

Vaissière, J. (2008), On acoustic salience of vowels and consonants predicted from articulatory models, Keynote paper , 8th Phonetic Conference of China and the International Symposium on Phonetic frontiers, Pékin, Chine, 18th-20th Avril 2008 published paper: halshs-00676253, v1

*On acoustic salience of  
vowels and consonants  
predicted from  
articulatory models*



Many thanks to  
Shinji Maeda  
And students

***Jacqueline Vaissière***

***Université Sorbonne Nouvelle - Paris***

***Laboratoire de Phonétique et de Phonologie***

PPC2008, Pékin, Chine



# A

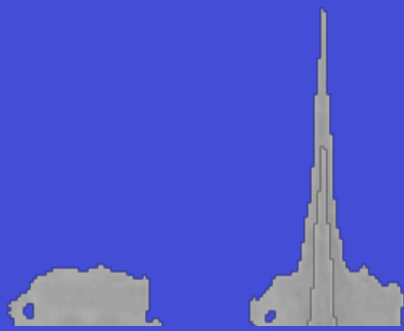
# Menu?



1. What is a model?
2. Why modelling?
3. Which models?
4. What is a useful model?
5. Demo (vowels consonants)
6. How done?
7. Other uses
  - speaker anatomy
  - speaker strategy
  - singing formant
  - inversion
  - etc

# B





# The questions



2

# What is a model?



What is a model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

vowels  
consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
investigation  
etc;



**A model is a simplified description of a complex system**

**A model is not a goal per se, but a tool**

- 1. to know what we know**
- 2. and what we don't know**
- 3. And where to look for ...**

# Why modelling?

What is a model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

vowels  
consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;



Modeling is needed to apprehend *complex* phenomena

Speech is *complex* enough to need modelisation

Modeling allows to evaluate the separate contribution of individual parameters and the acoustic and perceptual consequences of large and small movements

Modelisation in turn makes speech looking *less or sometimes more complex*

# Which models?

- What is a model?
- Why modelling?
- Which models?**
- What is a useful model?
- How done?
- Demo
- vowels
- Other
- ts
- atomy
- ategy
- rmant
- etc;



## Phonetics

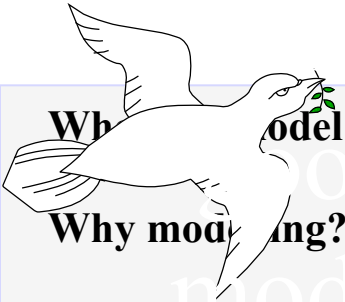
- *Prosodic modelling*
- *Source modelling*
- *Vocal tract*
  - *2D articulatory*
  - *3D articulatory*
- *Articulatory modelling*

## Psycholinguistics

- *perception and recognition*
  - *Phoneme*
  - *Word access*
  - *Sentences*
- **Etc.**



# Useful models?



Why model?

Why modeling?

Which models?

**What is a useful model?**

How done?

Demo

vowels


consonants


Other uses

speaker anatomy

category

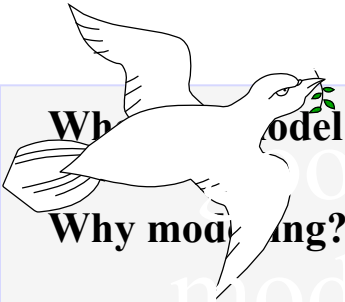
segment



- Be compatible with the observations*
  - Be testable*
  - Describes well the data and to be able to reproduce them from*
  - Few number of parameters*
  - As motivated as possible*
  - As independent as possible*
  - And for AM, react as a human vocal tract*  
*no stop without a burst, or no frication when humans do.*
  - To draw a direct link between articulatory, acoustic and perceptual observations*
- 



# How done?



Why model?

Why modeling?

Which models?

What is a useful model?

**How done?**

Demo

vowels


consonants

Other uses

speaker anatomy

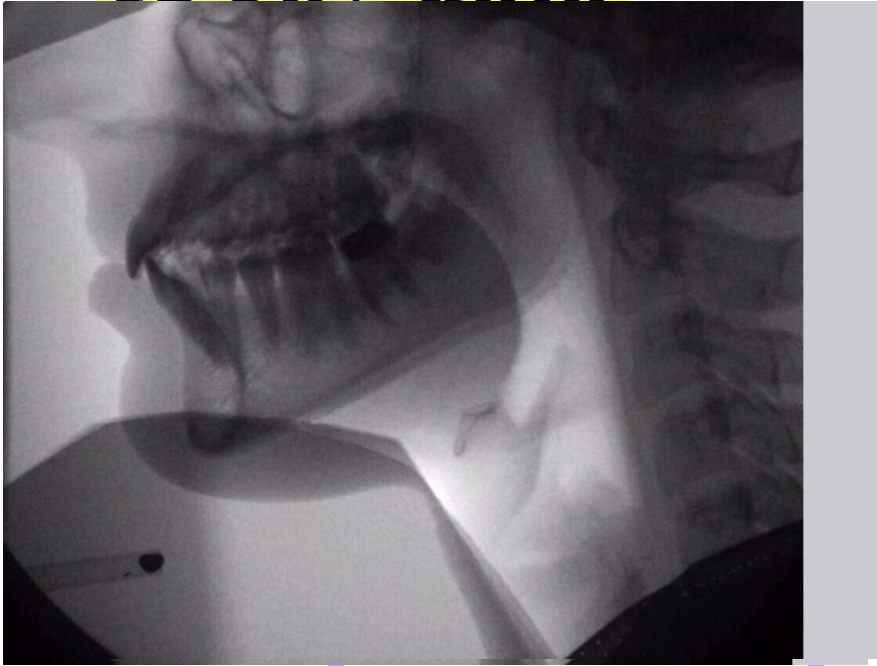
category

formant

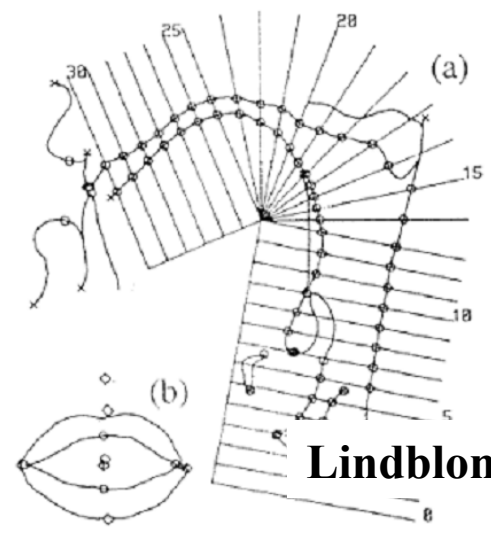


- *By Shinji Maeda*
- *From X-ray data for French (from Strasbourg)*
- *By statistical analysis*
- *The most widely used AM*
- *Used for pedagogical purpose in our lab and research in Grenoble, Nancy, UCLA, MIT, etc.*

# 1 X-ray data



2



Lindblom & al, 1971

3

## Guided PCA

4

7 parameters

5

One speaker  
10 sentences  
>1000 images

## VT shapes

First substract the effect of the jaw

alisation



# 4 first parameters Jaw + tongue

How done:

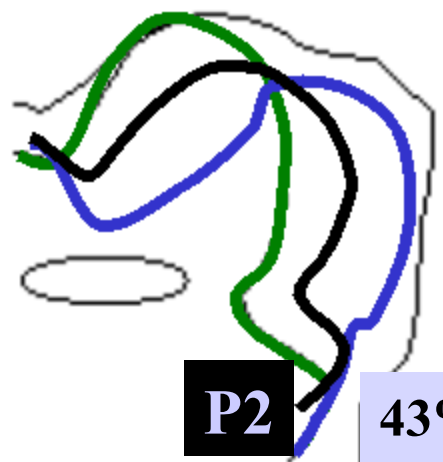
Demo

1+2= 58%, 56%

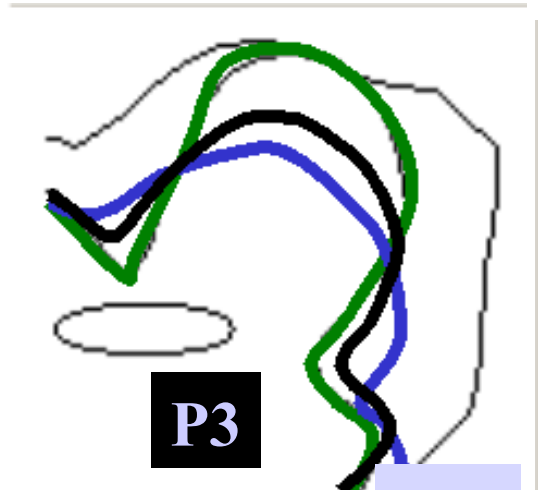
1+2+3= 81%, 79%

1+2+3+4= 88%, 86%

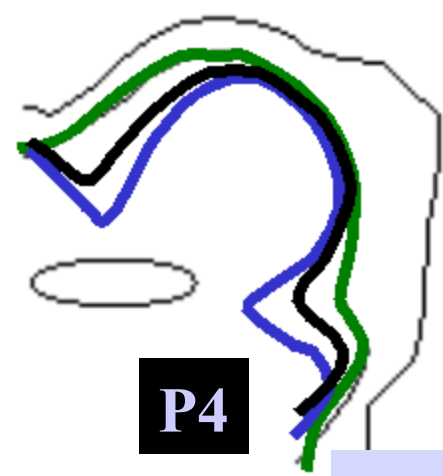
Other uses



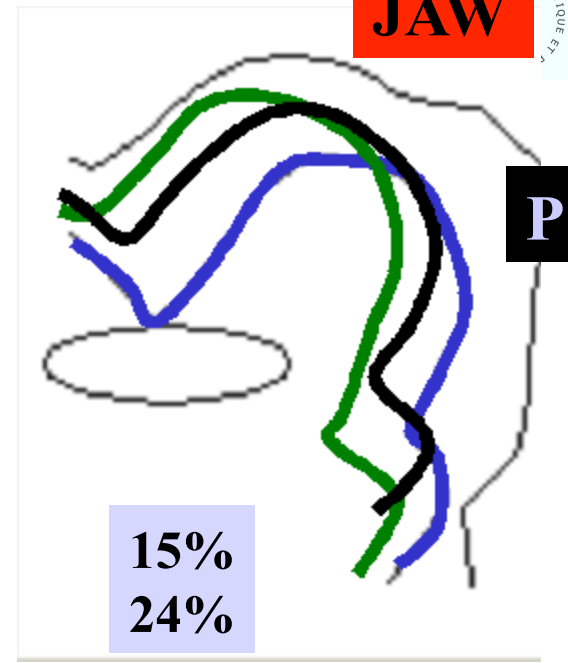
43%  
32%



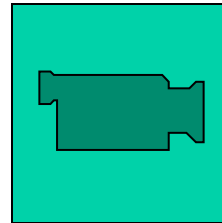
23%



7%

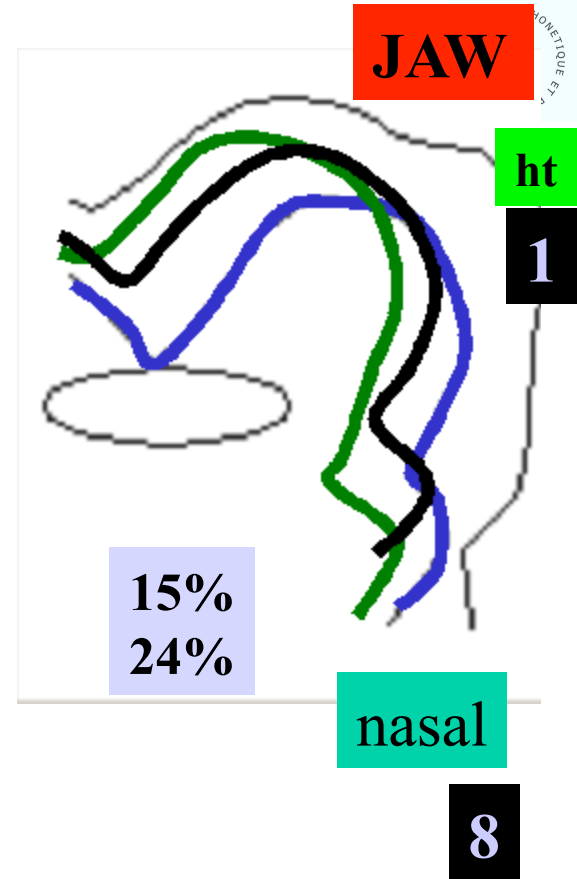


15%  
24%



4 3 more =  
Lip and larynx

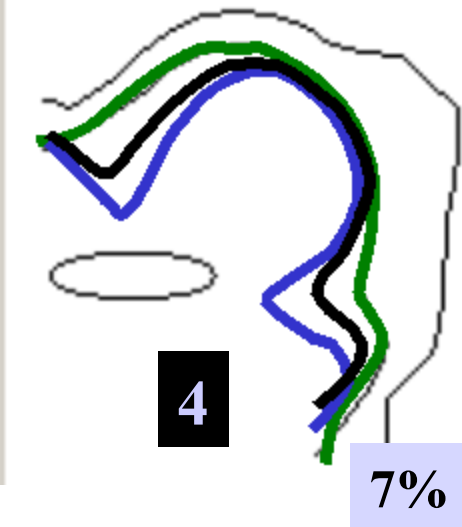
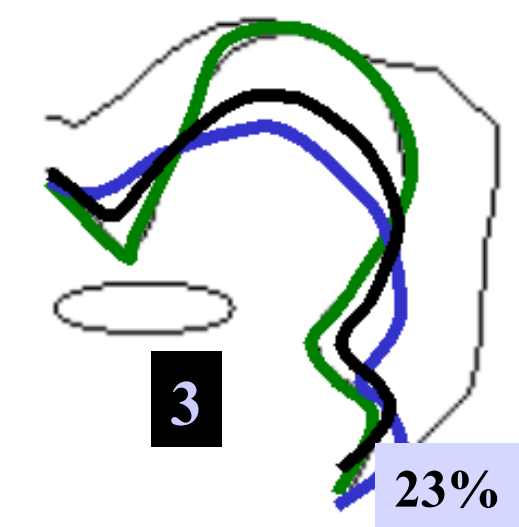
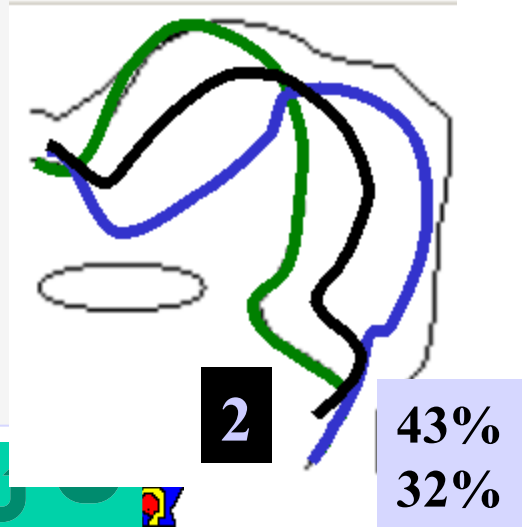
- LIP**
- ht 5
- pr 6
- larynx**
- ht 7



How done?

Demo  
vowels  
consonants

Other uses





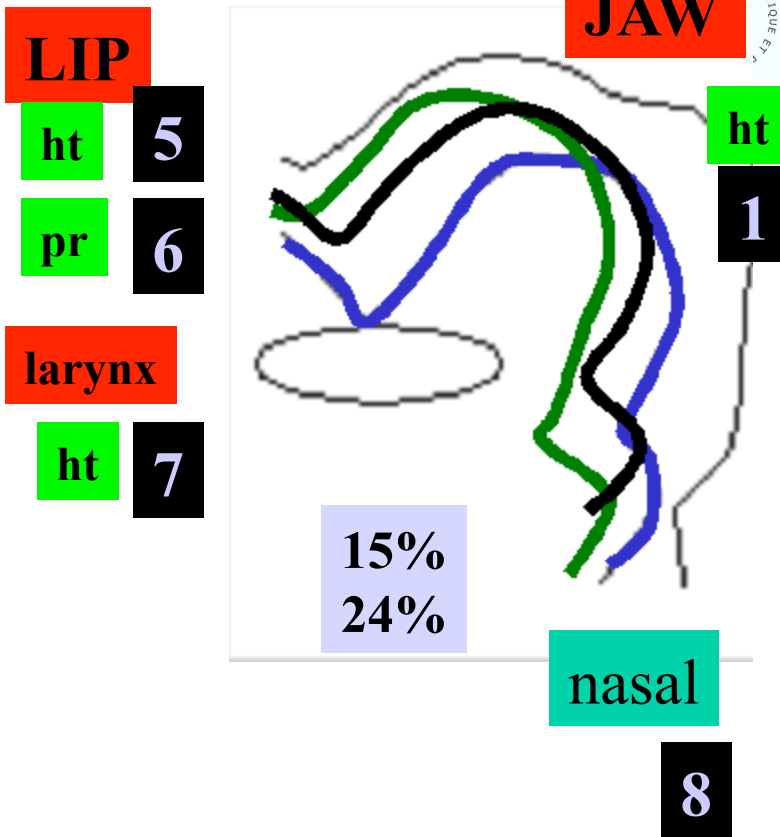
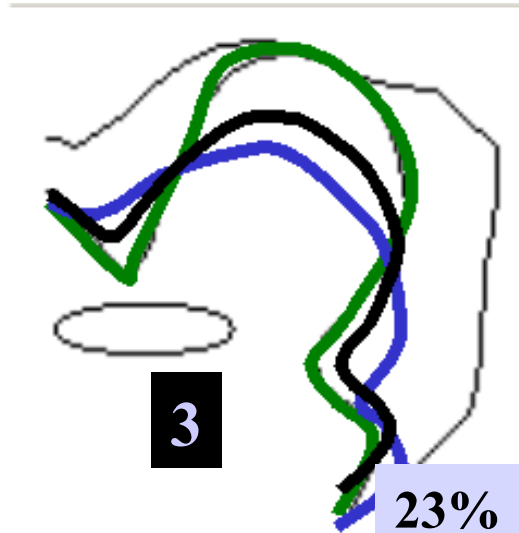
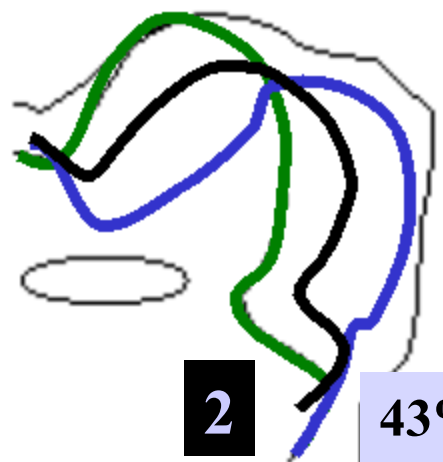
# 4 7 parameters = Speech organs

How done?

Demo

vowels  
consonants

Other uses



# Surprisingly or not

The 7 statistically found parameters could be interpreted in phonetic terms

2 for lips, 3 for tongue, 1 for jaw, 1 for larynx



4 7 commands

model?

How done?

Demo

vowels  
consonants

Other uses

body

TONGUE

shape

u, k

apex

i, e, dentals

8

LIP

ht

5

pr

6

larynx

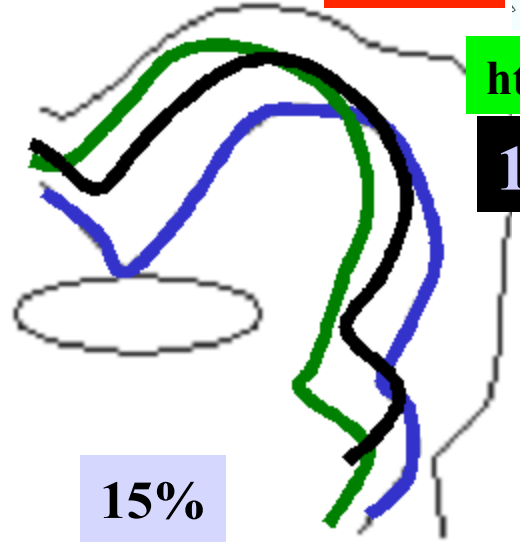
ht

7

JAW

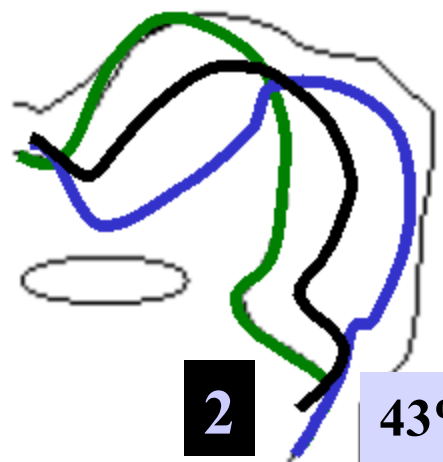
ht

1



15%  
24%

nasal



2

43%  
32%



3

23%



4

7%

Nasal cavity?  
Sublingual cavity  
Why modelling?

Which models?

What is a useful  
model?

How done?

Demo  
vowels  
consonants

Other uses  
speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

radiation load: RL\_CIRCUIT

wall: YIELDING

nasal tract: OFF

glottis: CLOSE

air density=0.00114

sound velocity=35000

wall resistance=1600

wall mass=1.5

wall compliance=300000



1

# Examples?

What is a model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

vowels  
consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

- 1) vowels
- 2) CVC sequences



1

# 1) vowels

What is a model?

Why modelling?

Which models?

What is a useful model?

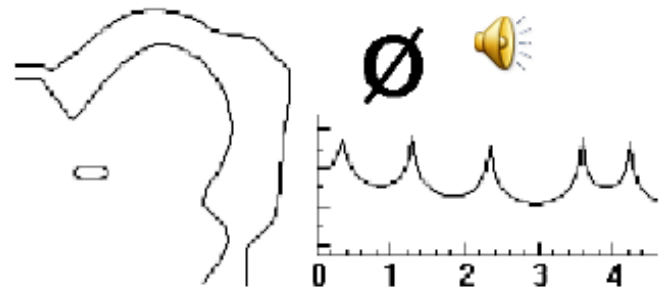
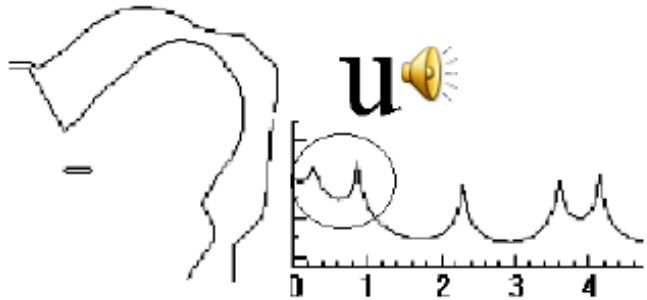
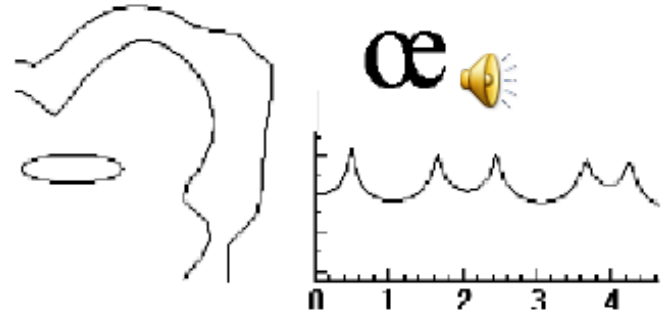
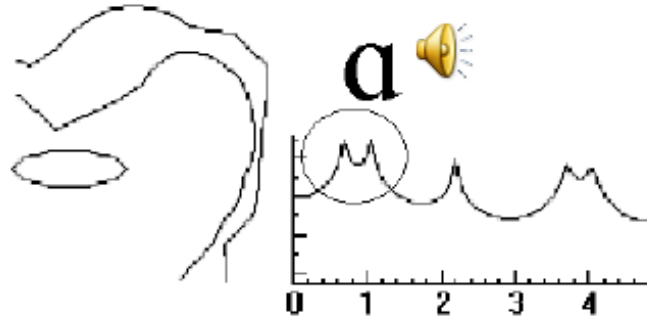
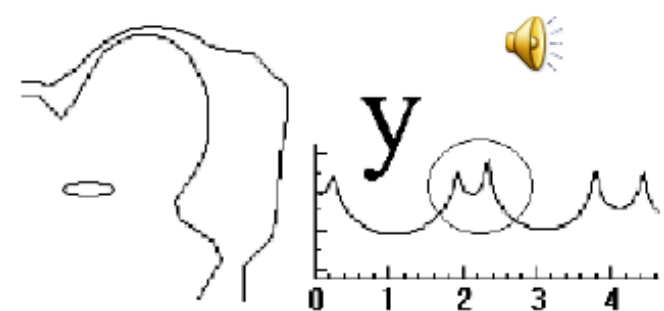
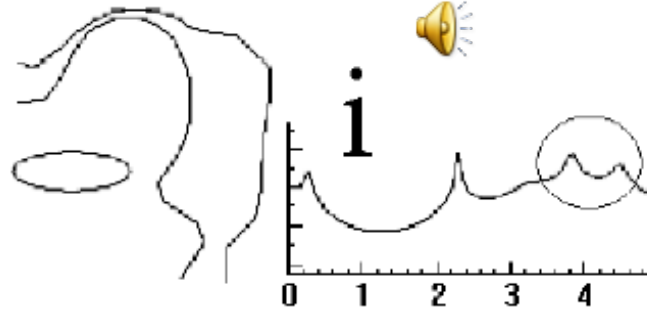
How done?

Demo

→ vowels  
consonants

Other uses

speaker anatomy  
speaker strategies  
singing forms  
inversion  
etc;



Vaissière, 2007

PPC2008, Pékin, Chine

**What is a model?**

**Why modelling?**

**Which models?**

**What is a useful  
model?**

**How done?**

**Demo**

→ vowels  
consonants

**Other uses**

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;



# 2) consonants

What is a model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

→ vowels  
→ consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;



F1 371  
F2 898  
F3 2136  
F4 3683  
F5 4024



F1 712  
F2 1145  
F3 2229  
F4 3745  
F5 4024



F1 186  
F2 1207  
F3 2321  
F4 3776  
F5 4457

closure 1



F1 248  
F2 2290  
F3 3838  
F4 4457

after release





# Beans on a string synthesis

1

# 2) consonants

What is a model?

Why modelling?

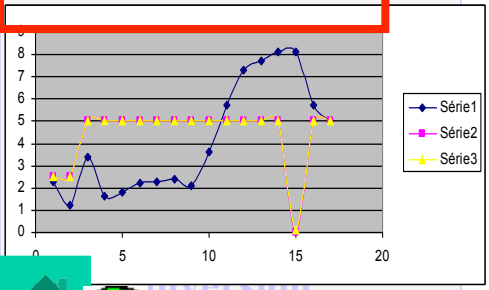
Which models?

What is a useful model?

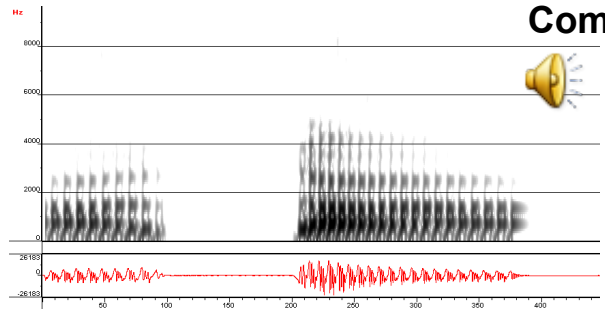
How done?

**Demo**

→ vowels  
→ consonants

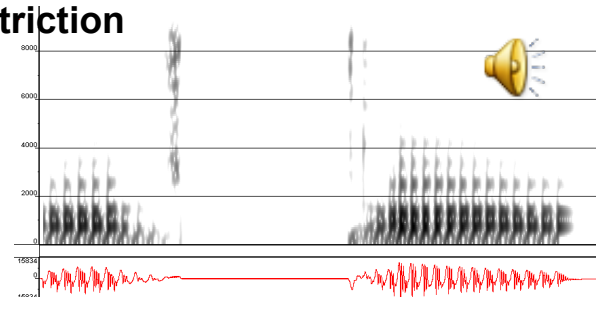


## Glottis close

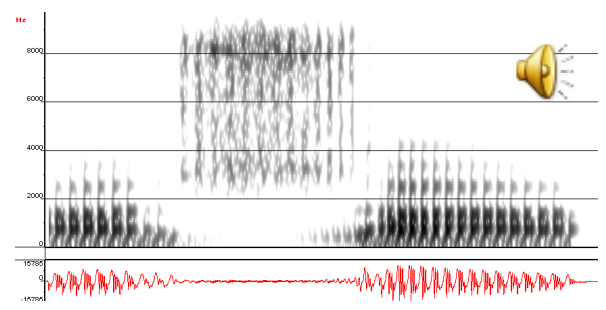


### Complete constriction

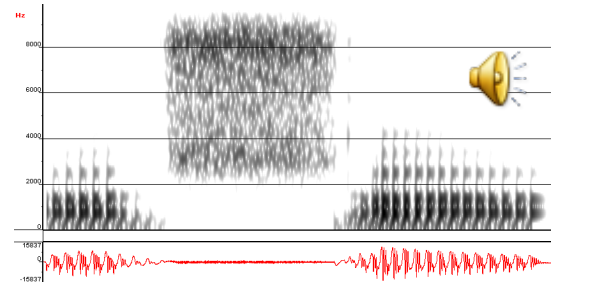
## Glottis open



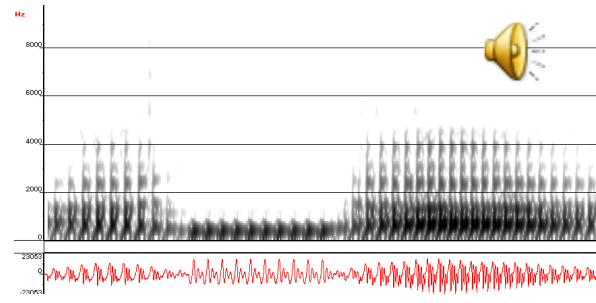
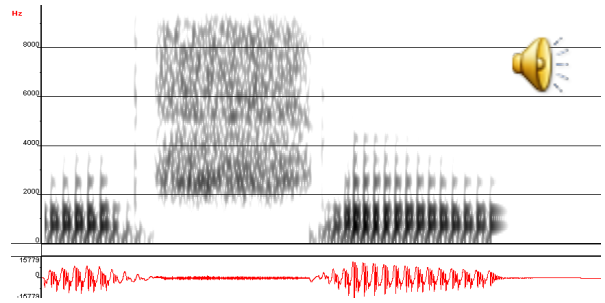
### Complete constriction



### Sublingual cavity



### Nasal port open



# Lip constriction

What is a model?

Why modelling?

Which models?

What is a useful model?

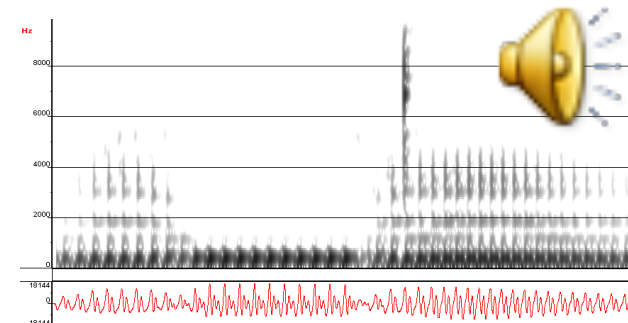
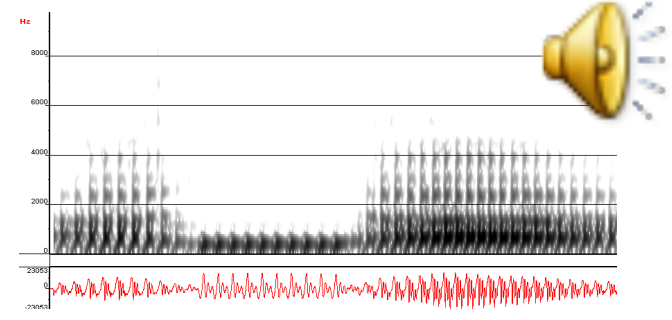
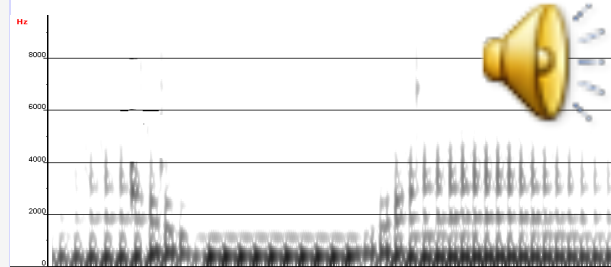
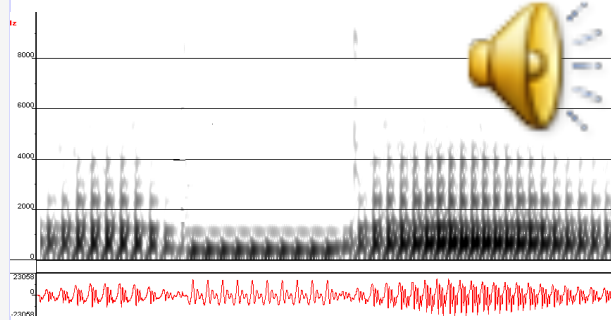
How done?

**Demo**

→ vowels  
consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;



• /a/ during 30 msec

cosine transition for 50 ms

/k/ closure for 100 ms

/k/ release for 17 msec

cosine transition for 17 ms

/u/ for 150 msec

What is a model

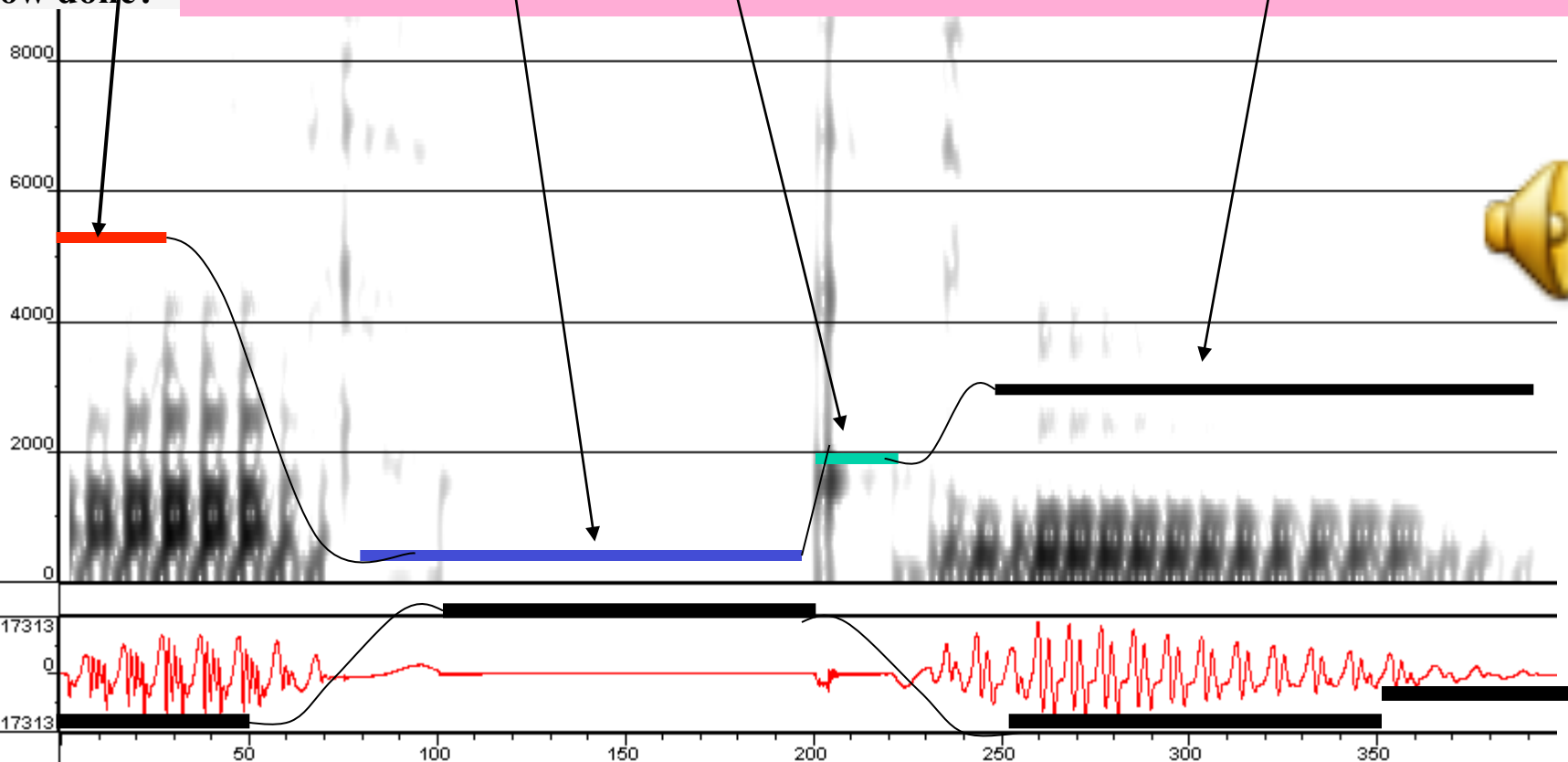
Why modelling

Which model

What is a useful model?

How done?

E



- **Start by close VF for 50 msec**

Then cosine transition to open glottis (0.4) at 100 msec

**Stay open for 100 msec**

Then cosine transition to close glottis at 250.

**Stay close for 100 msec**

Open slightly for the last 50 msec. (prepausal

opening)

What is a

Why mode

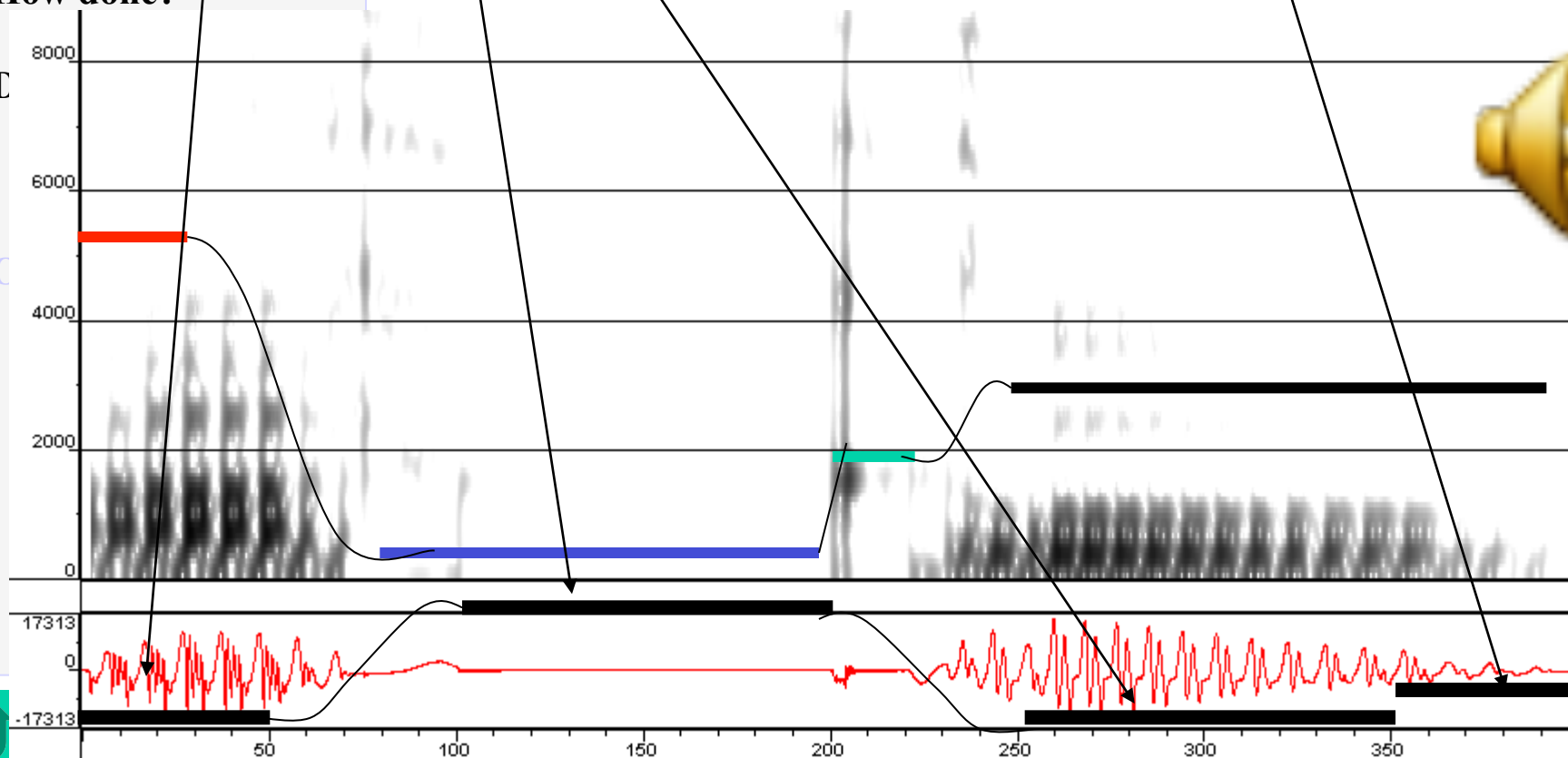
Which mo

What is a

mode

How done

E



What is a model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

→ vowels  
consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

# Glottis and the three sources in the model

- **A) closed**

- 

**voicing**

- **B) open**

- If less open than VT:

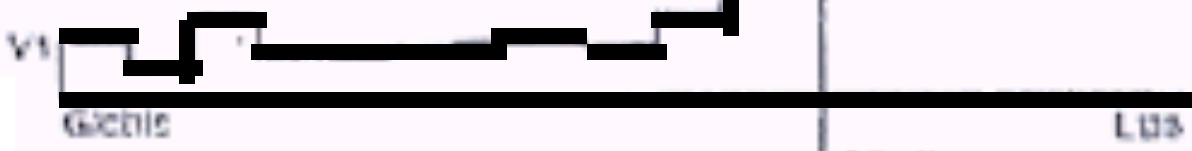
**aspiration**

- If VT less open than glottis: **frication**

1)

Invariant place and shape

V



C



V

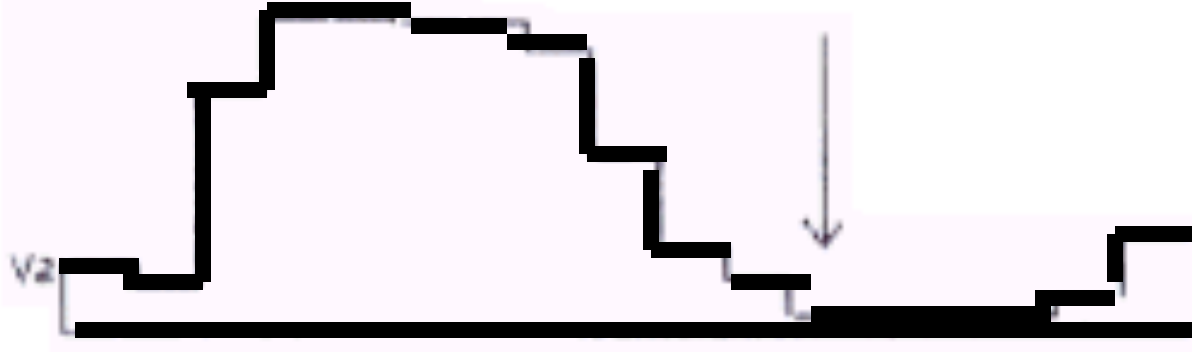


Figure 2: Interpolation of target area function

What is a model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo  
vowels  
consonants

Other uses  
speaker anatom  
speaker strategy  
singing forma  
inversion  
etc;



What is a model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

→ vowels  
→ consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

# Different degrees of sophistication

## 1) Simple concatenation

/asa/ = 3 targets

= /a/ + /s/ + /a/

/ata/ = 4 targets

= /a/ + /tclosure/ + /t release/ + /a/

+ Cosine transitions

## 2) Contextual allophones

Ease of articulation

Necessary for velar

**What is a model?**

**Why modelling?**

**Which models?**

**What is a useful model?**

**How done?**

**Demo**

→ vowels  
consonants

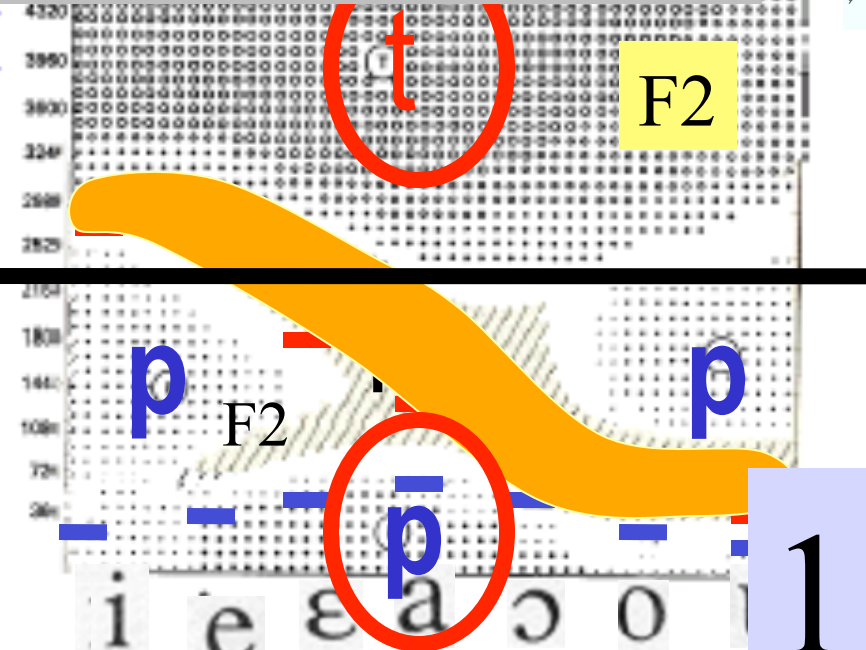
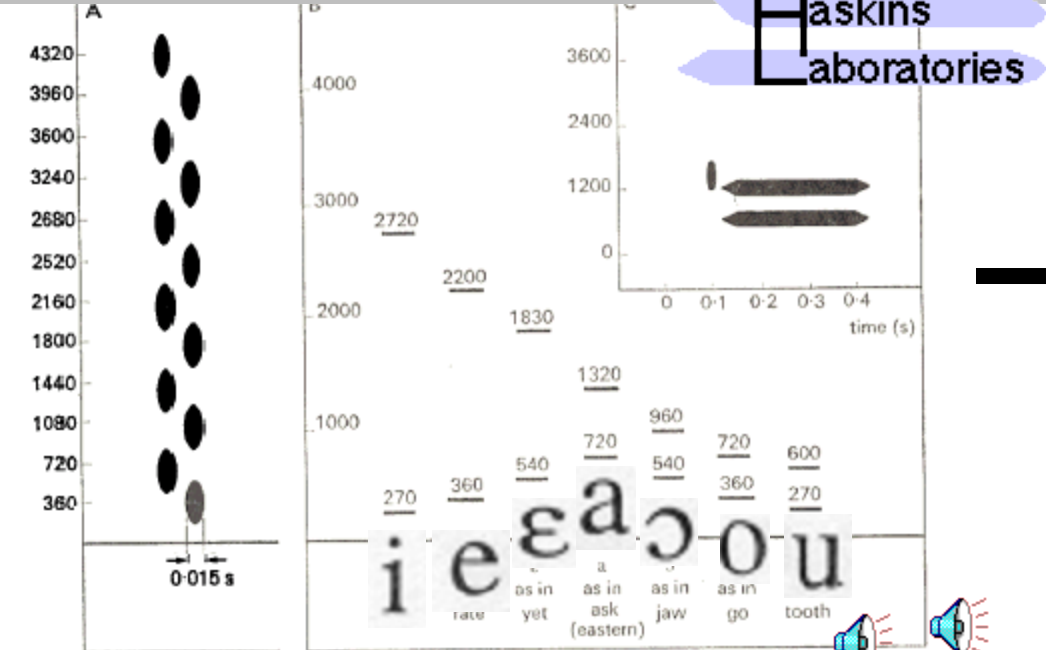
**Other uses**

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

- /k/ and /g/ require the burst to be in special position to the ‘main pitch’ of the vowel
- Different /k/ allophones
- Main pitch /i/= 3000 Hz, /y/= 2000 Hz, etc.
- Not articulatory ease, but perceptual requirement
- It seems to be not a linguistic but a purely psychoacoustic phenomena (temporal masking)

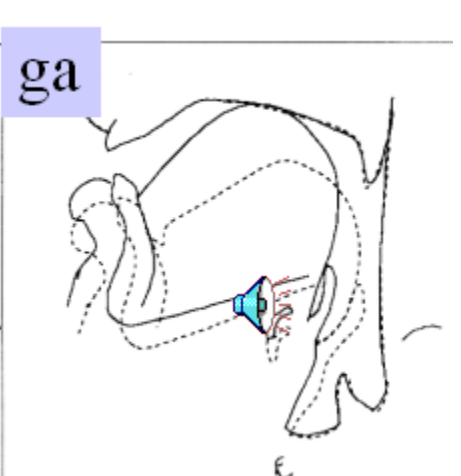
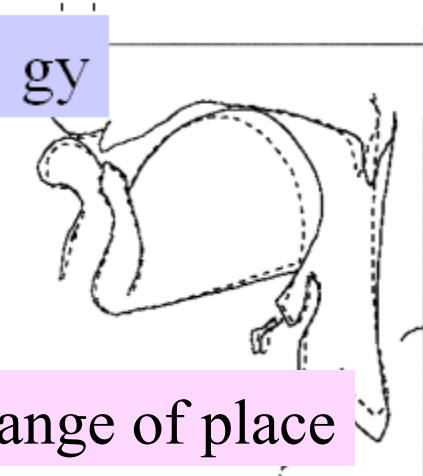
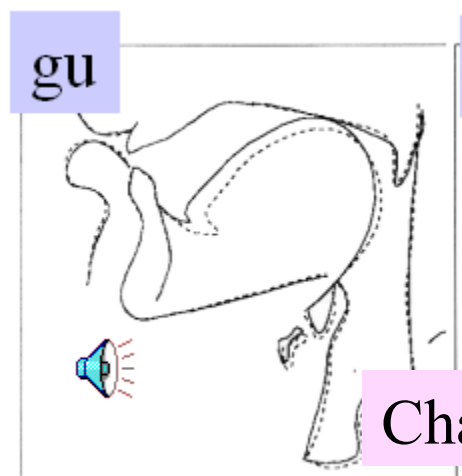
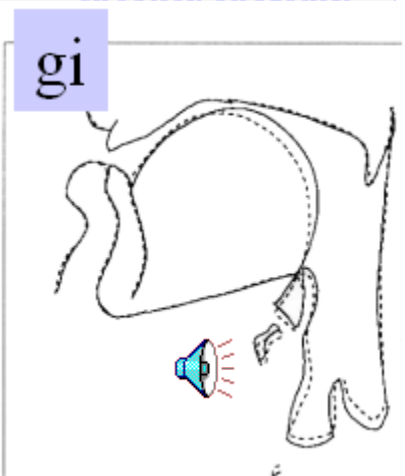


the reference to the vowel is not needed for discriminating [p] and [t]



velar  
Other uses

Perceptual requirement?



Change of place

What is a model?

Why modelling?

Which models?

What is a useful  
model?

How done?

Demo

→ vowels  
consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

# Different degrees of sophistication

**1) Simple concatenation**

**2) Contextual allophones**

**Velar**

**3) Anticipation and carry over  
phenomena**

**/aku/** model?

Why modelling?

Which models?

What is a useful model?

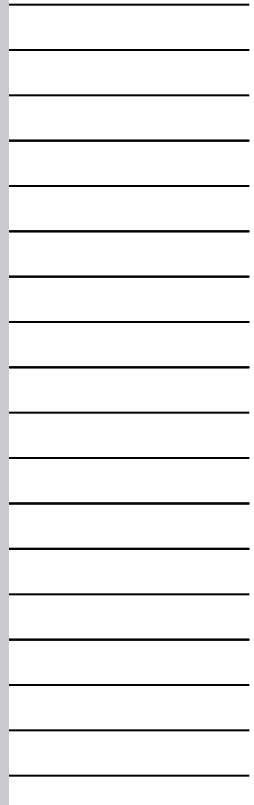
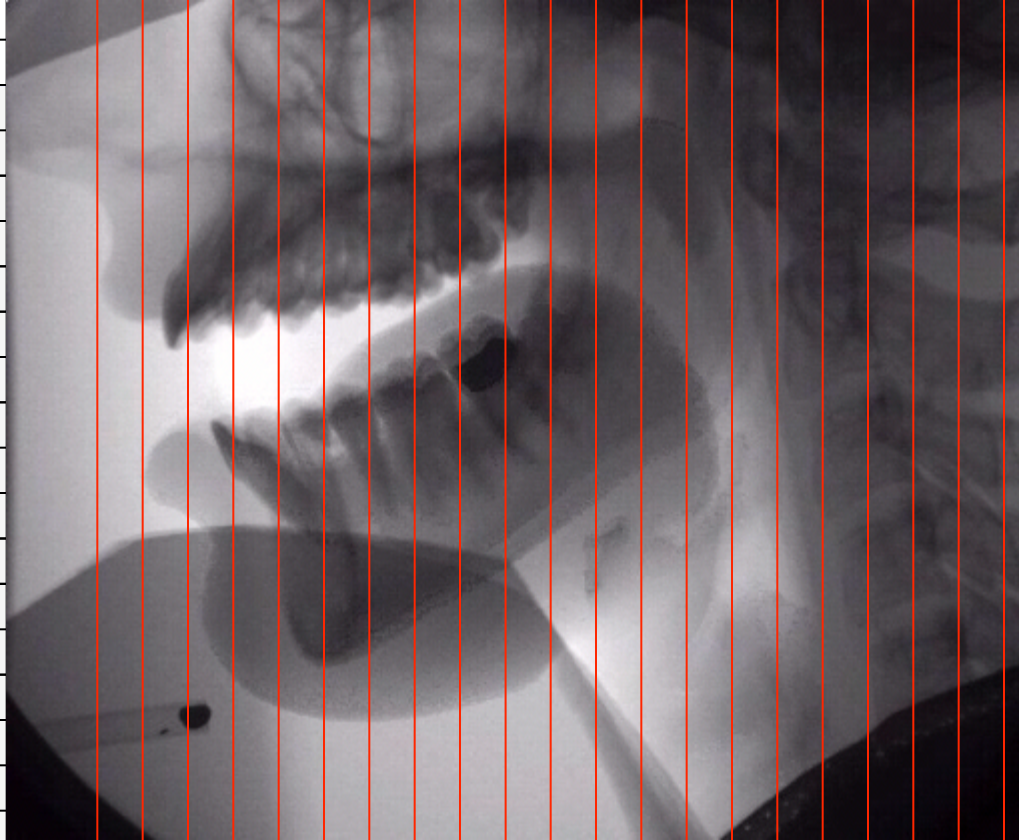
How done?

Demo

vowels

→ consonants

Other uses



**t0**

Mouth already open



**/aku/** model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

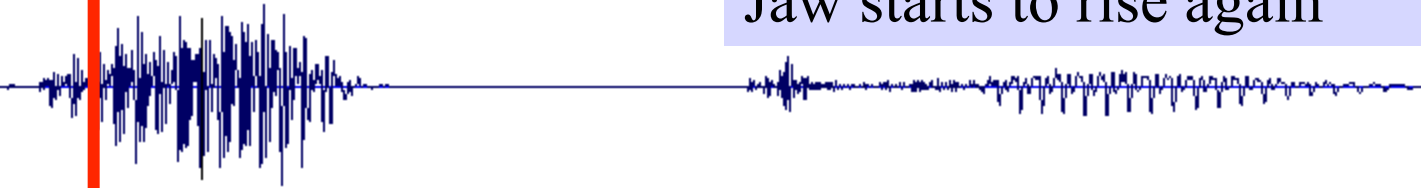
vowels

→ consonants

Other uses

**t1**

Maximum opening for /a/  
Jaw starts to rise again





**/aku/** model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

vowels

→ consonants

Other uses

t2

/k/ closing gesture starts  
Rounding starts



**/aku/** model?

Why modelling?

Which models?

What is a useful model?

How done?

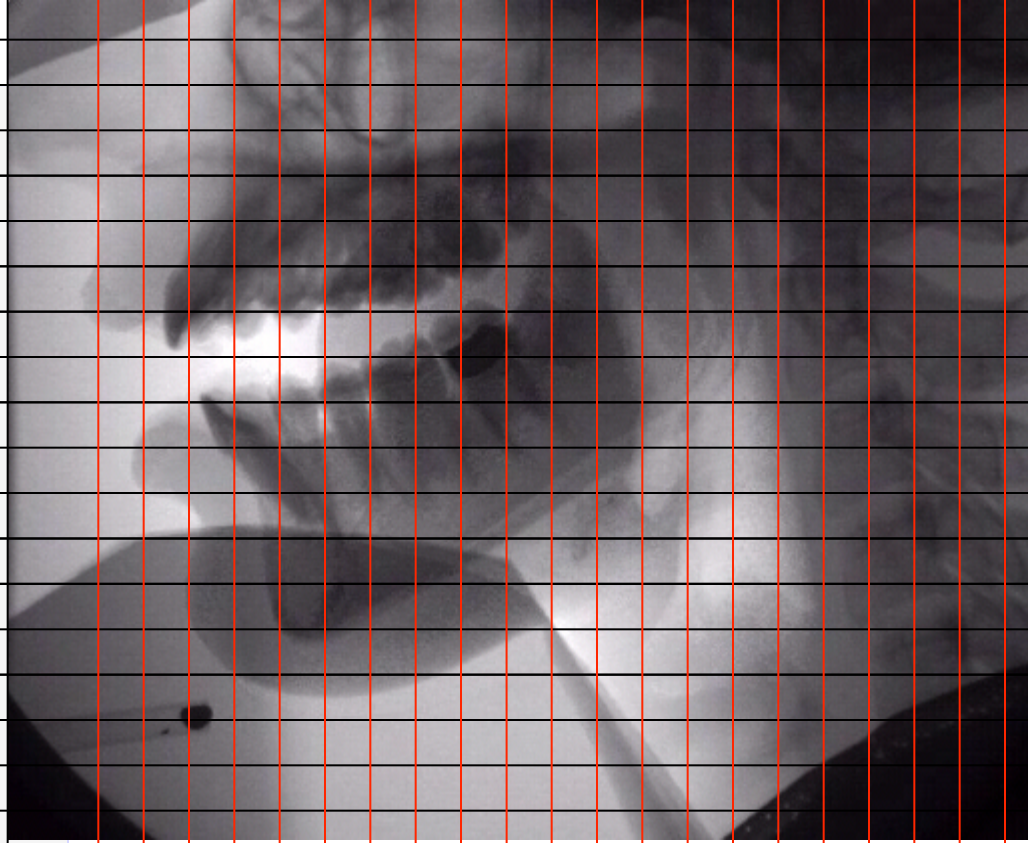
Demo

vowels

→ consonants

Other uses

t3



**/aku/** model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

vowels

→ consonants

Other uses

t4

Closure

Lip continue to round



**/aku/** model?

Why modelling?

Which models?

What is a useful model?

How done?

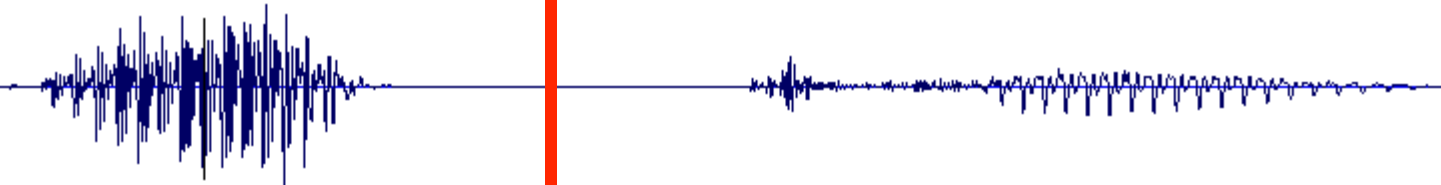
Demo

vowels

→ consonants

Other uses

t5





**/aku/** model?

Why modelling?

Which models?

What is a useful model?

How done?

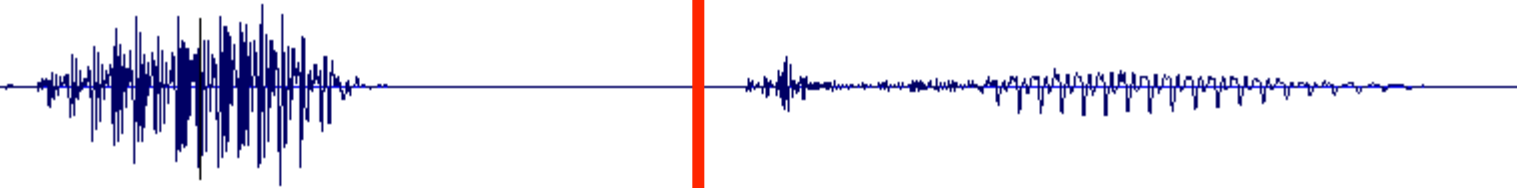
Demo

vowels

→ consonants

Other uses

t6



**/aku/** model?

Why modelling?

Which models?

What is a useful model?

How done?

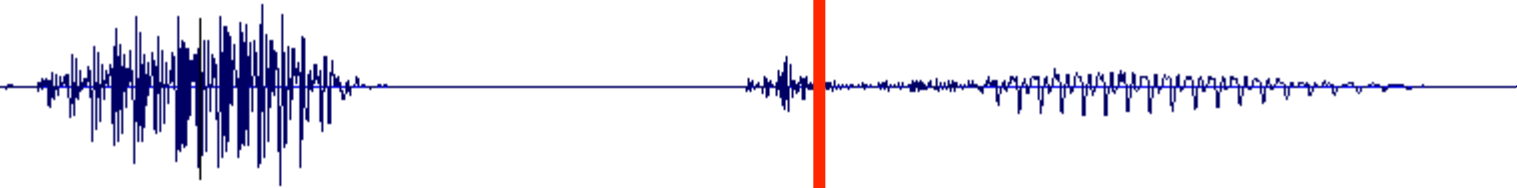
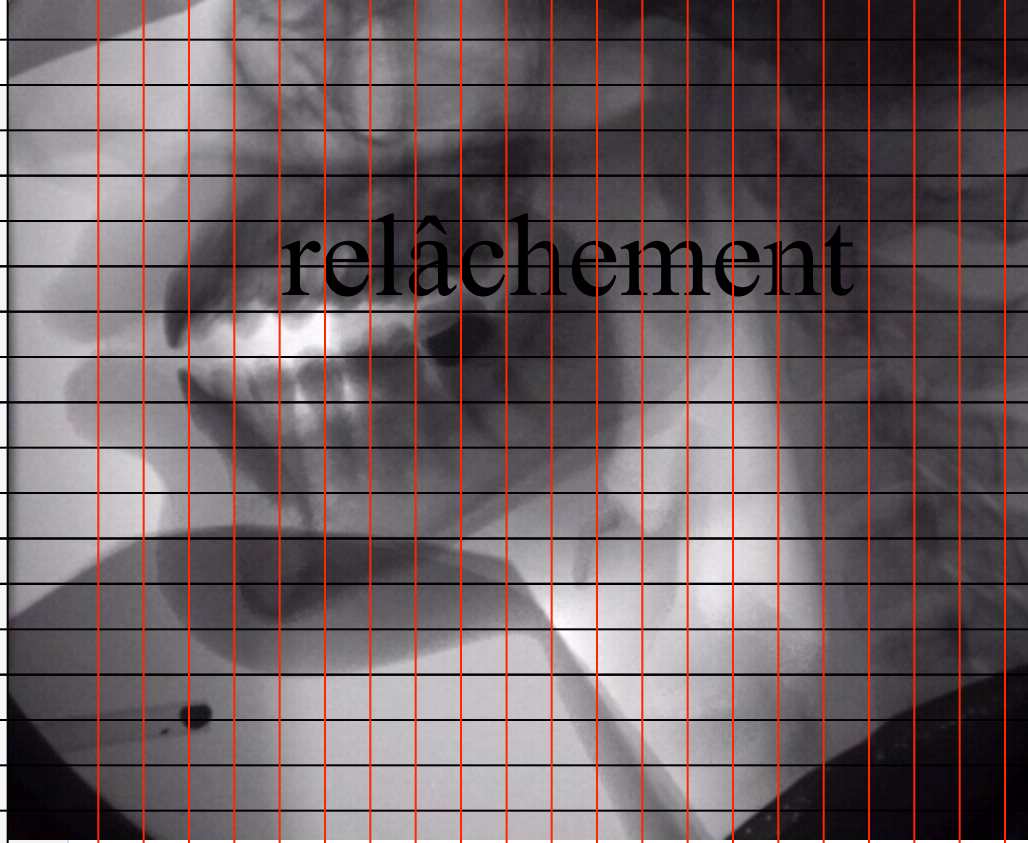
Demo

vowels

→ consonants

Other uses

t7



**/aku/** model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

vowels

→ consonants

Other uses

t8



**/aku/**

odel?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

vowels

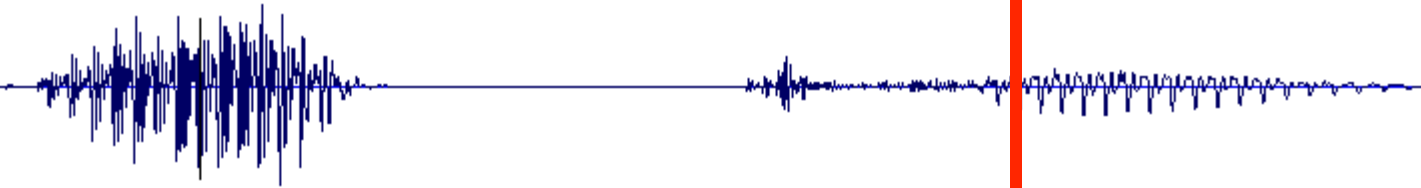
→ consonants

Other uses

ax anatomy

**t9**

Maximum rounding





**/aku/**

odel?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

vowels

→ consonants

Other uses

**t10**

Velum starts to lower  
Derounding starts



**/aku/**

odel?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

vowels



consonants

Other uses

**t11**

Larynx maximally low



**/aku/**

odel?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

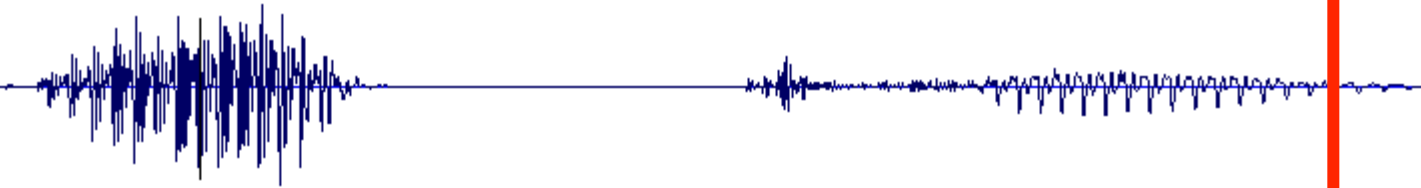
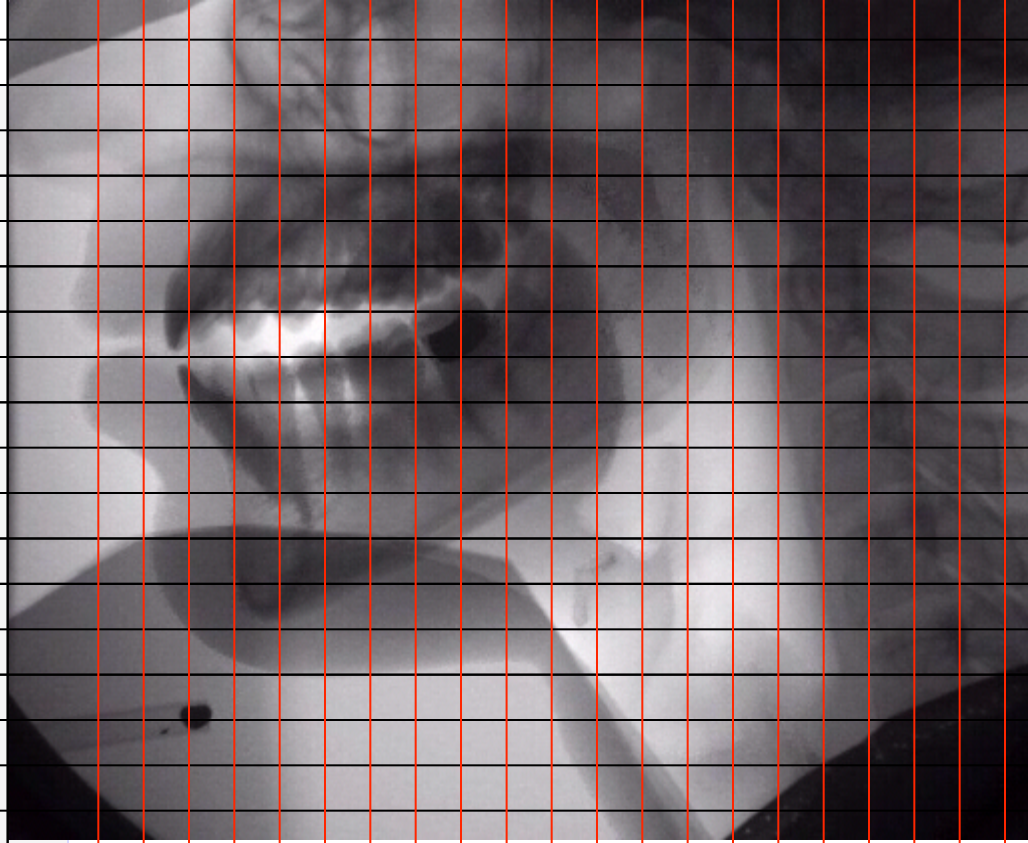
vowels



consonants

Other uses

**t12**



What is a model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

→ vowels  
→ consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

# Different degrees of sophistication

**1) Simple concatenation**

**2) Contextual allophones**

**Velar**

**3) Anticipation and carry over phenomena**

**4) Suprasegmental influences**

- Position in the syllable (nasal in preparation, masking)
- Position in the word or relative to stress (lenition and fortition) seen





# The points



# 1) Pedagogical point of view

*Draw the link between*

*-articulatory data,*

*-acoustic data*

*-and perception*

*>> Understanding for linguists*

*the basics of*

*Acoustic Theory of Speech production*

Ps, Pio

1) DATA

What is a model?

production

acoustics

perception

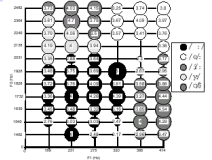
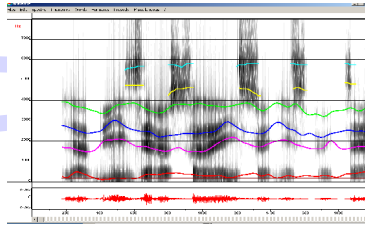


Figure 4: Swedish listeners' identification and goodness ratings.

What is a model?

How does it work?

Demo

phonetician

Other uses

speaker adaptation  
speaker adaptation  
singing formant  
inversion  
etc.

linguist

Language teacher

2) Coherent teaching  
Research tool

engineer

Speech therapist

Medical students

Modelisation



**What is a model?**

**Why modelling?**

**Which models?**

**What is a useful  
model?**

**How done?**

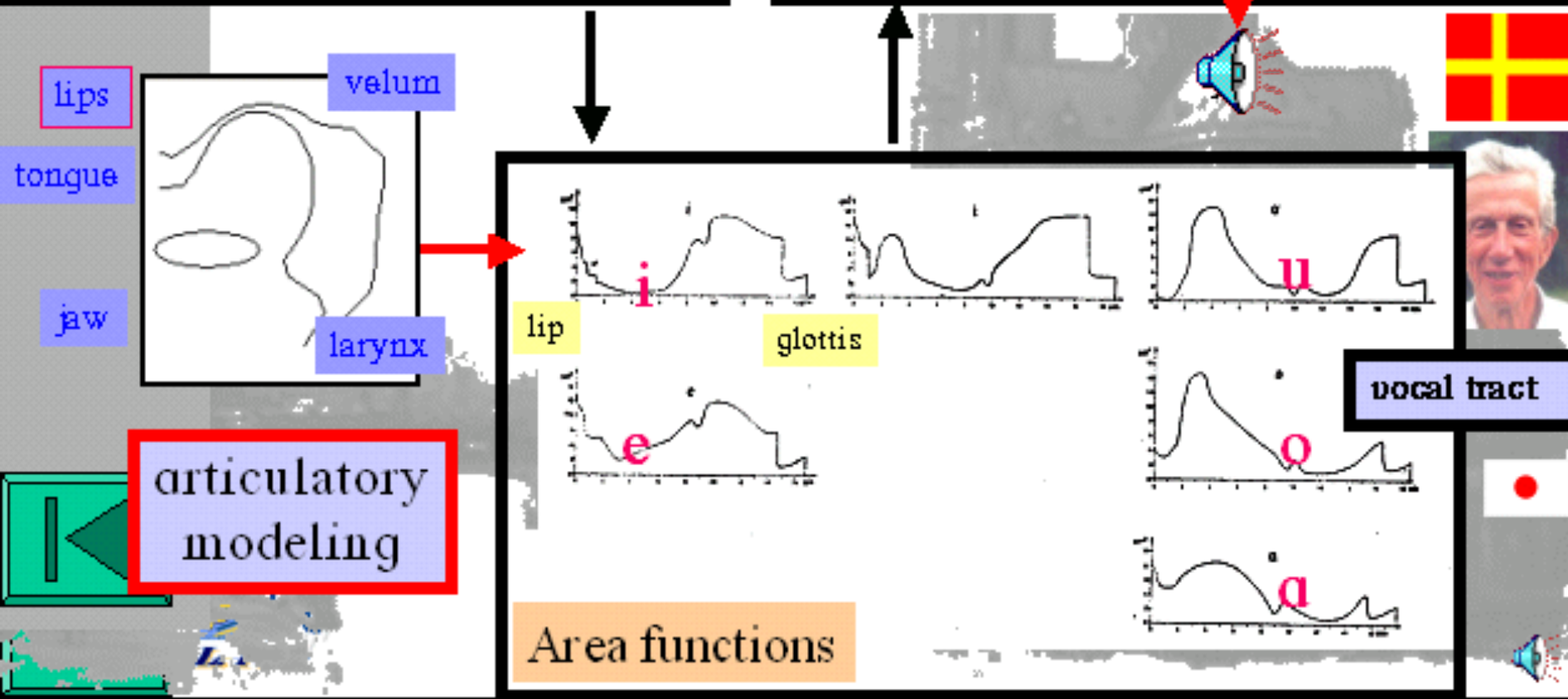
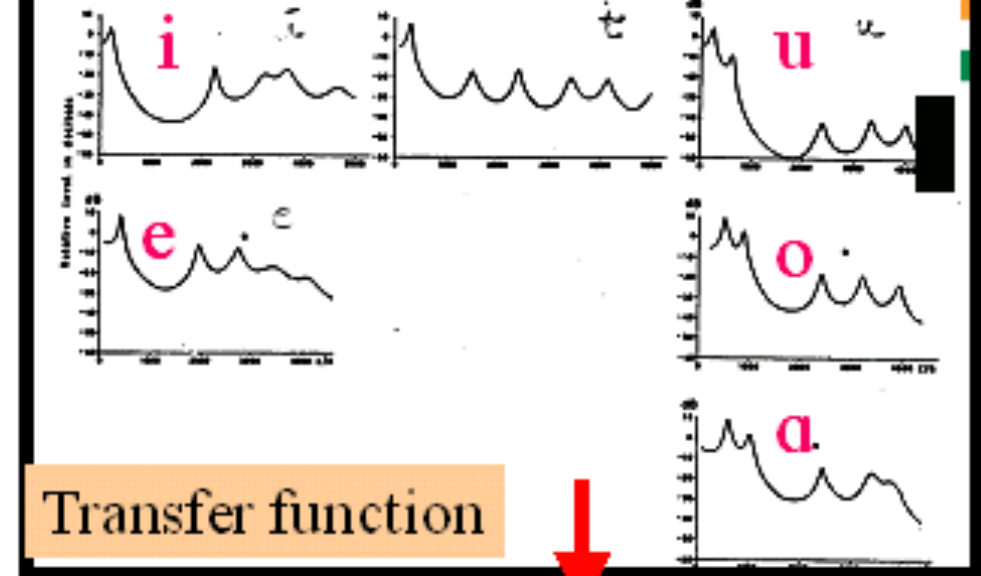
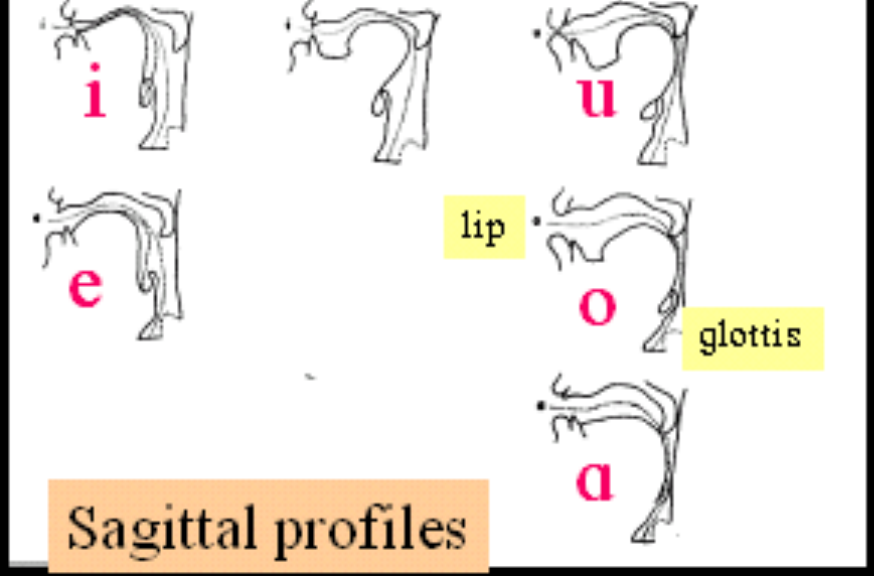
**Demo**

**vowels  
consonants**

**Other uses**

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;





## 2) For what concerns the vowels

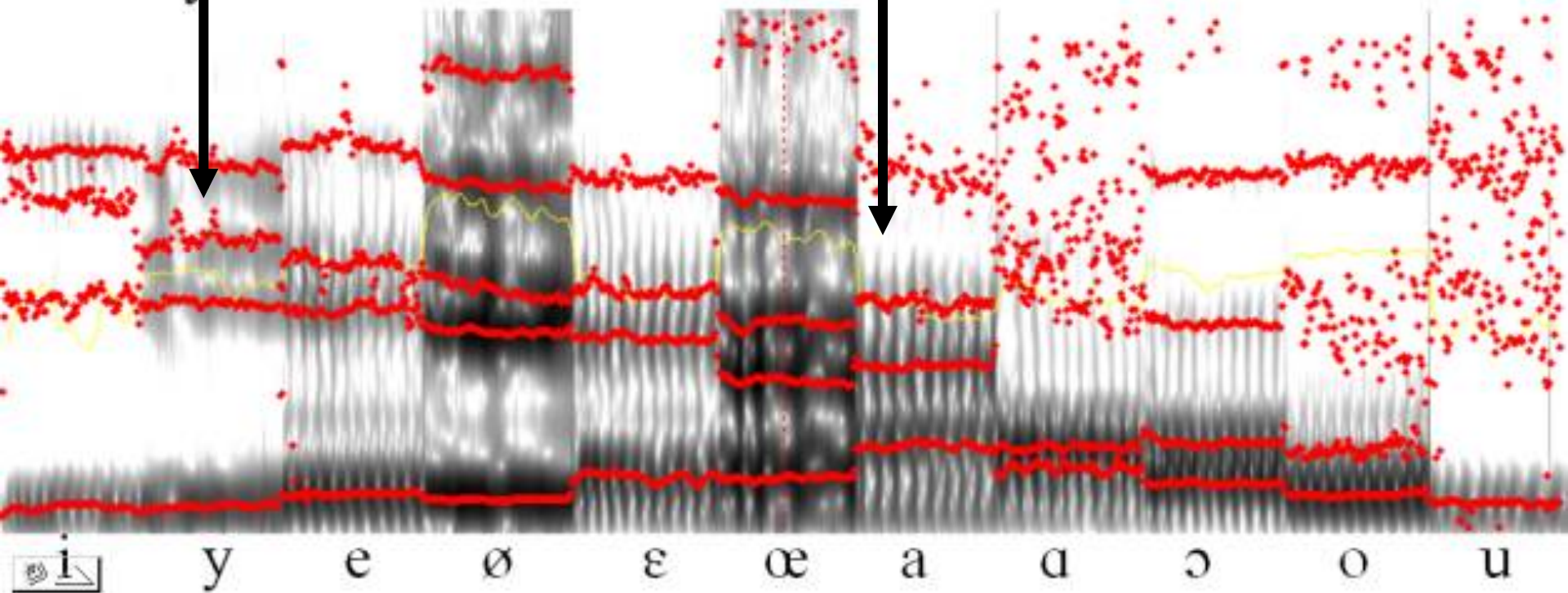
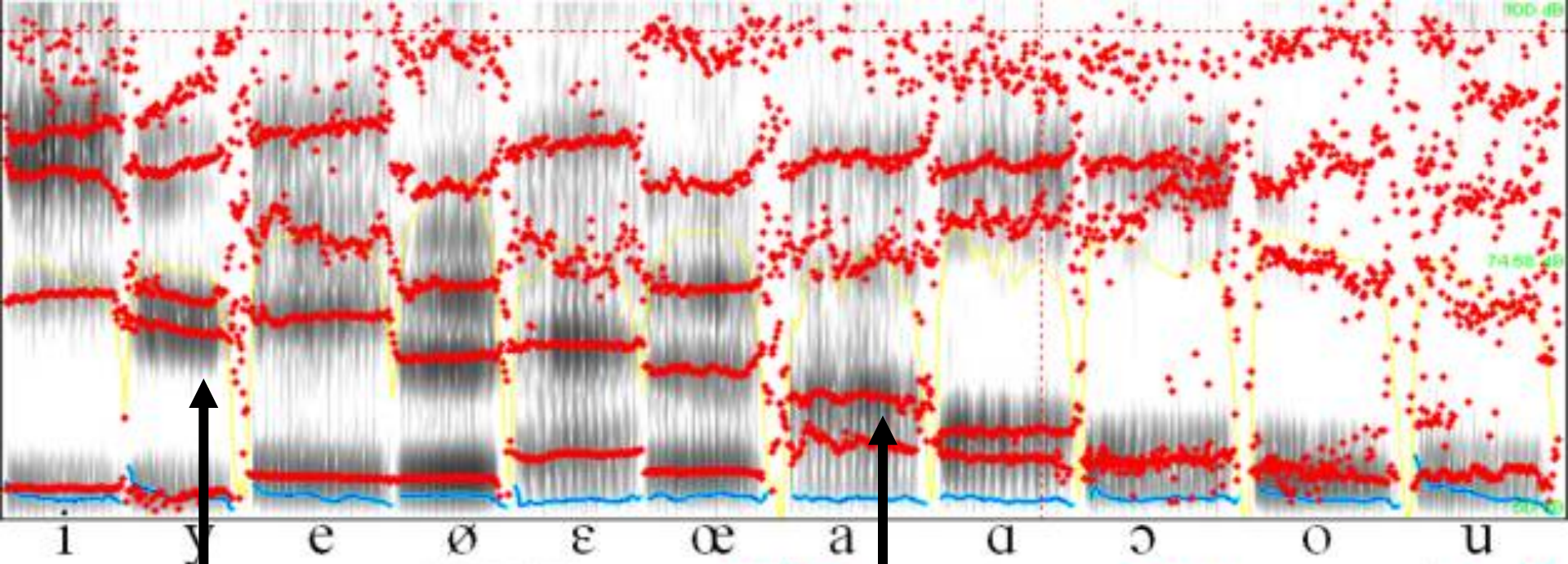
*a) Need for reference vowels*

*Cardinal vowels*

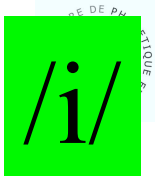
*Sensibility to formant amplitude and bandwidth*

*AM could help*





# observations



## articulatory

## acoustic

## perceptual

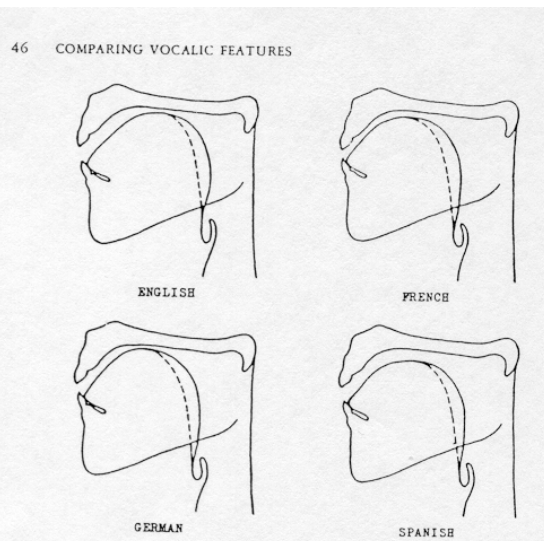


Figure 1. An example of what motion-picture X-ray can show which still X-ray cannot: the characteristic articulatory profiles of the vowel /i/ in English, German and Spanish, obtained by cineradiography, averaging several subjects in each language. Closure of the jaws (at the teeth) is narrower in French and German than in English and Spanish. Tongue constriction at the palate is narrower in English and German than in French and Spanish. The front cavity is characterized in English and German by a conical shape, in French and Spanish by a cylindrical shape – tongue parallel to palate. The back cavity is larger in French and German than in English and Spanish.

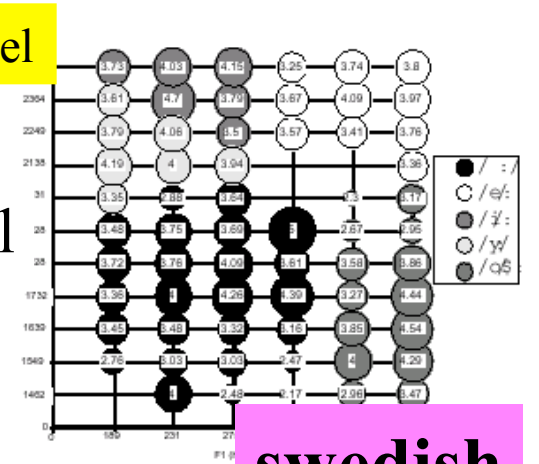
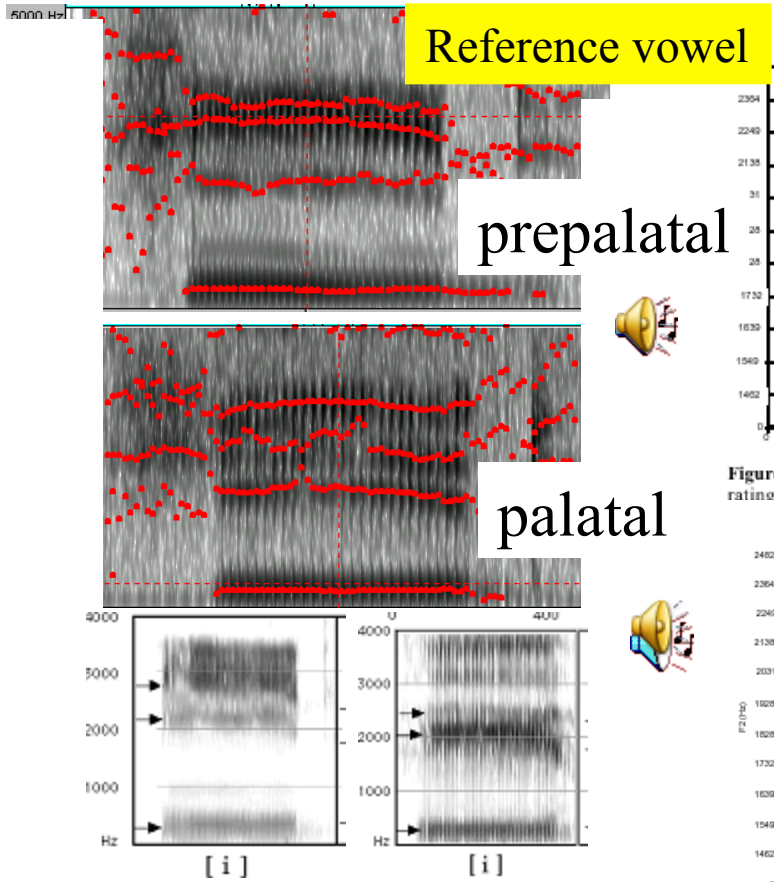


Figure 4: Swedish listener ratings.

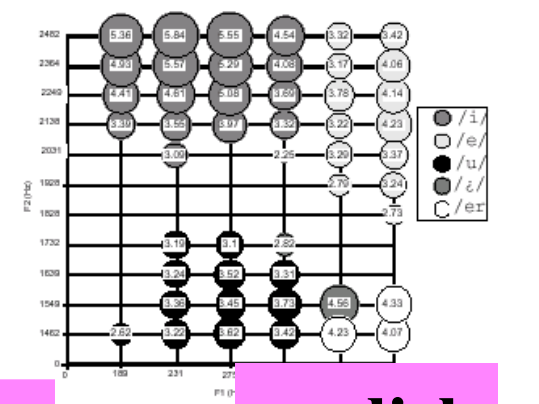


Figure 5: American English listener ratings.

1) Helps to represent in a comprehensive way the language-specific prototypes

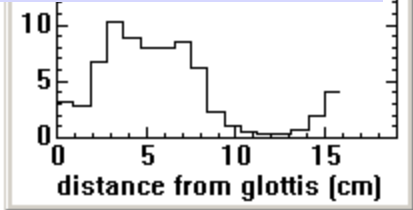
swedish

english

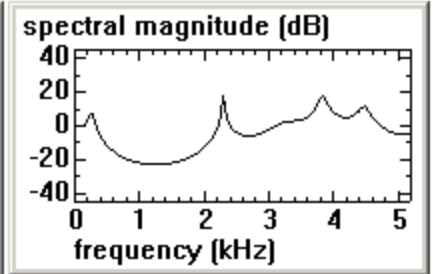
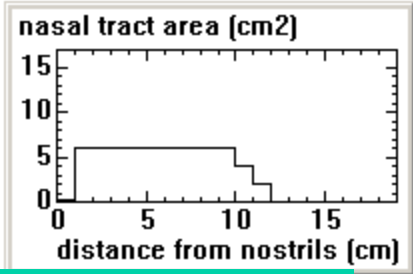
Willerman and Kuhl



# articulatory

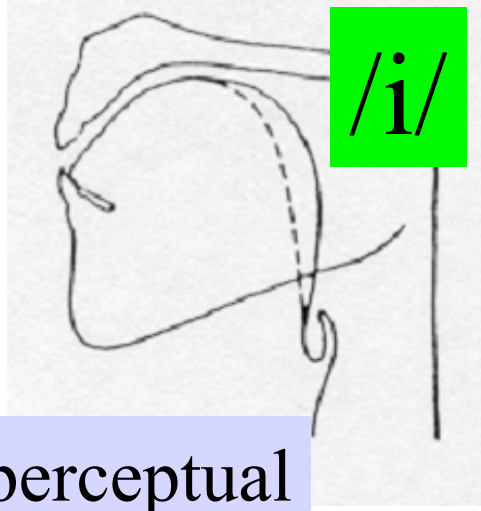


number of sections=17	<input type="checkbox"/>
nasal branch section=9	<input type="checkbox"/>
nasal coupling (cm2)=0	<input type="checkbox"/>
jaw=0.5	<input type="checkbox"/>
tongue=-2	<input type="checkbox"/>
shape=1	<input type="checkbox"/>
apex=-2	<input type="checkbox"/>
lip_ht=1	<input type="checkbox"/>
lip_pr=-1	<input type="checkbox"/>
larynx=0	<input type="checkbox"/>



Frq(Hz)	Bw(Hz)	A(dB)
F1	248	79
F2	2290	29
F3	3838	109
F4	4488	168

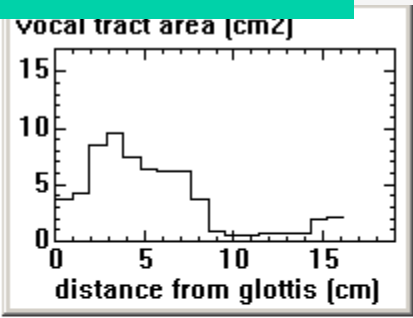
# acoustic



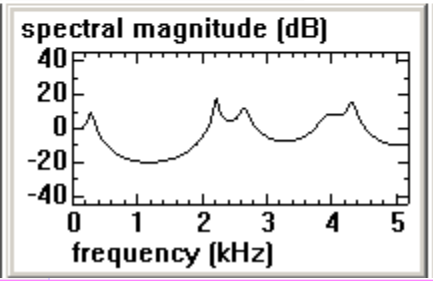
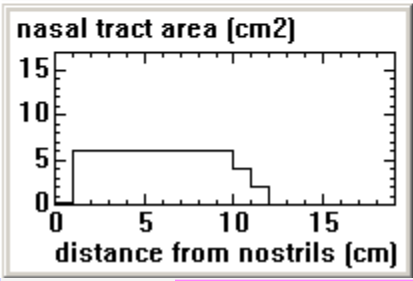
# perceptual

F1 (↑F3. F4)

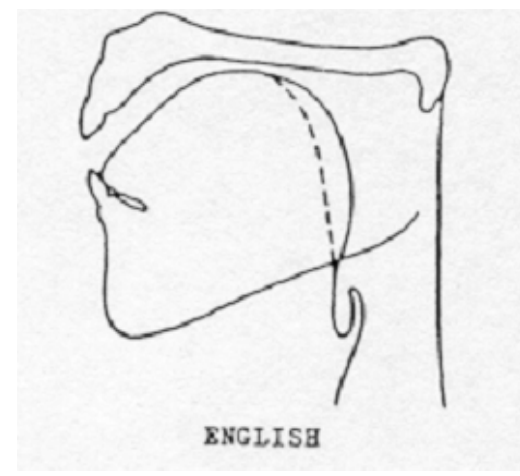
# simulation



number of sections=17	<input type="checkbox"/>
nasal branch section=9	<input type="checkbox"/>
nasal coupling (cm2)=0	<input type="checkbox"/>
jaw=4.4	<input type="checkbox"/>
tongue=0.4	<input type="checkbox"/>
shape=1.5	<input type="checkbox"/>
apex=-2.8	<input type="checkbox"/>
lip_ht=1	<input type="checkbox"/>
lip_pr=-1	<input type="checkbox"/>
larynx=0	<input type="checkbox"/>



Frq(Hz)	Bw(Hz)	A(dB)
F1	279	73
F2	2229	48
F3	2662	118
F4	3962	203
F5	4333	92



F1 ↑ F2

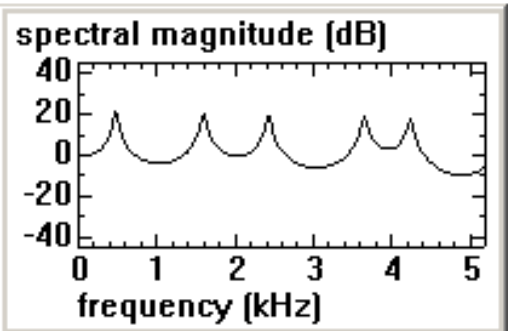
1) Helps to represent in a comprehensive way the language-specific prototypes



# neutral



s=17	<input type="checkbox"/>
ion=9	<input type="checkbox"/>
n2)=0	<input type="checkbox"/>
jaw=-1	<input type="checkbox"/>
lip_ht=0	<input type="checkbox"/>
tongue=-0.5	<input type="checkbox"/>
lip_pr=0	<input type="checkbox"/>
shape=0.5	<input type="checkbox"/>
larynx=0	<input type="checkbox"/>
apex=-2	<input type="checkbox"/>



	Frq(Hz)	Bw(Hz)	A(dB)
F1	464	47	22
F2	1610	64	20
F3	2414	56	20
F4	3652	57	19
F5	4240	58	18

vowels  
consonants

## Other uses

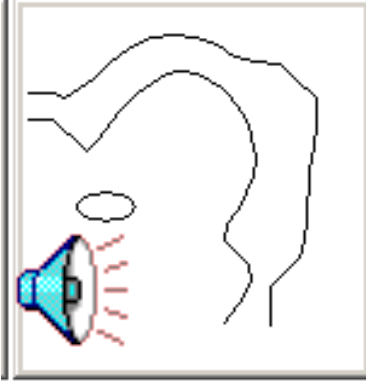
- speaker anatomy
- speaker strategy
- singing formant
- inversion
- etc;



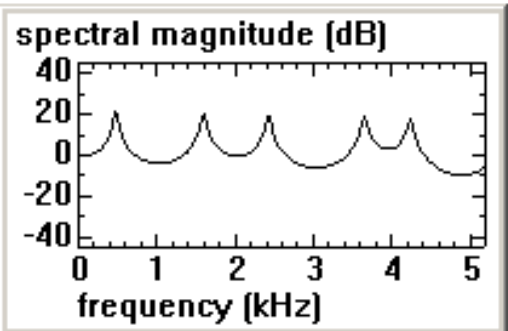


**neutral**

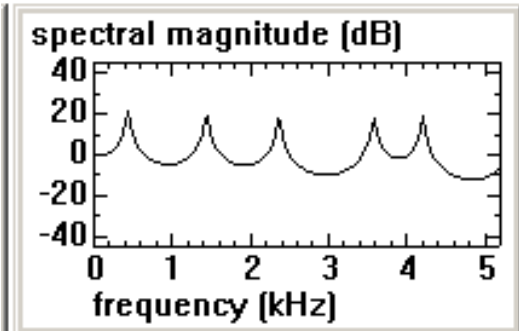
s=17	I
ion=9	I
n2]=0	F
jaw=-1	F
lip_ht=0	F
tongue=-0.5	F
lip_pr=0	F
shape=0.5	F
larynx=0	F
apex=-2	F



number of sections=17	I
nasal branch section=9	I
nasa	F
lip_pr	lip_pr .11
jaw=-1	F
lip_ht=0	F
tongue=-0.5	F
lip_pr	-5 <3< 5
shape=0.5	F
larynx=0	F
apex=-2	F



Frq(Hz)	Bw(Hz)	A(dB)
F1 464	47	22
F2 1610	64	20
F3 2414	56	20
F4 3652	57	19
F5 4240	58	18



Frq(Hz)	Bw(Hz)	A(dB)
F1 433	37	21
F2 1455	46	19
F3 2352	47	18
F4 3590	46	18
F5 4210	34	19

vowels  
consonants

Other uses

- speaker anatomy
- speaker strategy
- singing formant
- inversion
- etc;

( F2 ↓ ↓ ↓ F3 )

ion



**neutral**

number of sections=17  
nasal branch section=9  
nasal coupling (cm2)=0  
jaw=-1  
tongue=-0.5  
shape=0.5  
apex=-2

lip\_ht=0  
lip\_pr=0  
larynx=0

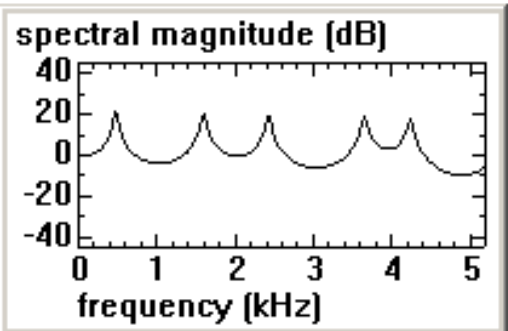
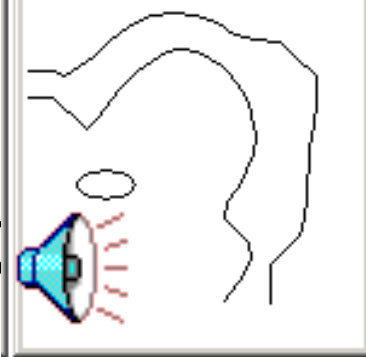


and

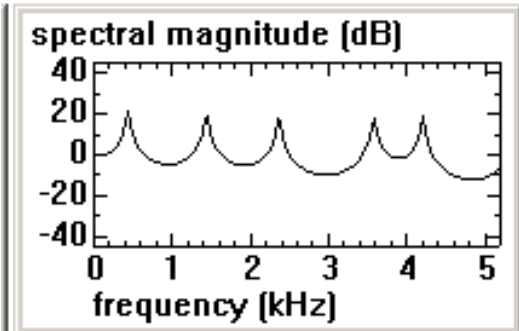
**+ protrusion**

number of sections=17  
nasal branch section=9  
nasal coupling (cm2)=0  
jaw=-1  
tongue=-0.5  
shape=0.5  
apex=-2

lip\_ht=0  
lip\_pr=1.1



Frq(Hz)	Bw(Hz)	A(dB)
F1	464	22
F2	1610	20
F3	2414	20
F4	3652	19
F5	4240	18

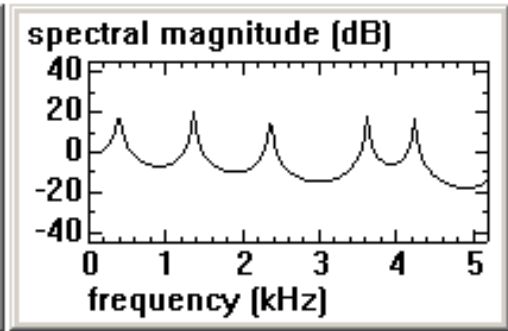


Frq(Hz)	Bw(Hz)	A(dB)
F1	433	21
F2	1455	19
F3	2352	18
F4	3590	18
F5	4210	19

**+ compression**

number of sections=17  
nasal branch section=9  
nasal coupling (cm2)=0  
jaw=-1  
tongue=-0.5  
shape=0.5  
apex=-2

lip\_ht=1.1



Frq(Hz)	Bw(Hz)	A(dB)
F1	402	16
F2	1362	21
F3	2352	14
F4	3621	18
F5	4240	17

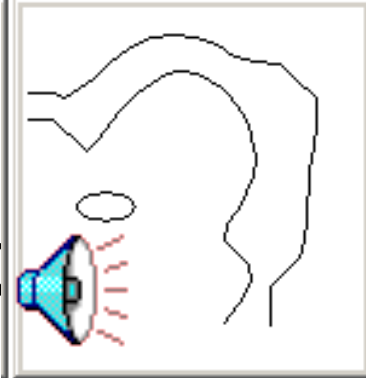
odelisation

neutral



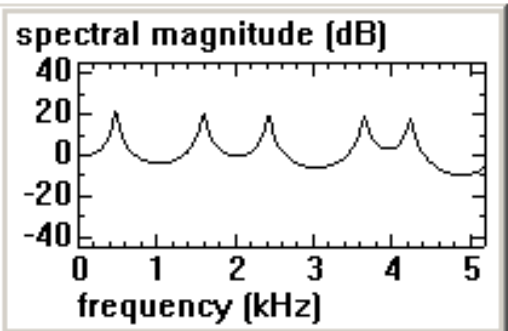
number of sections=17	<input type="checkbox"/>
nasal branch section=9	<input type="checkbox"/>
nasal coupling (cm <sup>2</sup> )=0	<input type="checkbox"/>
jaw=-1	<input type="checkbox"/>
tongue=-0.5	<input type="checkbox"/>
shape=0.5	<input type="checkbox"/>
apex=-2	<input type="checkbox"/>
lip_ht=0	<input type="checkbox"/>
lip_pr=0	<input type="checkbox"/>
larynx=0	<input type="checkbox"/>

and



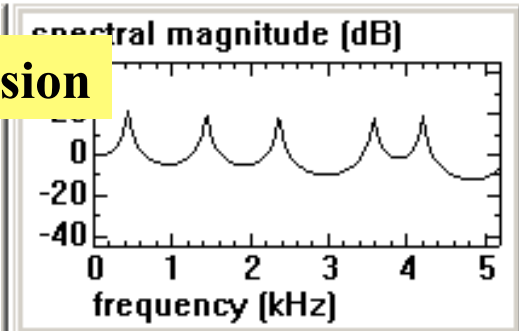
number of sections=17	<input type="checkbox"/>
nasal branch section=9	<input type="checkbox"/>
nasal coupling (cm <sup>2</sup> )=0	<input type="checkbox"/>
jaw=-1	<input type="checkbox"/>
tongue=-0.5	<input type="checkbox"/>
shape=0.5	<input type="checkbox"/>
apex=-2	<input type="checkbox"/>
lip_ht=0	<input type="checkbox"/>
lip_pr=11	<input type="checkbox"/>
larynx=0	<input type="checkbox"/>

+ protrusion



Frq(Hz)	Bw(Hz)	A(dB)
F1 464	4	20
F2 1610	64	20
F3 2414	56	20
F4 3652	57	19
F5 4240	58	18

+ compression



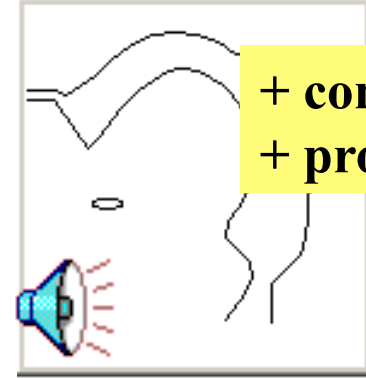
Frq(Hz)	Bw(Hz)	A(dB)
F1 433	37	21
F2 1455	46	19
F3 2352	47	18
F4 3590	46	18
F5 4210	34	19

LIP

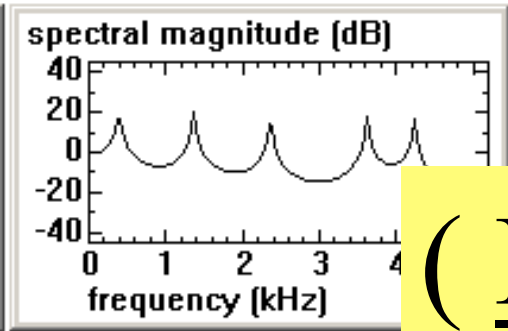


number of sections=17	<input type="checkbox"/>
nasal branch section=9	<input type="checkbox"/>
nasal coupling (cm <sup>2</sup> )=0	<input type="checkbox"/>
jaw=-1	<input type="checkbox"/>
tongue=-0.5	<input type="checkbox"/>
shape=0.5	<input type="checkbox"/>
apex=-2	<input type="checkbox"/>
lip_ht=11	<input type="checkbox"/>
lip_pr=3	<input type="checkbox"/>
larynx=0	<input type="checkbox"/>

+ compression  
+ protrusion

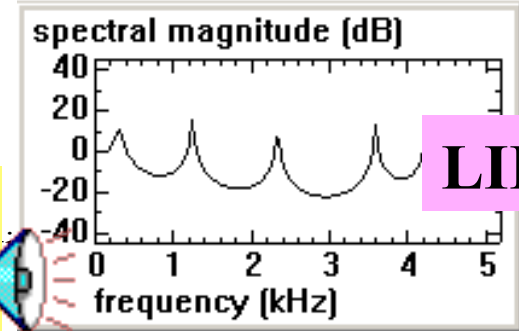


number of sections=17	<input type="checkbox"/>
nasal branch section=9	<input type="checkbox"/>
nasal coupling (cm <sup>2</sup> )=0	<input type="checkbox"/>
jaw=-1	<input type="checkbox"/>
tongue=-0.5	<input type="checkbox"/>
shape=0.5	<input type="checkbox"/>
apex=-2	<input type="checkbox"/>
lip_ht=11	<input type="checkbox"/>
lip_pr=3	<input type="checkbox"/>
larynx=0	<input type="checkbox"/>



Frq(Hz)	Bw(Hz)	A(dB)
F1 402	59	16
F2 1362	23	21
F3 2352	42	14

( F2 ↓ ↓ F3 )



Frq(Hz)	Bw(Hz)	A(dB)
F1 310	68	11
F2 1238	21	16
F3 2352	20	11
F4 3652	57	19
F5 4240	58	18

LIP + TONGUE

neutral



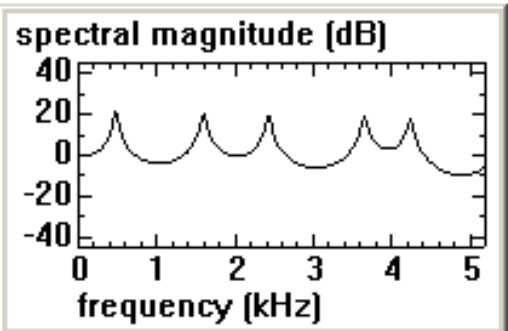
number of sections=17	<input type="checkbox"/>
nasal branch section=9	<input type="checkbox"/>
nasal coupling (cm2)=0	<input type="checkbox"/>
jaw=-1	<input type="checkbox"/>
tongue=-0.5	<input type="checkbox"/>
shape=0.5	<input type="checkbox"/>
apex=-2	<input type="checkbox"/>
lip_ht=0	<input type="checkbox"/>
lip_pr=0	<input type="checkbox"/>
larynx=0	<input type="checkbox"/>

and



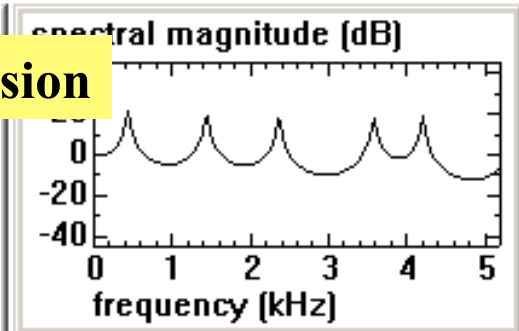
number of sections=17	<input type="checkbox"/>
nasal branch section=9	<input type="checkbox"/>
nasal coupling (cm2)=0	<input type="checkbox"/>
jaw=-1	<input type="checkbox"/>
tongue=-0.5	<input type="checkbox"/>
shape=0.5	<input type="checkbox"/>
apex=-2	<input type="checkbox"/>
lip_ht=0	<input type="checkbox"/>
lip_pr=11	<input type="checkbox"/>
larynx=0	<input type="checkbox"/>

+ protrusion



Frq(Hz)	Bw(Hz)	A(dB)
F1 464	4	20
F2 1610	64	20
F3 2414	56	20
F4 3652	57	19
F5 4240	58	18

+ compression



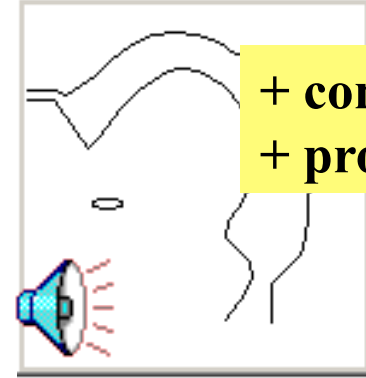
Frq(Hz)	Bw(Hz)	A(dB)
F1 433	37	21
F2 1455	46	19
F3 2352	47	18
F4 3590	46	18
F5 4210	34	19

LIP

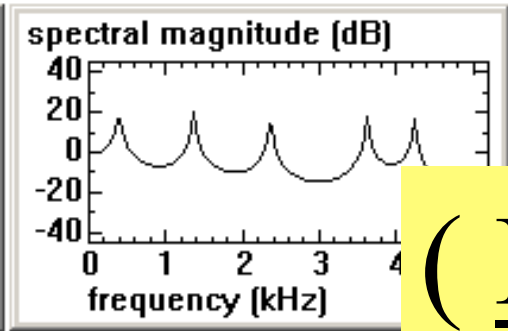


number of sections=17	<input type="checkbox"/>
nasal branch section=9	<input type="checkbox"/>
nasal coupling (cm2)=0	<input type="checkbox"/>
jaw=-1	<input type="checkbox"/>
tongue=-0.5	<input type="checkbox"/>
shape=0.5	<input type="checkbox"/>
apex=-2	<input type="checkbox"/>
lip_ht=11	<input type="checkbox"/>
lip_pr=3	<input type="checkbox"/>
larynx=0	<input type="checkbox"/>

+ compression  
+ protrusion

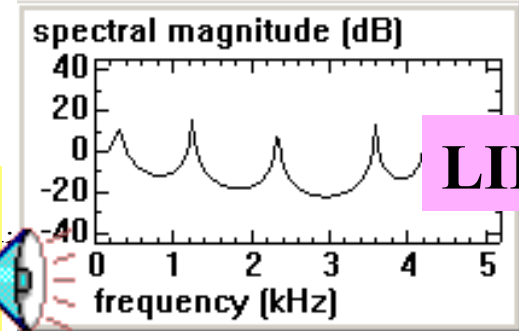


number of sections=17	<input type="checkbox"/>
nasal branch section=9	<input type="checkbox"/>
nasal coupling (cm2)=0	<input type="checkbox"/>
jaw=-1	<input type="checkbox"/>
tongue=-0.5	<input type="checkbox"/>
shape=0.5	<input type="checkbox"/>
apex=-2	<input type="checkbox"/>
lip_ht=11	<input type="checkbox"/>
lip_pr=3	<input type="checkbox"/>
larynx=0	<input type="checkbox"/>



Frq(Hz)	Bw(Hz)	A(dB)
F1 402	59	16
F2 1362	23	21
F3 2352	42	14

( F2 ↓ ↓ F3 )



Frq(Hz)	Bw(Hz)	A(dB)
F1 310	68	11
F2 1238	21	16
F3 2352	42	14
F4 3590	46	18
F5 4240	20	11

LIP + TONGUE

<b>Location of the constriction</b>			
<b>Lowest possible formant</b>		<b>Highest possible formant</b>	
↓F1	anterior part of the VT	↑F1	posterior part of the VT
↓F2	velar region + <i>lip rounding</i>	↑F2	mid-palatal region + glottal region
↓F3	pharyngeal region + <i>bunching of the tongue,</i> <u><i>retroflexion</i></u> + <i>lip rounding and lip protrusion</i>	↑F3	apical and <u>prepalatal</u> regions + lip spreading + glottal region ( <i>larynx lowering</i> )

Table 3: Tongue constriction position, lip configuration, tongue shape and F-pattern values: a summary.

## 2) For what concerns the vowels

*a) Need for references vowels*

*b) The rôle of anatomical details*

*- the laryngeal cavity for the vowels*

*- rôle of pharyngeal volume, etc.*



# Role of anatomical details

What is a model?

Why modelling?

Which models?

What is a useful model?

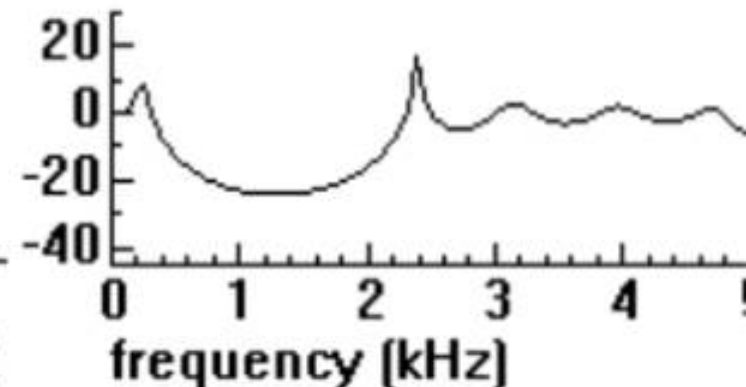
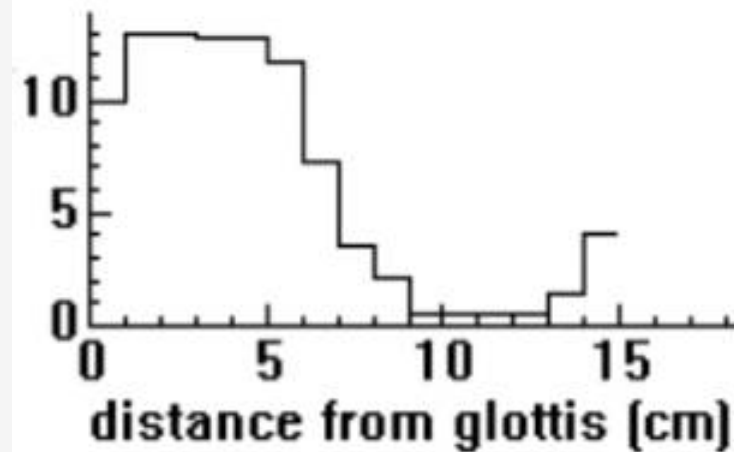
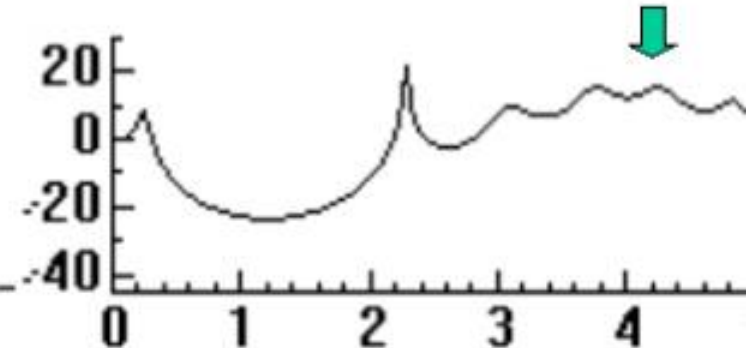
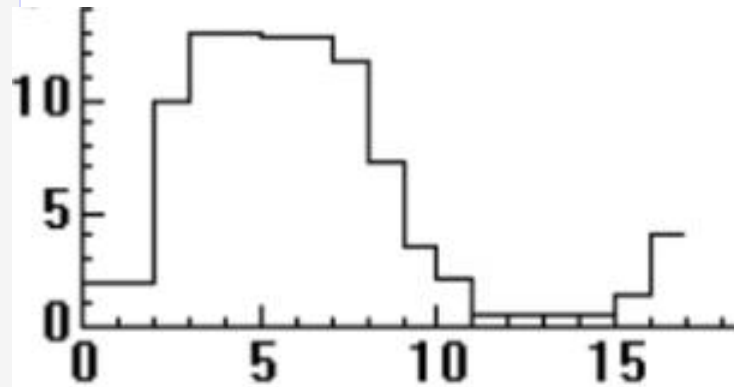
How done?

Demo

vowels  
consonants

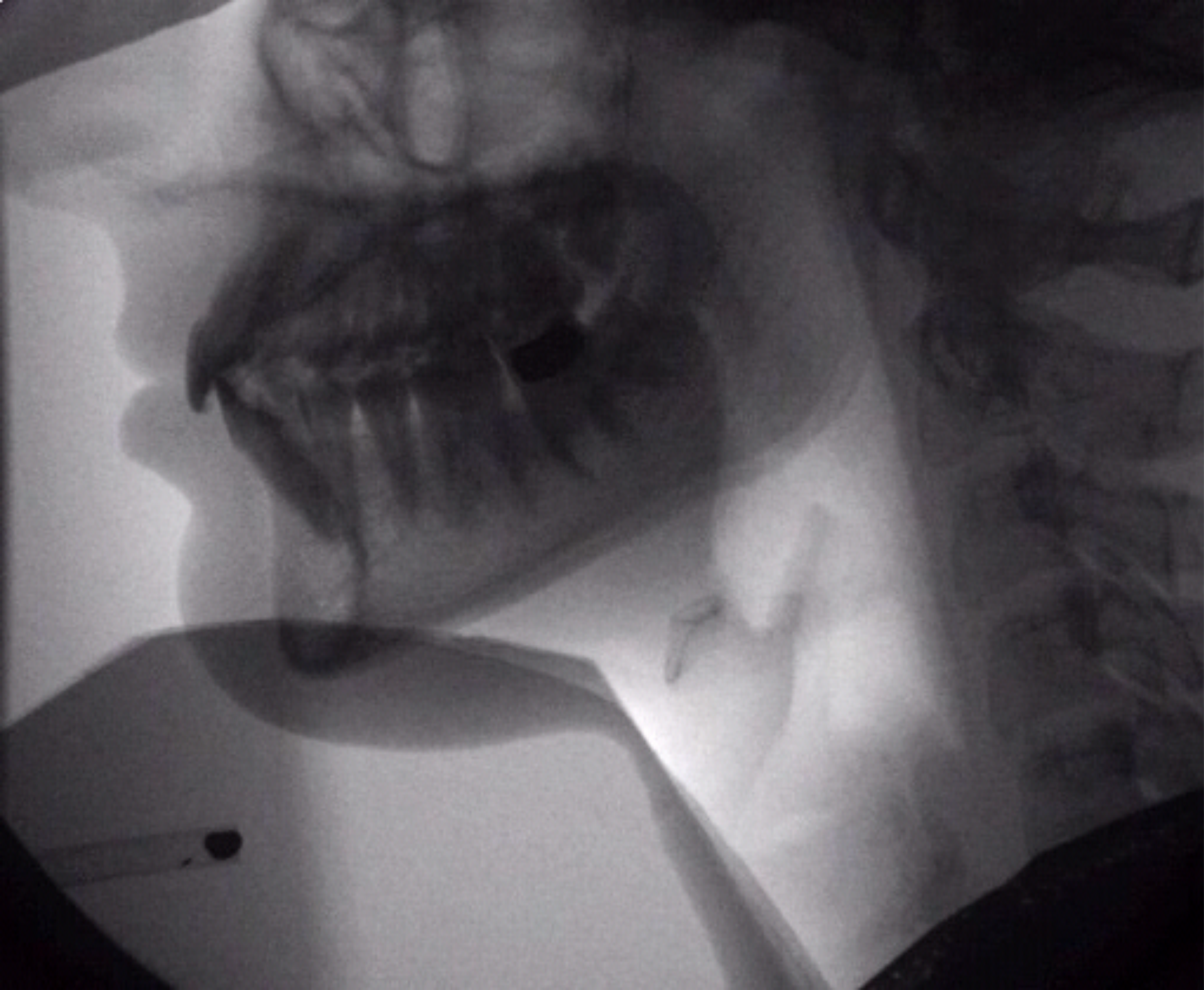
Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;



Modélisation





**2) For what concerns the vowels**

*a) Need for references vowels*

*b) The rôle of anatomical details*

*c) The acoustic properties largely  
Prevail over articulation*

# Perception prevails over articulation

What is a model?

Why modelling?

Which models?

What is a useful model?

How done?

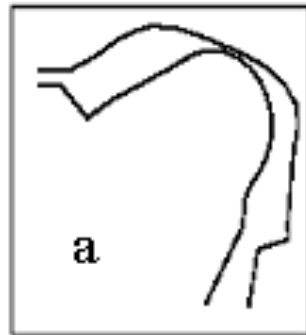
Demo

vowels  
consonants

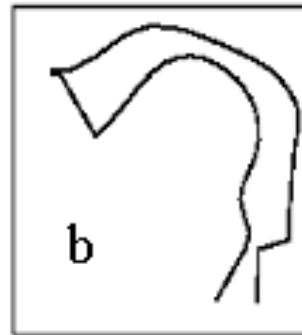
Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

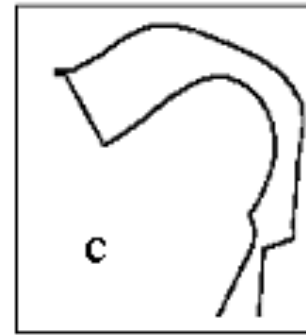
About the same sound (except the focal and extreme sounds can be obtained by Fairly different VT configuration



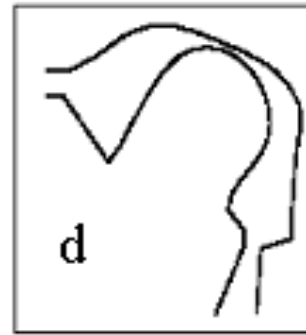
335 904 2150



341 986 2113



346 953 2118



352 949 2112



## 2) For what concerns the vowels

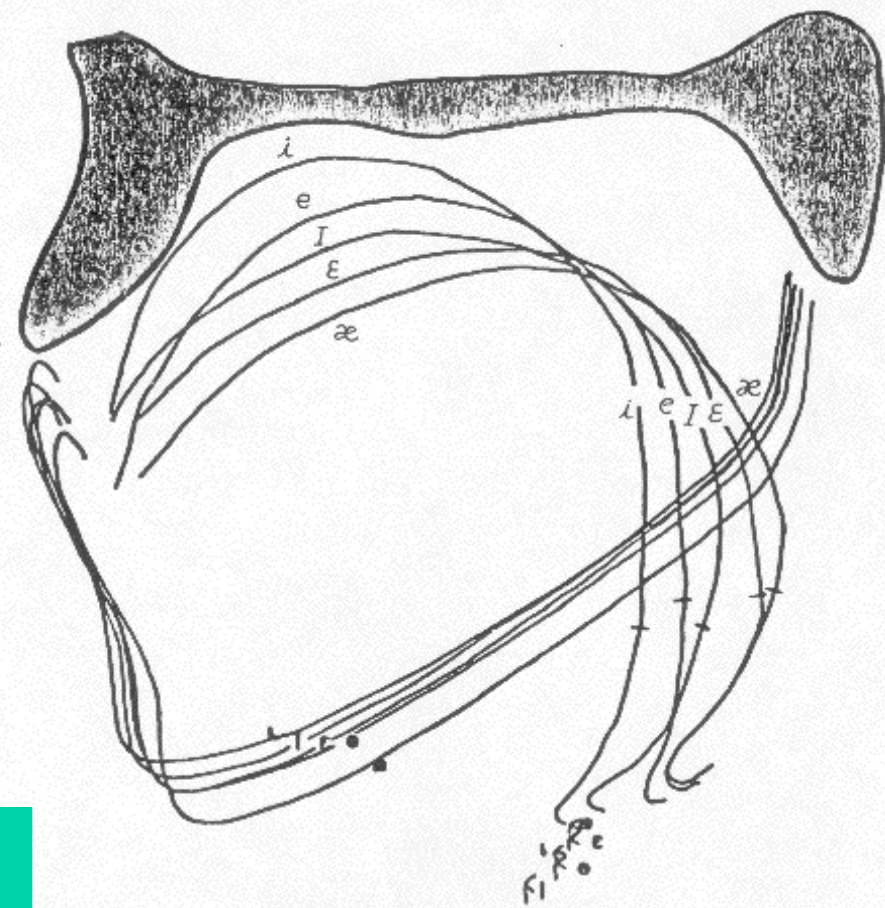
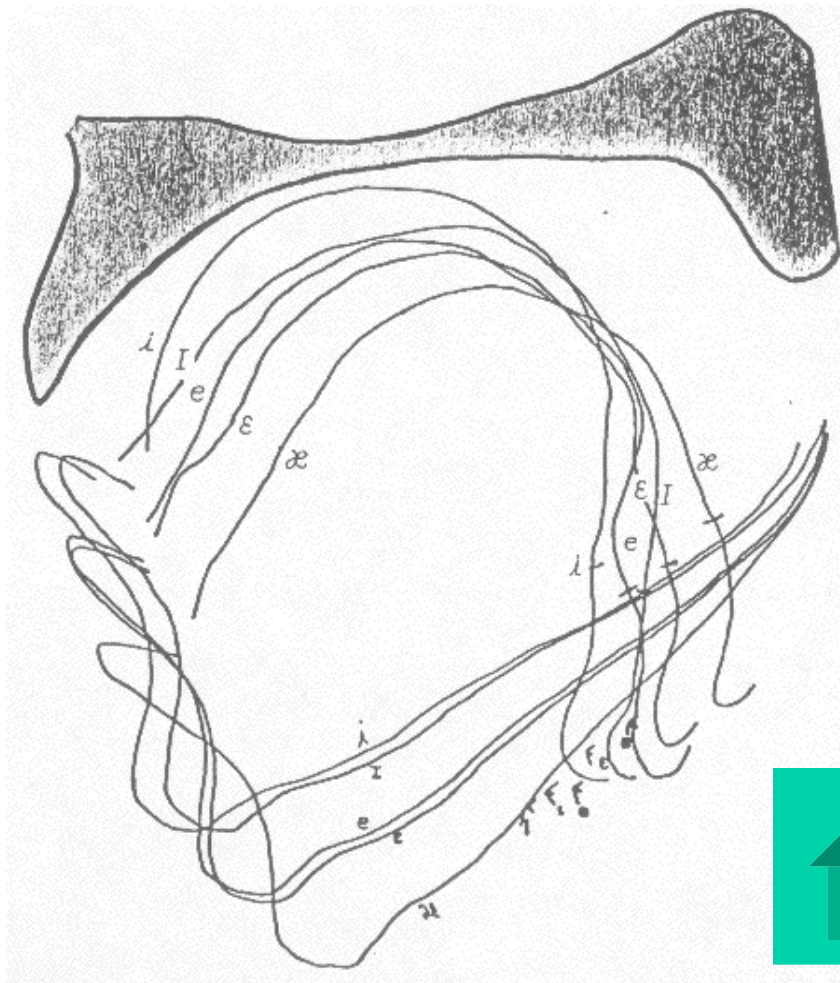
*a) Need for reference vowels*

*b) The rôle of anatomical details*

*c) The acoustic properties largely  
Prevail over articulation*

*d) Different strategies can be modelled*





inversion

# Speaker different strategies



## **2) For what concerns the consonants**

*a) Simple concatenation works better than expected*



## 2) For what concerns the consonants

*a) Simple concatenation work than expected*

*b) The rôle of anatomical details: sublingual cavity*

**What is a model?**

**Why modelling?**

**Which models?**

**What is a useful model?**

**How done?**

**Demo**

vowels  
consonants

**Other uses**

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

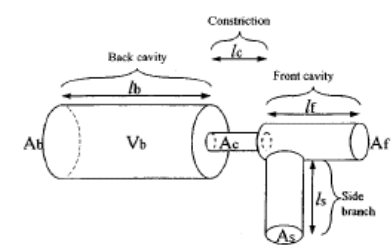
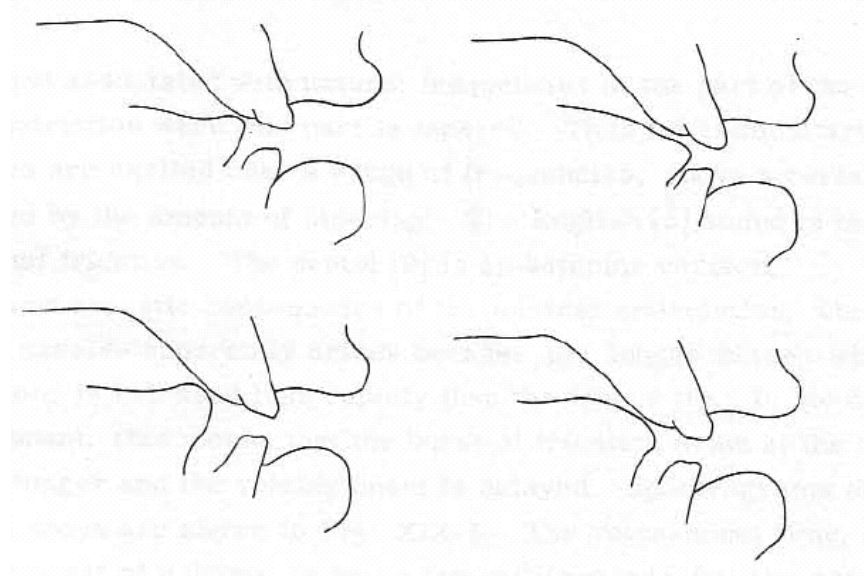
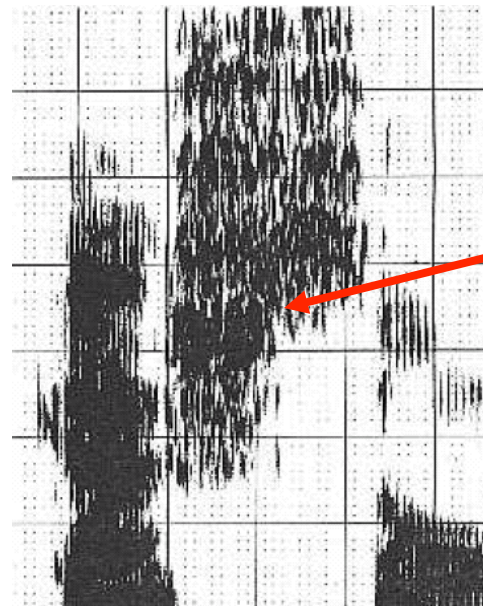


FIG. 5. The tube configuration simulated in the "weakly coupled" models.



sublingual cavity  
Necessary  
To recreate existing  
Acoustic discontinuity

Modelisation



## 2) For what concerns the consonants

*a) Simple concatenation work than expected*

*b) The rôle of anatomical details: sublingual cavity*

*c) The acoustic properties largely  
Prevail over articulation: the velar*

## 2) For what concerns the consonants

*a) Simple concatenation work than expected*

*b) The rôle of anatomical details: sublingual cavity*

*c) The acoustic properties largely  
Prevail over articulation: the velars*

*d) Exotic consonants may be easy to model*

What is a model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

vowels  
consonants

Other uses

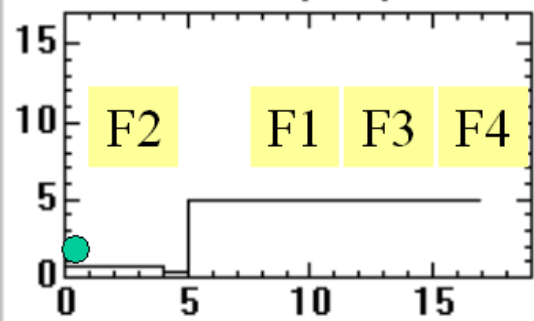
speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

## Non French consonants

- Start from articulatory model for the flanking vowels
- Tube or area function for the consonant
- Examples: posterior arabic consonants (Yeou' s thesis)



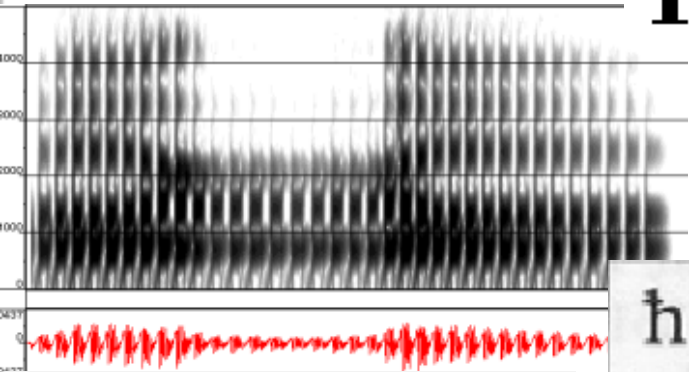
vocal tract area (cm<sup>2</sup>)



95%, 69\*

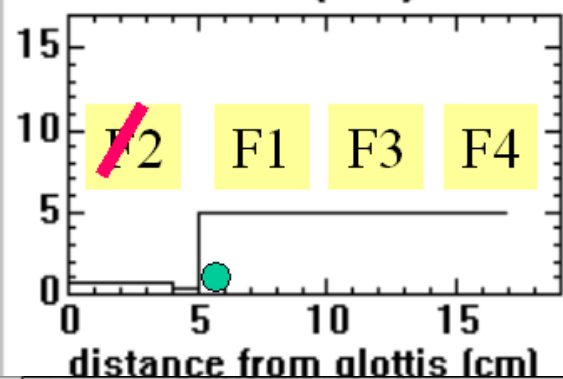


h



h

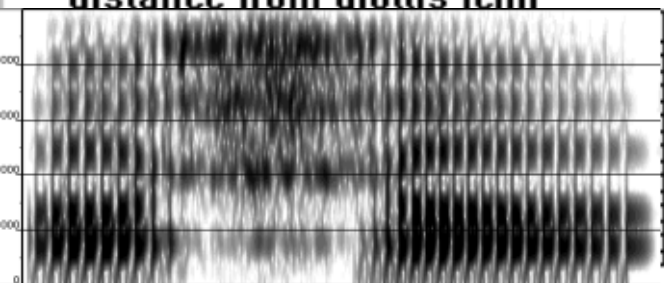
vocal tract area (cm<sup>2</sup>)



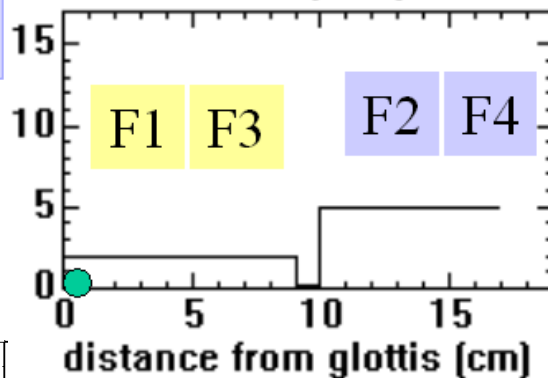
100%, 72



h



vocal tract area (cm<sup>2</sup>)



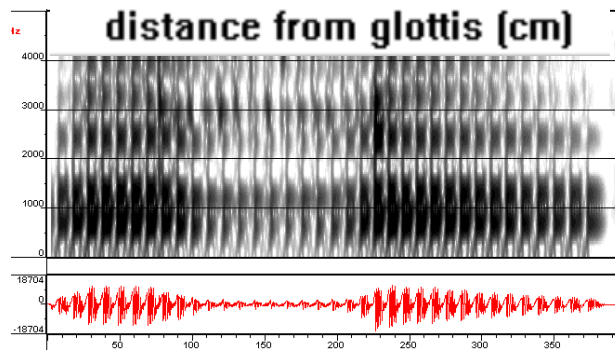
100%, 76



k

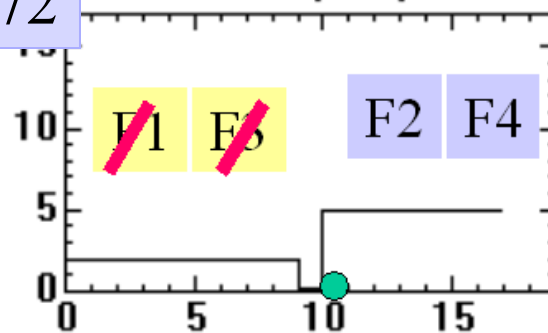
0.35

Glottis more open?



χ k

vocal tract area (cm<sup>2</sup>)

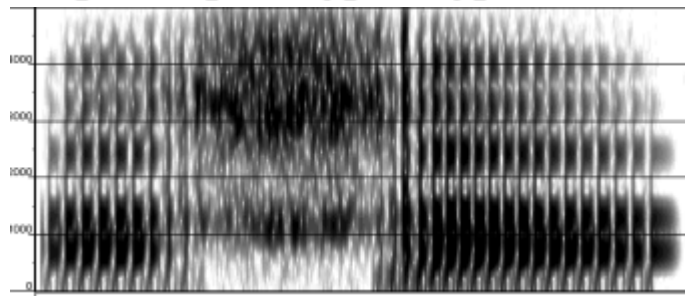


100%, 79



χ

0.20



10 listeners  
10 repetitions

\*Alwan, 1989



73

W  
h  
W  
S  
H  
X  
K

• Pharyngeal sourde

• 16.5

• 0.7 0.7 0.7 0.7 0.30 5.0 5.0 5.0 5.0 5.0 5.0 5.0 5.0 5.0 5.0 5.0 5.0 5.0



• Pharyngeale sonore

• 16.5

• 0.7 0.7 0.7 0.7 0.35 5.0 5.0 5.0 5.0 5.0 5.0 5.0 5.0 5.0 5.0 5.0 5.0



• [uvulaire sourde]

• 17

• 2.0 2.0 2.0 2.0 2.0 2.0 2.0 2.0 2.0 0.2 5.0 5.0 5.0 5.0 5.0 5.0 5.0



• [uvulaire sonore]

• 17

• 2.0 2.0 2.0 2.0 2.0 2.0 2.0 2.0 2.0 0.35 5.0 5.0 5.0 5.0 5.0 5.0 5.0



Ag = 0 cm2

Ag = 0.20-0.25 cm2

+ timing





## 2) For what concerns the consonants

*a) Simple concatenation work than expected*

*b) The rôle of anatomical details: sublingual cavity*

*c) The acoustic properties largely  
Prevail over articulation: the velars*

*d) Exotic consonants may be easy to model*

*e) Lenition and fortition not difficult to model*

# lenition

What is a model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

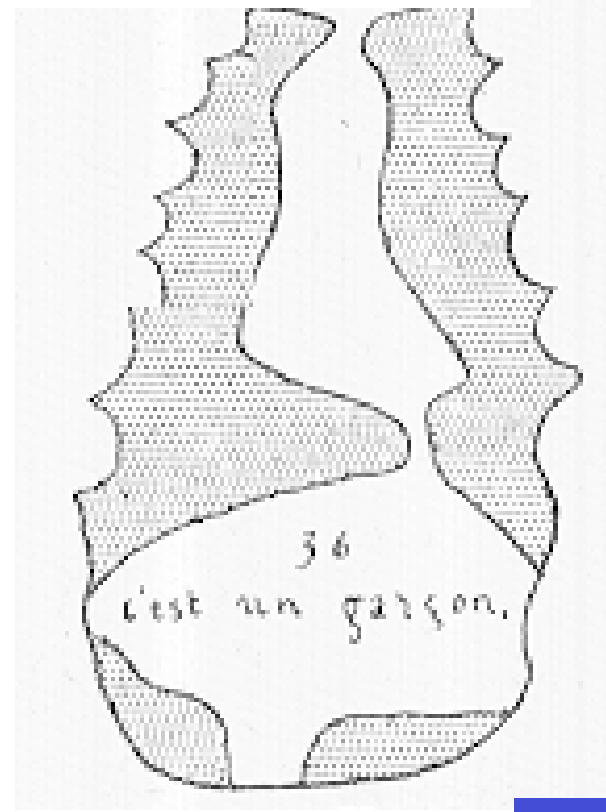
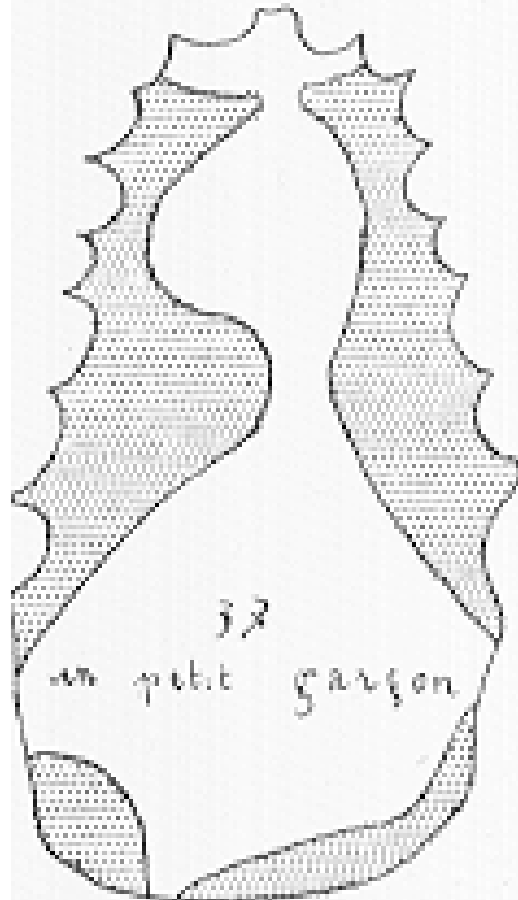
vowels  
consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

- Less strong constriction
  - More close glottis
- ↓
- Ration between the two areas may change, and to the category of the consonant
  - But the underlying F-pattern is the same





speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

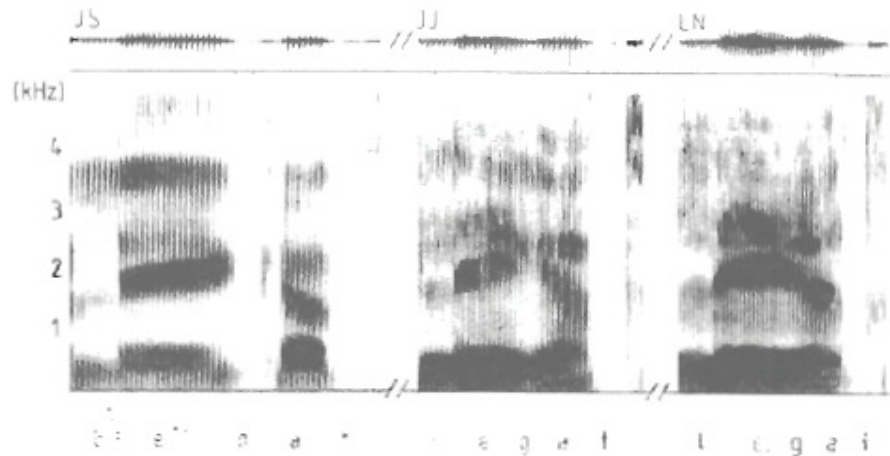


Figure 4.4.2. Normal and two degrees of target undershoot of a voiced stop.

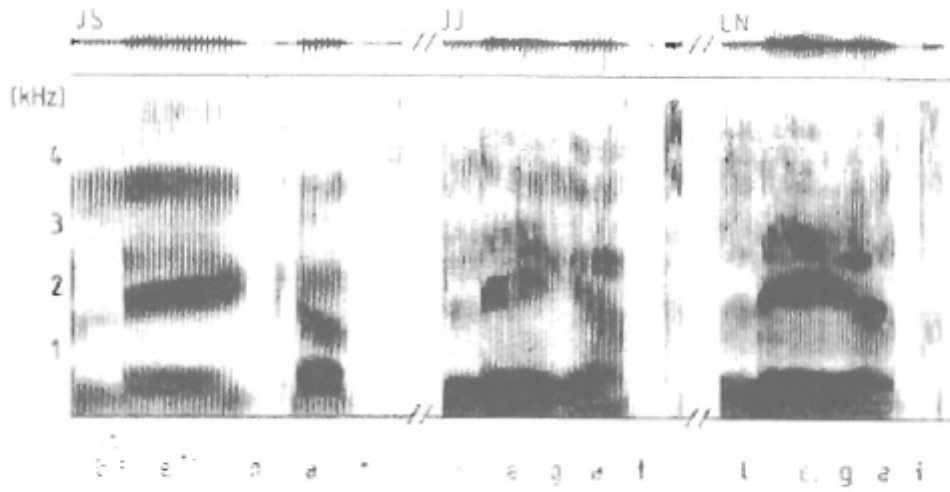
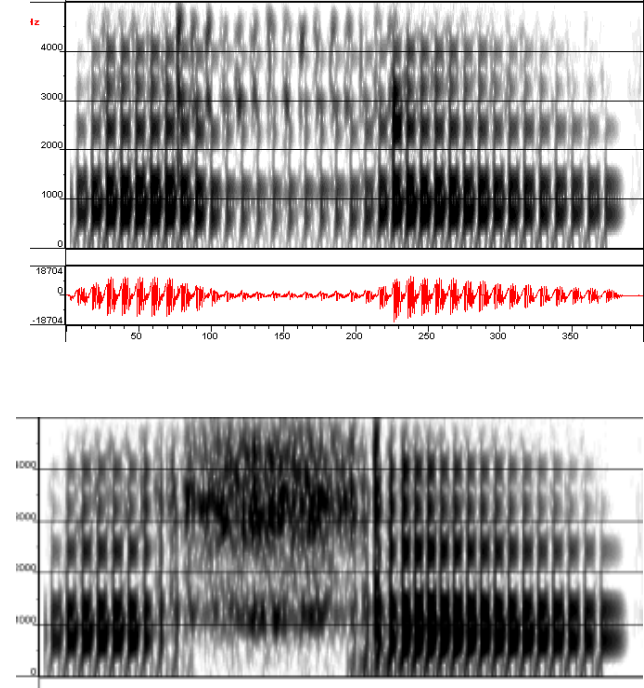


Figure 4.4.2. Normal and two degrees of target undershoot of a voiced stop.



Vt more close than vf

Less vf opening

Voiceless > voiced

Noise created at the narrowest

How done?

Demo

vowels  
consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

Less strong closure  
in the middle

Stops > fricatives  
> sonantes



## 2) For what concerns the consonants

*a) Simple concatenation work than expected*

*b) The rôle of anatomical details: sublingual cavity*

*c) The acoustic properties largely  
Prevail over articulation: the velars*

*d) Exotic consonants may be easy to model*

*e) Lenition and fortition not difficult to model*

*f) Shape of the tongue counts  
more than exact place of constriction*

What is a mo

Why modelli

Which model

What is a use  
model?

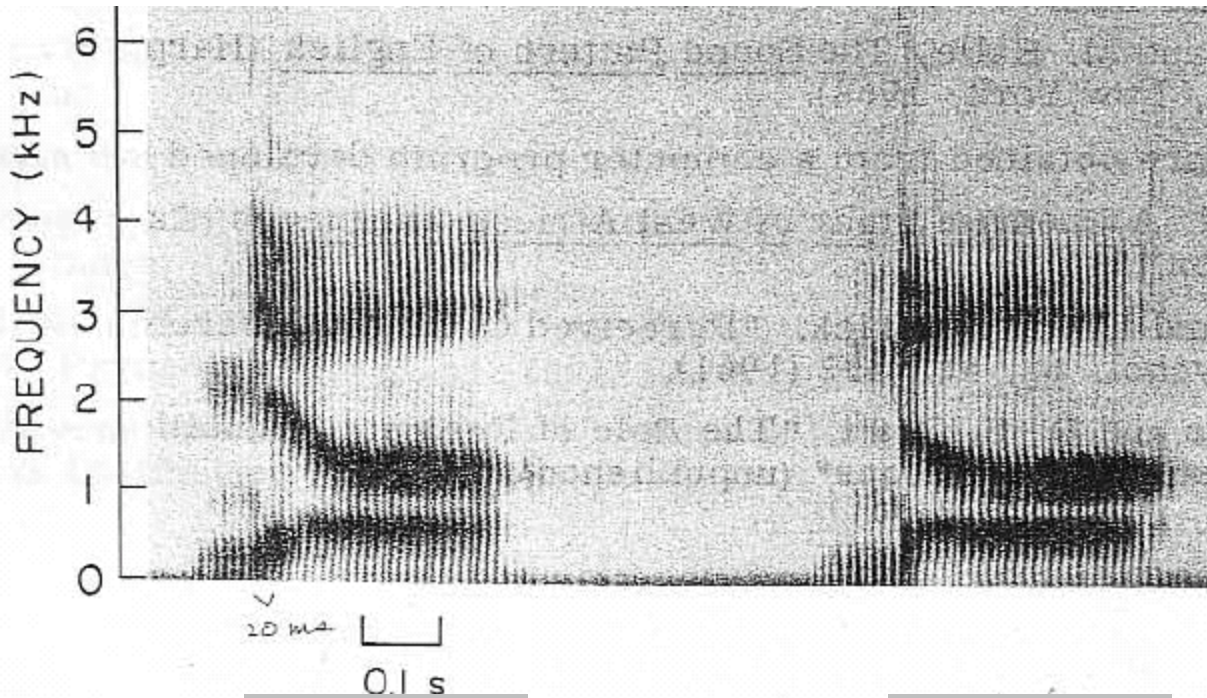
How done?

Demo

vowels  
consona

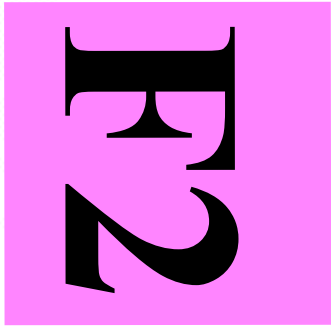
Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;



laminal

apical



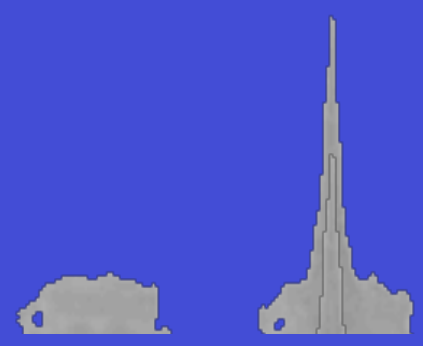
And shape of the tongue is influenced by the vowels



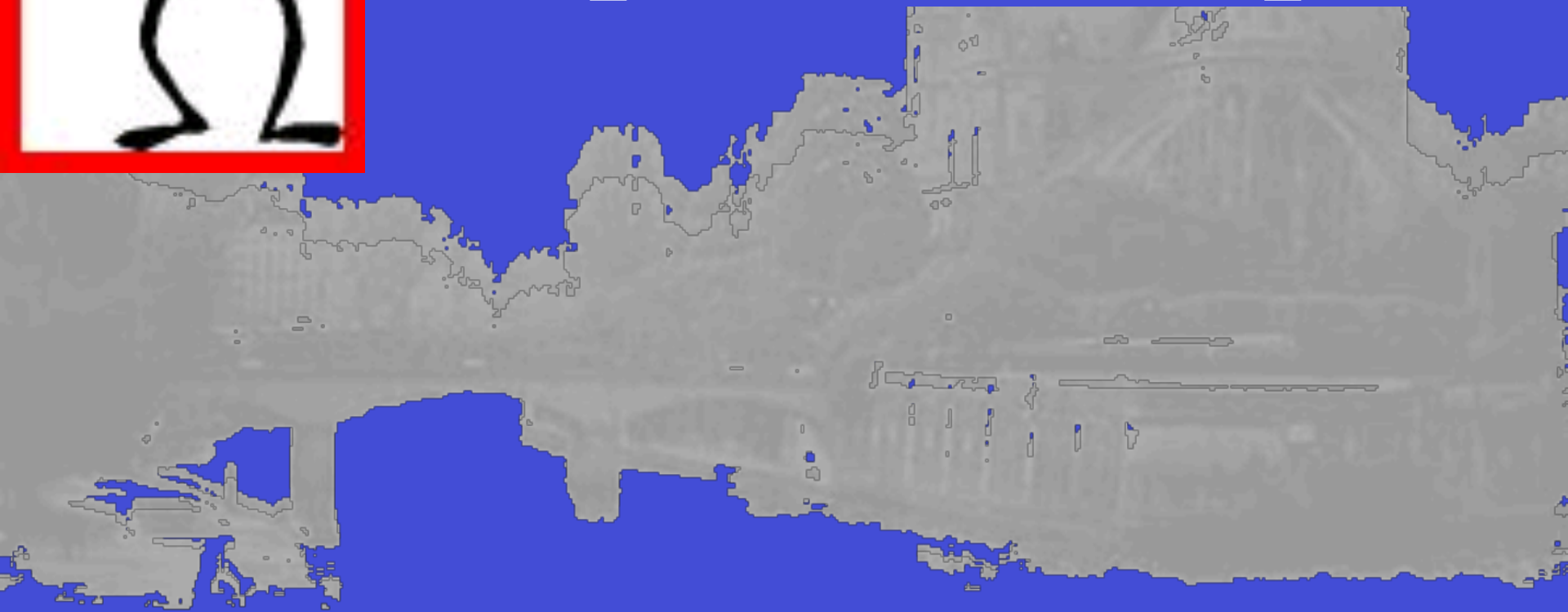
## 2) For what concerns the consonants

- a) Simple concatenation work than expected*
- b) The rôle of anatomical details: sublingual cavity*
- c) The acoustic properties largely  
Prevail over articulation: the velars*
- d) Exotic consonants may be easy to model*
- e) Lenition and fortition not difficult to model*
- f) Shape of the tongue counts  
more that exact place of constriction*
- g) Suppressing the coarticulation effect allows  
to distangle the perceptual effects C to V and V to V*





# Speaker adaptation



# VLAM (Maeda, 1994)

## Variable Linear Articulatory Model

lip, jaw, tongue, larynx

What is a model?

Why modelling?

Which models?

What is a useful model?

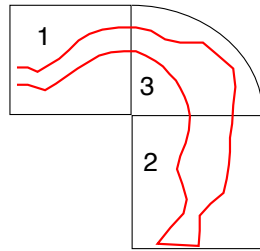
How done?

Demo

vowels  
consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;



The Larynx Height Index variation is introduced by modifying the longitudinal dimension of the vocal-tract according to 2 scaling factors and 3 zones:

- (1) for the anterior part of the vocal-tract
- (2) for the pharynx
- (3) interpolating the zone in-between

Modelisation

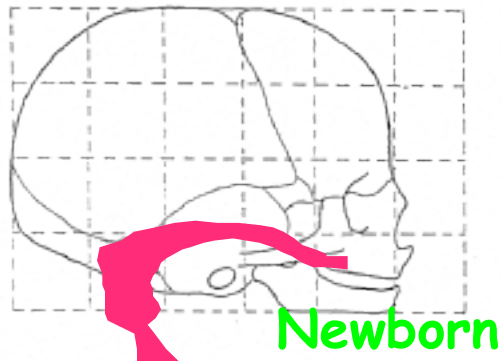
Babies  
Neanderthal  
monkey

# Vowel prototypes for ontogenesis

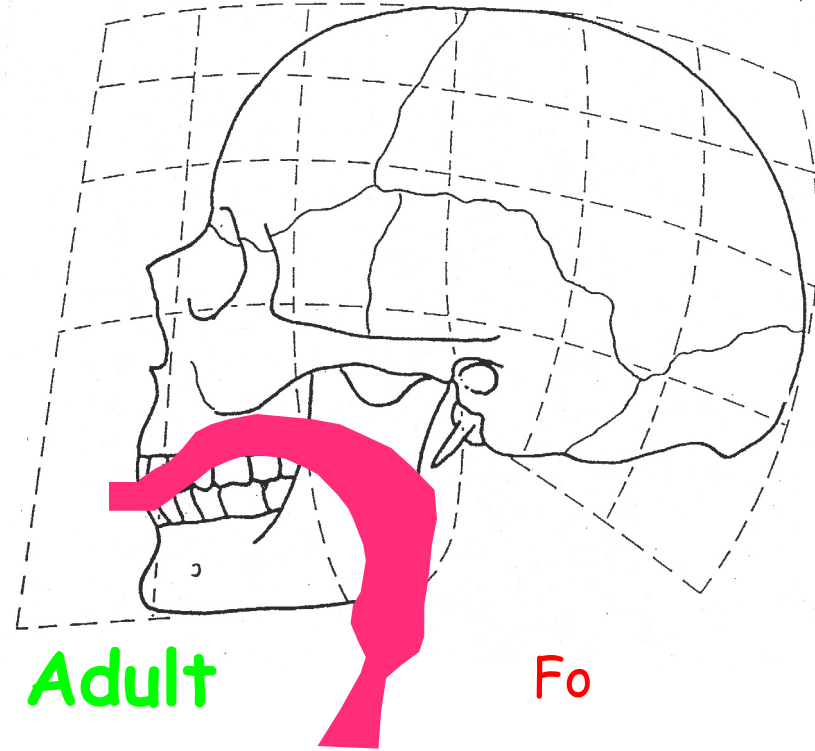


Which models?

What is a useful model?



Ontogenesis



speaker anatomy  
speaker strategy  
**Fo** formant  
singling  
inversion  
etc;

Fo  
Larynx

Maeda + ICP-Grenoble

Modelisation

Babies  
Neanderthal  
monkey





# Larynx Height Index

What is a model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

Vowels  
consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

1.00    0.95    0.88    0.79    0.61    0.55

Man    Woman    Child (10y 4y)    Newborn

Ape  
Newborn Adult

Neandertal

*La Ferrassie La Chapelle-aux-Saints*

Maeda + ICP-Grenoble

Babies  
Neanderthal  
monkey

Modelisation





# Vowel prototypes (adult man)

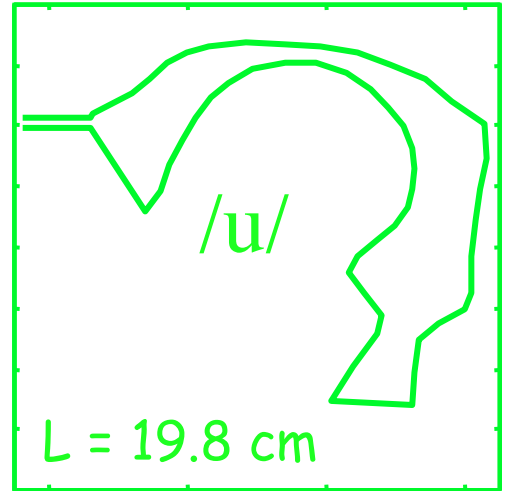
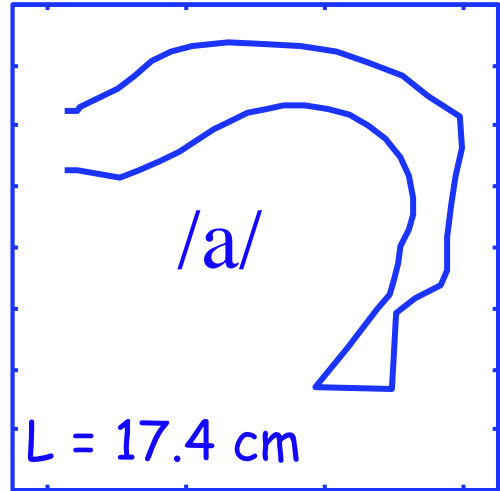
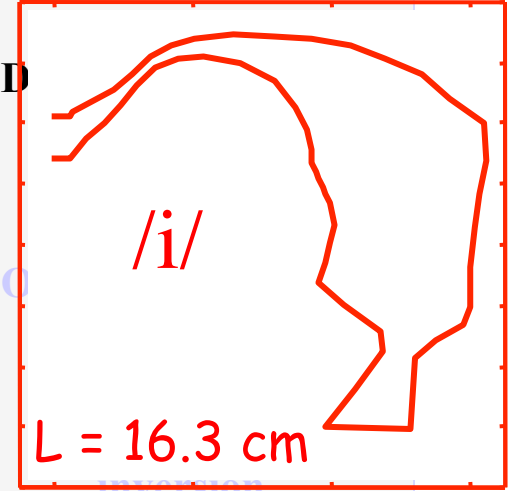
What is a model?

Why modelling?

Which models?

What is a useful model?

How done?



inversion  
etc;



Maeda + ICP-Grenoble

Modelisation



Babies  
Neanderthal  
monkey





t

Why r

Which

What

model?

How done?

Demo

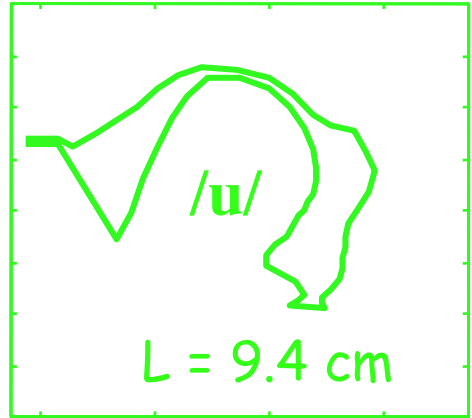
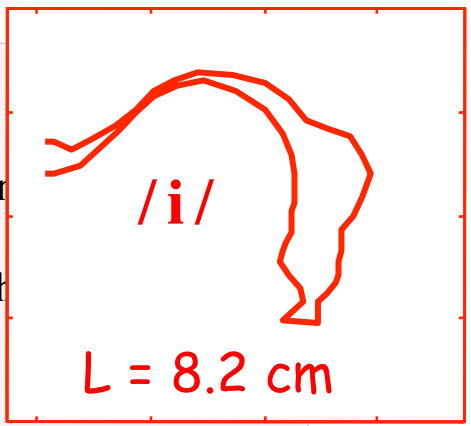
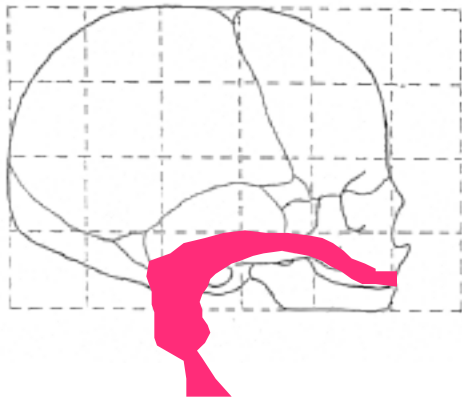


vowels  
consonants

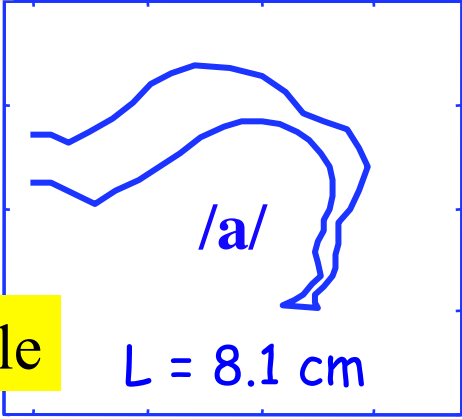
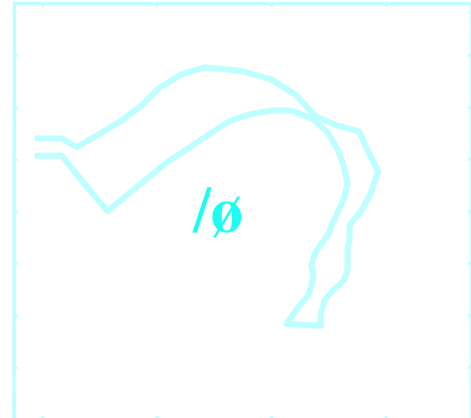
Other uses

- speaker anatomy
- speaker strategy
- singing formant
- inversion
- etc;

Maeda + ICP-Grenoble



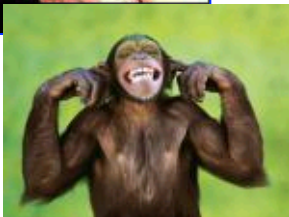
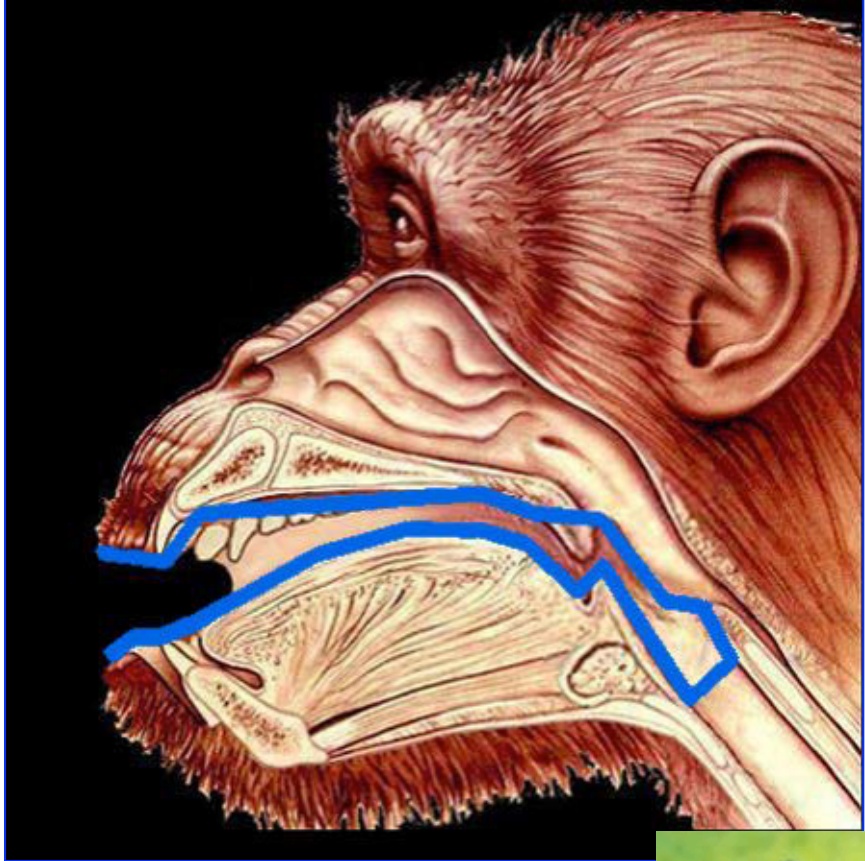
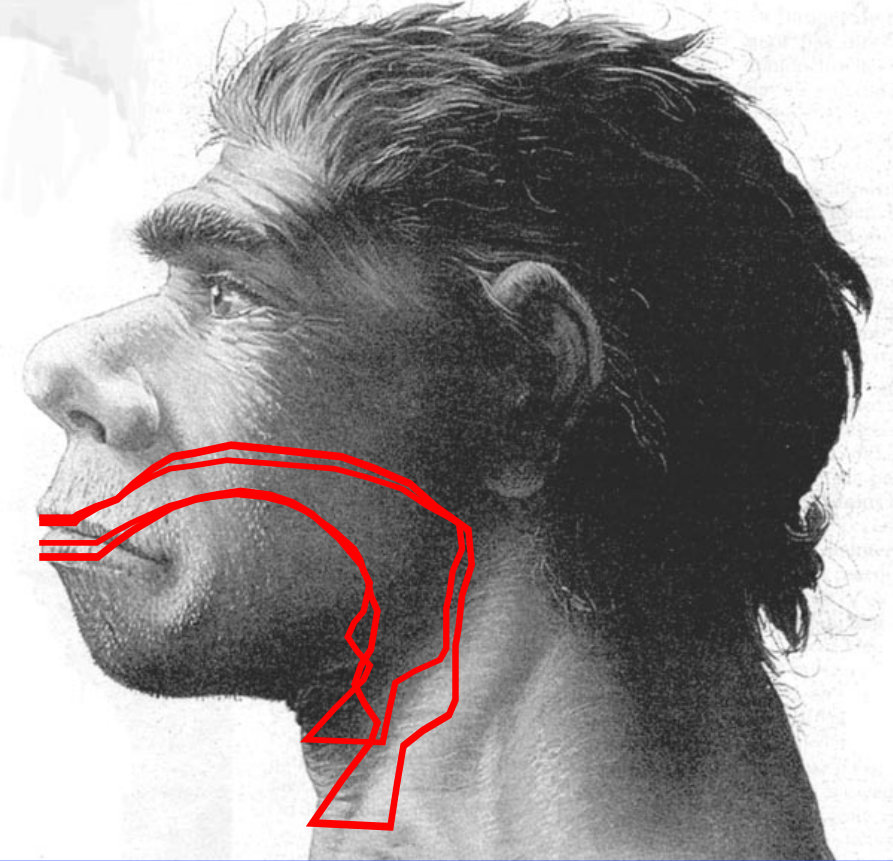
4 months



Model on







Other uses

- speaker anatomy
- speaker strategy
- singing formant
- inv...
- etc:

Contra Lieberman & Crelin

Modelisation of the differences suggests that the neandertals were no more vocally handicapped than children at birth are

Maeda + ICP-Grenoble

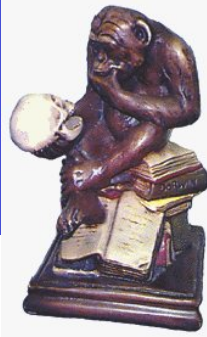
Modelisation



Babies  
Neanderthal  
monkey



# Conclusions?



What

Why

Which models?

What is a useful model?

How done?

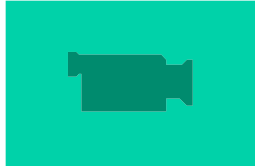
Demo

vowels  
consonants

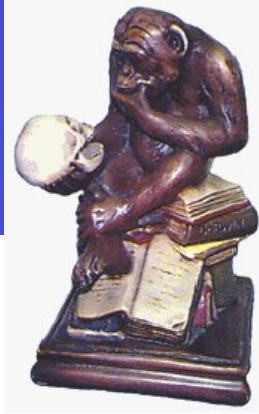
Other uses

speaker anatom  
speaker strategy  
singing forma  
inversion  
etc;

- AM is a very useful research and teaching tool
  - Essential as a platform for discussions between phoneticians and engineers
- As a complement of other instrumentation such as EMMA, echograph, IRM, etc.
- Still in progress (a third cavity)
- New thesis coming in fricatives in different languages (Toda/Maeda), MRI, ATR



# Conclusions?



What

Why

Which models?

What is a useful model?

How done?

Demo

vowels  
consonants

Other uses

speaker anatom  
speaker strategy  
singing forma  
inversion  
etc;

- **limits**

1. /l/ and many other sounds not yet done (as presented in this congress)
1. Register/Phonation types/laryngeal region area and Fo/quality interaction not included (could replace the actual 3 parameters)
2. Aerodynamic constrains increases when VT very close
3. mid-sagittal profile conversion to area function not straightful process (3D)
4. New X-ray difficult to obtain (MRI and new instrumentation)











PPC2008, Pékin, Chine