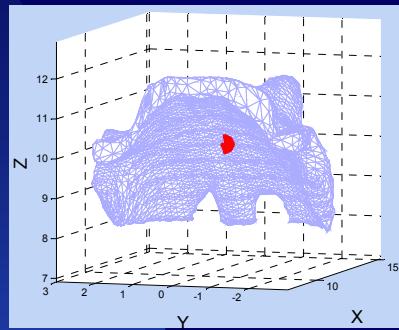
2nd parameter : VS



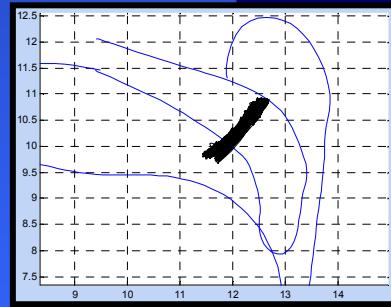
VL vs. VS



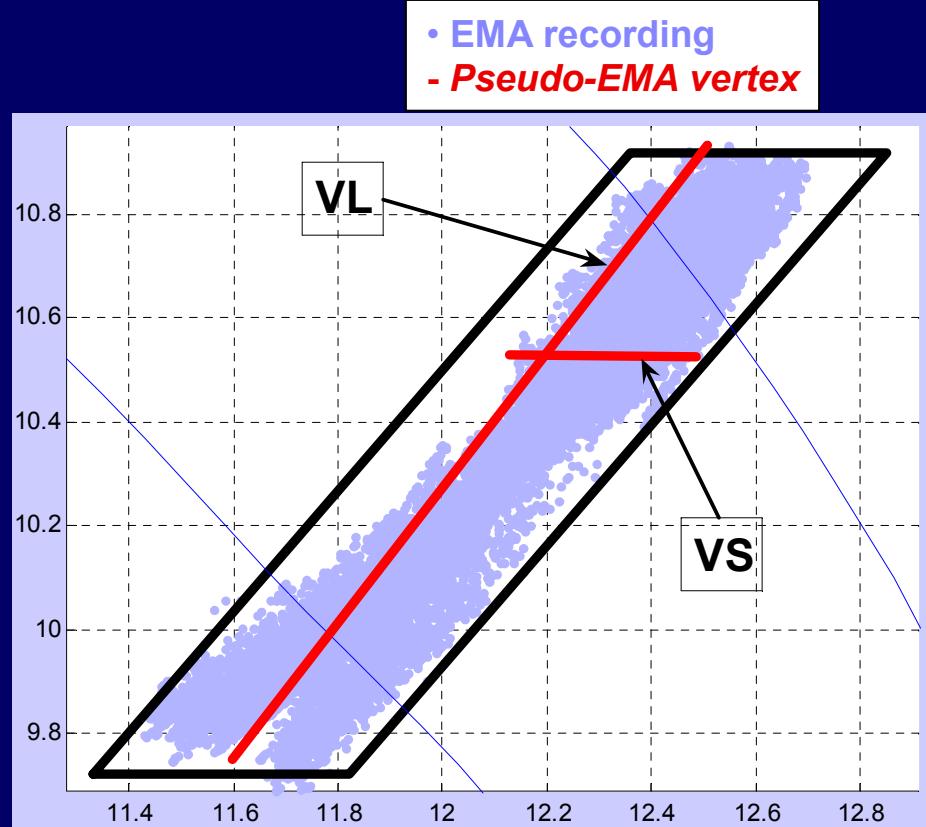
Coherence 3D articulatory model - EMA



- Articulatory model
- Pseudo-EMA vertex



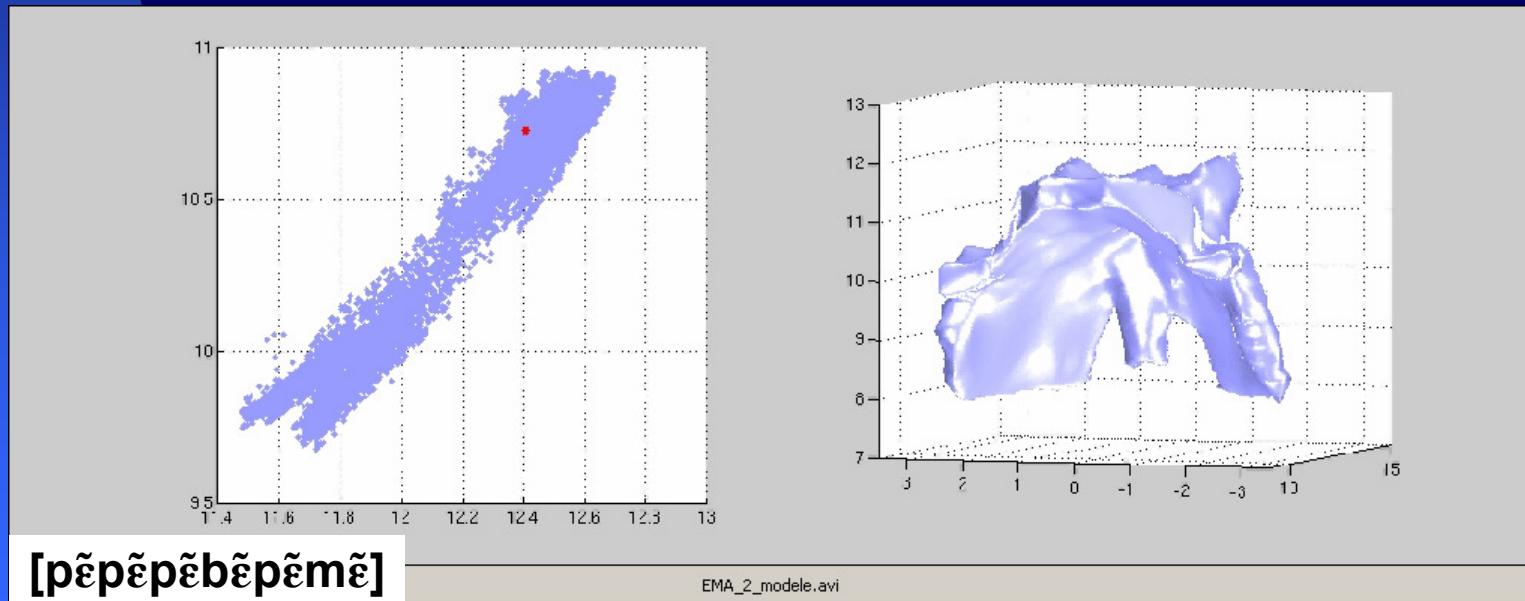
EMA recording



⇒ Coherence between the two data type

Control of the model from EMA data

- Inversion of VL et VS parameters from the *Pseudo-EMA vertex* → 3D RMS error = 0.07 cm
- Control of the model from EMA recording



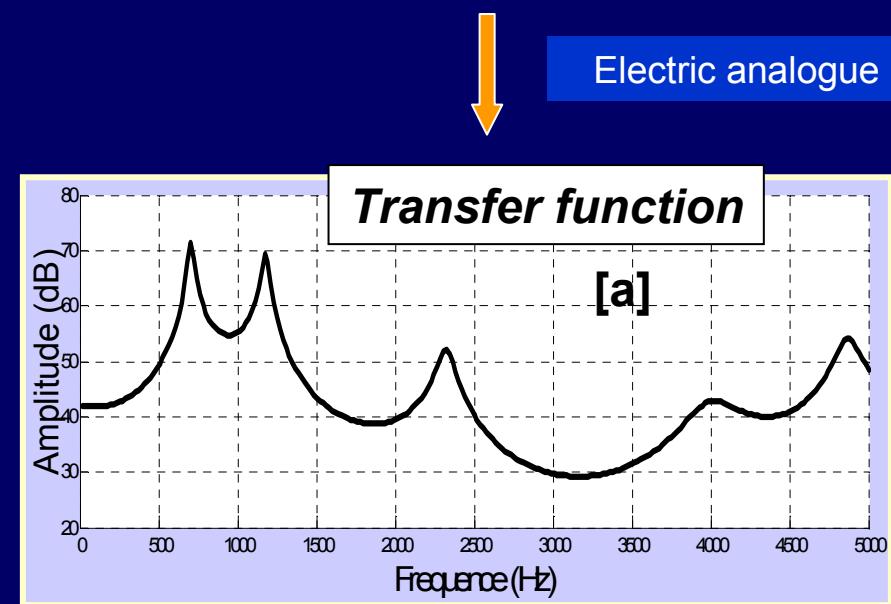
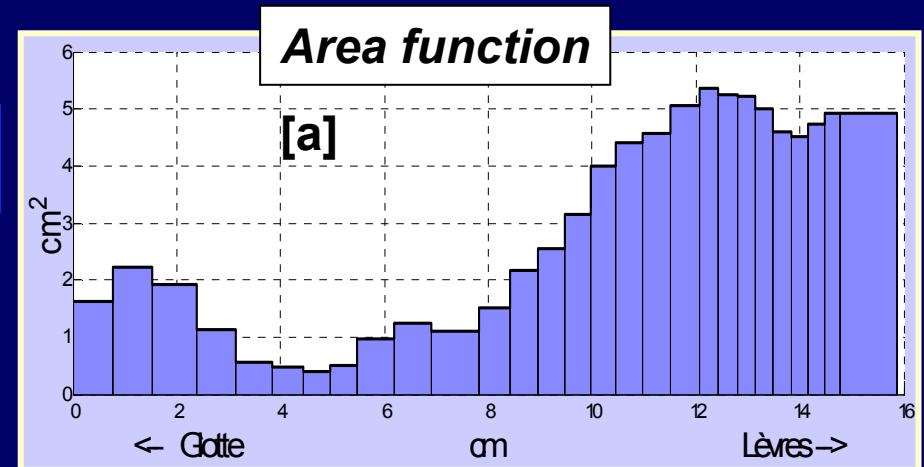
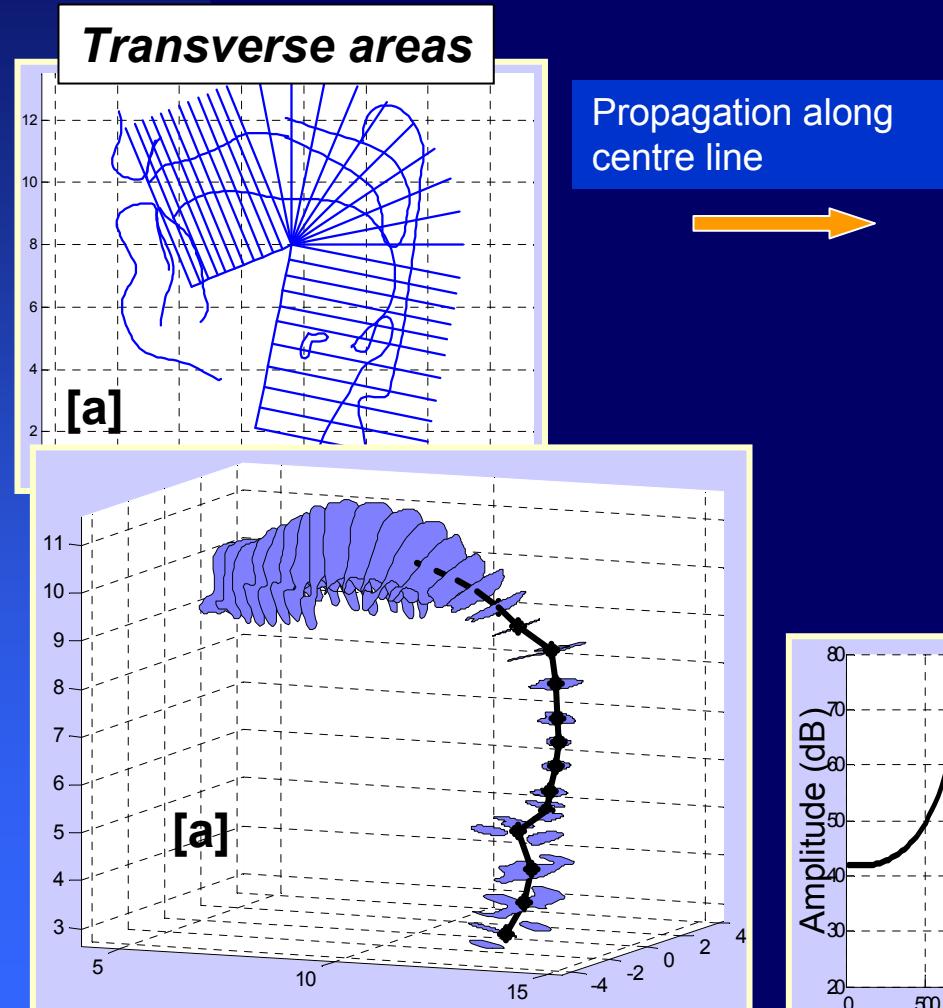
⇒ 3D reconstruction from 2D EMA dynamic recording

Three-dimensional model : conclusion

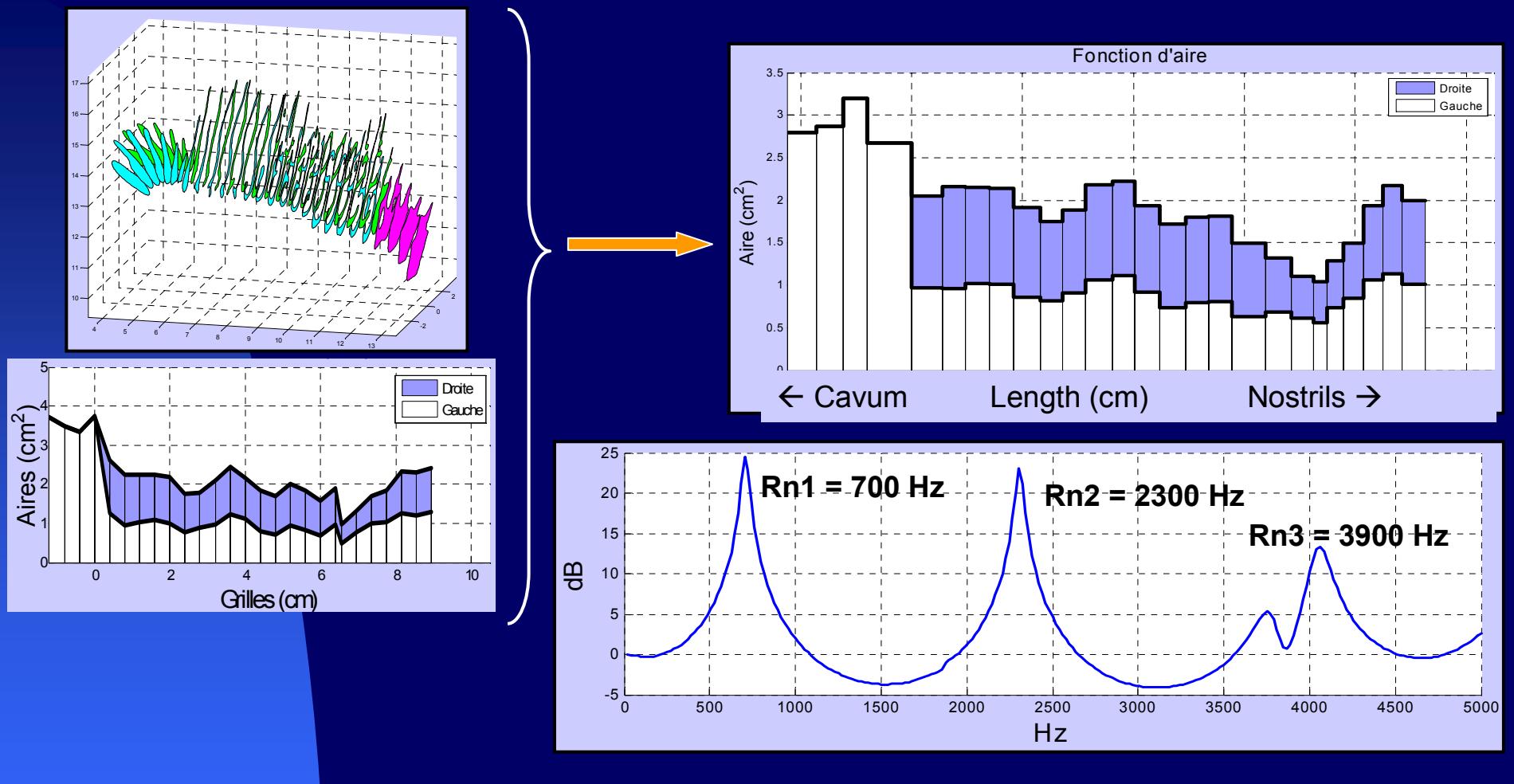
- Articulatory model with 2 parameters:
 - VL : Velum (83 %) + Nasopharyngeal wall
 - VS : Velum (6 %)
 - Accuracy < 1 mm
- Coherence EMA data + Model
- 2D / 3D correlation : RMS reconstruction error of 3D from 2D < 1 mm

Acoustic model

Hypothesis: $F < 5000 \text{ Hz} \Leftrightarrow$ Plane propagation



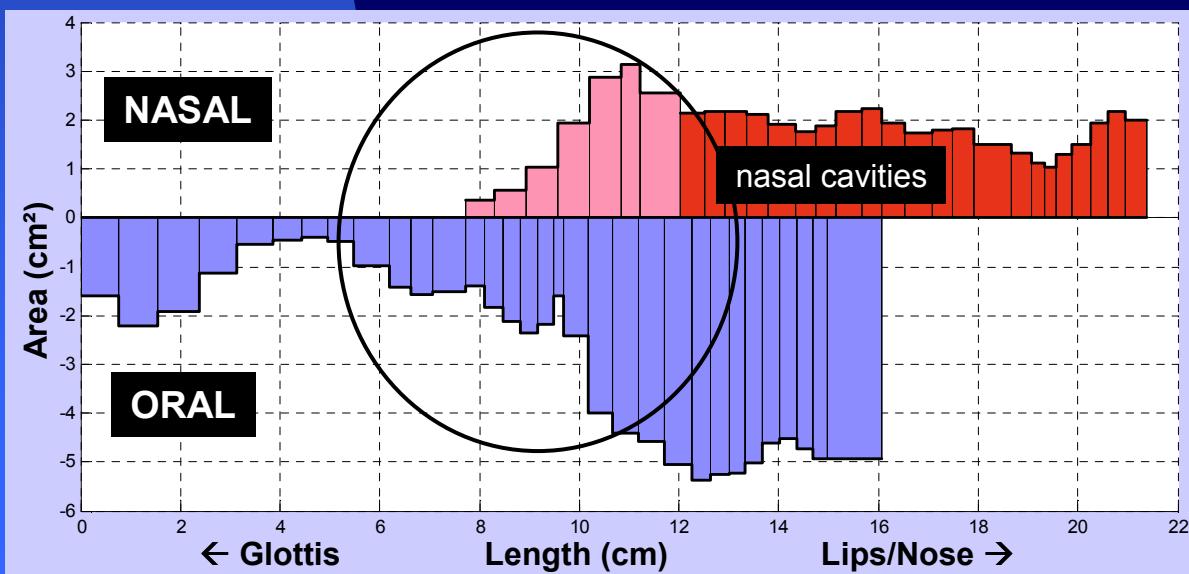
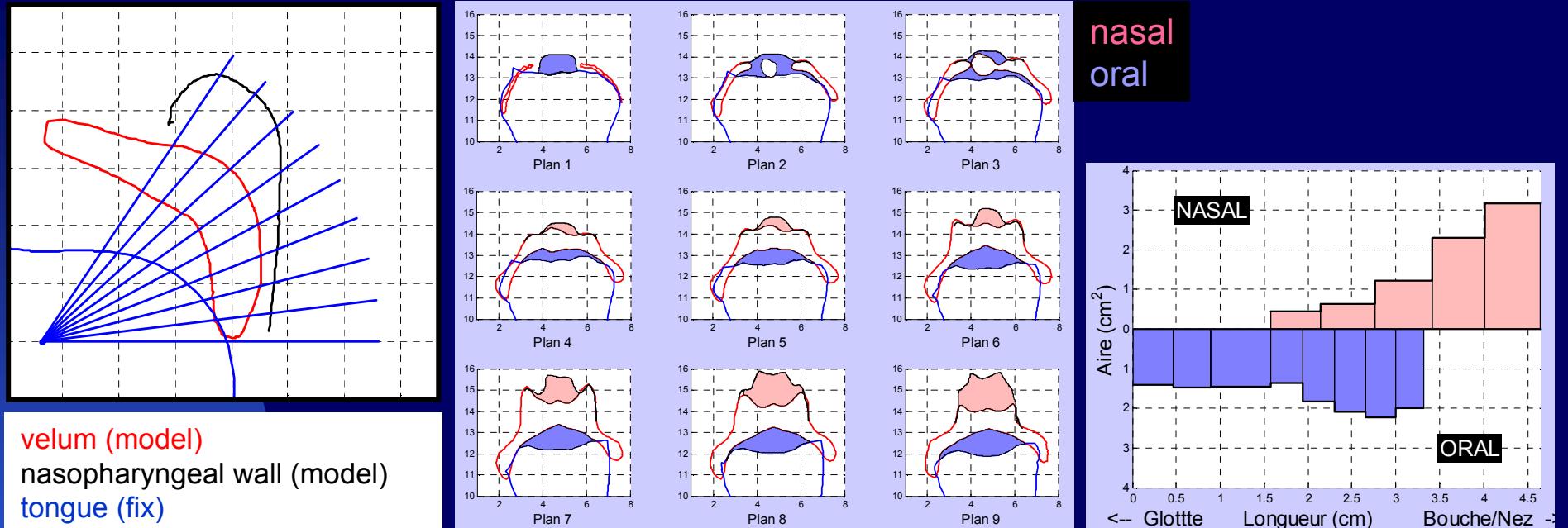
Nasal cavities



- Nostril constriction $\approx 1 \text{ cm}^2$
- Approximately $\frac{1}{4}$ wave length resonances
- Small asymmetry (R3)

**Nostril constriction
assumption not verified**

Velar region : area function

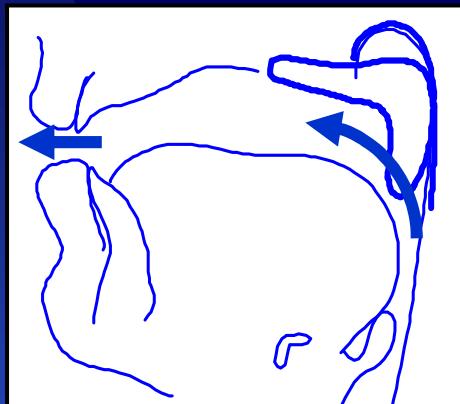


Inclusion in a fix
oral tract :

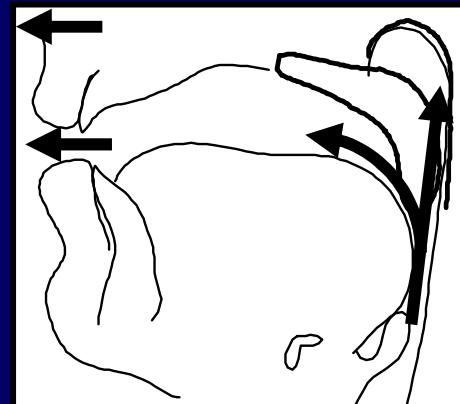
[**a, i, u, ã, ë, œ, õ**]
 cardinal vowels nasal vowels

Articulatory limits of nasality

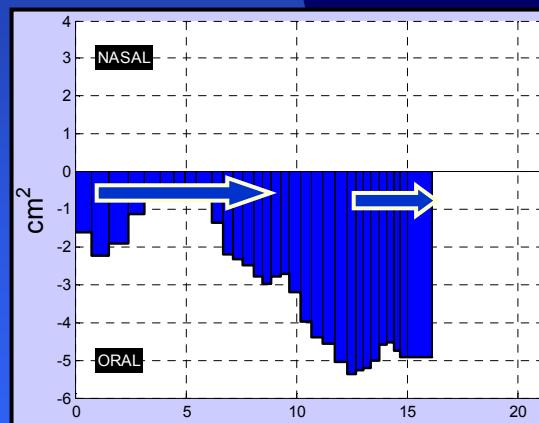
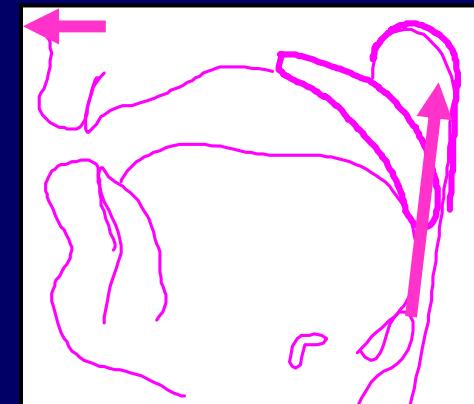
Oral configuration



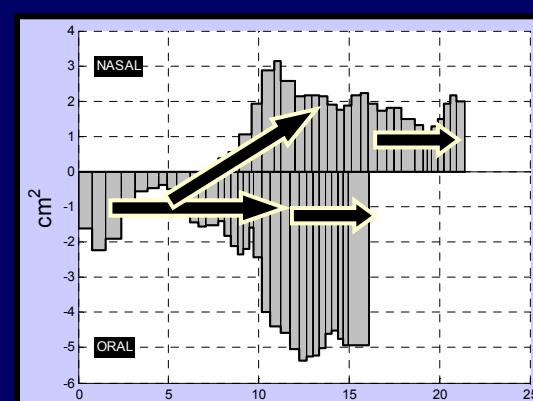
Nasal configuration



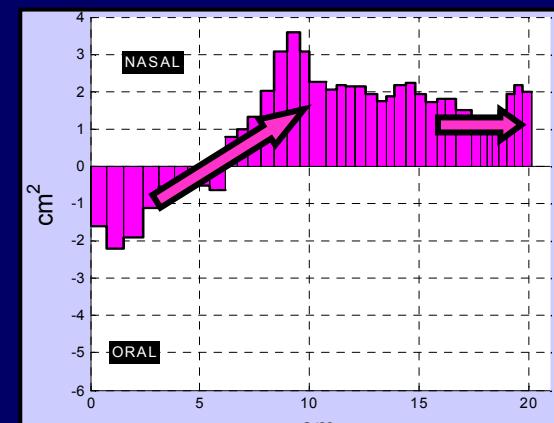
Nasopharyngeal configuration



$$TRFC = \frac{D_L}{D_G}$$

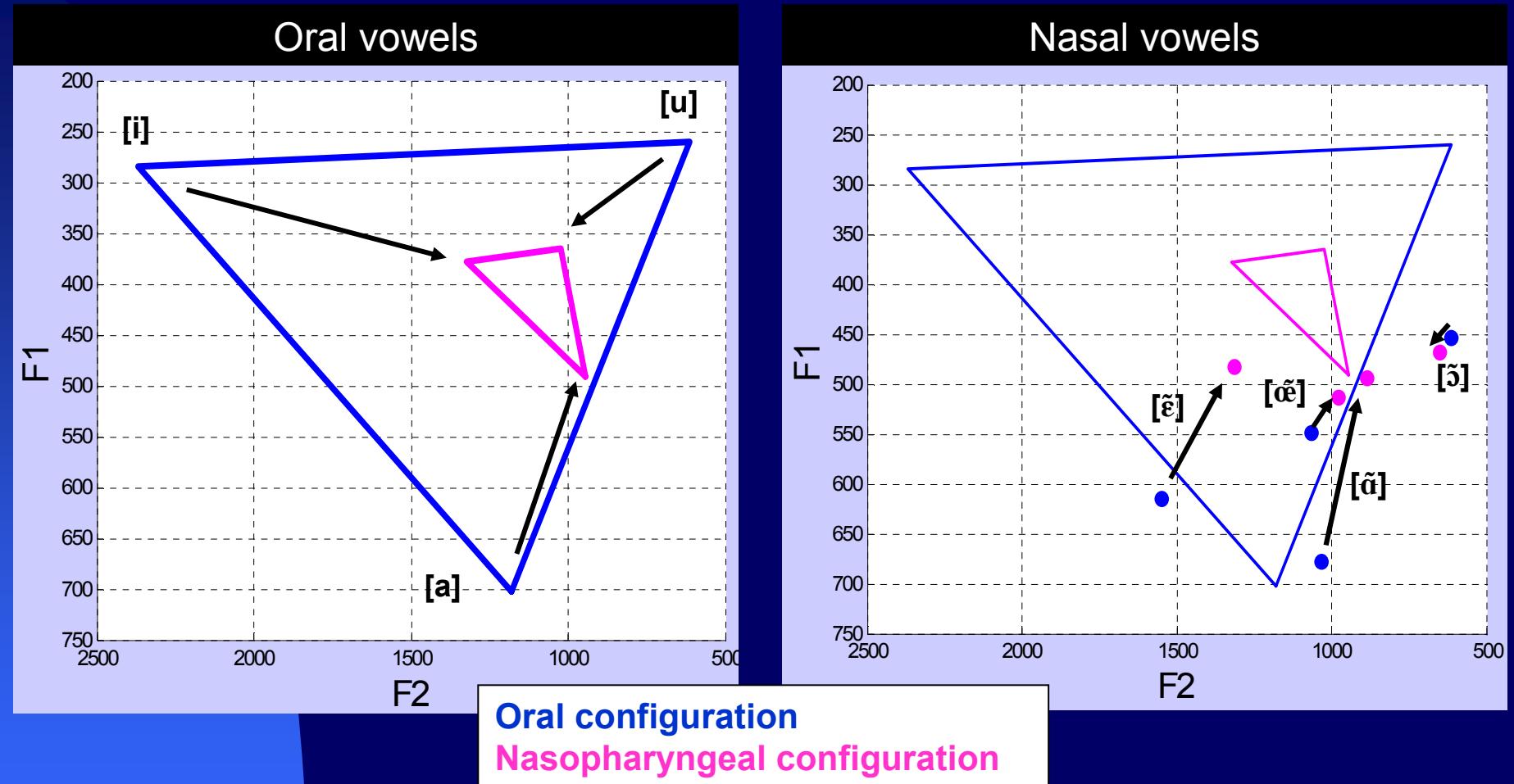


$$TRFC = \frac{D_L + D_N}{D_G}$$



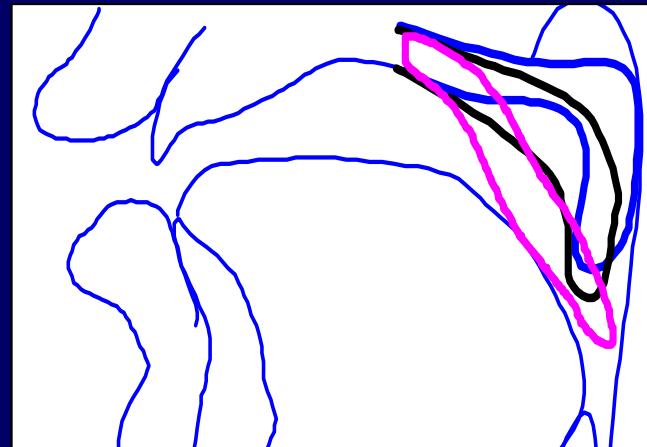
$$TRFC = \frac{D_N}{D_G}$$

Extreme articulations : low frequencies

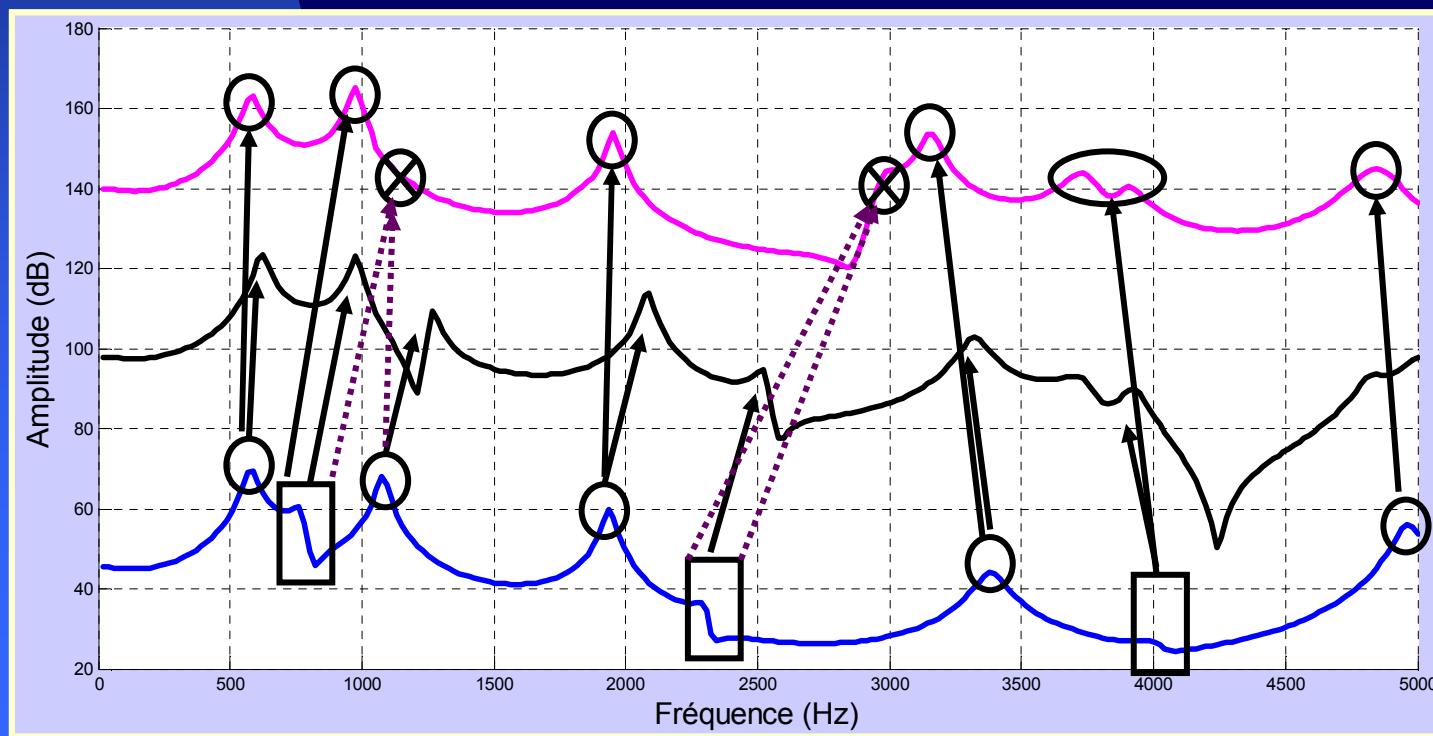


⇒ **Nasal gap** around 450 Hz – 1100 Hz

Acoustic coupling

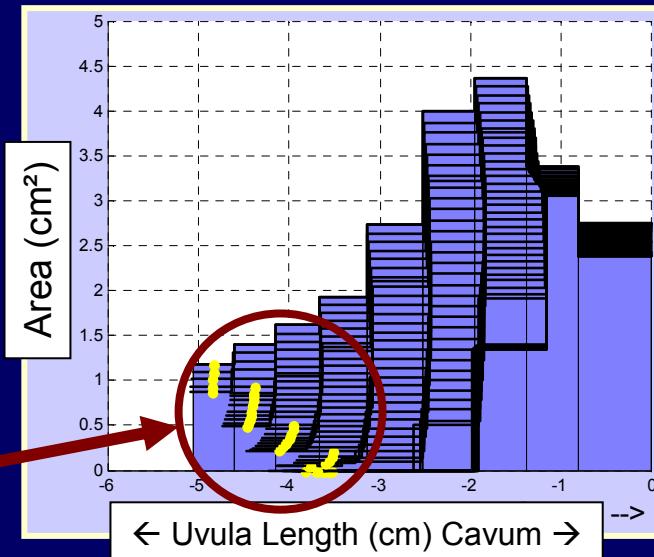
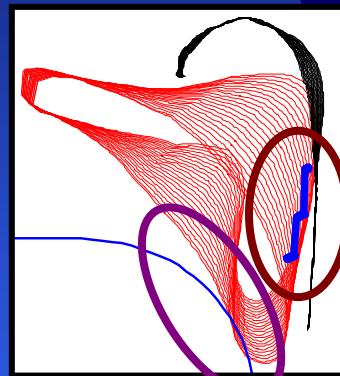


- ⇒ Resonances coming from coupling
- ⇒ Resonances coming from oral formants



VL : area function

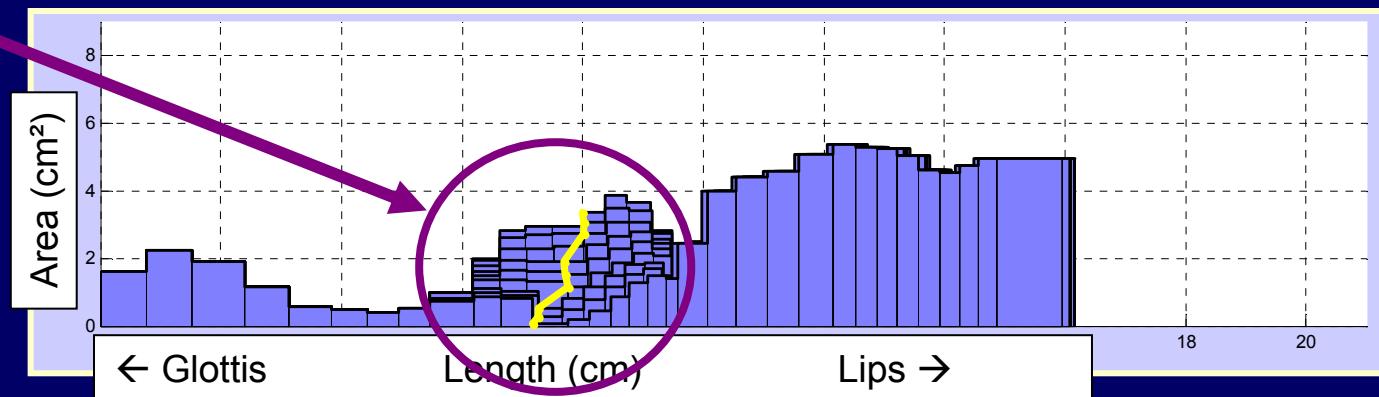
High variations of the velopharyngeal port



$\Delta\text{area} = 0 \text{ cm}^2 \rightarrow 0.8 \text{ cm}^2$

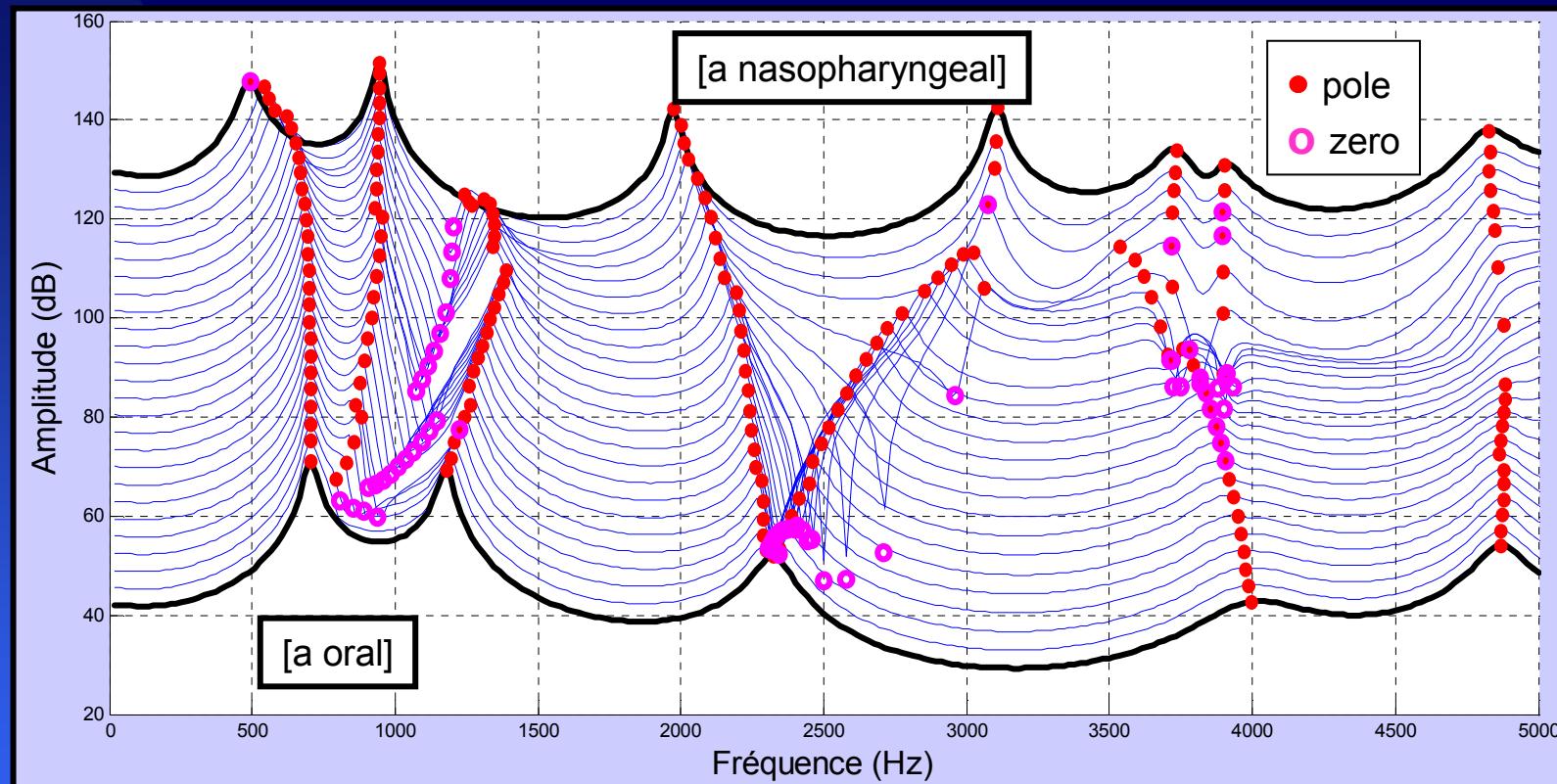
$\Delta\text{position} > 1 \text{ cm}$

High oral variations



VL : acoustic

nasopharyngeal

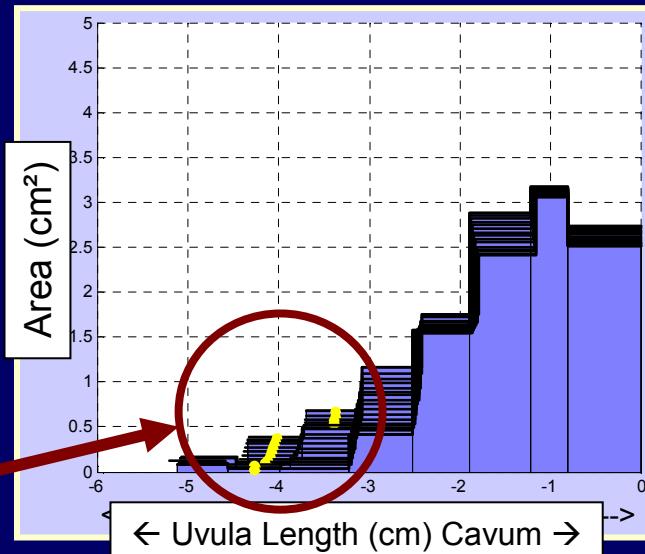
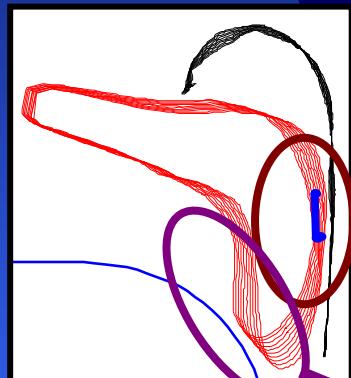


oral

⇒ Variations of the resonances coming from oral formants until 30 % depending of the vowel

VS : area function

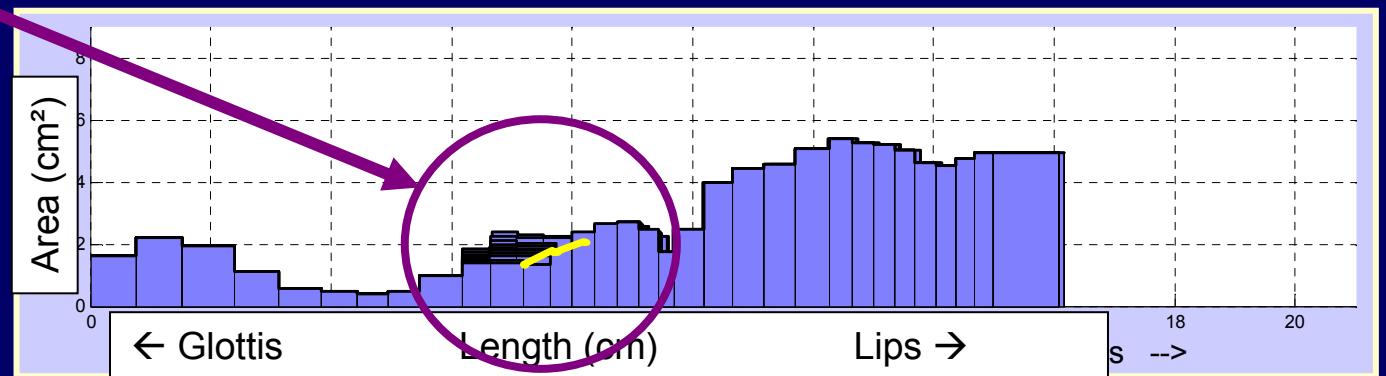
High variations of the nasal coupling area



$$\Delta \text{area} = 0.1 \text{ cm}^2 \rightarrow 0.6 \text{ cm}^2$$

$$\Delta \text{position} = 0.9 \text{ cm}$$

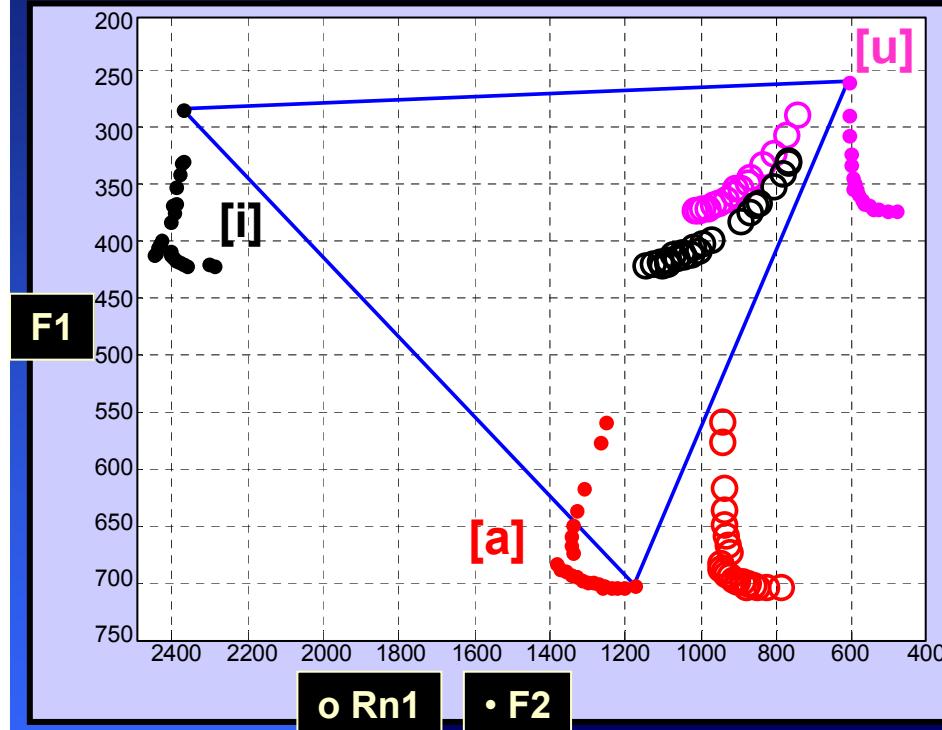
Smaller oral variations



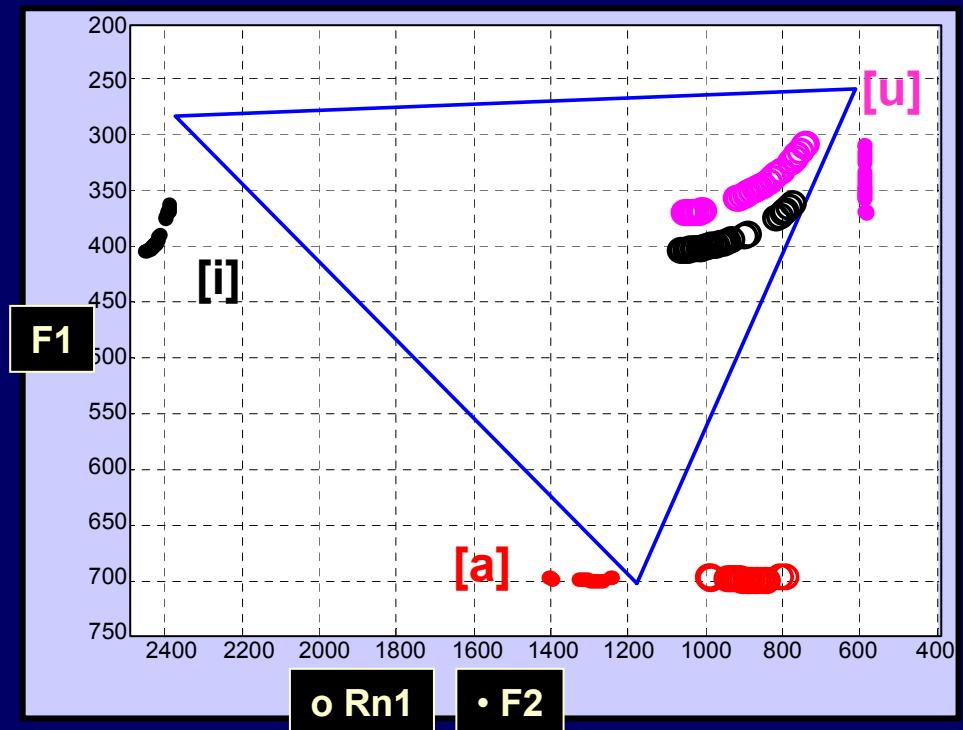
Low frequency nomograms of VL and VS

VL and VS variations in the data range around [a] [i] [u]

VL



VS



Variations of the nasal area function

Variations of the oral area function

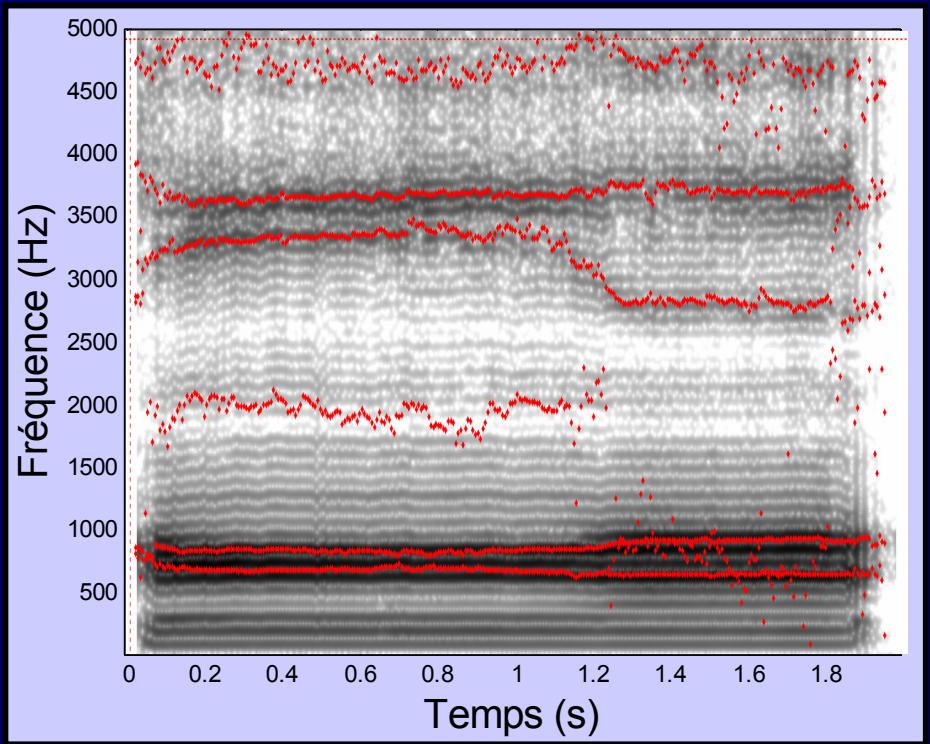
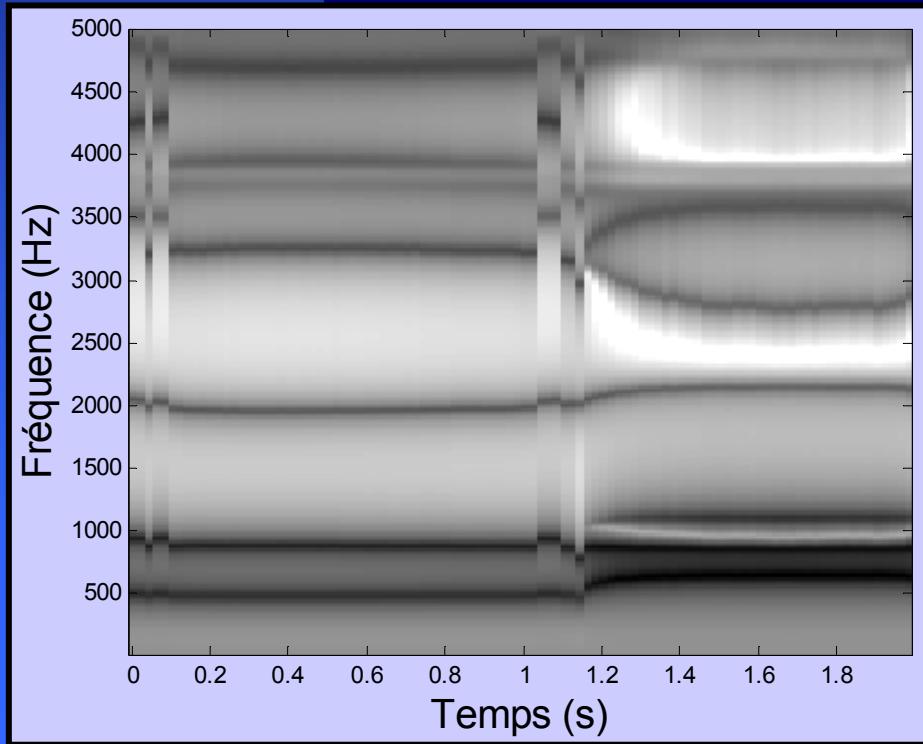
Simulations - recording comparison

EMA recording + synchronous Acoustic

Control of the articulatory model
Simulated nomograms

Recorded nomograms

$[\tilde{a}] \rightarrow [a]$



Articulatory-acoustic modelling : conclusion

- Realist Area Functions of the nasal tract
 - Nasal cavities : $\frac{1}{4}$ wave length resonances
 - Velopharyngeal port : VL ($\pm 0.8 \text{ cm}^2$) + VS ($\pm 0.5 \text{ cm}^2$)
 - Velar region of the oral tract : VL
- Coupling influence on low frequencies
 - Concentration in the nasal gap around 450 Hz – 1100 Hz
 - Variations of the poles sensitive to VL and VS
- Coherence between the model and the recording in high frequencies
- *1st nasal formant* not obtained in simulation

Conclusion

Articulatory data

- Cross use of different nature data
- 3D geometry of the nasal cavities and the velopharyngeal port

3D articulatory model of the velopharyngeal port : 2 parameters

- VL : dominant parameter (*Velum Levator*)
- VS : lower amplitude parameter
- High correlation between 2D movements and 3D shapes

Acoustic Characterization of the articulatory gesture

- Realist Area Functions of the nasal tract
- Nasal cavities : nostril constriction assumption not verified
- Influence of VL and VS on low poles

Perspectives

Acoustics

- 3D acoustic propagation
- Origin of nasal F1? Sinuses ?
- Comparisons with acoustic recording + EMA
- Possible external coupling consideration

Articulatory synthesis

- Inclusion in a complete articulatory model
- Study of the coordination velum / tongue / lips
- Control of the model from real-time MRI
- Articulatory synthesis and perceptual tests

- Extension to other subjects
- Text-to-Speech articulatory synthesis

Thank you!
