



**HAL**  
open science

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

Jacqueline Vaissière

► **To cite this version:**

Jacqueline Vaissière. On acoustic salience of vowels and consonants predicted from articulatory models. 8th Phonetic Conference of China and the International Symposium on Phonetic Frontiers, Apr 2008, Pékin, Chine. halshs-00678785

**HAL Id: halshs-00678785**

**<https://shs.hal.science/halshs-00678785>**

Submitted on 31 Mar 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PPC2008, Pékin, Chine

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



Many thanks to  
Shinji Maeda  
And LPP students  
In Paris

**Jacqueline Vaissière**  
*Université Sorbonne Nouvelle - Paris*  
*Laboratoire de Phonétique et de Phonologie*

PPC2008, Pékin, Chine



# A



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

# Menu?

# 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 ...**

3

# 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*

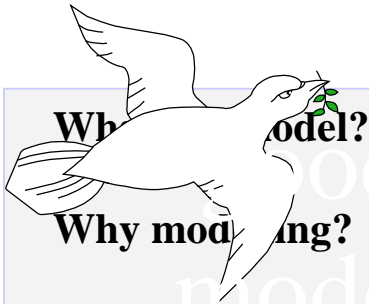
## Psycholinguistics

- *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

formant



*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*

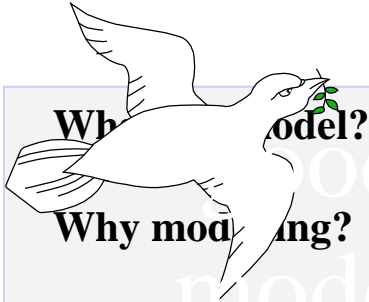
*no stop without a burst, or no frication when humans do.*

*To draw a direct link between articulatory, acoustic and perceptual observations*



6

# How done?



Why model?

Why modeling?

Which models?

What is a useful model?

**How done?**

Demo

vowels


consonants

Other uses

speaker anatomy

strategy

armant



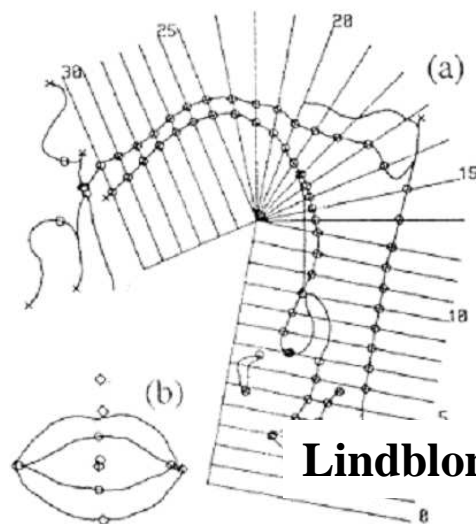
- *By Shinji Maeda*
- *From X-ray data for French (from Strasbourg)*
- *By statistical analysis*
- *The most widely used AM*
- *Used for pedagogical purpose in our lab and research in Grenoble, Nancy, UCLA, MIT, etc.*



# 1 X-ray data



2



Lindblom & al, 1971

# 3 Guided PCA

4 7 parameters

5

One speaker

10 sentences  
>1000 images

VT shapes

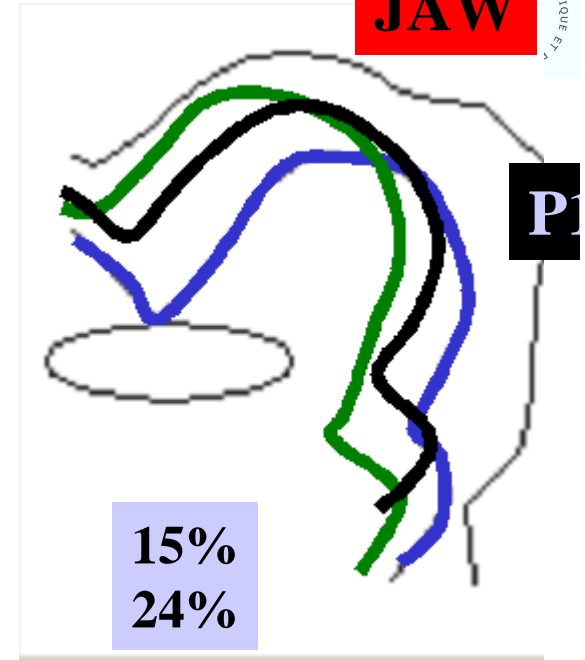
First substract the effect of the jaw

alisation



# 4 first parameters Jaw + tongue

**JAW**



How done:

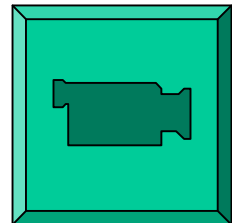
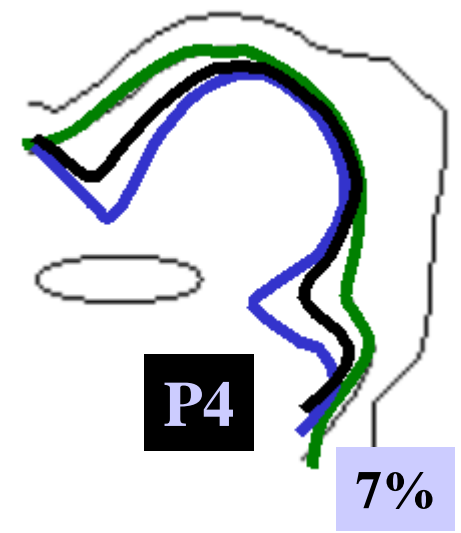
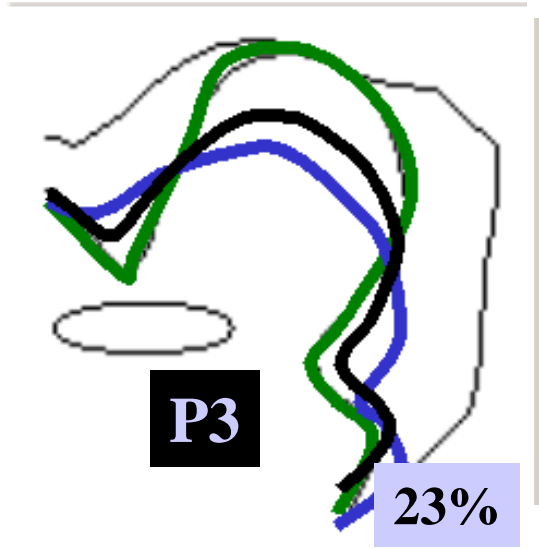
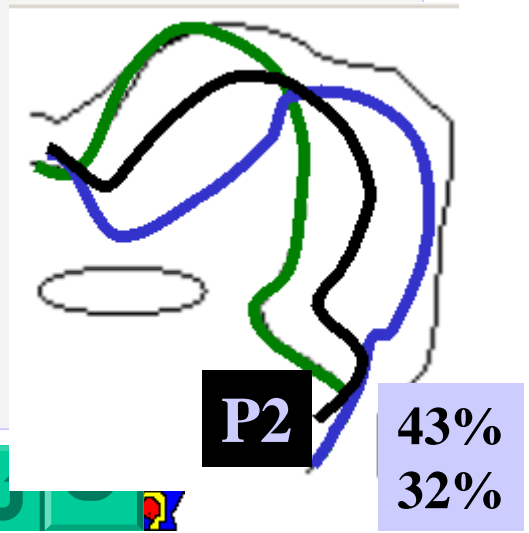
Demo

**1+2= 58%, 56%**

**1+2+3= 81%, 79%**

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

Other uses



v  
v  
v  
v

4

3 more =  
Lip and larynx

LIP

ht

5

pr

6

larynx

ht

7

JAW

ht

1

nasal

8

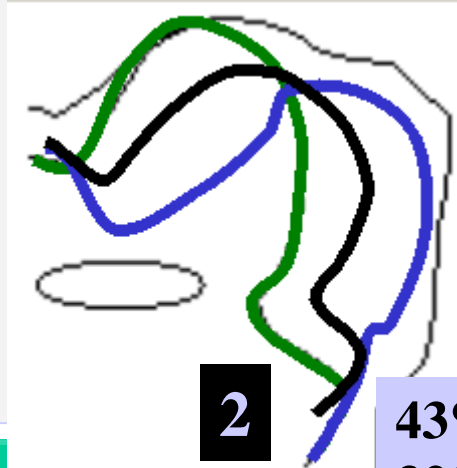
15%  
24%

How done?

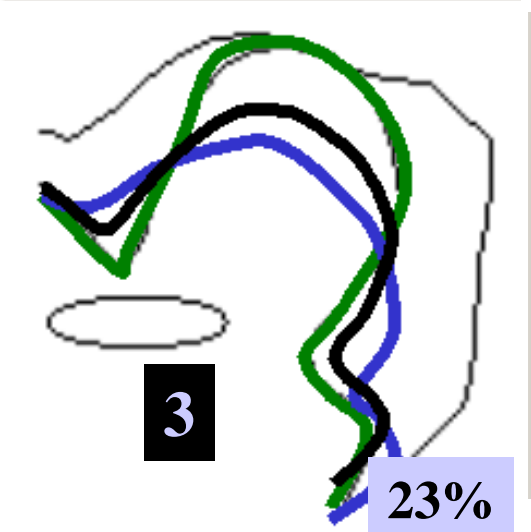
Demo

vowels  
consonants

Other uses



43%  
32%



23%



7%

4

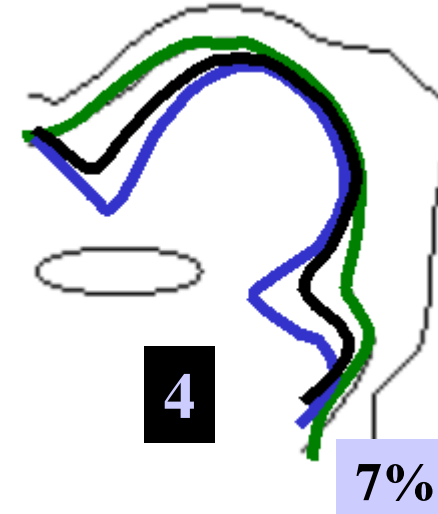
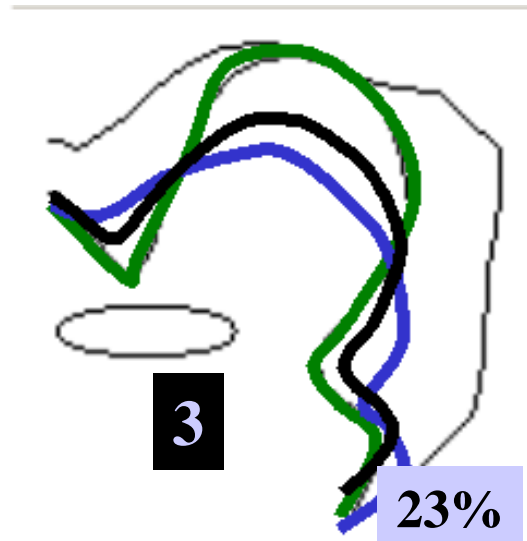
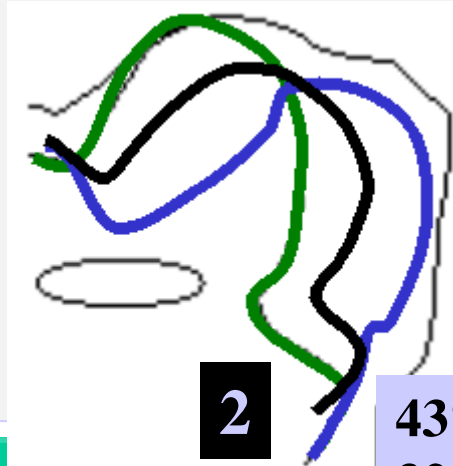
7 parameters =  
Speech organs

How done?

Demo

vowels  
consonants

Other uses



LIP

ht 5

pr 6

larynx

ht 7

JAW

ht

1

15%  
24%

nasal

8

# 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

body

Other uses

2

43%  
32%

3

23%

4

7%

LIP

ht 5

pr 6

larynx

ht 7

JAW

ht 1

nasal

8

TONGUE

shape

u, k

apex

i, e, dentals

15%  
24%



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?

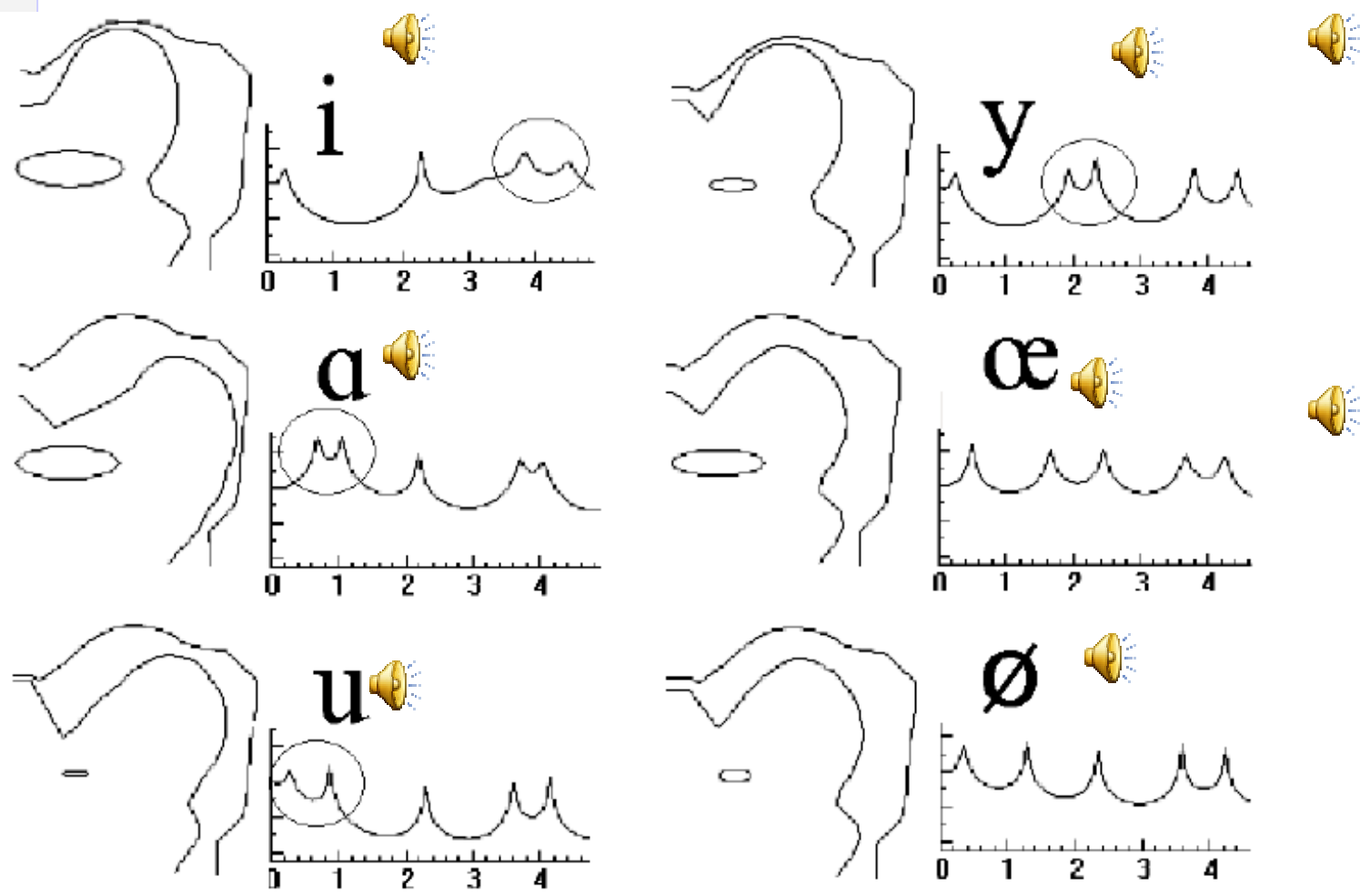
How done?

Demo

→ vowels  
consonants

Other uses

speaker anatom  
speaker strateg  
singing forma  
inversion  
etc;



Vaissière, 2007

PPC2008, Pékin, Chine



**What is a model?**

**Why modelling?**

**Which models?**

**What is a useful  
model?**

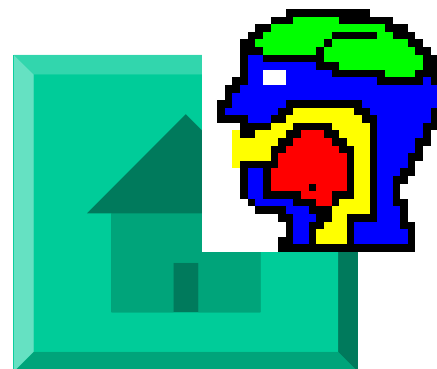
**How done?**

**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?

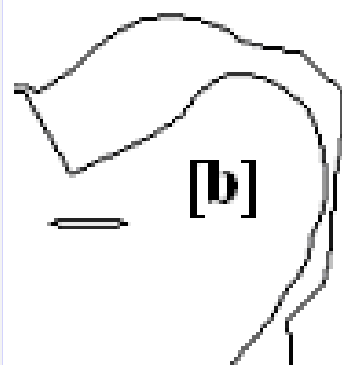
How done?

Demo

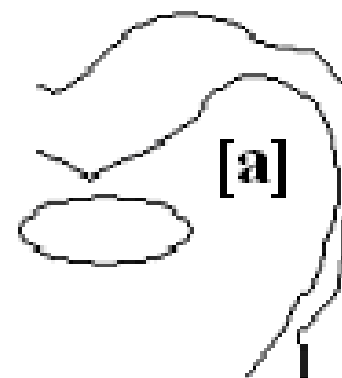
→ vowels  
→ consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;



F1	371
F2	898
F3	2136
F4	3683
F5	4024

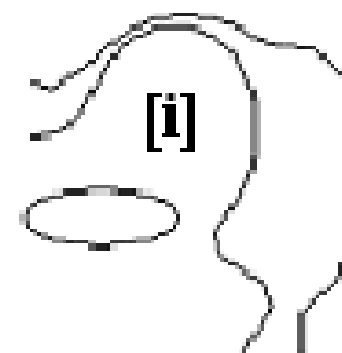


F1	712
F2	1145
F3	2229
F4	3745
F5	4024



F1	186
F2	1207
F3	2321
F4	3776
F5	4457

closure ]



F1	248
F2	2290
F3	3838
F4	4457

after release



# Beans on a string synthesis

1

# 2) consonants

What is a model?

Why modelling?

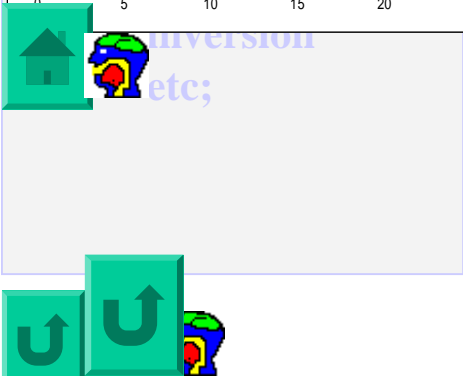
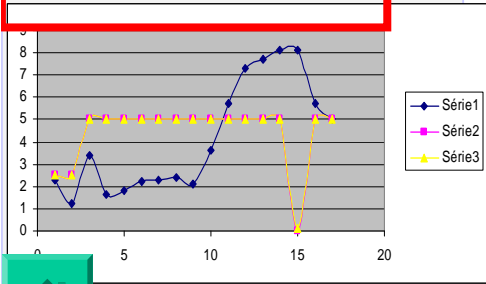
Which models?

What is a useful model?

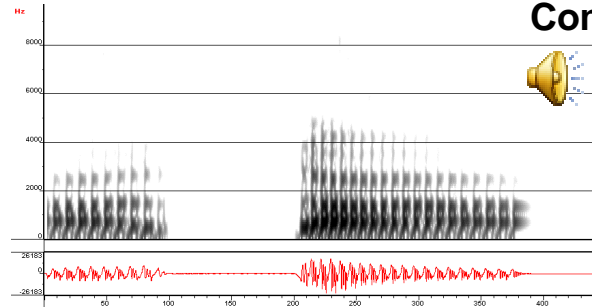
How done?

**Demo**

→ vowels  
→ consonants

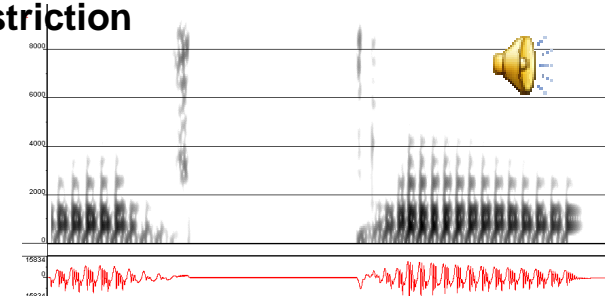


Glottis close

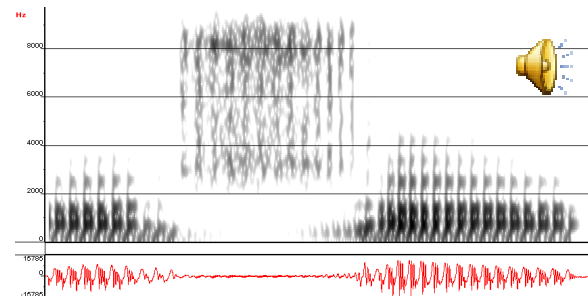


Complete constriction

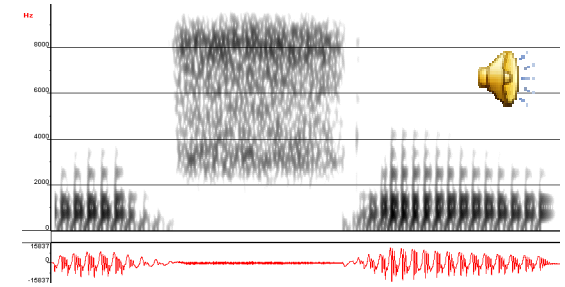
Glottis open



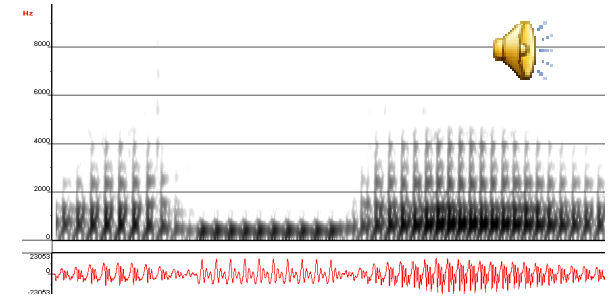
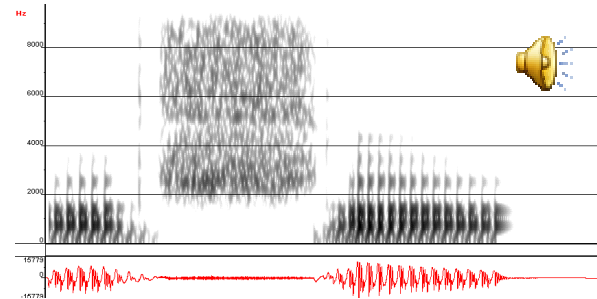
Complete constriction



Sublingual cavity



Nasal port open



# Lip constriction

What is a model?

Why modelling?

Which models?

What is a useful model?

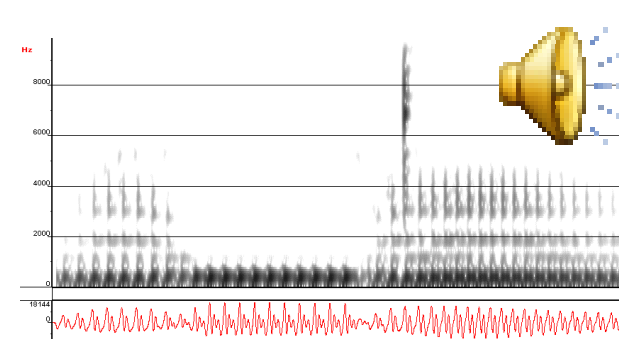
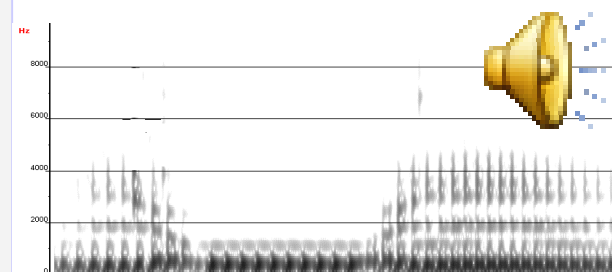
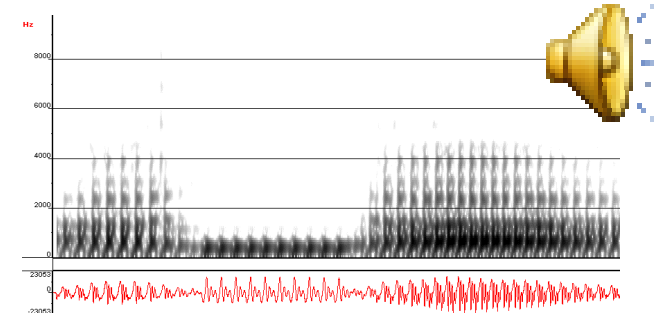
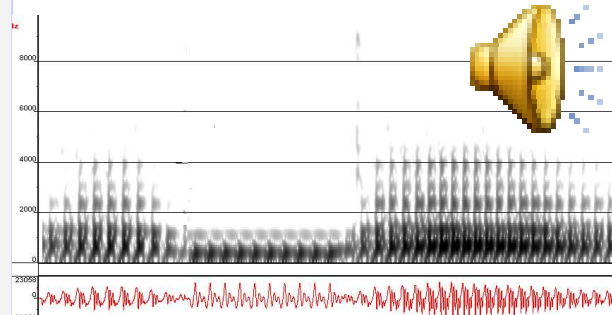
How done?

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**

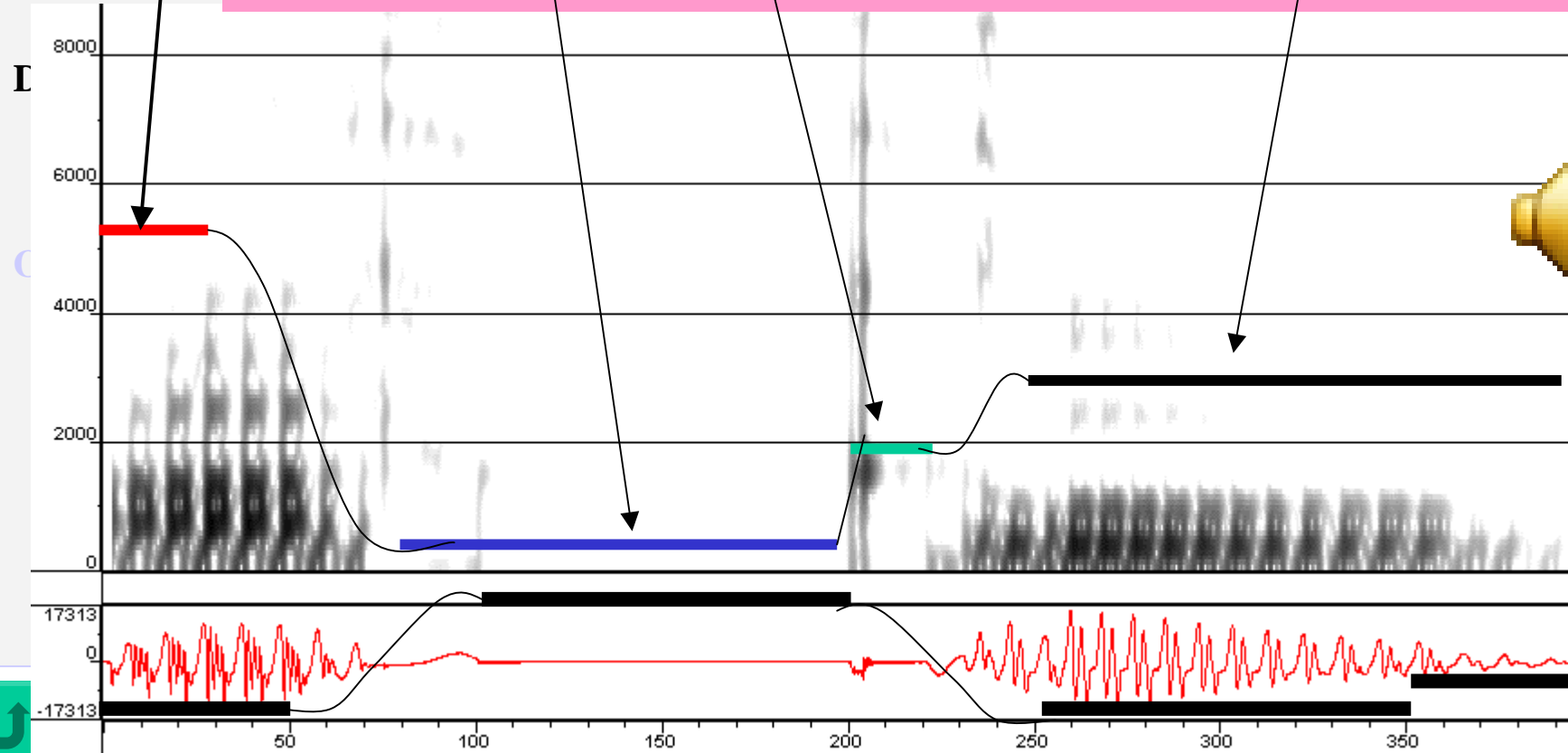
What is a model

Why modelling

Which model

What is a useful model?

How done?





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

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

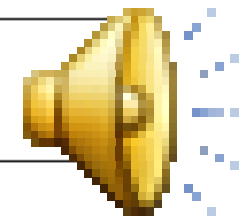
**Stay open for 100 msec**

Then cosine transition to close glottis at 250.

**Stay close for 100 msec**

Open slightly for the last 50 msec. (prepausal

opening)



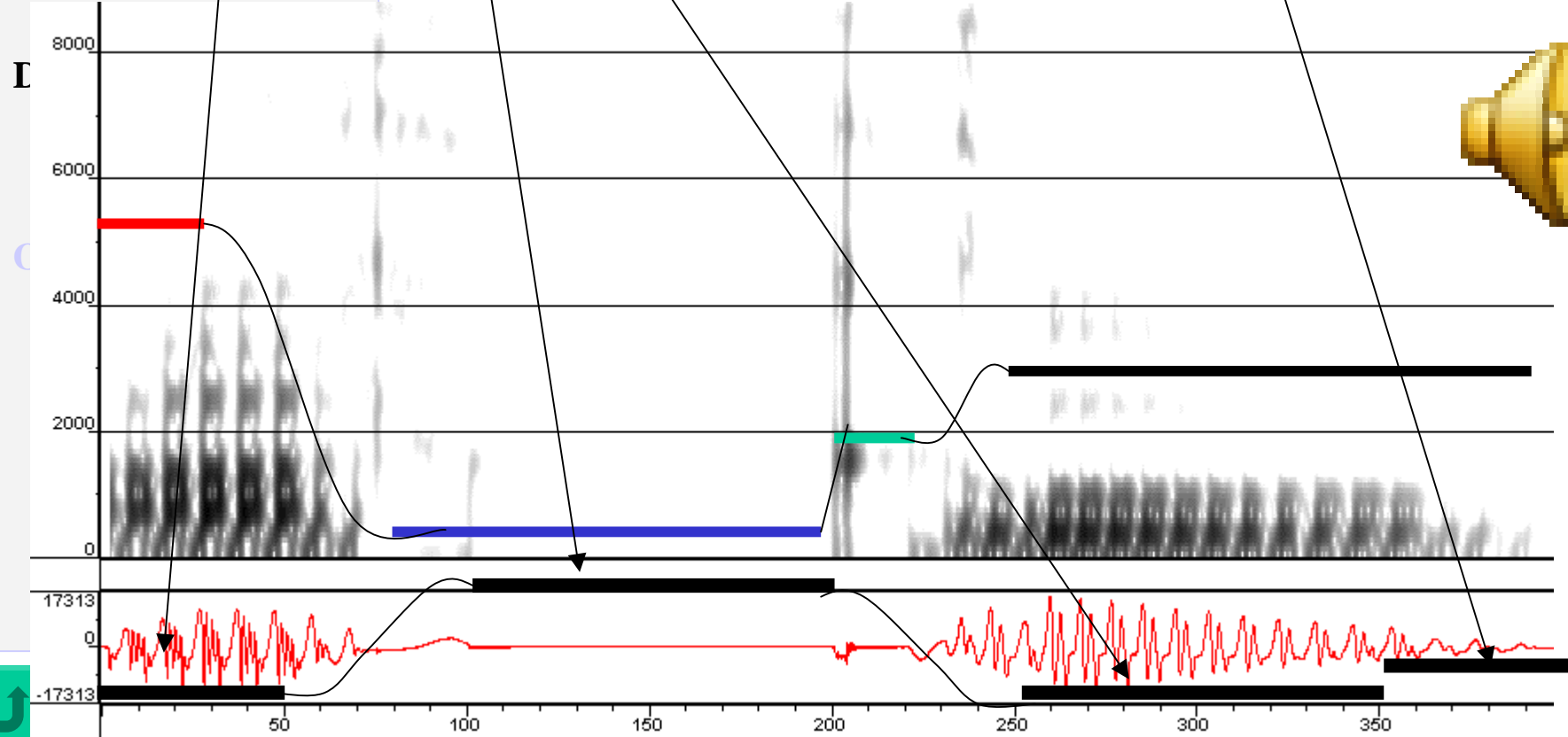
What is a

Why mode

Which mo

What is a  
mode

How done:



What is a model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

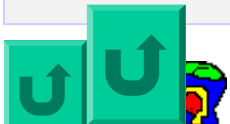
→ vowels  
consonants

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

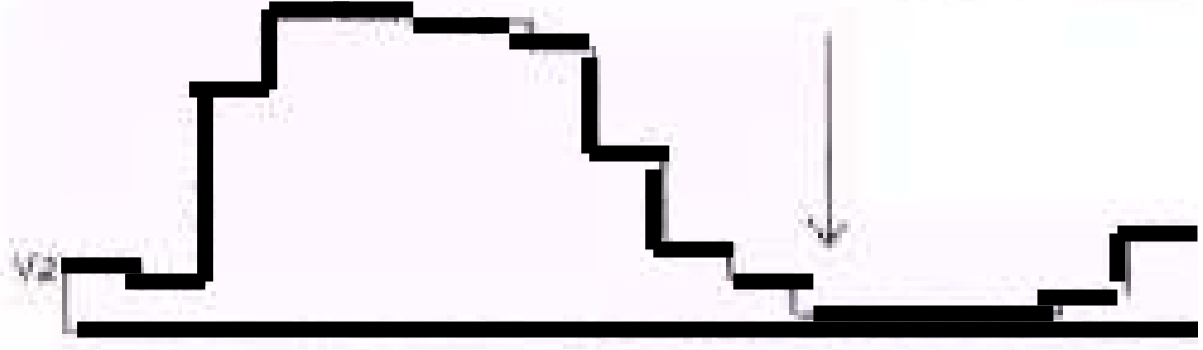


Figure 2: Interpolation of target area function

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

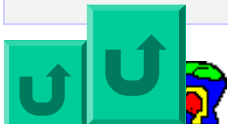
= /a/ + /tclosure/ + /t release/ + /a/

+ 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?**

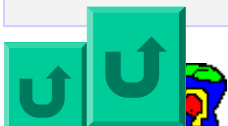
**Demo**

→ vowels  
consonants

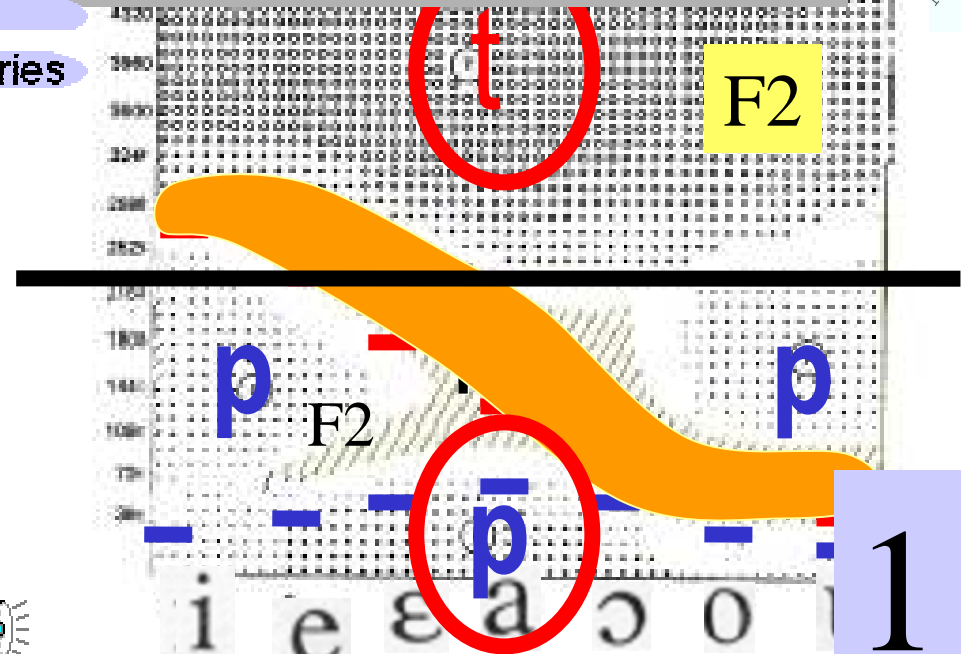
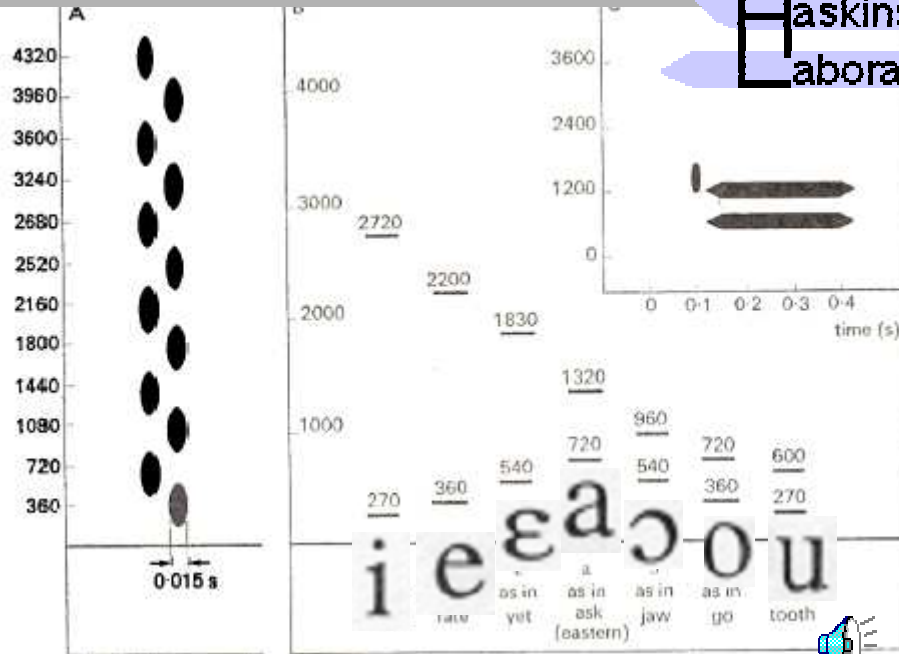
**Other uses**

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)



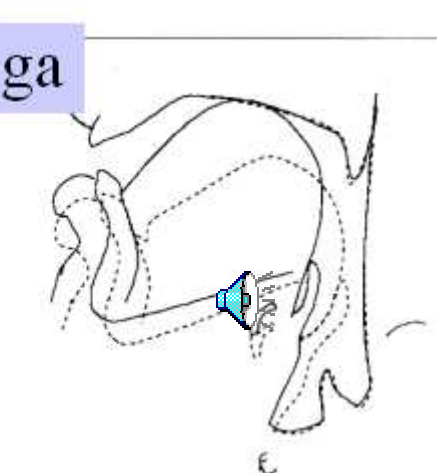
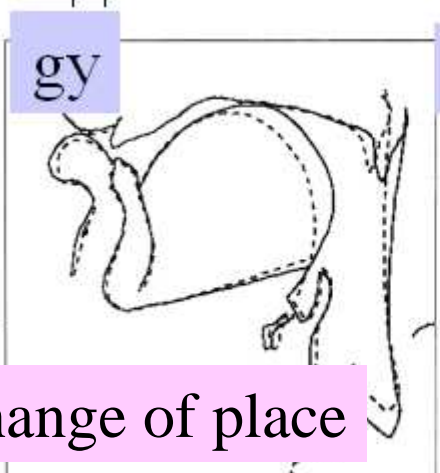
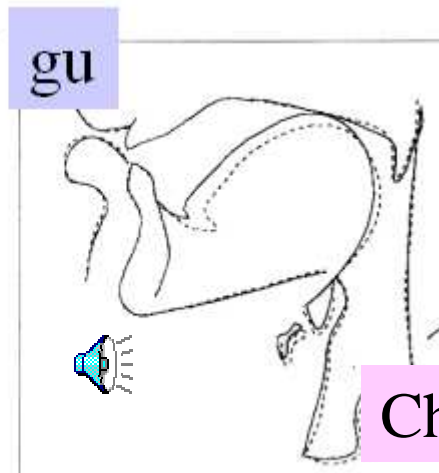
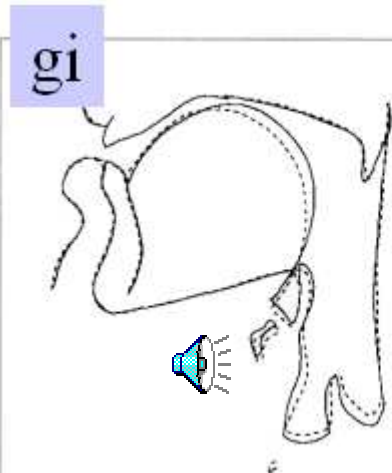
the reference to the vowel is not needed for discriminating [p] and [t]



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

**What is a model?**

**Why modelling?**

**Which models?**

**What**

**How**

**Demo**

**Other**

- The X-ray data are due to the University in Strasbourg
- Laboratoire de phonétique
- R. Sock
- Frame by frame
- Illustration of anticipation

See the work at IPS Strasbourg (web)





**/aku/**

model?

Why modelling?

Which models?

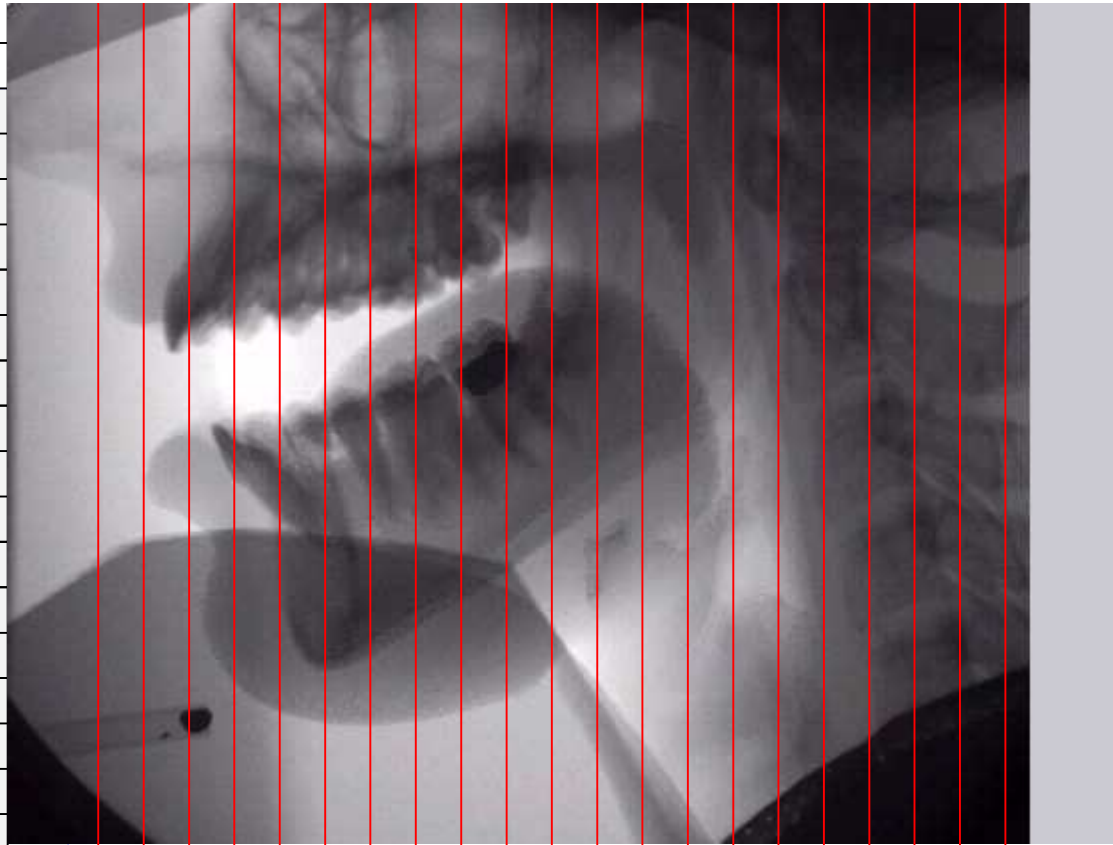
What is a **useful**  
model?

How done?

Demo

vowels

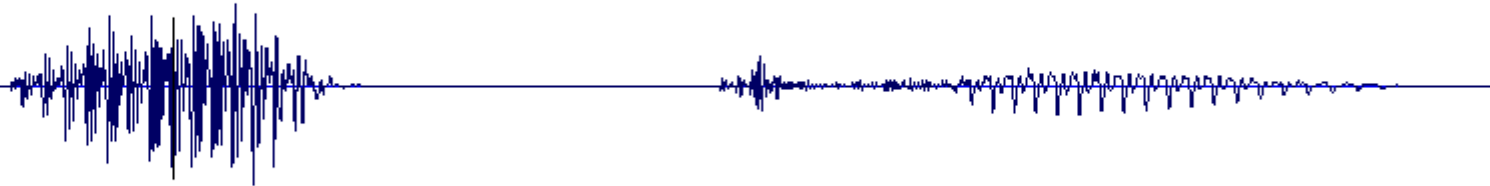
→ consonants



Other uses

**t0**

Mouth already largely open



Before /a/

**/aku/**

Model?

Why modelling?

Which models?

What is a **useful** model?

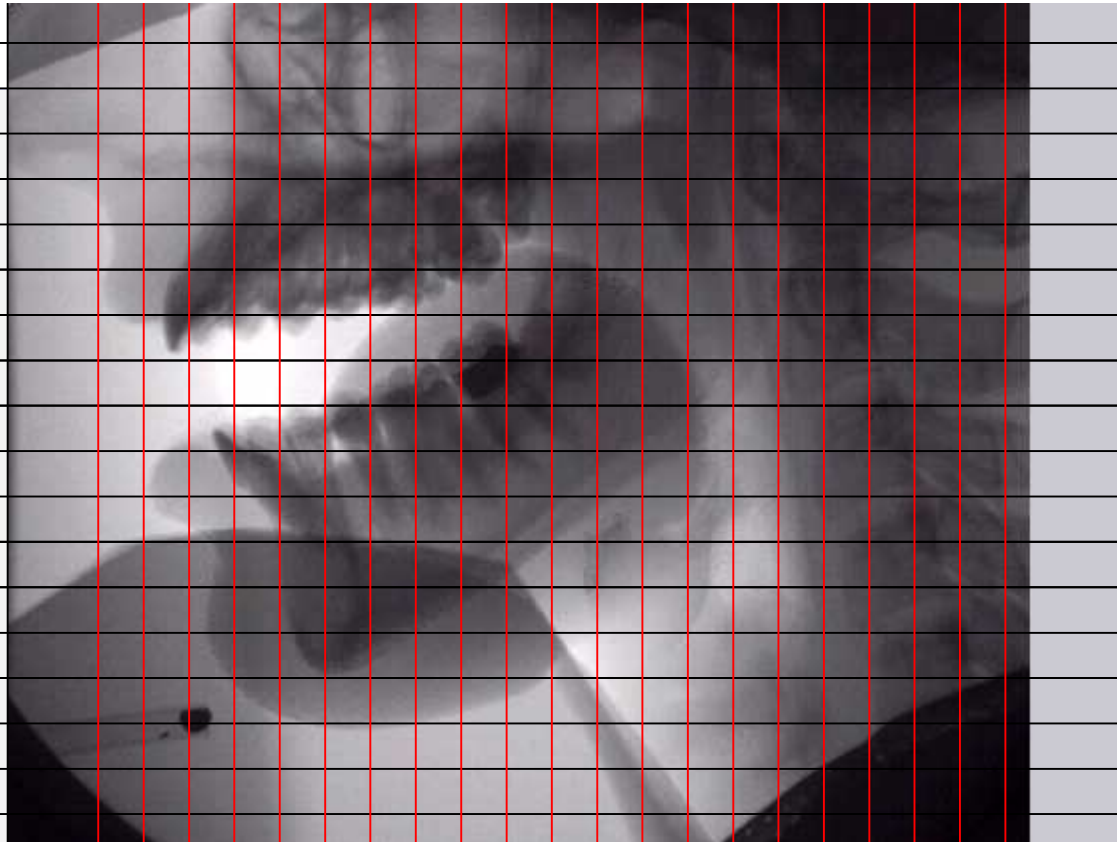
How done?

Demo

vowels



consonants



Other uses

**t1**

Maximum opening for /a/  
Jaw starts to rise again



**/a/ onset**

**/aku/**

model?

Why modelling?

Which models?

What is a **useful** model?

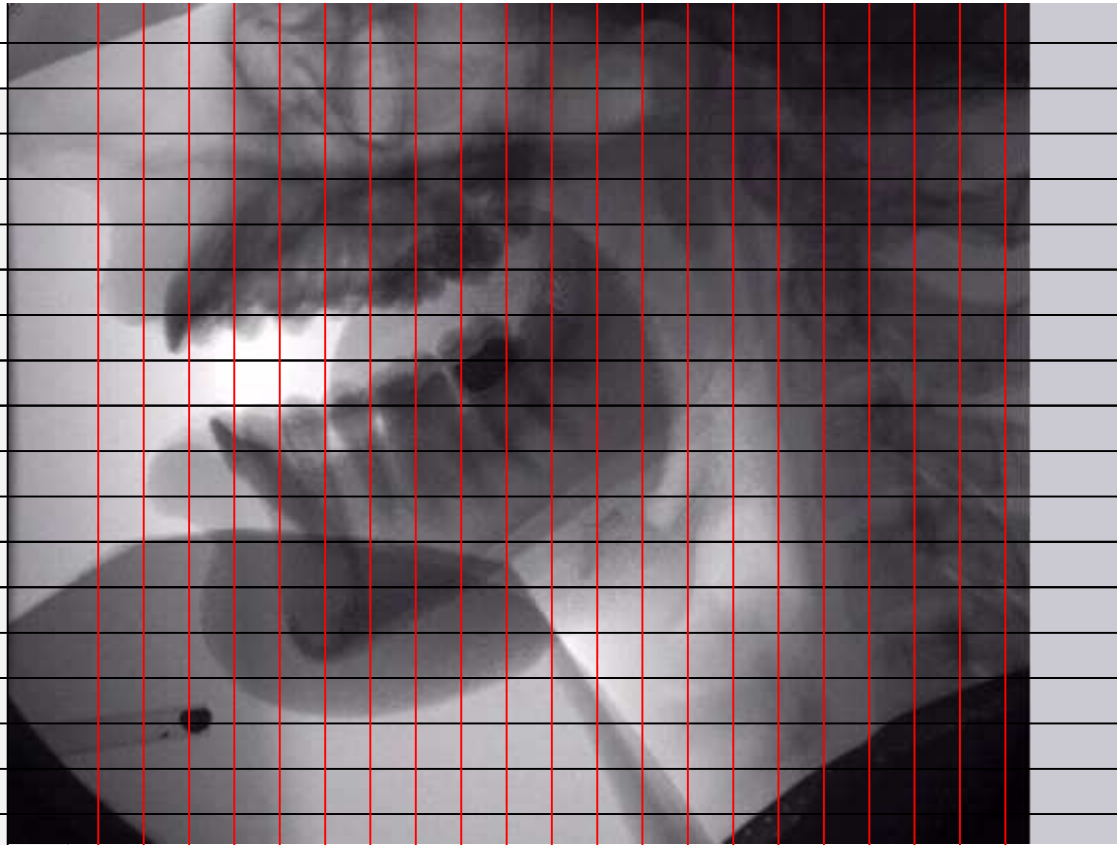
How done?

Demo

vowels



consonants



Other uses

**t2**

**/k/ closing gesture starts**  
**Rounding starts**



**/a/ midle**

**/aku/**

odel?

Why modelling?

Which models?

What is a **useful** model?

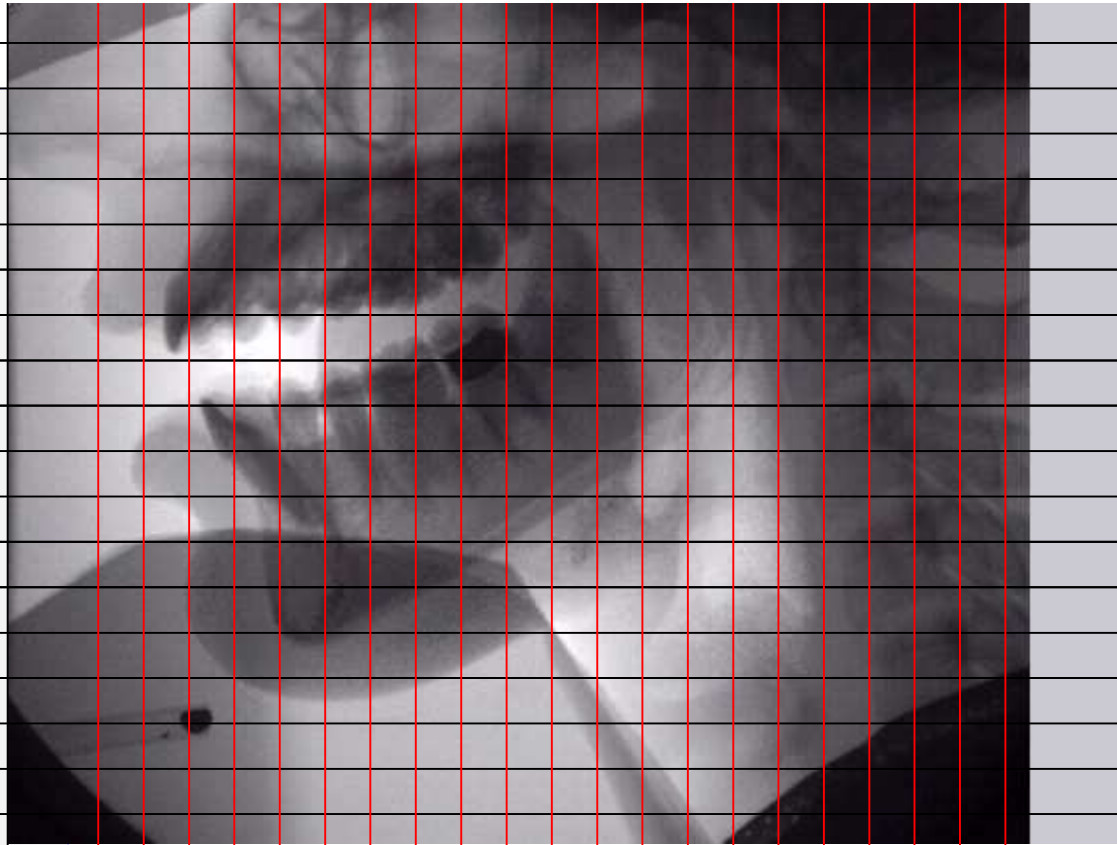
How done?

Demo

vowels



consonants



Other uses

**t3**



**/a/ offset**

**/aku/**

model?

Why modelling?

Which models?

What is a **useful** model?

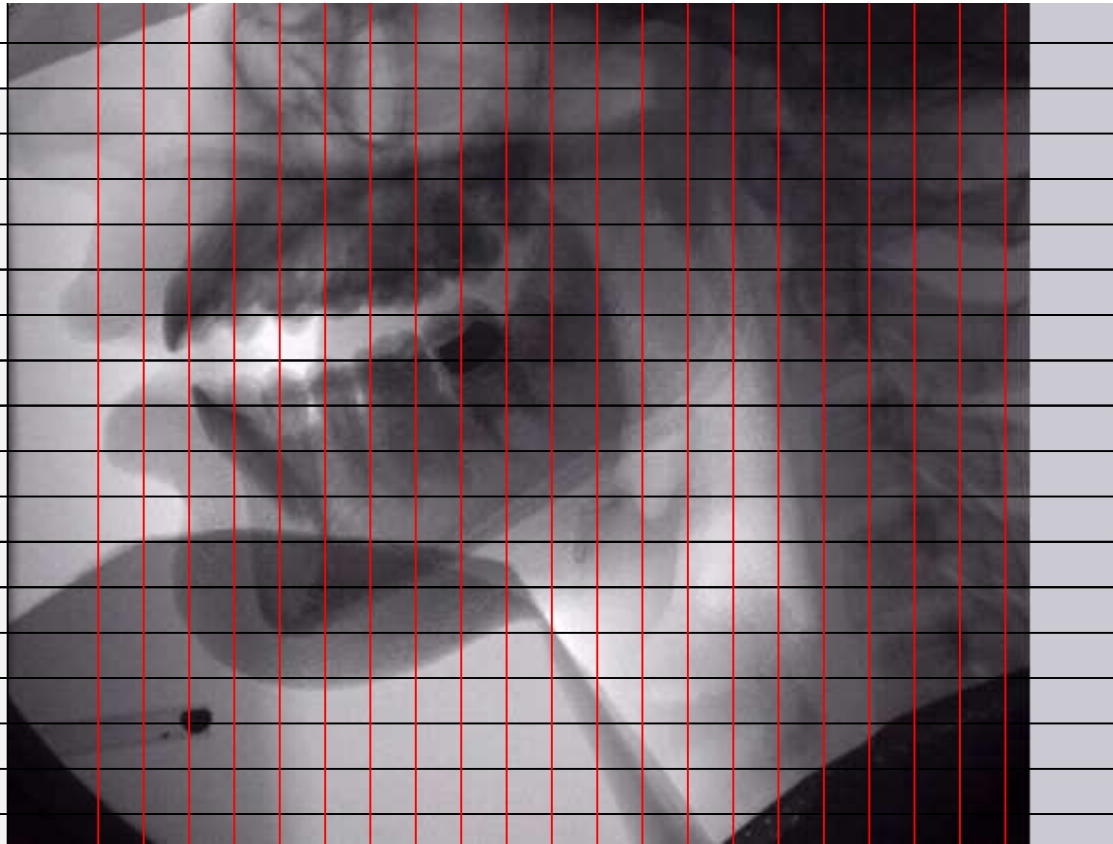
How done?

Demo

vowels



consonants



Other uses

**t4**

Closure

Lip continue to round



**/k/ first part**

**/aku/**

odel?

Why modelling?

Which models?

What is a **useful** model?

How done?

Demo

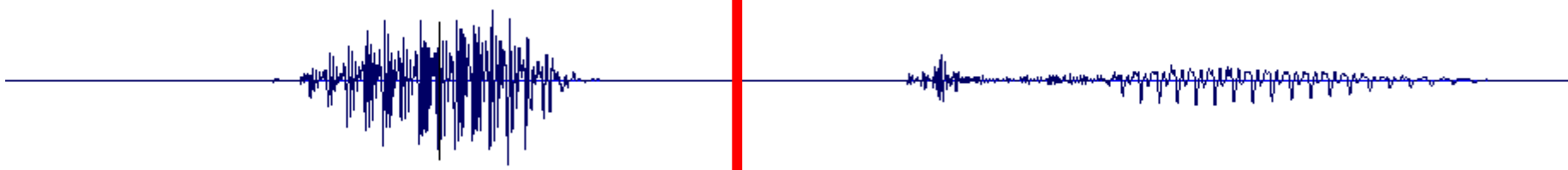
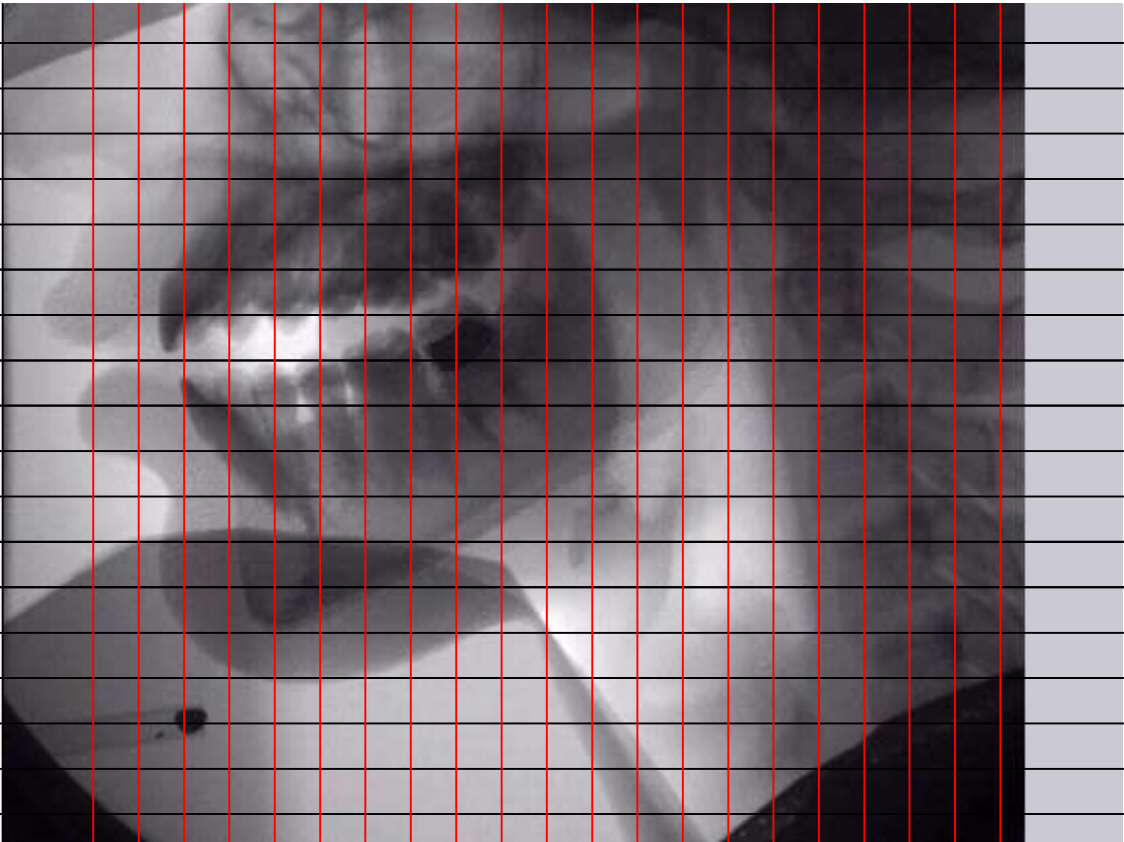
vowels



consonants

Other uses

t5



**/k/** midle

**/aku/**

odel?

Why modelling?

Which models?

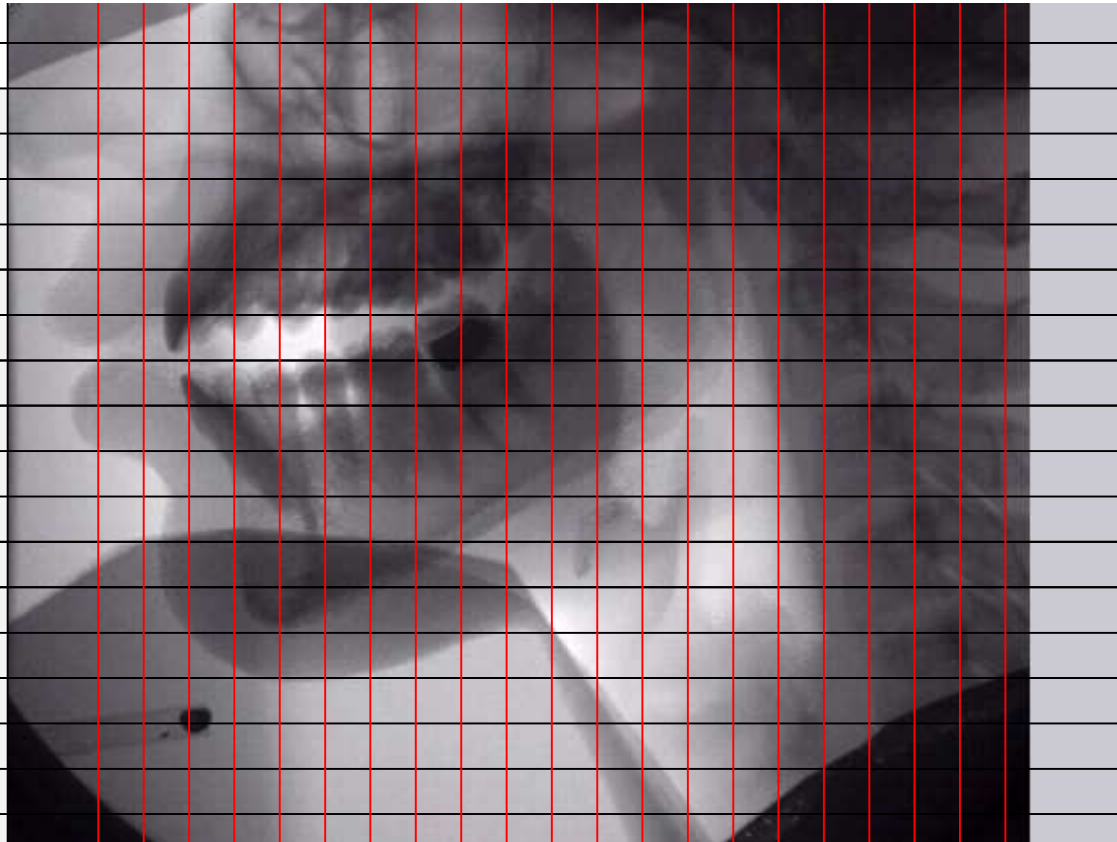
What is a **useful** model?

How done?

Demo

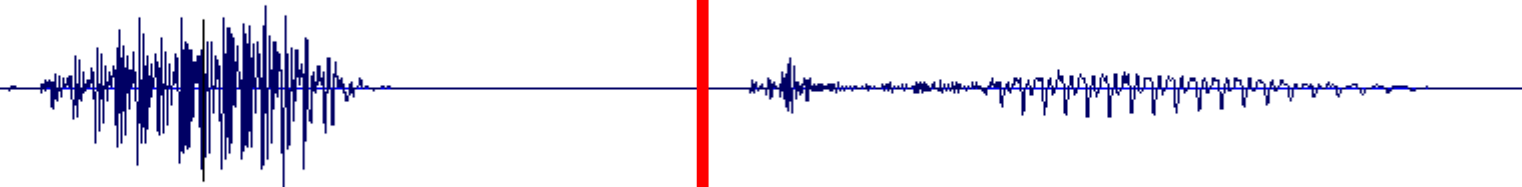
vowels

→ consonants



Other uses

t6



**/k/ before release**

**/aku/**

model?

Why modelling?

Which models?

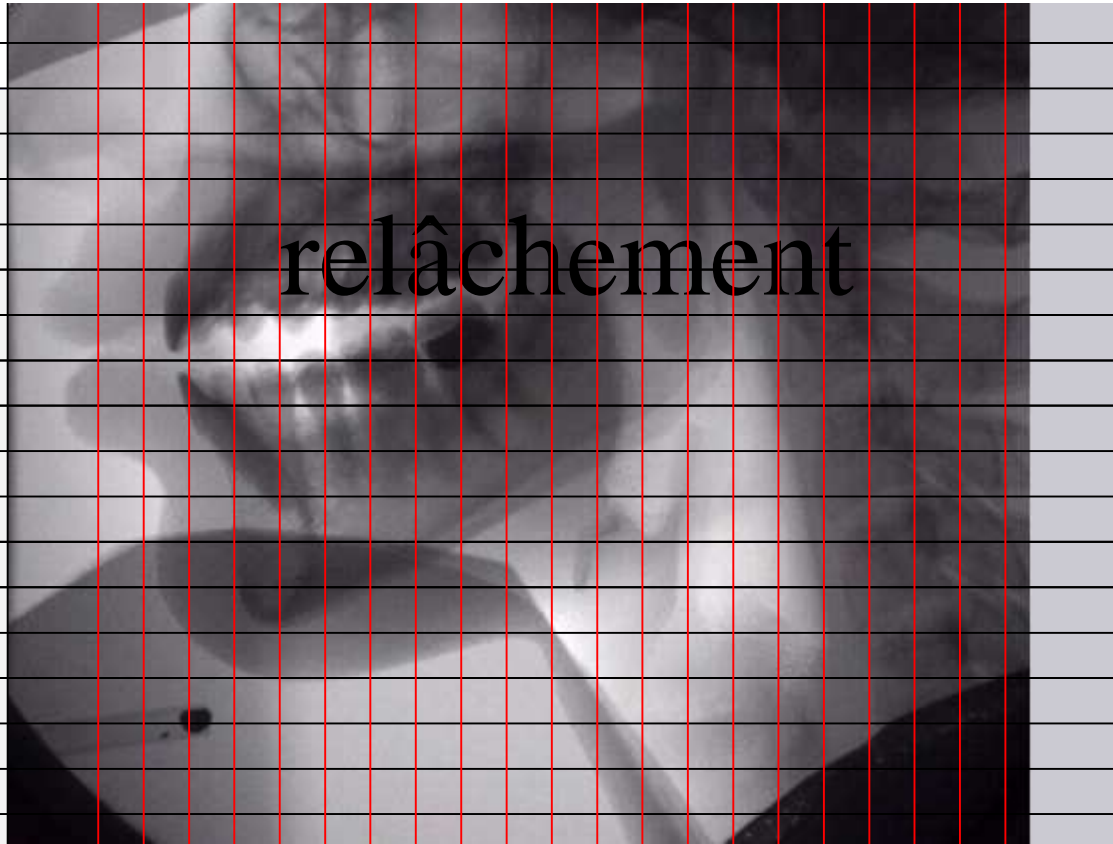
What is a **useful**  
model?

How done?

Demo

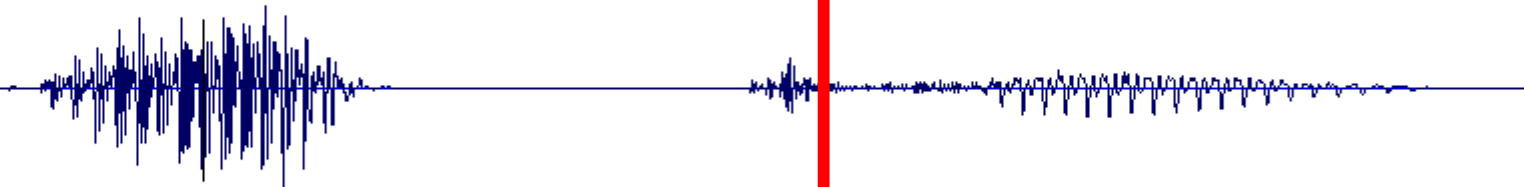
vowels

→ consonants



Other uses

**t7**



**/k/** after release



**/aku/**

odel?

Why modelling?

Which models?

What is a **useful** model?

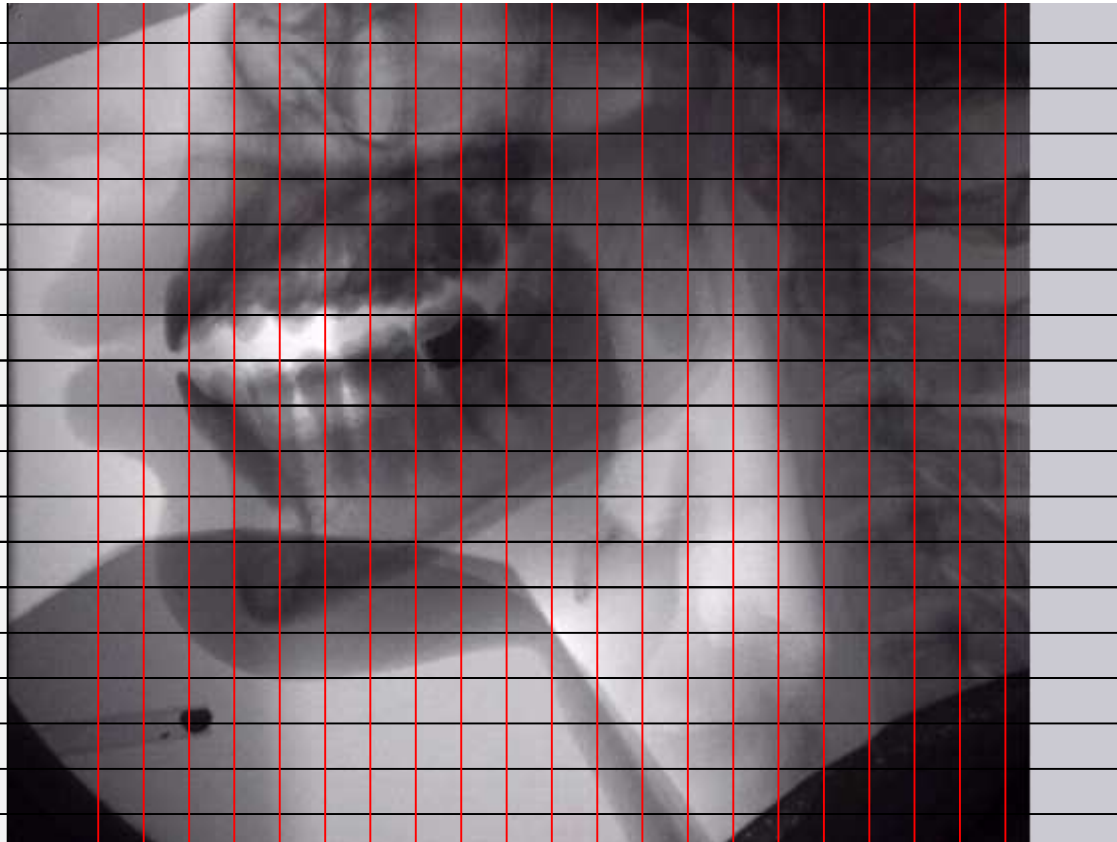
How done?

Demo

vowels



consonants



Other uses

t8



**/k/ friction**

**/aku/**

Model?

Why modelling?

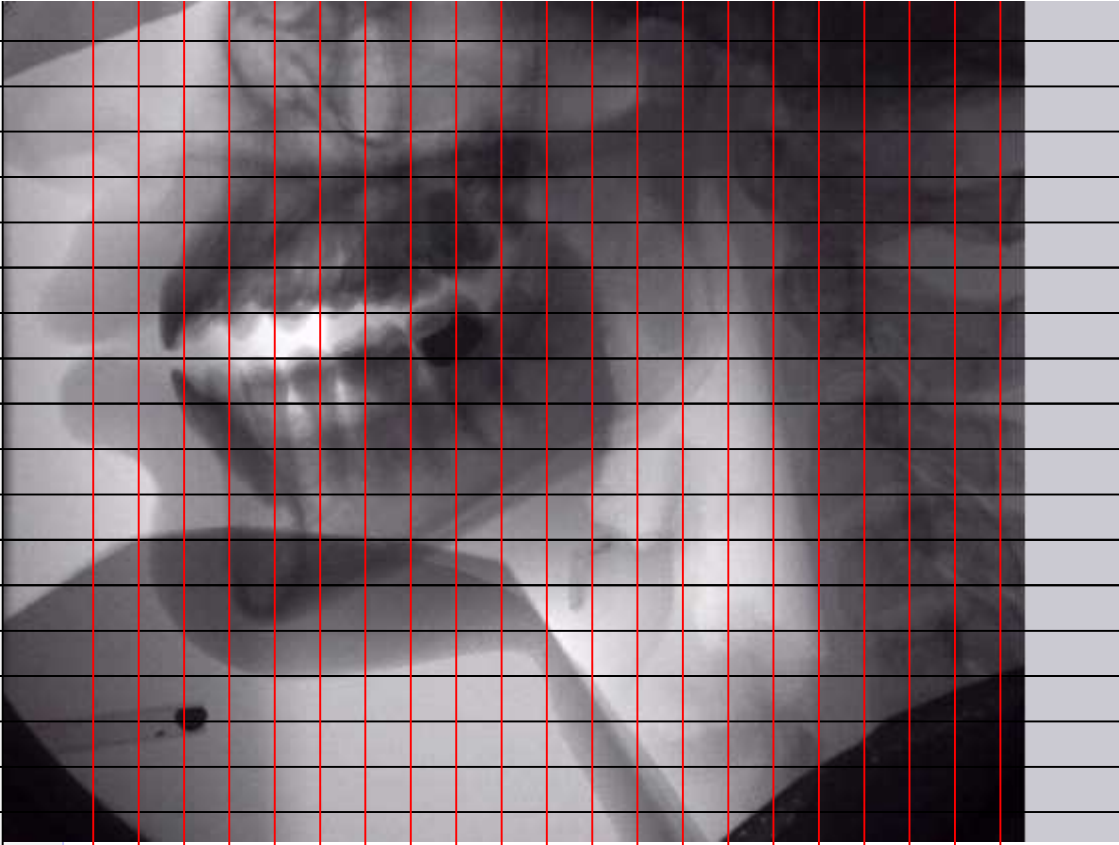
Which models?

What is a **useful** model?

How done?

**Demo**  
vowels  
→ consonants

Other uses



**t9**

Maximum rounding



**/u/ onset**

**/aku/**

Model?

Why modelling?

Which models?

What is a **useful** model?

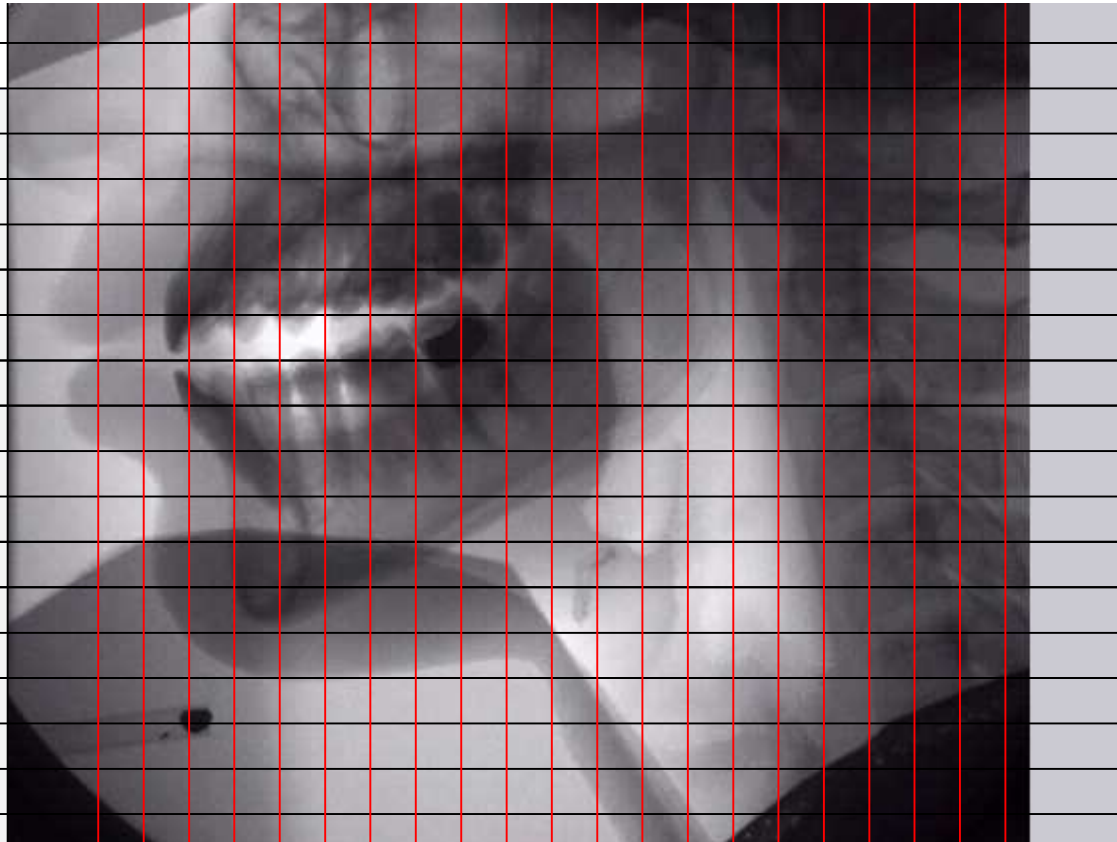
How done?

Demo

vowels



consonants



Other uses

**t10**

Velum starts to lower  
Derounding starts

**/u/ middle**

**/aku/**

odel?

Why modelling?

Which models?

What is a **useful**  
model?

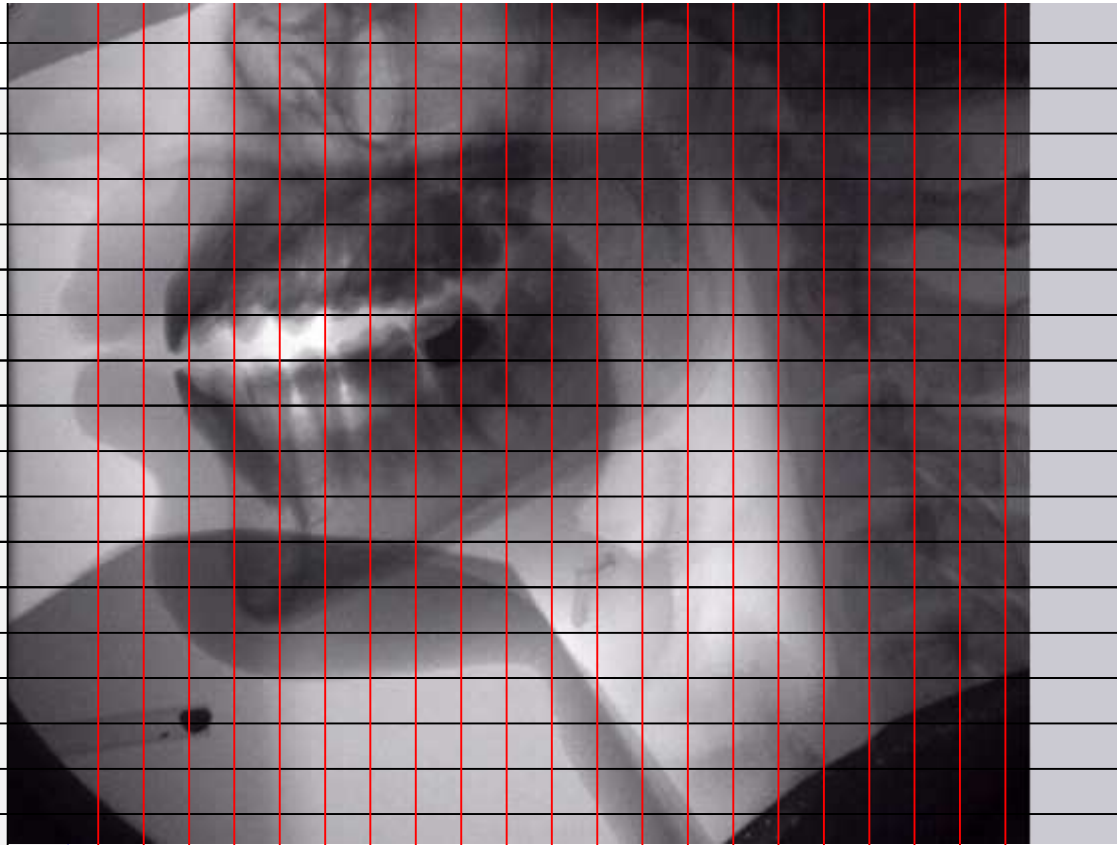
How done?

Demo

vowels



consonants



Other uses

**t11**

Larynx maximally low



**/u/ end**

**/aku/**

Model?

Why modelling?

Which models?

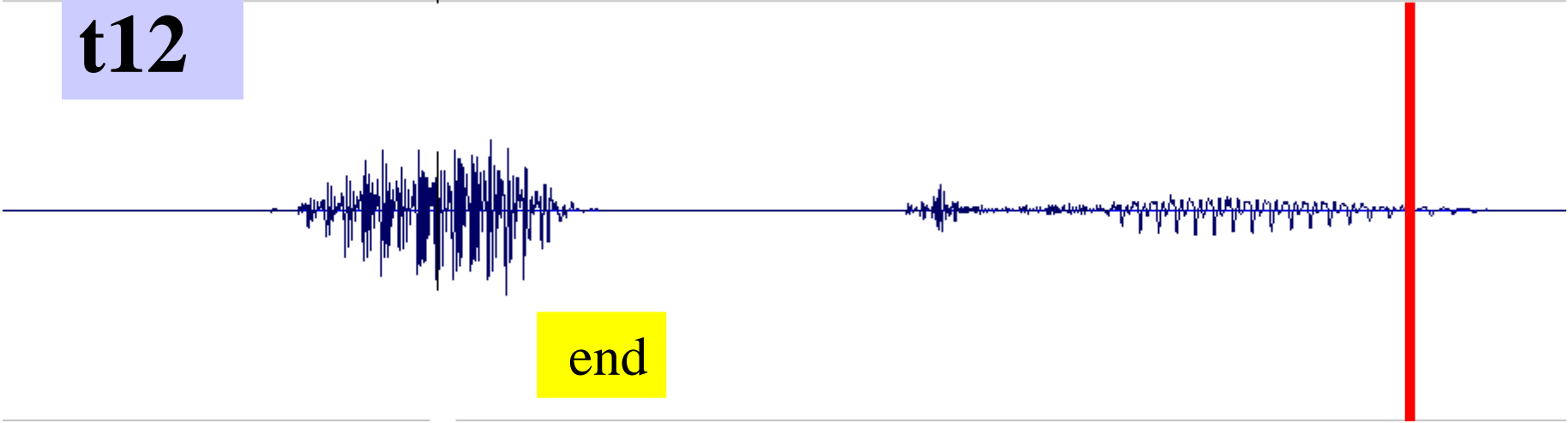
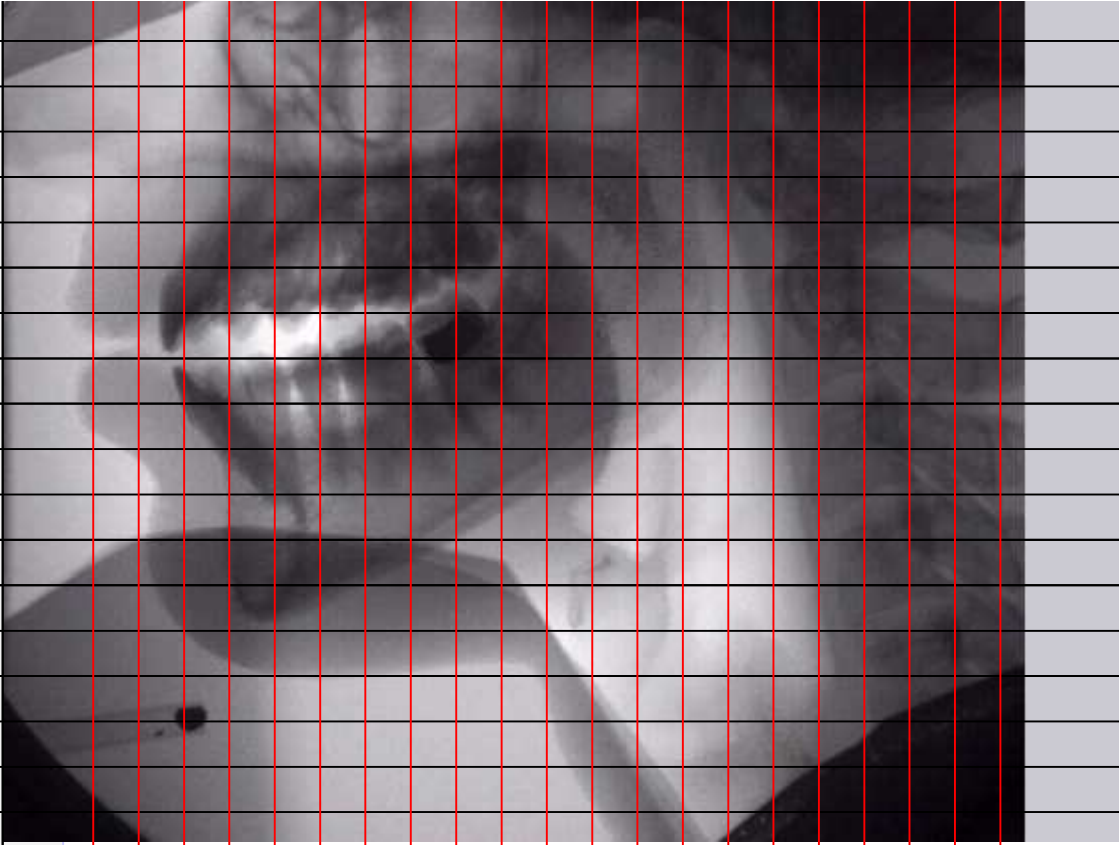
What is a **useful** model?

How done?

**Demo**  
vowels  
→ consonants

Other uses

**t12**



**end**

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*

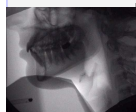
*Acoustic Theory of Speech production*



Ps, Pio

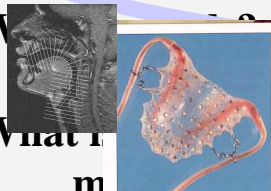
1) DATA

What is a model?




**production**

What is a model?



What is a model?



How does it work?

Demo

phonetician

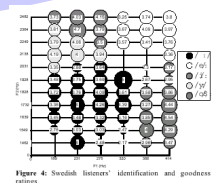
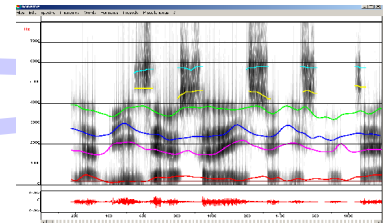
Other uses

- speaking
- speaking
- singing
- formant inversion
- etc.

**production**

**acoustics**

**perception**



linguist

Language teacher

2) Coherent teaching  
Research tool

engineer

Speech therapist

Medical students

Modellisation



**What is a model?**

**Why modelling?**

**Which models?**

**What is a useful  
model?**

**How done?**

**Demo**

vowels

consonants

**Other uses**

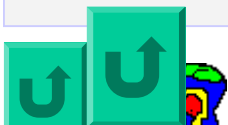
speaker anatomy

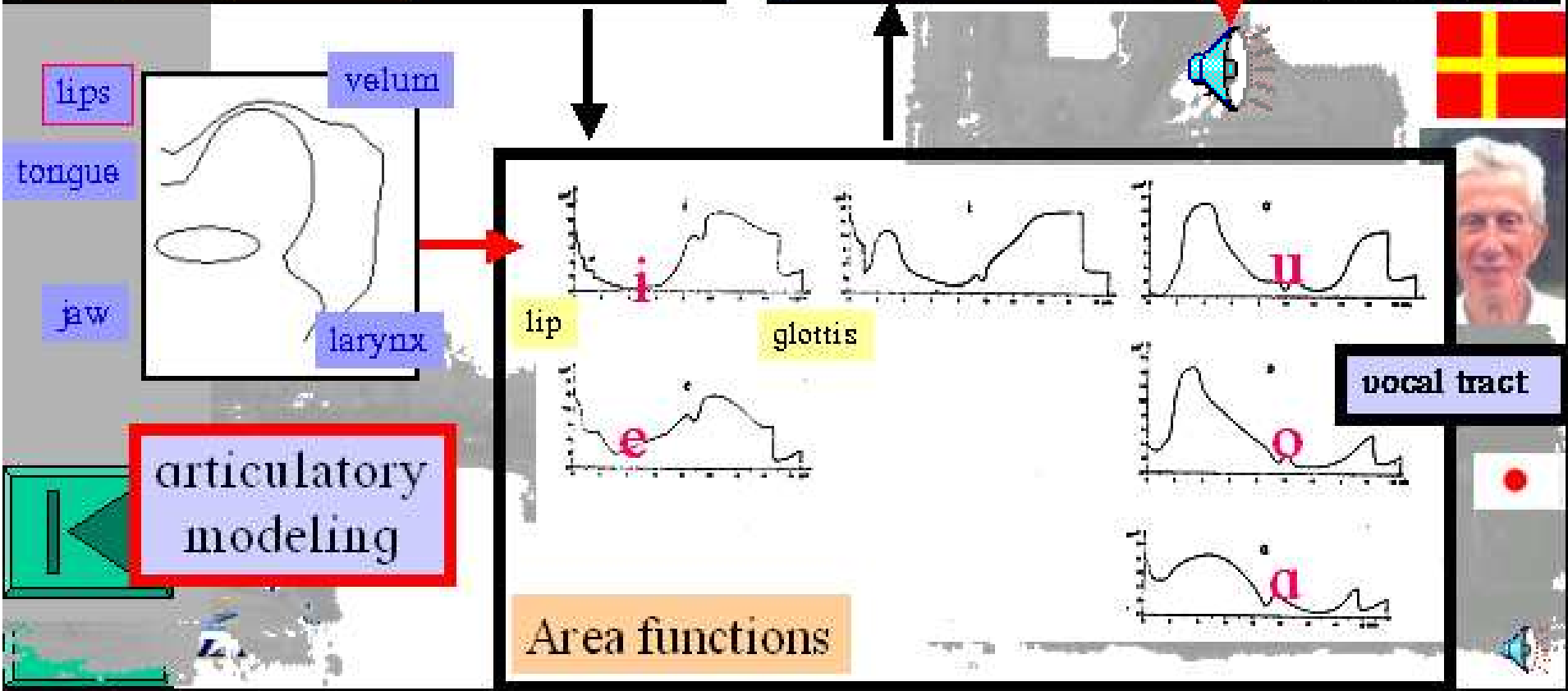
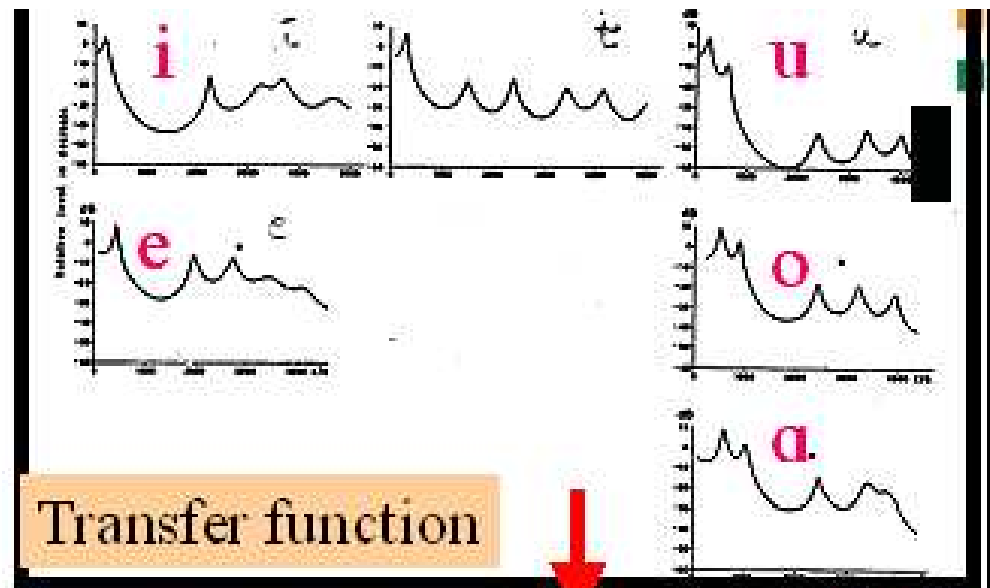
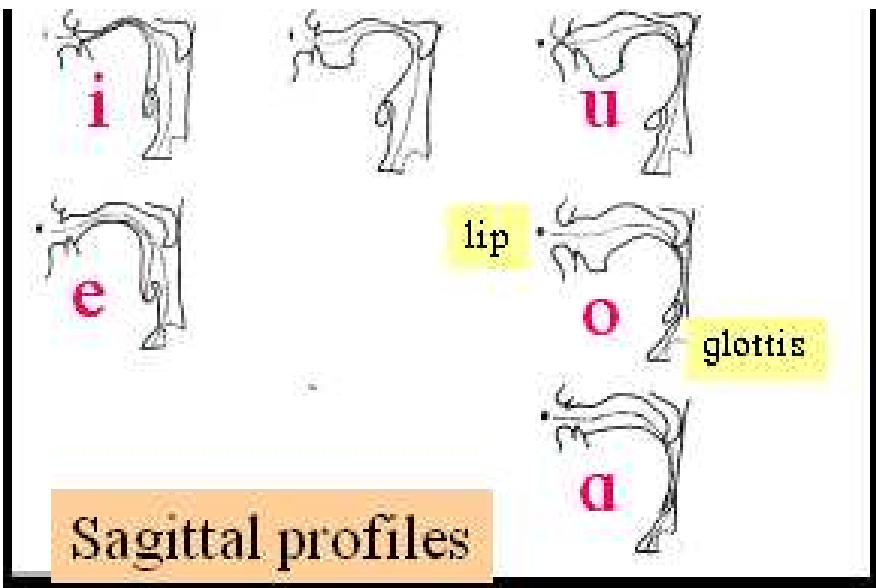
speaker strategy

singing formant

inversion

etc;





## 2) For what concerns the vowels

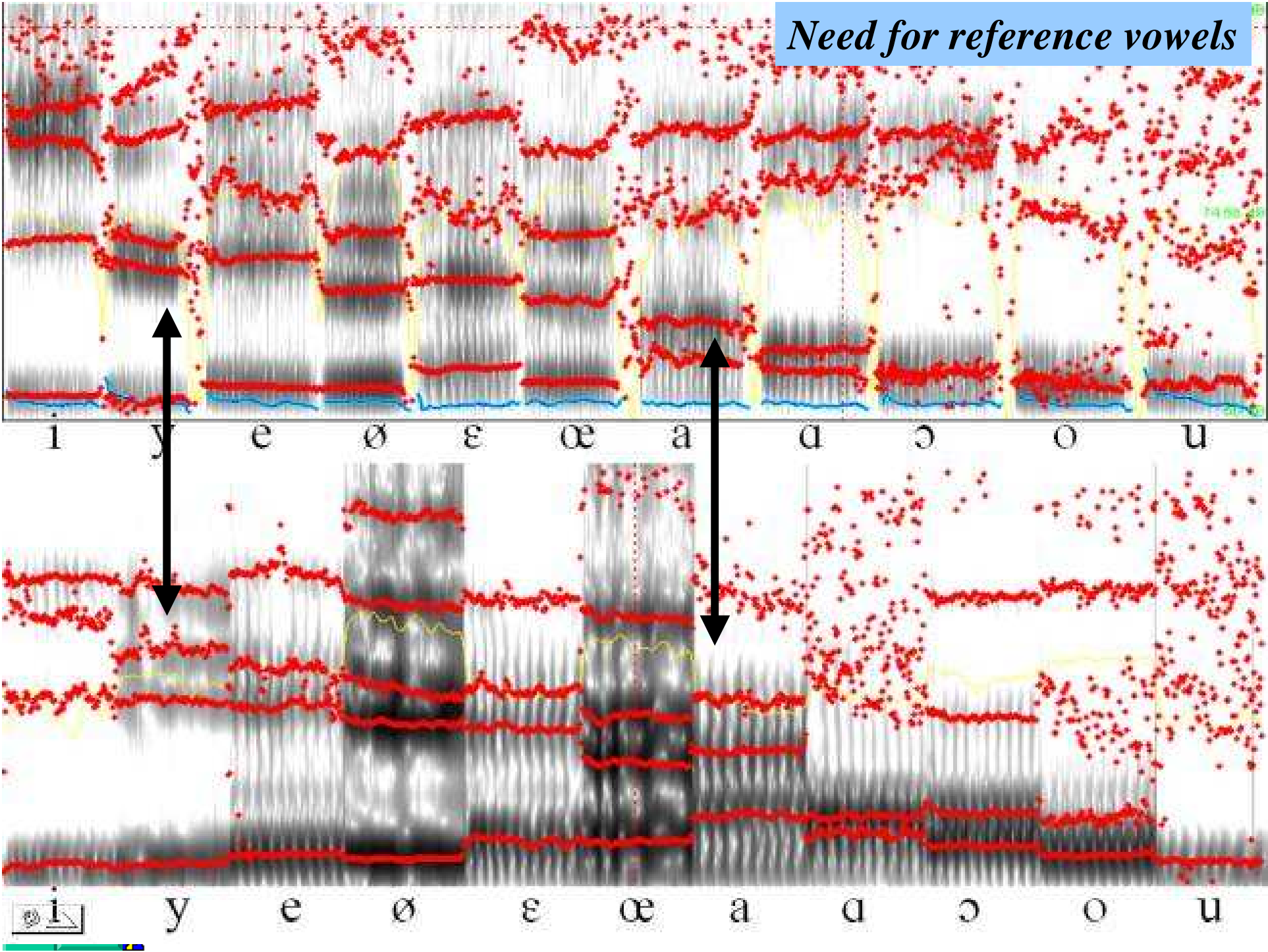
*a) Need for reference vowels*

*Cardinal vowels*

*Sensibility to formant amplitude and bandwidth*

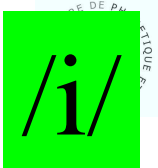
*AM could help*

*Need for reference vowels*



observations

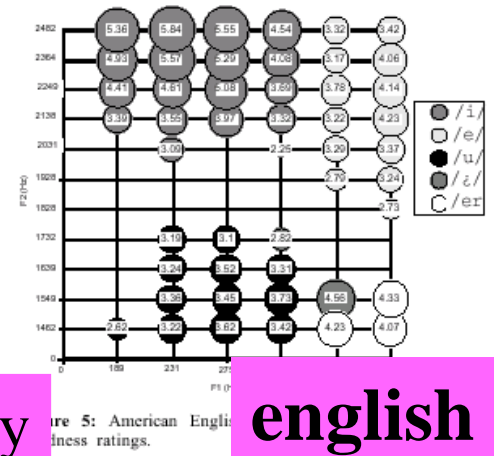
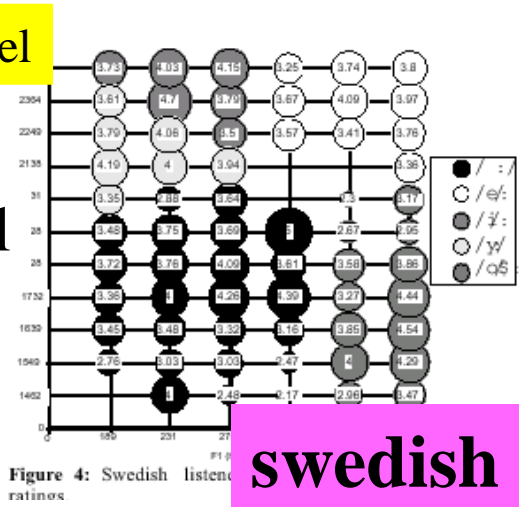
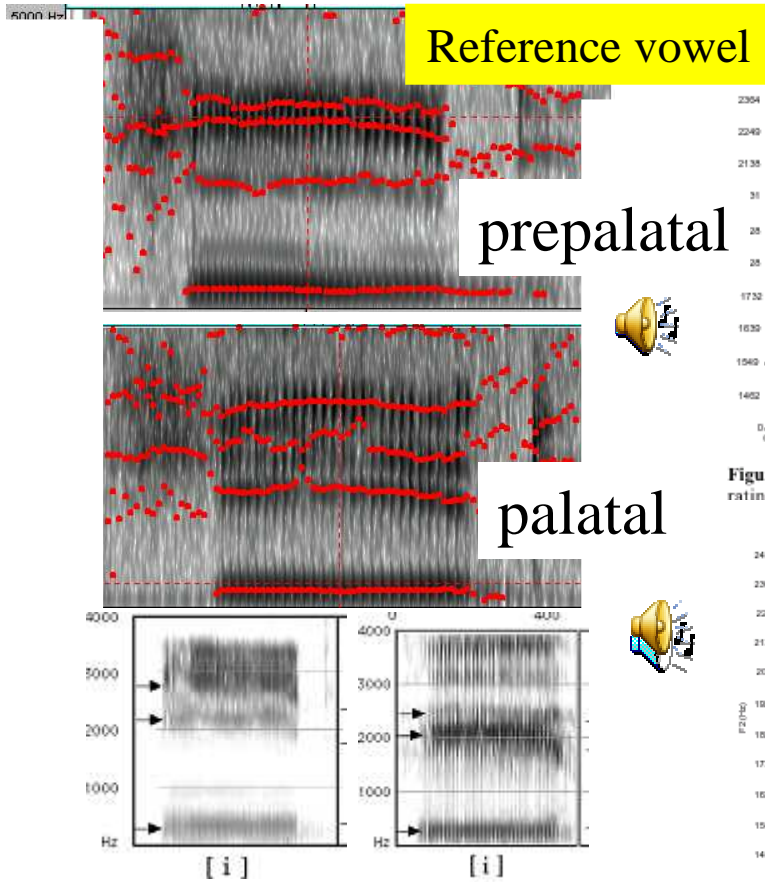
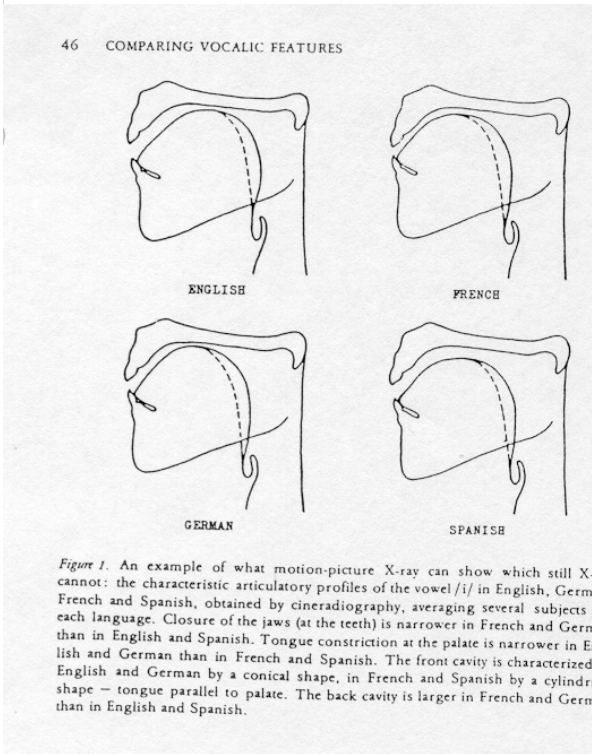
Need for reference vowels



articulatory

acoustic

perceptual



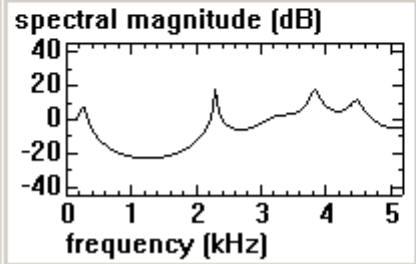
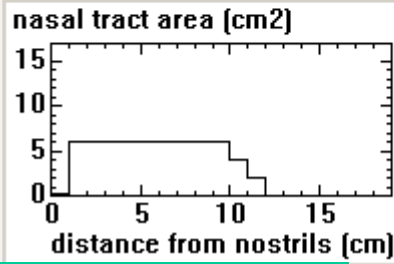
1) Helps to represent in a comprehensive way the language-specific prototypes



# articulatory



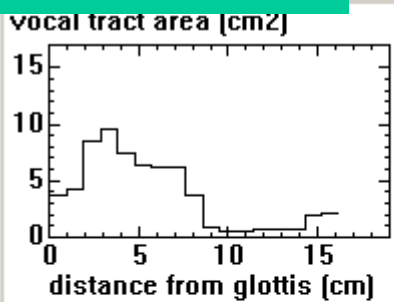
number of sections=17	<input type="checkbox"/>
nasal branch section=9	<input checked="" type="checkbox"/>
nasal coupling (cm <sup>2</sup> )=0	<input type="checkbox"/>
jaw=0.5	<input type="checkbox"/>
tongue=-2	<input type="checkbox"/>
shape=1	<input type="checkbox"/>
apex=-2	<input type="checkbox"/>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



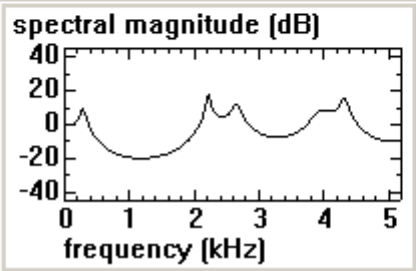
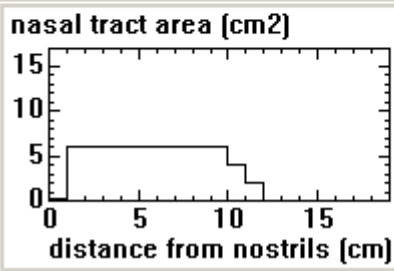
Frq(Hz)	Bw(Hz)	A(dB)
F1	248	79
F2	2290	29
F3	3838	109
F4	4488	168

# acoustic

# simulation

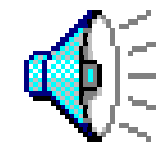
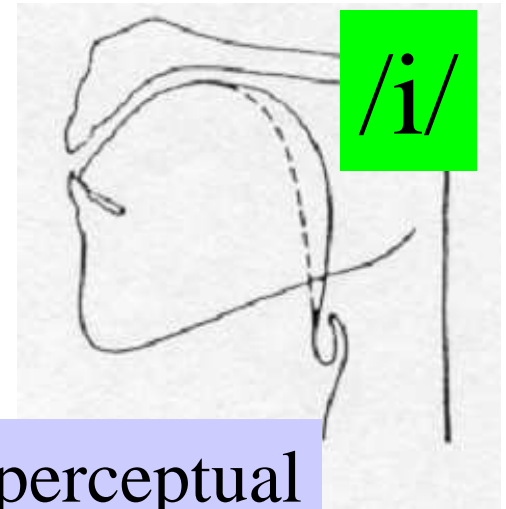


number of sections=17	<input type="checkbox"/>
nasal branch section=9	<input checked="" type="checkbox"/>
nasal coupling (cm <sup>2</sup> )=0	<input type="checkbox"/>
jaw=4.4	<input type="checkbox"/>
tongue=0.4	<input type="checkbox"/>
shape=1.5	<input type="checkbox"/>
apex=-2.8	<input type="checkbox"/>
lip_ht=1	<input type="checkbox"/>
lip_pr=-1	<input type="checkbox"/>
larynx=0	<input type="checkbox"/>



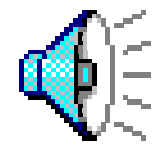
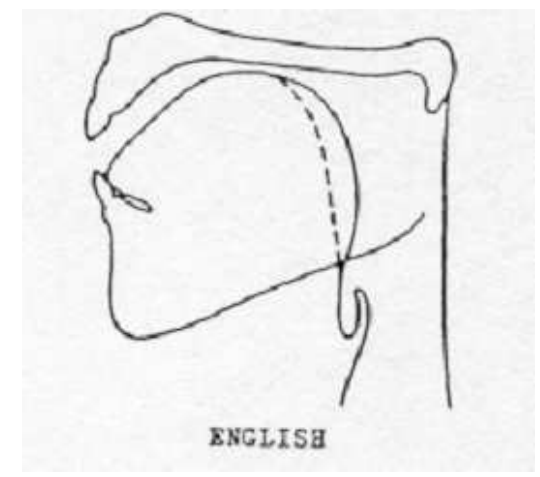
Frq(Hz)	Bw(Hz)	A(dB)
F1	279	73
F2	2229	48
F3	2662	118
F4	3962	203
F5	4333	92

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



# perceptual

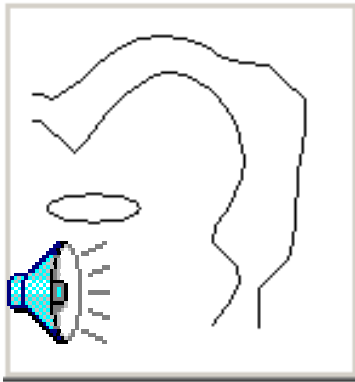
F1 (↑F3. F4)



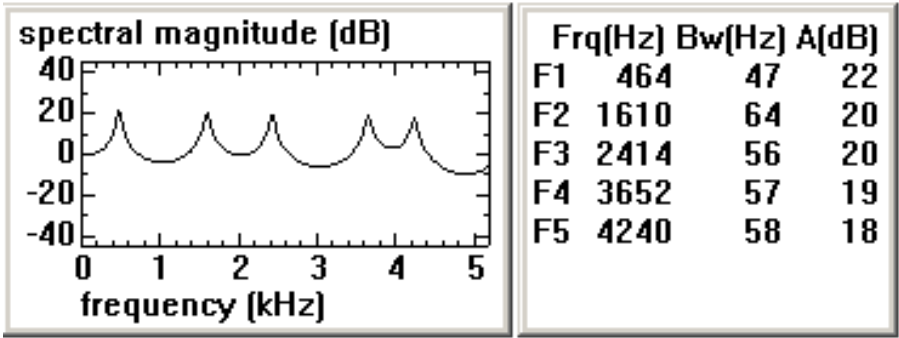
F1 ↑ F2



*Need for reference vowels*



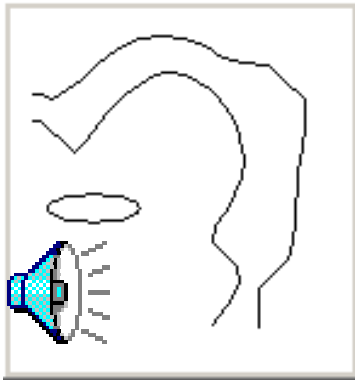
<b>neutral</b>		s=17	<input type="checkbox"/>
		ion=9	<input type="checkbox"/>
		(h2)=0	<input type="checkbox"/>
jaw=-1	<input type="checkbox"/>	lip_ht=0	<input type="checkbox"/>
tongue=-0.5	<input type="checkbox"/>	lip_pr=0	<input type="checkbox"/>
shape=0.5	<input type="checkbox"/>	larynx=0	<input type="checkbox"/>
apex=-2	<input type="checkbox"/>		



- vowels
- consonants
- Other uses
  - speaker anatomy
  - speaker strategy
  - singing formant
  - inversion
  - etc;

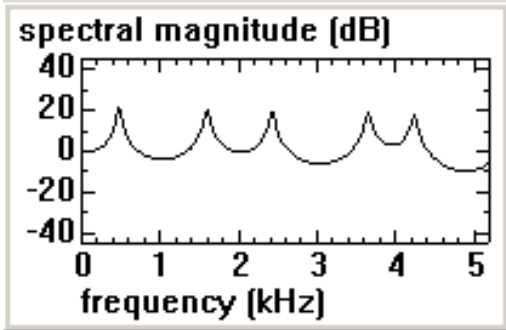




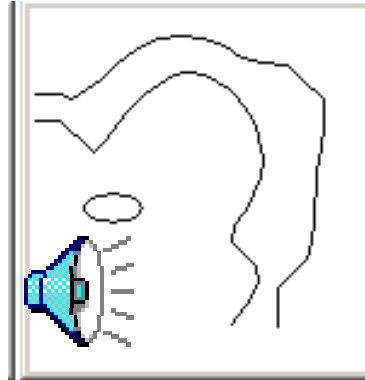


neutral

number of sections=17	<input type="checkbox"/>
nasal branch section=9	<input type="checkbox"/>
nasal protrusion=0	<input type="checkbox"/>
lip_ht=0	<input type="checkbox"/>
lip_pr=0	<input type="checkbox"/>
larynx=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"/>

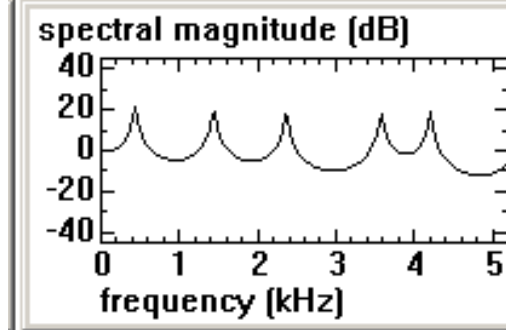


Frq(Hz)	Bw(Hz)	A(dB)
F1	464	47
F2	1610	64
F3	2414	56
F4	3652	57
F5	4240	58



+ protrusion

number of sections=17	<input type="checkbox"/>
nasal branch section=9	<input type="checkbox"/>
nasal protrusion=1	<input type="checkbox"/>
lip_ht=0	<input type="checkbox"/>
lip_pr=1.1	<input type="checkbox"/>
larynx=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"/>



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

vowels  
consonants

Other uses

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

Need for reference vowels

( F2 ↓ ↓ ↓ F3 )



ion



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"/>

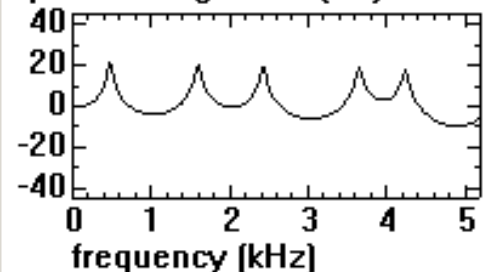
and

+ 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=0	<input type="checkbox"/>
lip_pr=11	<input type="checkbox"/>
lip_pr .11	<input type="checkbox"/>
-5 <3< 5	<input type="checkbox"/>

+ compression

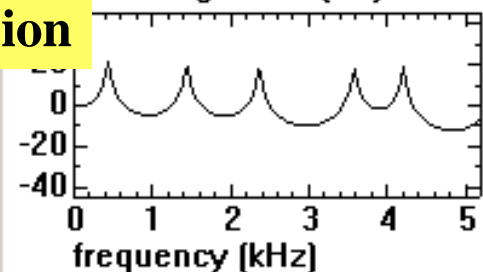
spectral magnitude (dB)



Frq(Hz) Bw(Hz) A(dB)

Frq(Hz)	Bw(Hz)	A(dB)
F1 464	44	20
F2 1610	64	20
F3 2414	56	20
F4 3652	57	19
F5 4240	58	18

spectral magnitude (dB)



Frq(Hz) Bw(Hz) A(dB)

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

+ compression  
+ protrusion

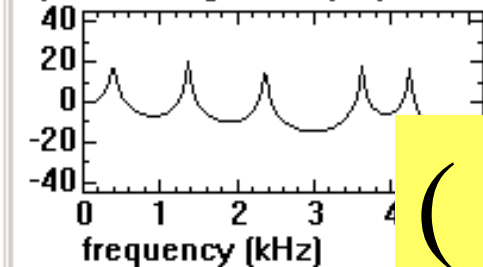
number of sections=17	<input type="checkbox"/>
nasal branch section=9	<input type="checkbox"/>
nasal coupling (cm <sup>2</sup> )=0	<input type="checkbox"/>
jaw=-1	<input type="checkbox"/>
tongue=-0.5	<input type="checkbox"/>
shape=0.5	<input type="checkbox"/>
apex=-2	<input type="checkbox"/>
lip_ht=11	<input type="checkbox"/>
lip_pr=3	<input type="checkbox"/>
lip_pr .11	<input type="checkbox"/>
-5 <-0.9< 5	<input type="checkbox"/>

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=-0.9	<input type="checkbox"/>
lip_pr=3	<input type="checkbox"/>
lip_pr .11	<input type="checkbox"/>
-5 <-0.9< 5	<input type="checkbox"/>

LIP + TONGUE

( F2 ↓ ↓ F3 )

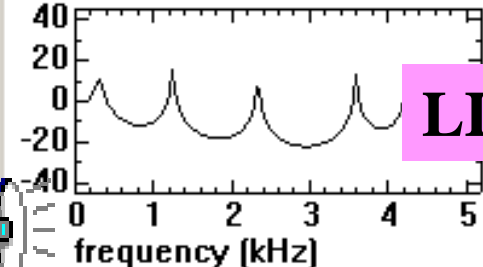
spectral magnitude (dB)



Frq(Hz) Bw(Hz) A(dB)

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

spectral magnitude (dB)



Frq(Hz) Bw(Hz) A(dB)

Frq(Hz)	Bw(Hz)	A(dB)
F1 310	68	11
F2 1238	21	16
F3 2352	42	14
F4 3652	57	19
F5 4240	20	11

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"/>

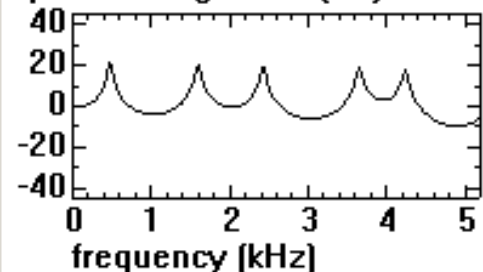
and

+ 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=0	<input type="checkbox"/>
lip_pr=11	<input type="checkbox"/>
lip_pr_min=-5	<input type="checkbox"/>
lip_pr_max=5	<input type="checkbox"/>

+ compression

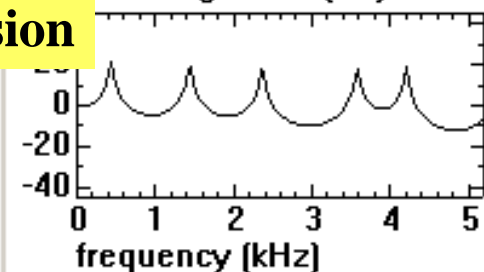
spectral magnitude (dB)



Frq(Hz) Bw(Hz) A(dB)

Frq(Hz)	Bw(Hz)	A(dB)
F1 464	44	20
F2 1610	64	20
F3 2414	56	20
F4 3652	57	19
F5 4240	58	18

spectral magnitude (dB)



Frq(Hz) Bw(Hz) A(dB)

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

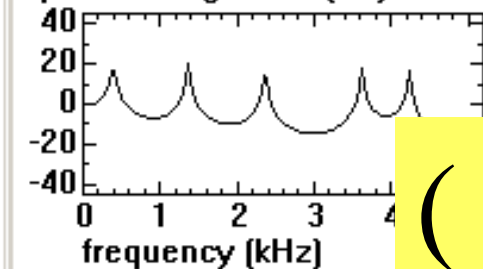
LIP

+ 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_min=-5	<input type="checkbox"/>
lip_pr_max=5	<input type="checkbox"/>

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=-0.9	<input type="checkbox"/>
lip_pr=3	<input type="checkbox"/>
larynx=0	<input type="checkbox"/>

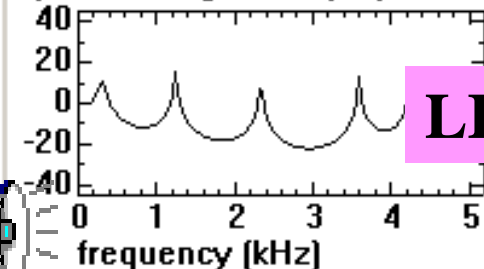
spectral magnitude (dB)



Frq(Hz) Bw(Hz) A(dB)

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

spectral magnitude (dB)



Frq(Hz) Bw(Hz) A(dB)

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

LIP + TONGUE

( F2 ↓ ↓ F3 )

Location of the constriction			
Lowest possible formant		Highest possible formant	
↓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?

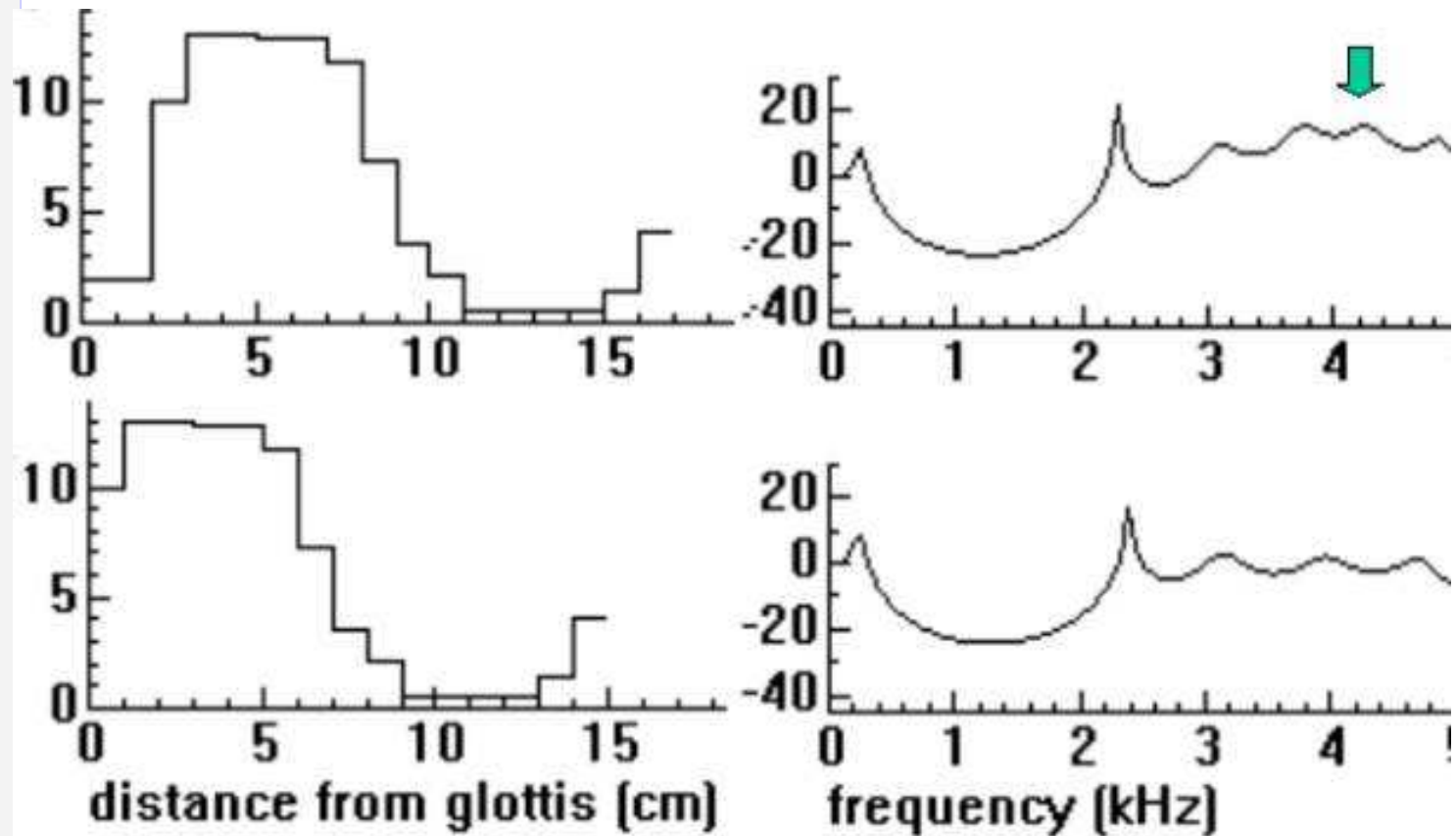
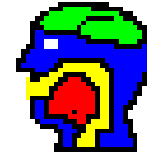
How done?

Demo

vowels  
consonants

Other uses

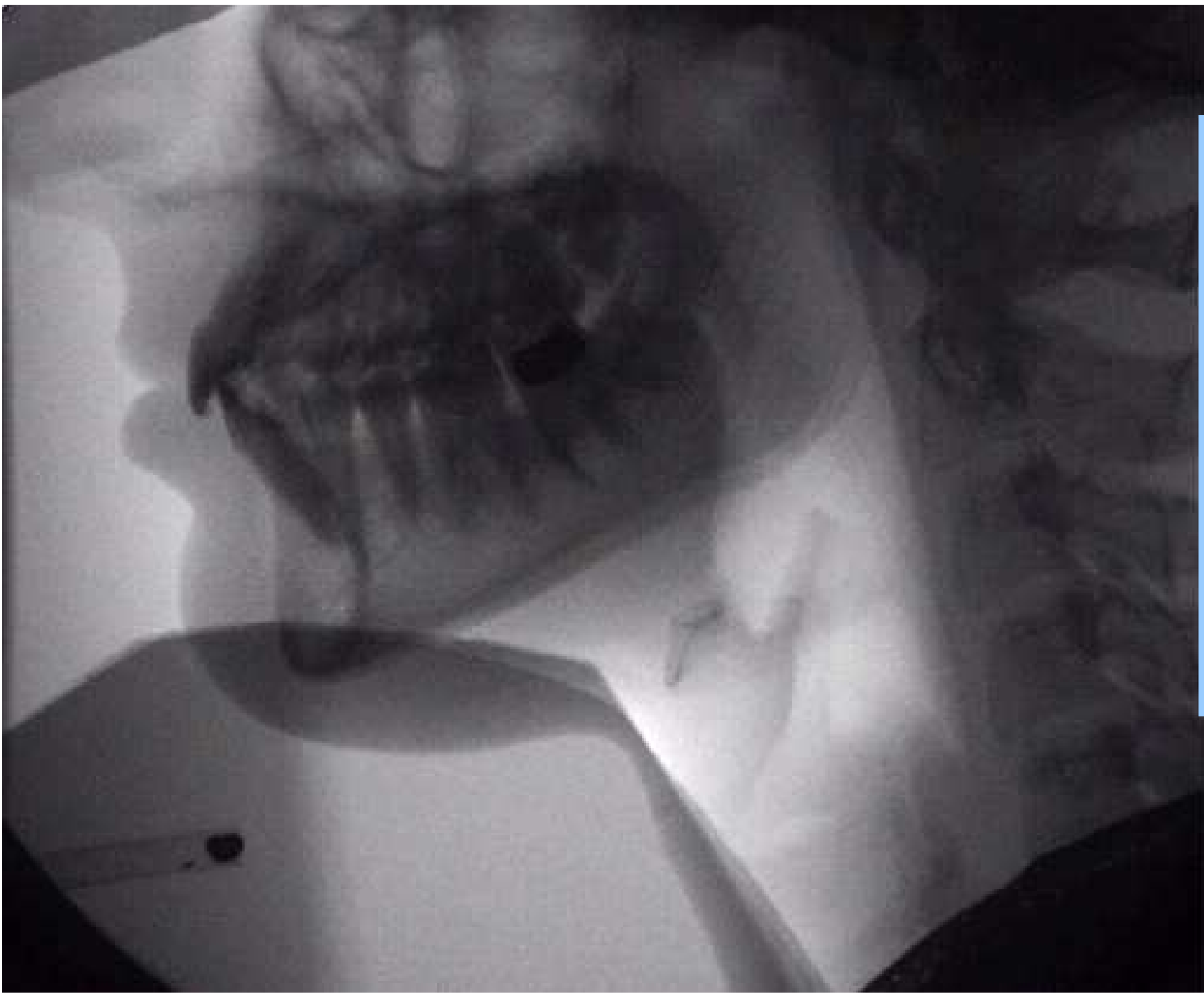
speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;



MODELSATION

02





Role of anatomical details



## 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

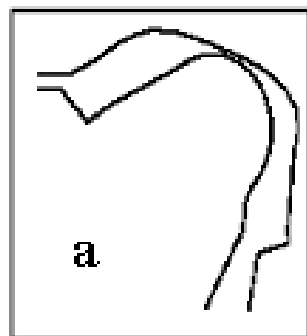
speaker strategy

singing formant

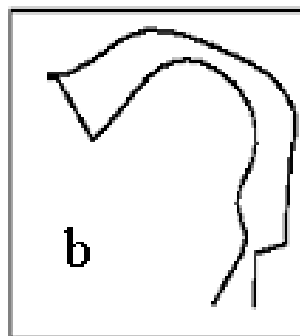
inversion

etc;

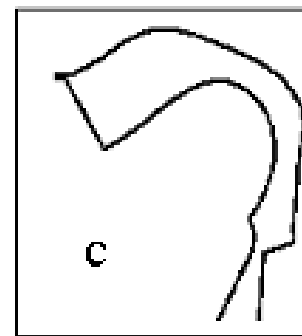
About the same sound (except the focal and extreme sounds can be obtained by Fairly different VT configuration



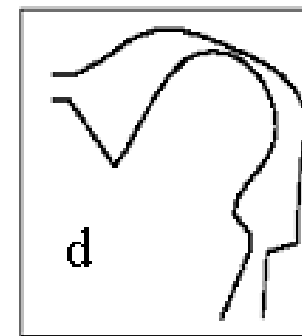
335 904 2150



341 986 2113



346 953 2118



352 949 2112



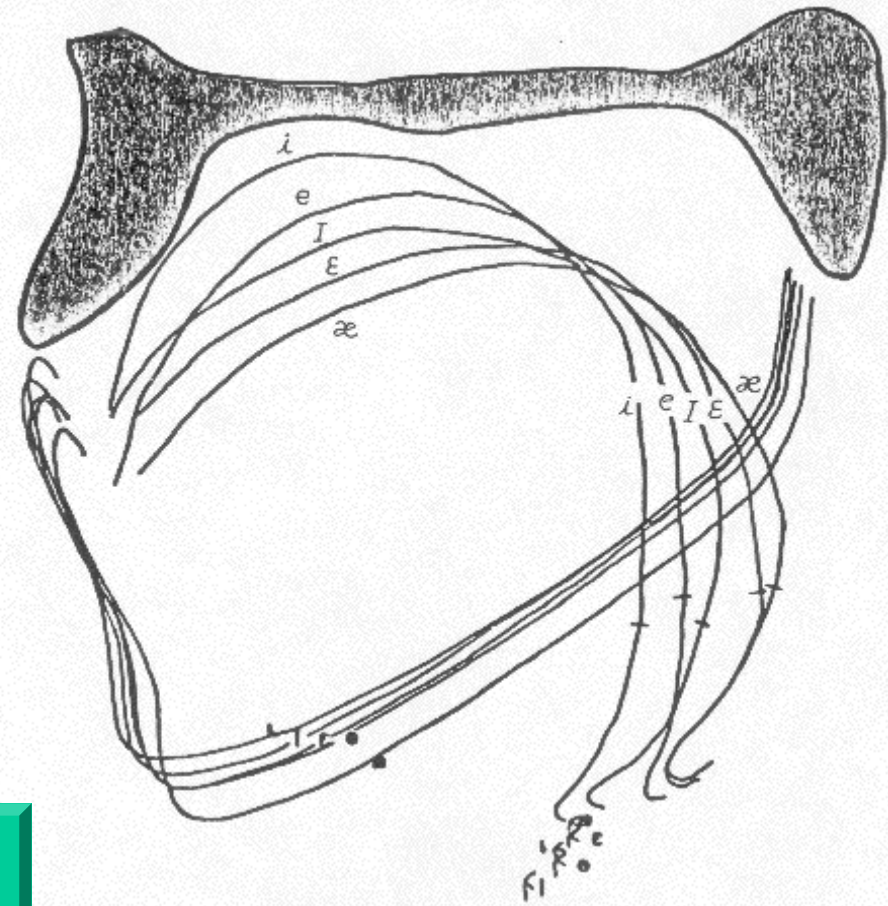
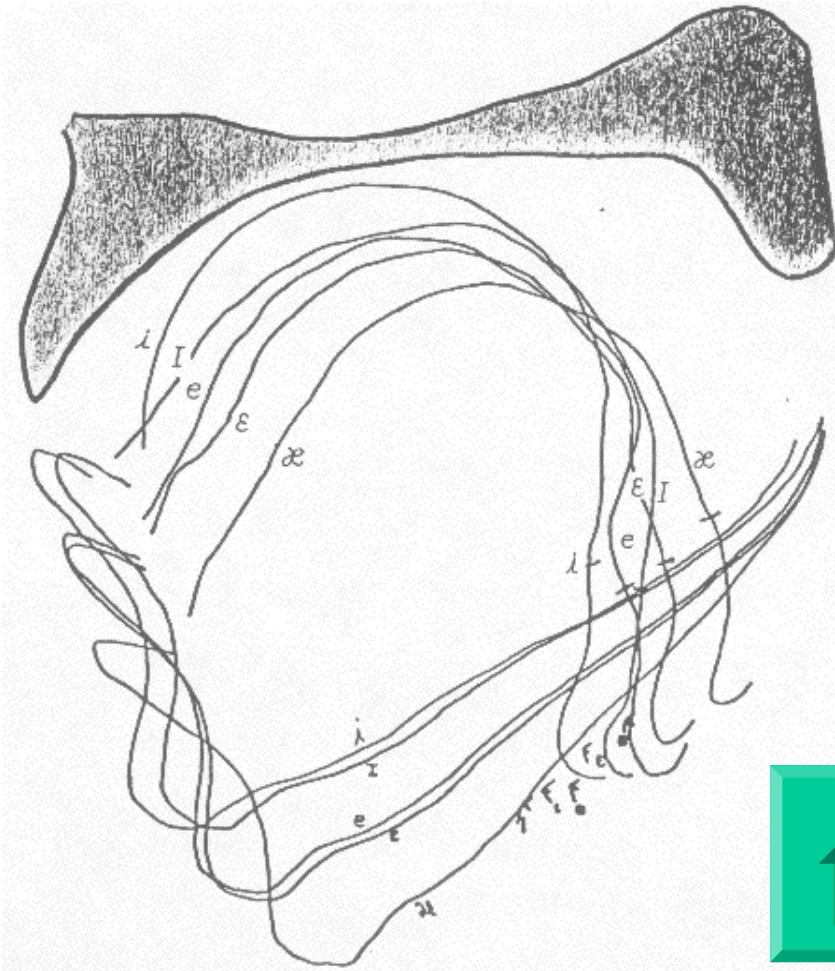
## 2) For what concerns the vowels

*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;

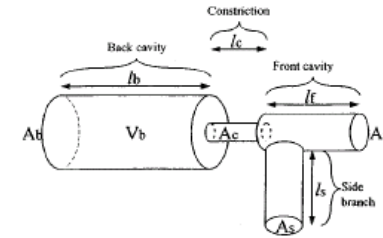
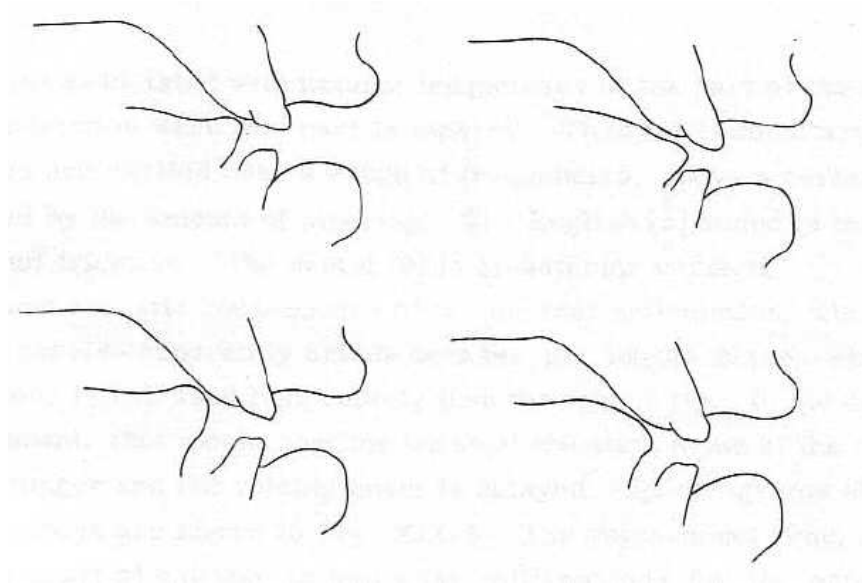
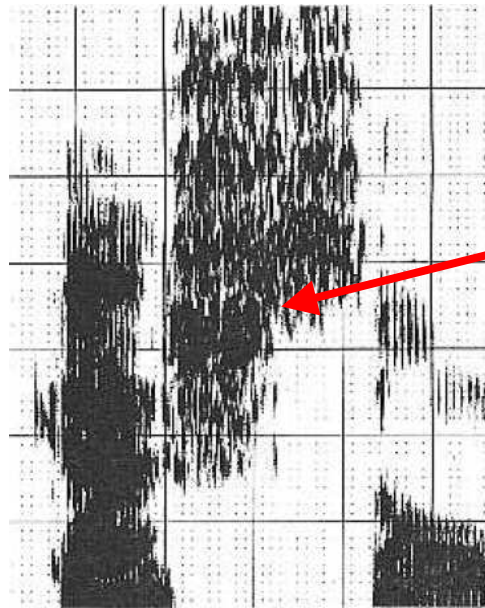


FIG. 5. The tube configuration simulated in the "weakly coupled" models.



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*

# Beans on a string synthesis

What is a model?

Why modelling?

Which models?

What is a useful model?

How done?

Demo

vowels  
consonants

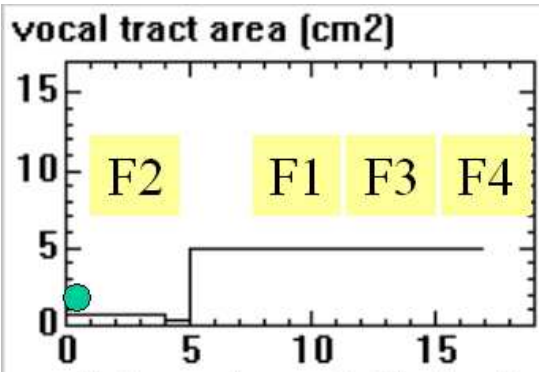
Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

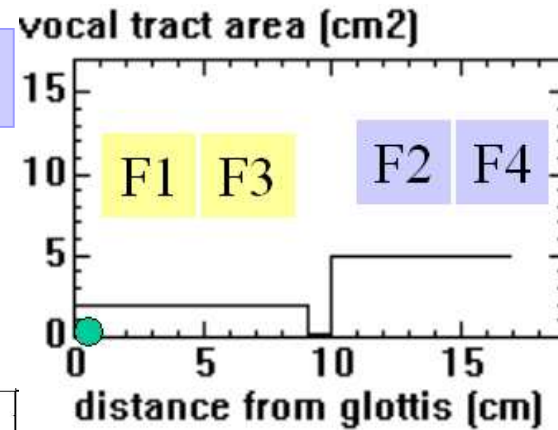
## Non French consonants

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





95%, 69\*



100%, 76

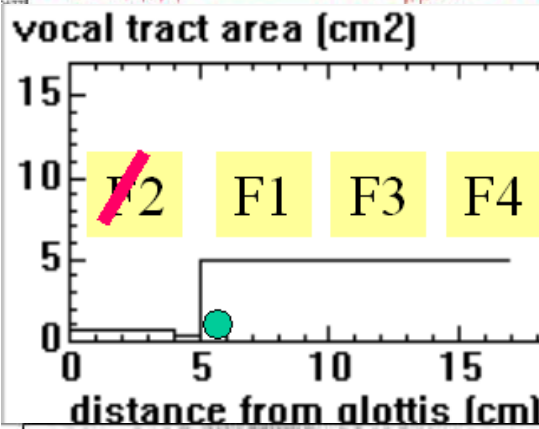


0.35

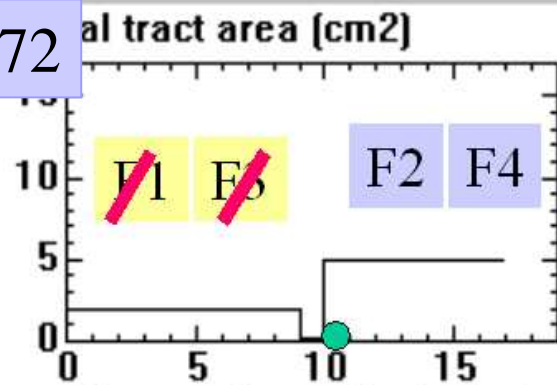
Glottis more open?



χ k



100%, 72



100%, 79



0.20

10 listeners  
10 repetitions

\*Alwan, 1989



74

# Beans on a string synthesis

W	• Pharyngeal sourde		
h	• 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
W	• Pharyngeale sonore		
ʕ	• 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
H	• [uvulaire sourde]		
χ	• 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 5.0
	• [uvulaire sonore]		
B	• 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 5.0
	• [uvulaire sourde]		

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**

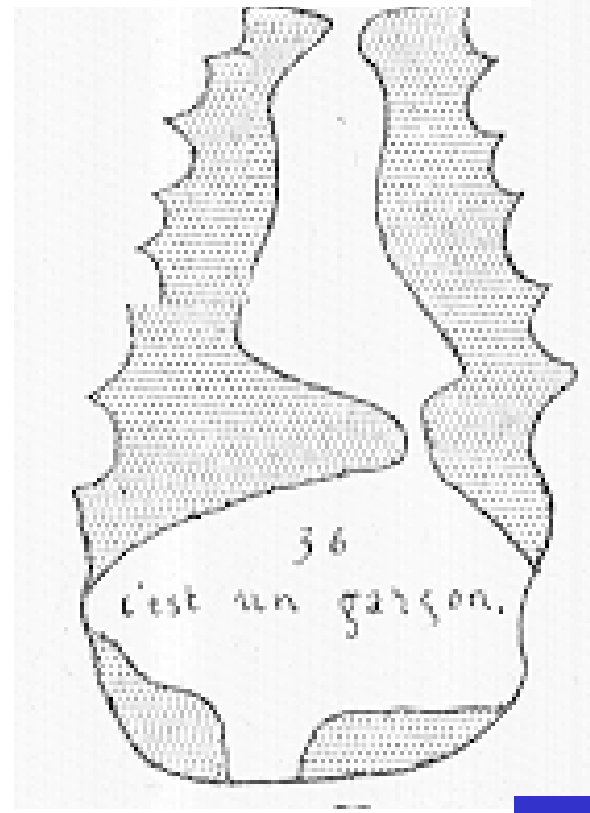
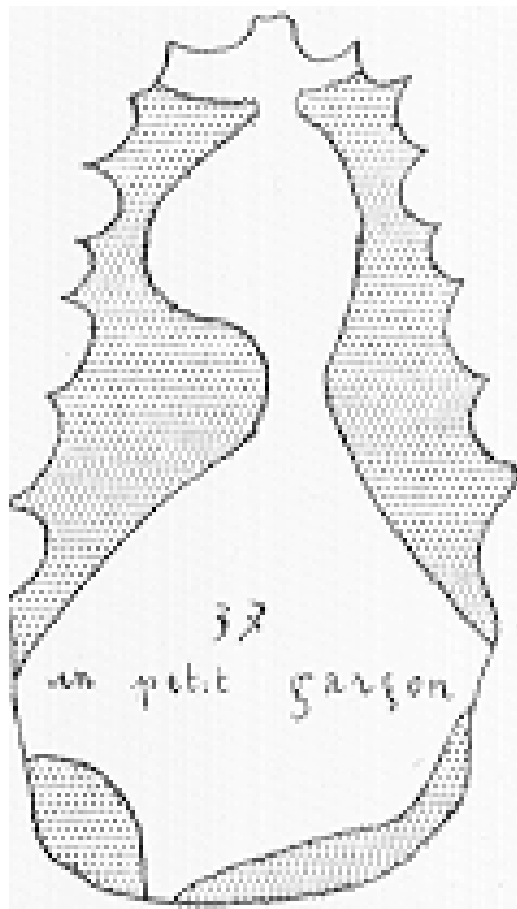
- **More close glottis**



- Ration between the two areas may change, and to the category of the consonant
- But the underlying F-pattern is the same



# Marguerite Durand



speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;

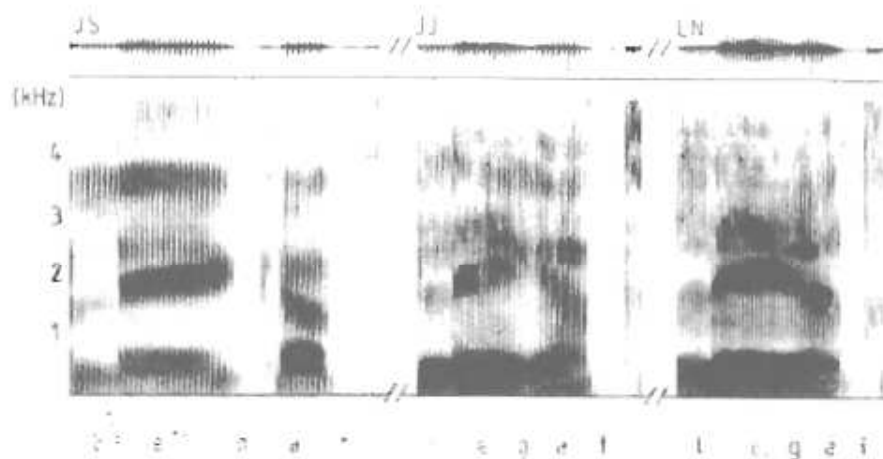


Figure 4.4.2. Normal and two degrees of target undershoot of a voiced stop.



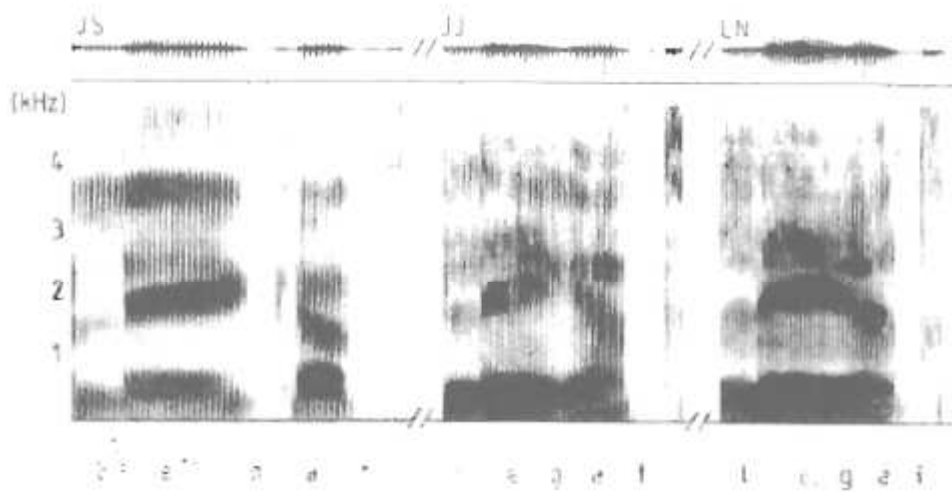
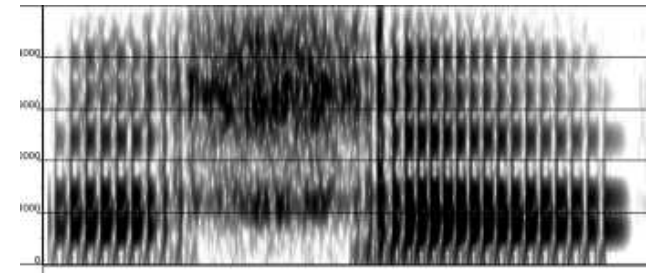
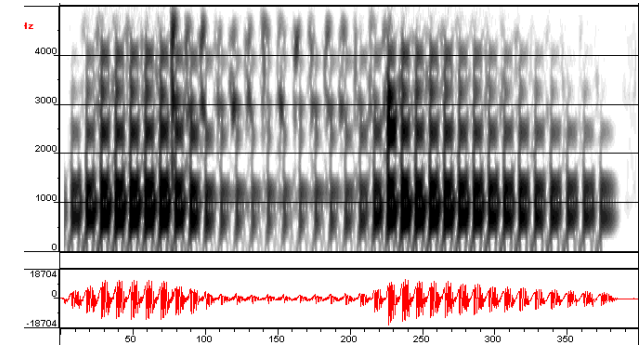


Figure 4.4.2. Normal and two degrees of target undershoot of a voiced stop.



Vt more close than vf

Less vf opening

Voiceless > voiced

Noise created at the narrowest

**How done?**

**Demo**

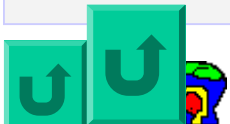
vowels  
consonants

**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 model?

Why model?

Which model?

What is a useful model?

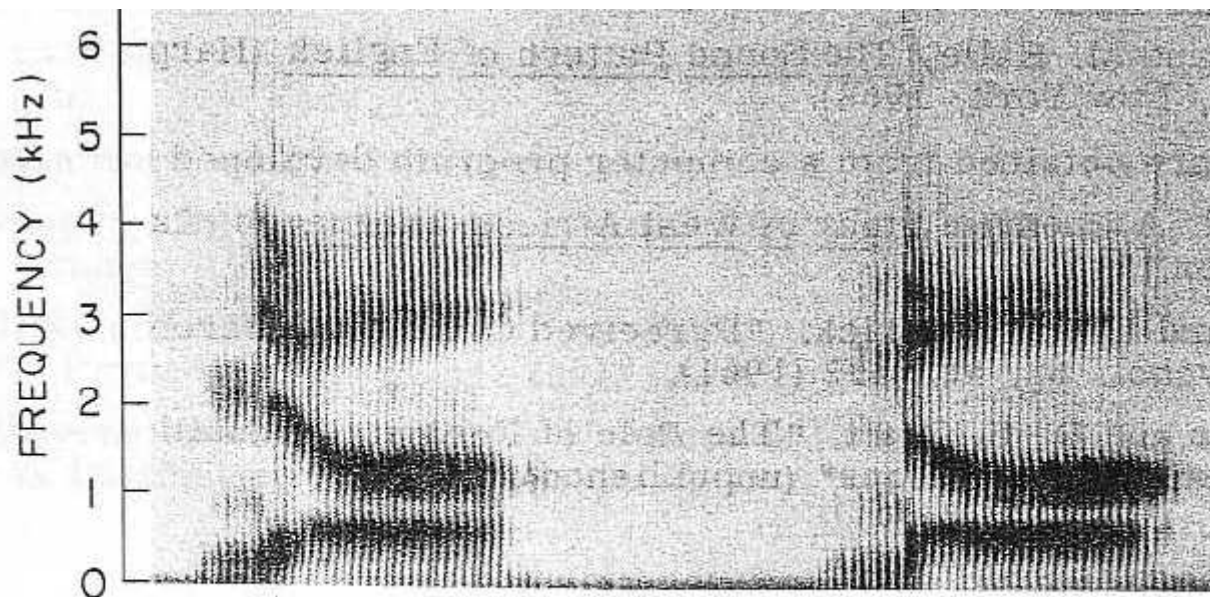
How does it work?

Demo

vowels  
consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;



laminal

apical

**F2**

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*

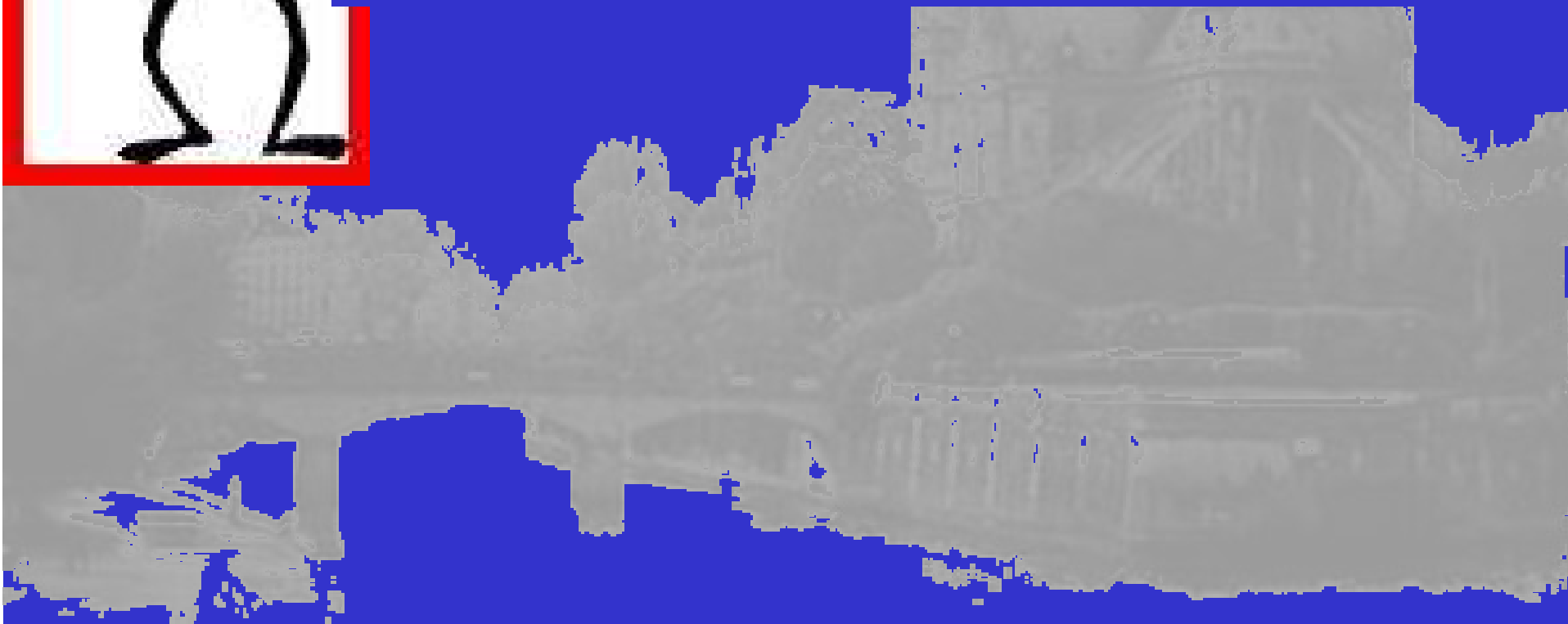
*e) Lenition and fortition not difficult to model*

*f) Shape of the tongue counts  
more than 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?

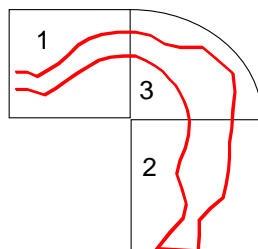
Demo

vowels

consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc:



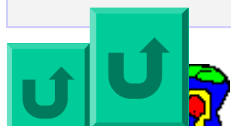
The Larynx Height Index variation is introduced by modifying the longitudinal dimension of the vocal-tract according to 2 scaling factors and 3 zones:

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

See the work at ICP Grenoble (web)

Modelisation

→ Babies  
Neanderthal  
monkey

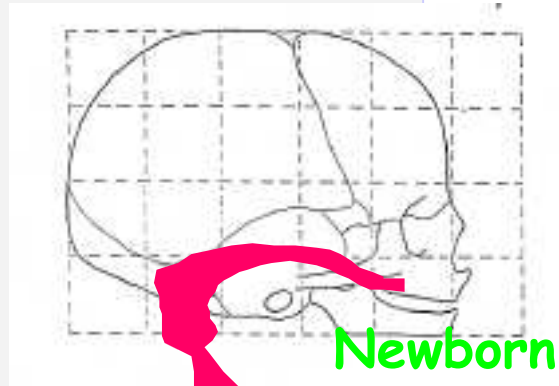


# Vowel prototypes for ontogenesis



Which models?

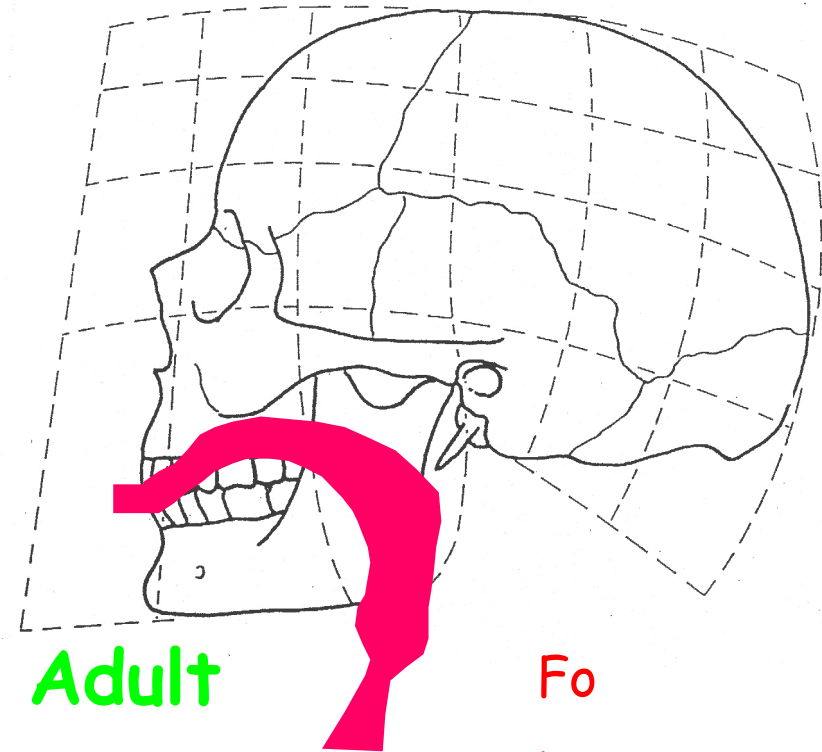
What is a useful model?



speaker anatomy  
speaker strategy  
singling formant  
inversion  
etc;

Fo

Ontogenesis



Adult

Fo

Larynx

See the work at ICP Grenoble (web) with Maeda

Babies  
Neanderthal  
monkey





# Larynx Height Index

What is a model?

Why modelling?

Which models?

What is a useful model?

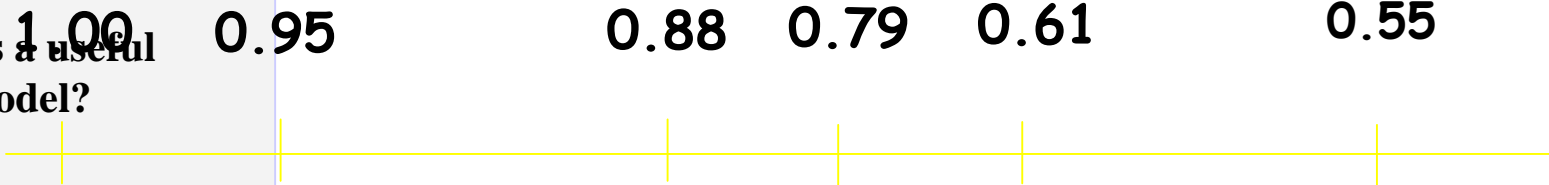
How done?

Demo

Vowels  
consonants

Other uses

speaker anatomy  
speaker strategy  
singing formant  
inversion  
etc;



Man Woman Child (10y 4y) Newborn

Ape  
Newborn Adult

Neandertal

*La Ferrassie La Chapelle-aux-Saints*

See the work at ICP Grenoble (web), Boé, etc. with Maeda

Babies  
Neanderthal  
monkey





# Vowel prototypes (adult man)

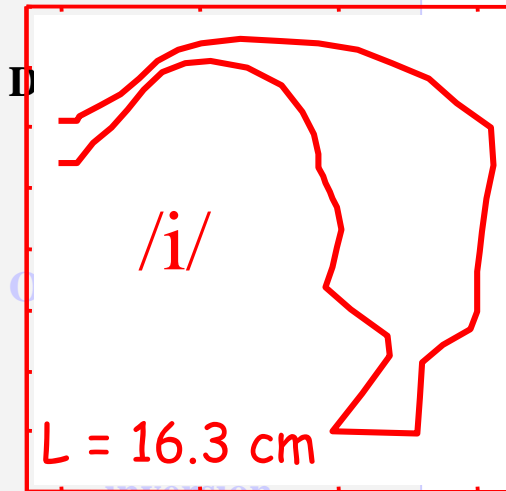
What is a model?

Why modelling?

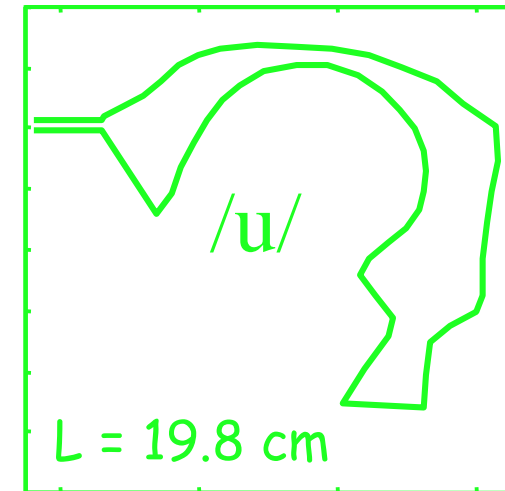
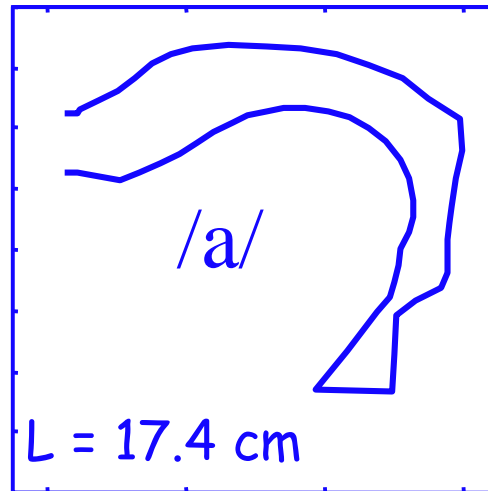
Which models?

What is a useful model?

How done?



inversion  
etc;

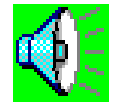


See the work at ICP Grenoble (web) with Maeda

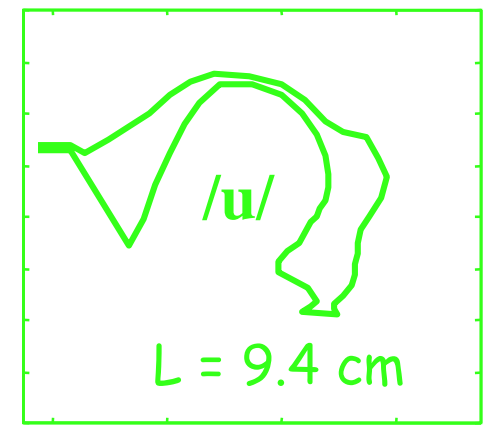
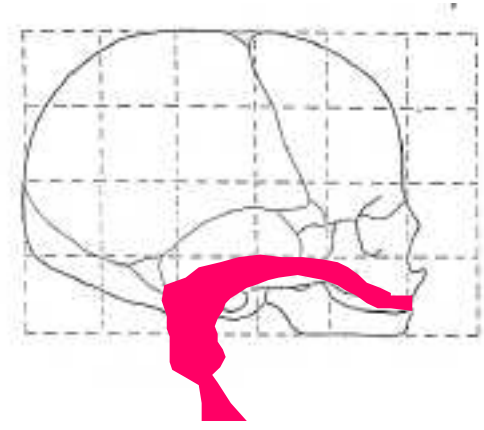
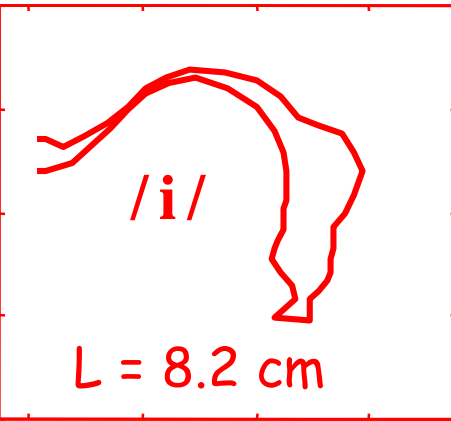
Babies  
Neanderthal  
monkey







Why i  
Which  
What



model?

How done?

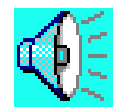
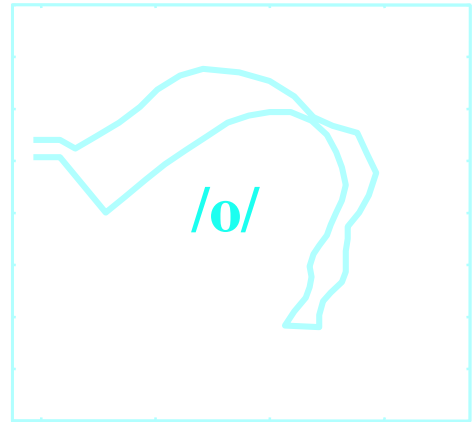
Demo



vowels  
consonants



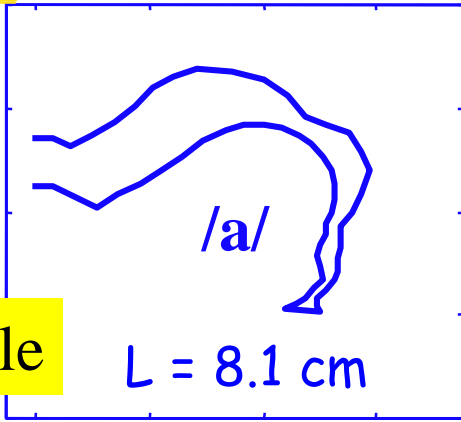
4 months



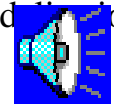
Other uses

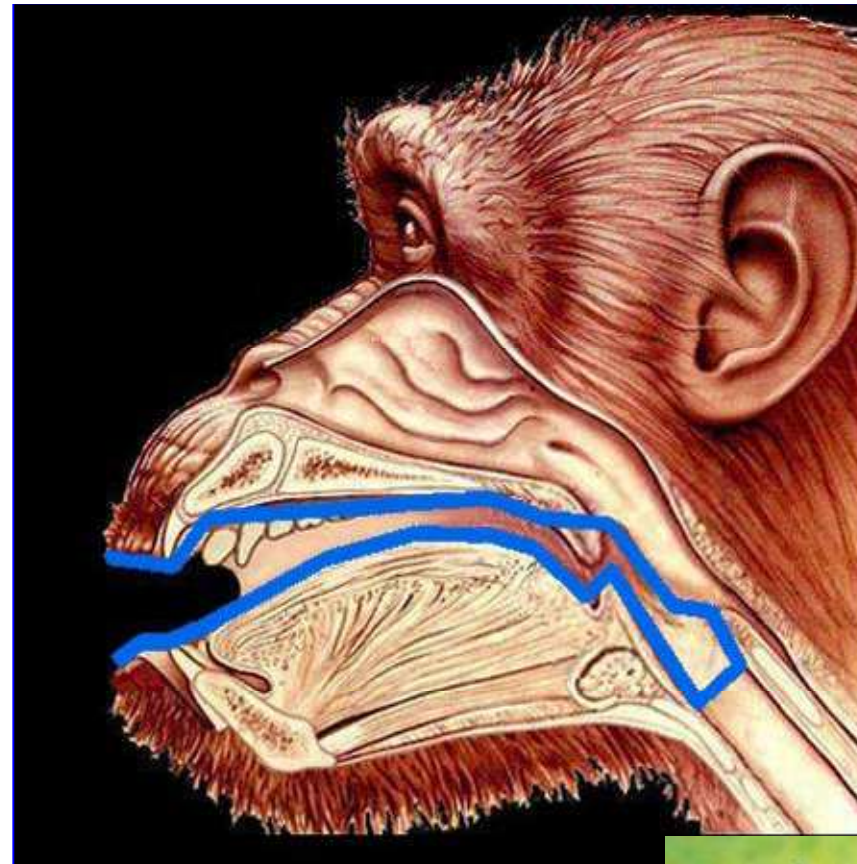
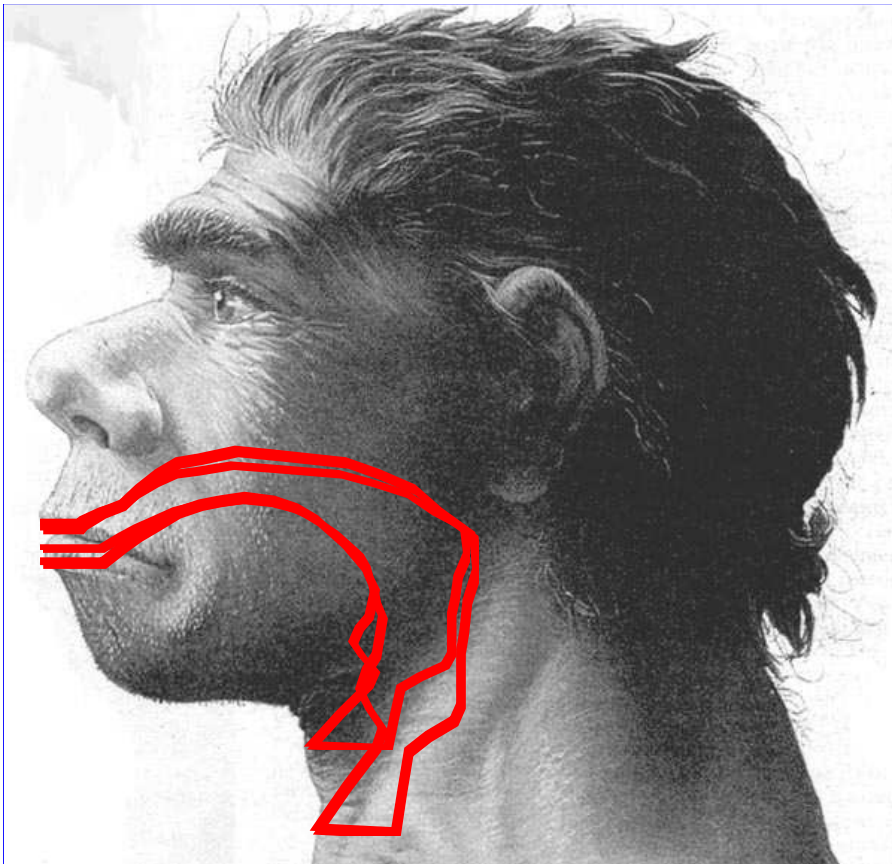
- speaker anatomy
- speaker strategy
- singing formant
- inversion
- etc;

Maeda + ICP-Grenoble



Model on





Other uses

- speaker anatomy
- speaker strategy
- singing formant
- investigative
- 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

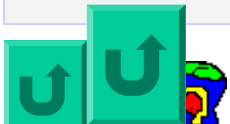
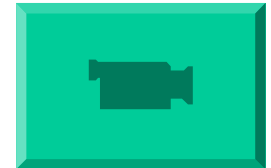
speaker strategy

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 instrumentations 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?



- **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 increase 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)

What

Why

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**

- The corresponding paper with references are available.

**How**

**Demo**

**Other**



Bon anniversaire !

