What is phonetics?

‘Telle que nous l’entendons aujourd’hui, la phonétique est la science des sons de langage. C’est une branche de l’acoustique, des sciences naturelles, psychologiques et sociales. Cette complexité n’a pu qu’en retarder les progrès’.

Rousselot leçon d’ouverture au Collège de France 1924

What is a relevant speech sound?

1. Describe and understand the diversity of sounds in languages.

2. Describe the sound system structures of languages, their dynamics and complexity.

3. Understand the biological, physiological and physical mechanisms of speech.
   ⇒ *development of experimental methods.*

4. Origin and evolution of sound production and perception in primates.

The sound system of language is an open system.

What are the theoretical consequences of this claim?
⇒ open systems evolve towards a complexification of their structures.

What are the limits of languages sound systems?

Take into account auditory and perceptual constraints.

Are there invariants? If they exist what are they?

What is a relevant speech unit? C, V, tone, syllable, feature, gesture?

1. Describe and understand the diversity of sounds in languages

What do we learn from the study of the sounds in languages?

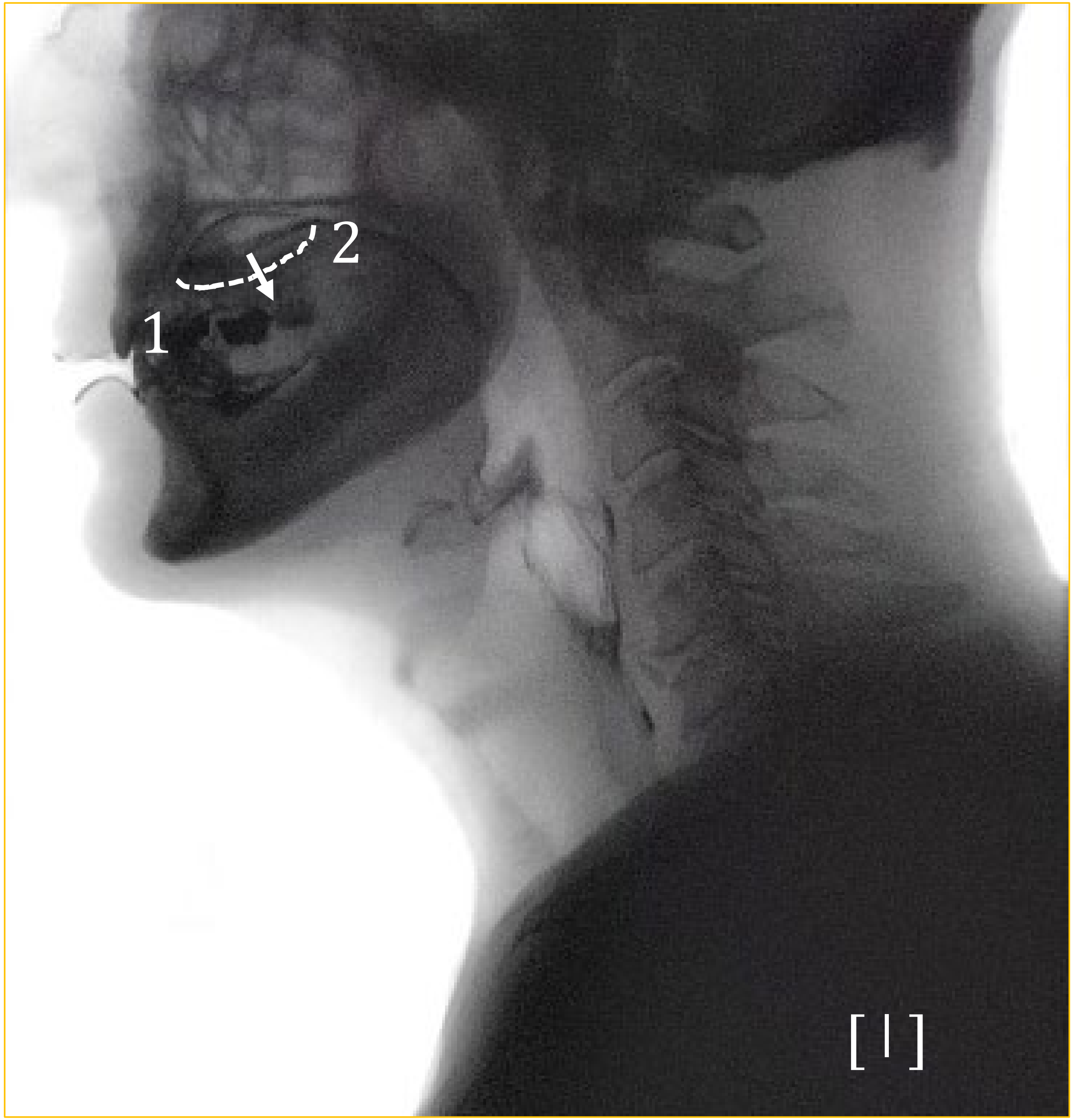
How are speech sounds made and controlled?

Acoustic theory and speech models

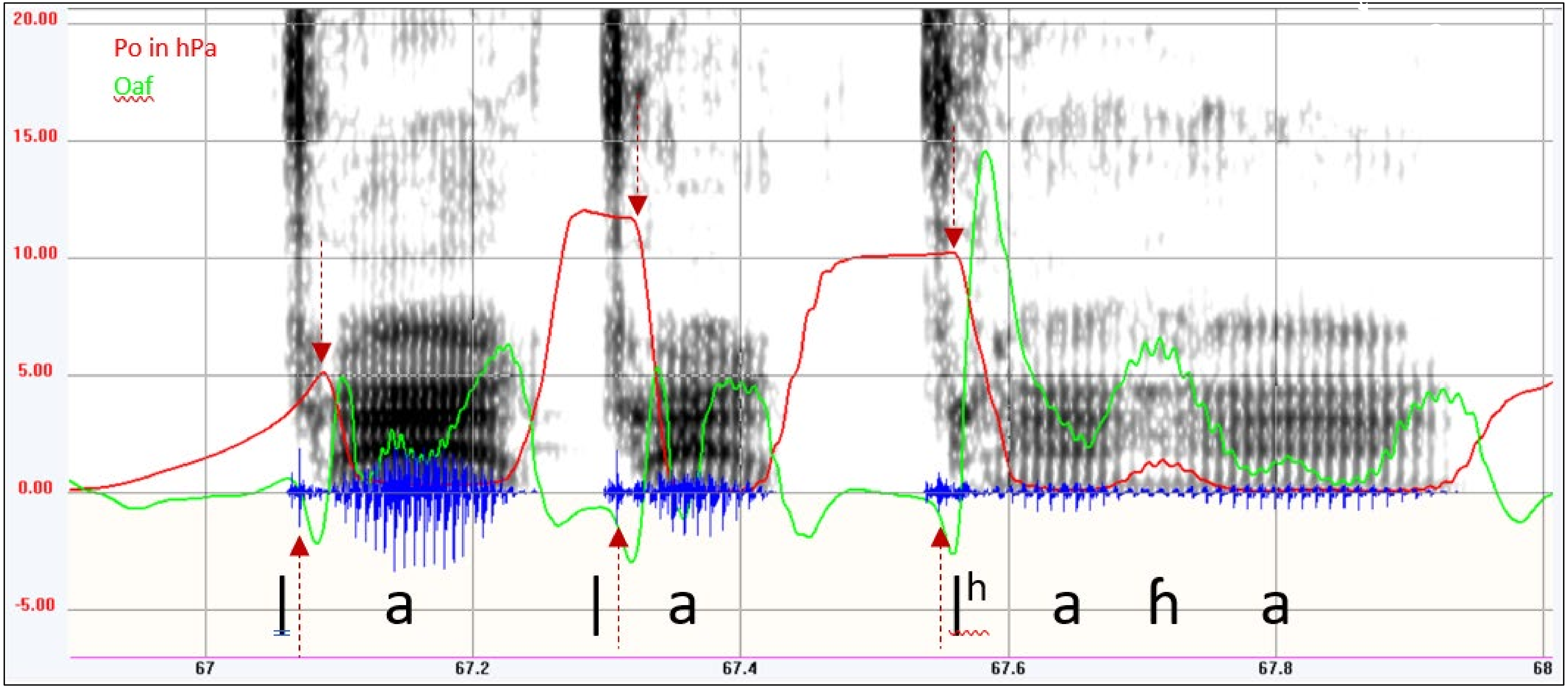
The basis of experimental science is direct observation

Lessons from unwritten languages (90% of the world's languages)

Dental click [ǀ]

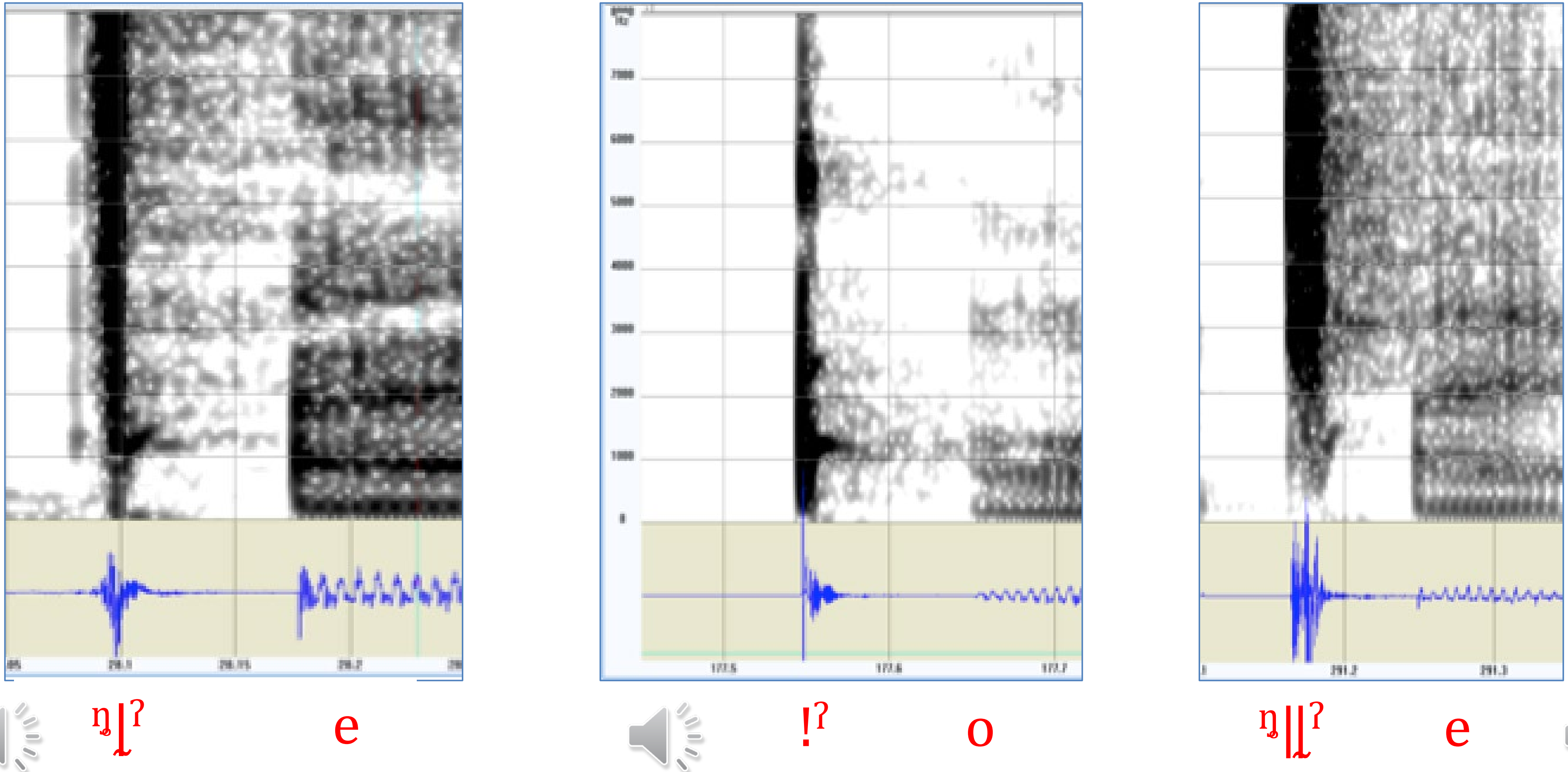


Timing of front and back closures releases in clicks - *Hadza*



How is this controlled?

The auditory system does a frequency, duration and intensity analysis to make the distinction between the different clicks.



Clicks exploit fundamental sensitivities of the auditory system.

The speed of the noise damping enhance clicks perceptual salience.

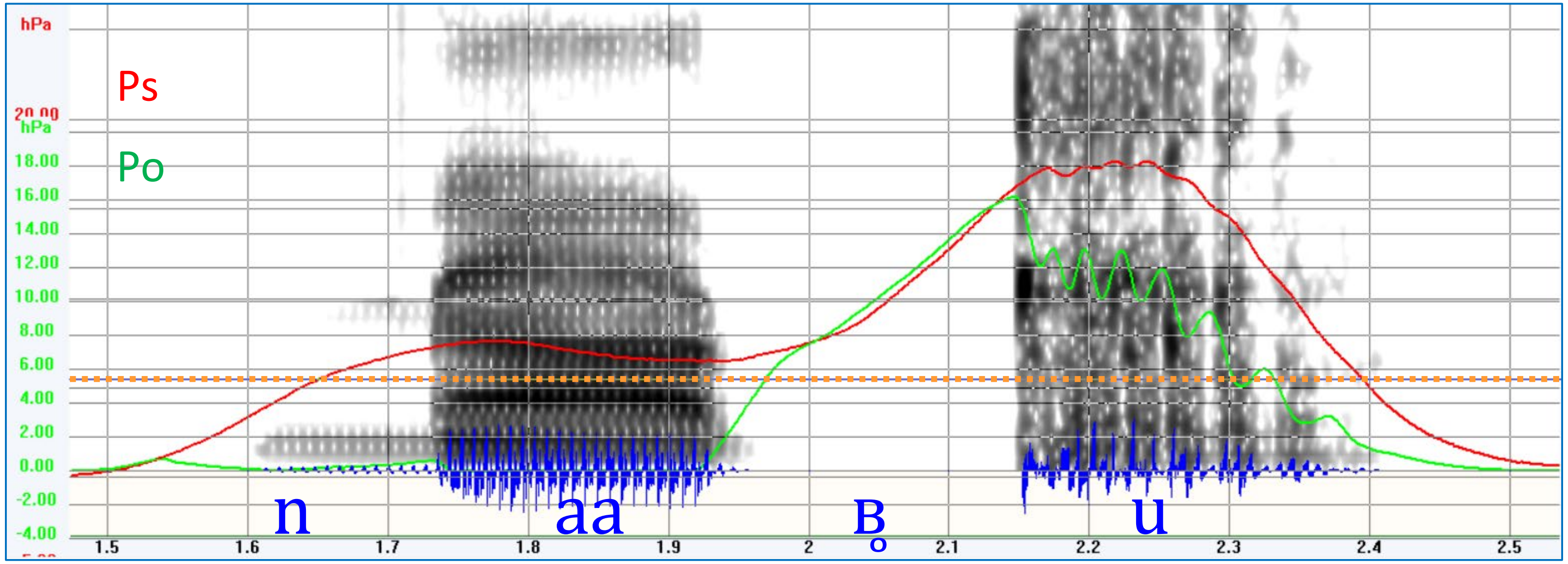
Some auditory theresholds are reached allowing the different clicks accurate identification.

The timing specifications and the time to produce some gestures are not directly related (one to one) with the phonological features.

Hadza, as most of the languages where clicks are found, remain a challenge to give adequate answers to this question.

Aerodynamic theresholds for trills - *Mangbetu, French*

Mangbetu [ʙ]

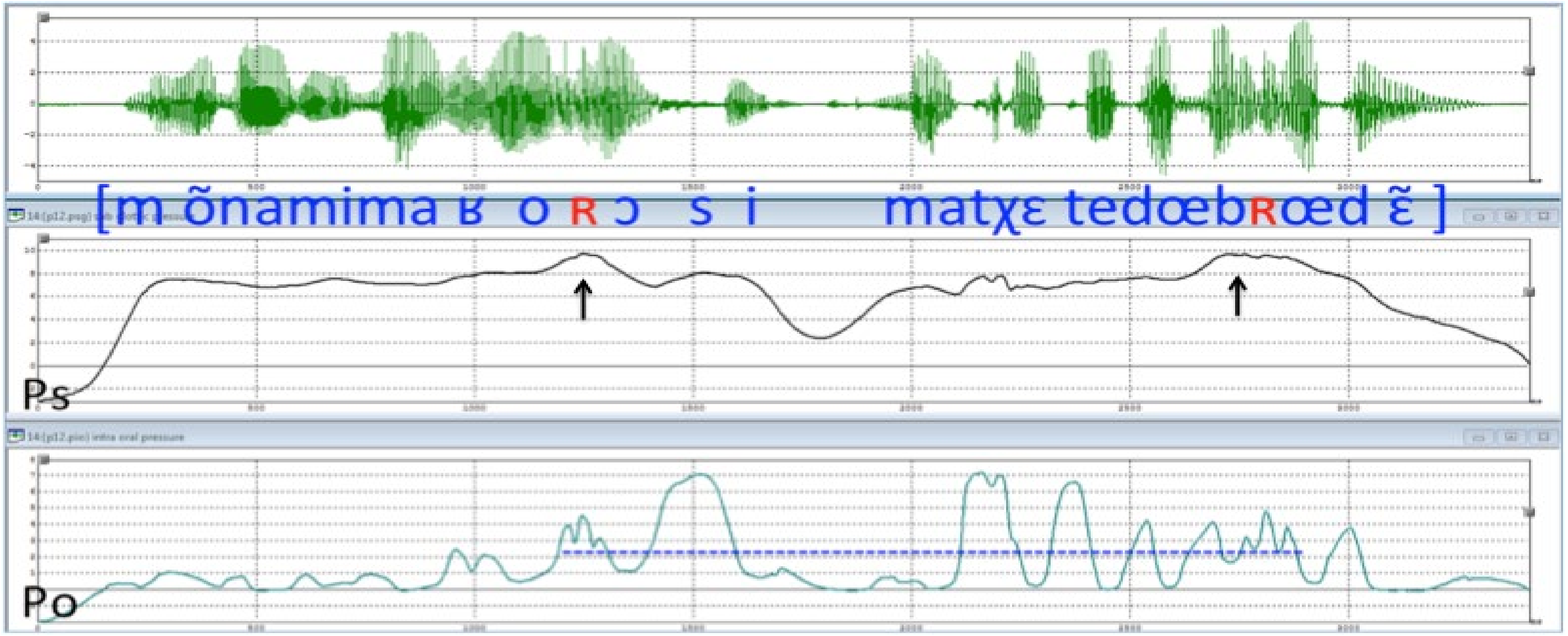


ΔPo/Pa > 5 hPa

C&V gestures overlap

Cheeks and VT compliance

French [ʀ]

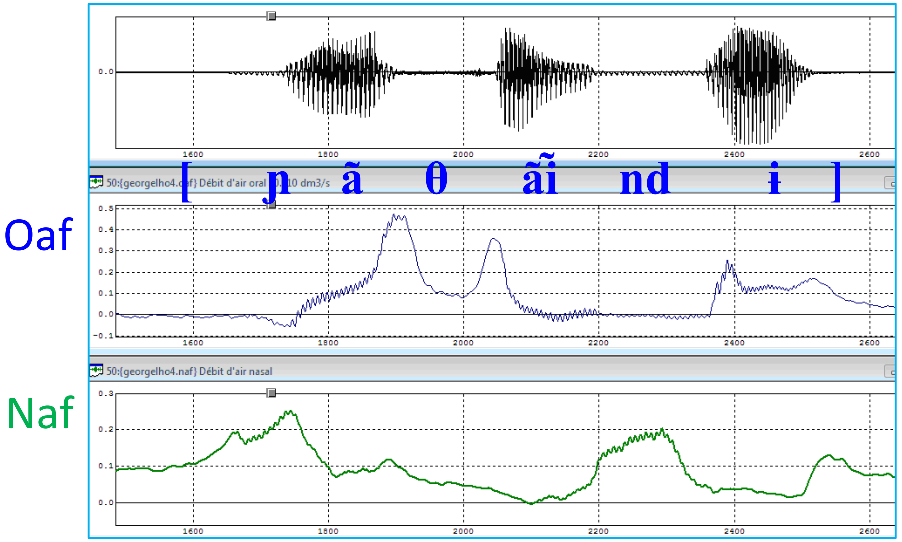


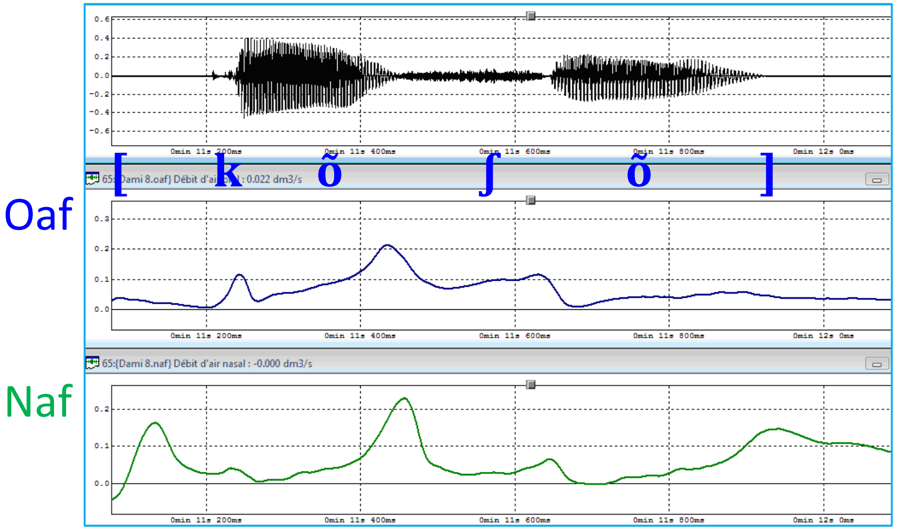
ΔPo/Pa > 3 hPa

How are these theresholds controlled?

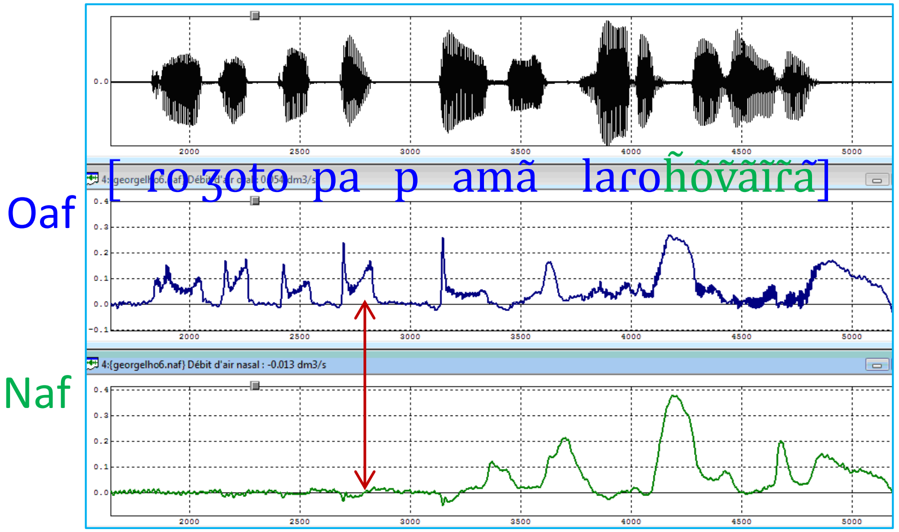
Guarani nasalized fricatives & aerodynamic theresholds ⇒ for acoustics

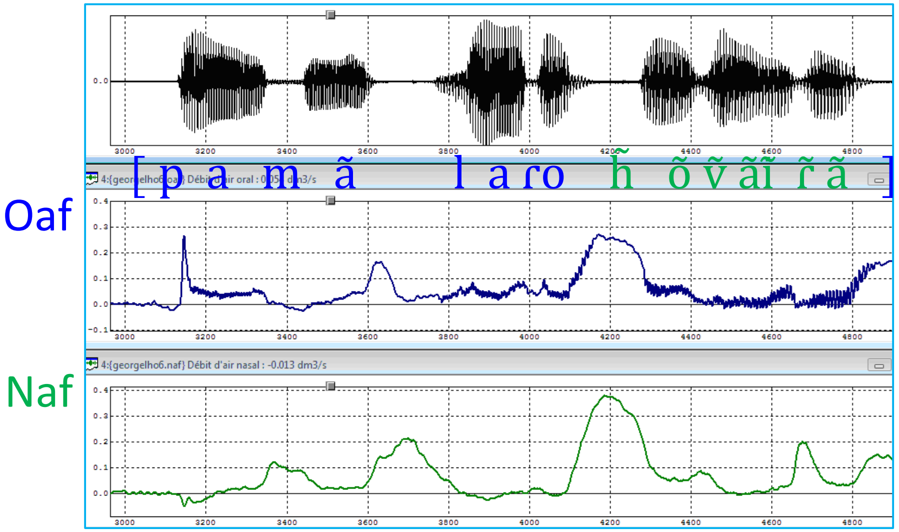
[s̃ > θ]   /ɲãsãĩndɨ/ > [ɲãθãĩndɨ]   /kõʃõ/ ∫ Naf < critical thereshold = frication



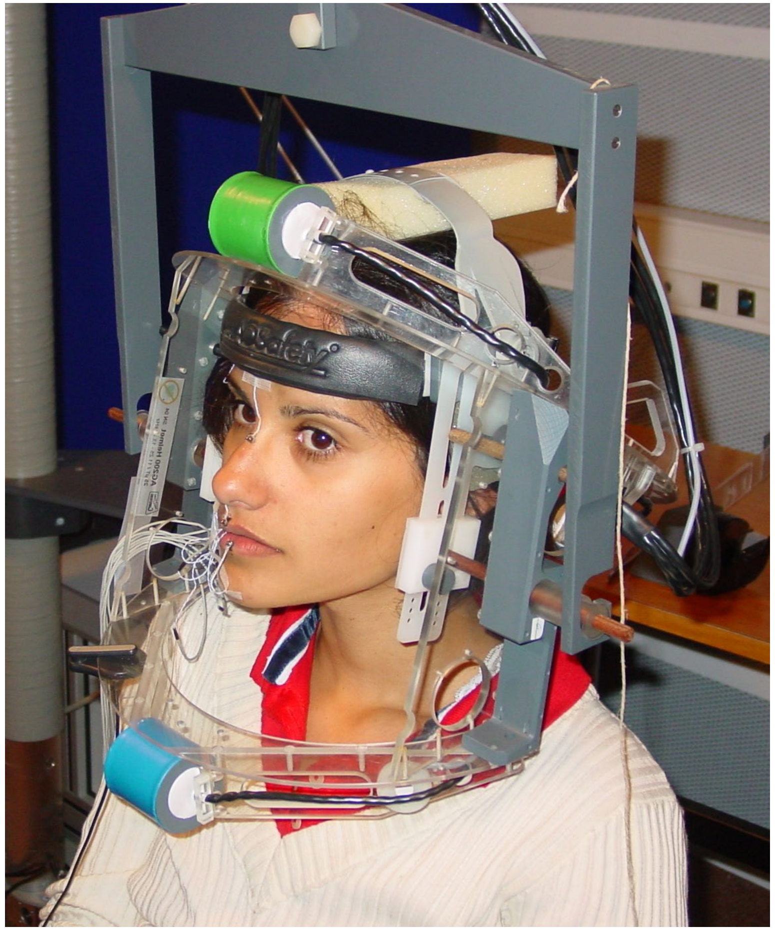


Nasal harmony





Velum mouvement in BP nasal diphtongs



Demasi 2023

[pãw̃]

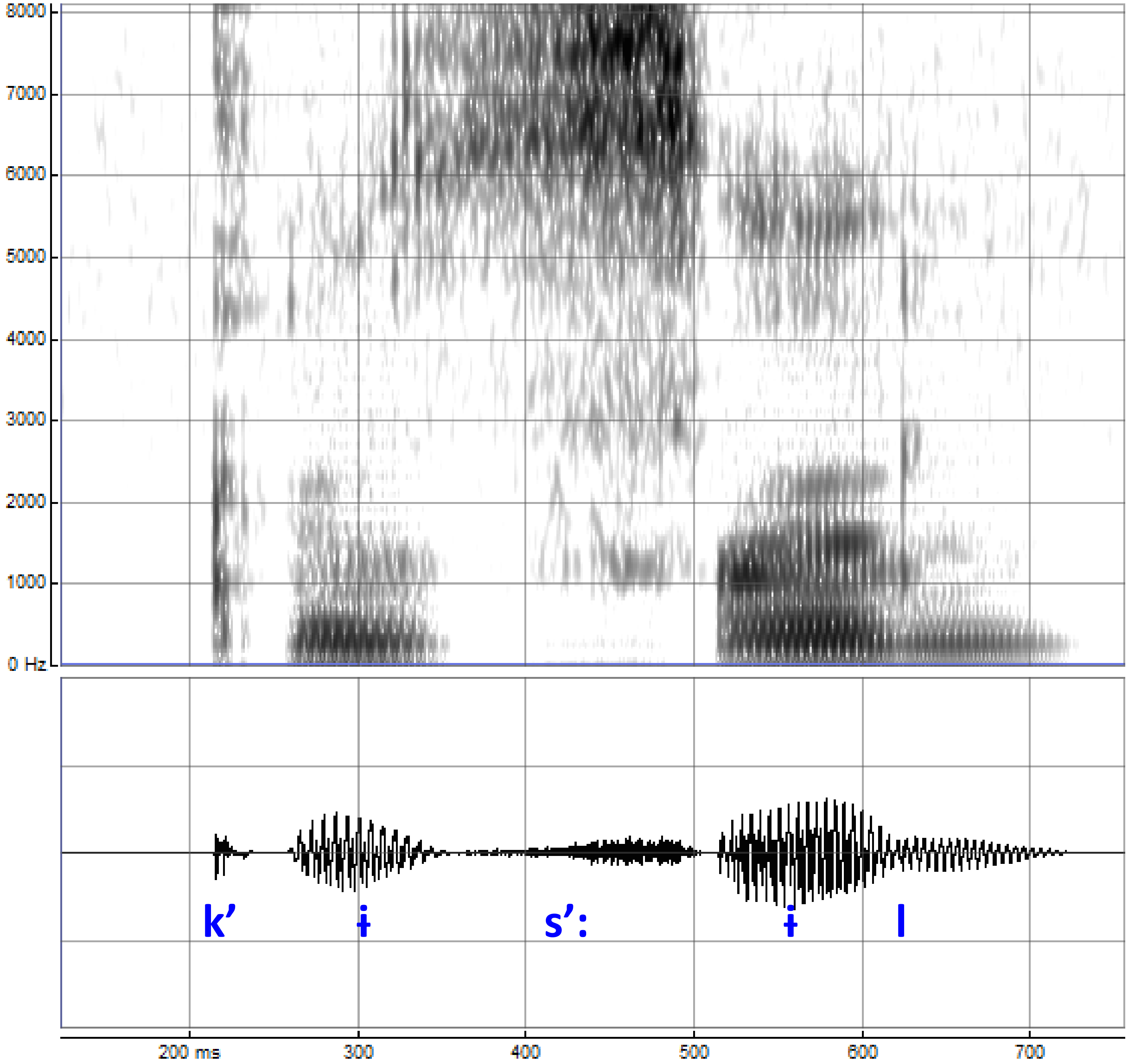
Amharic

geminated alveolar fricatives **s’:**

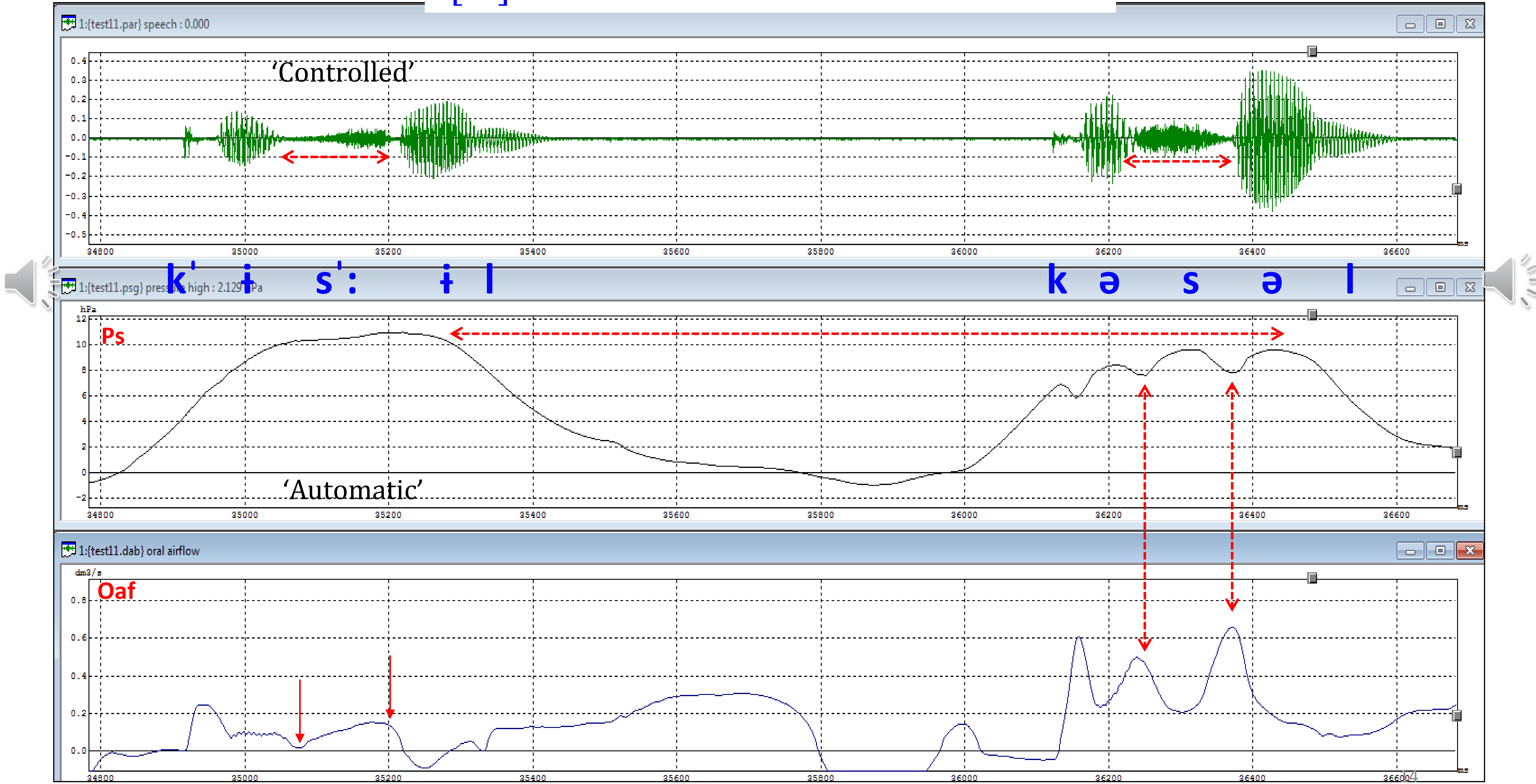
Possible?

The elevation of the larynx is slowed down to produce the frication noise with the limited aerodynamic resources.

⇒ this is controlled to produce the acoustic target.



[s’:] in *Amharic* controlled vs automatic

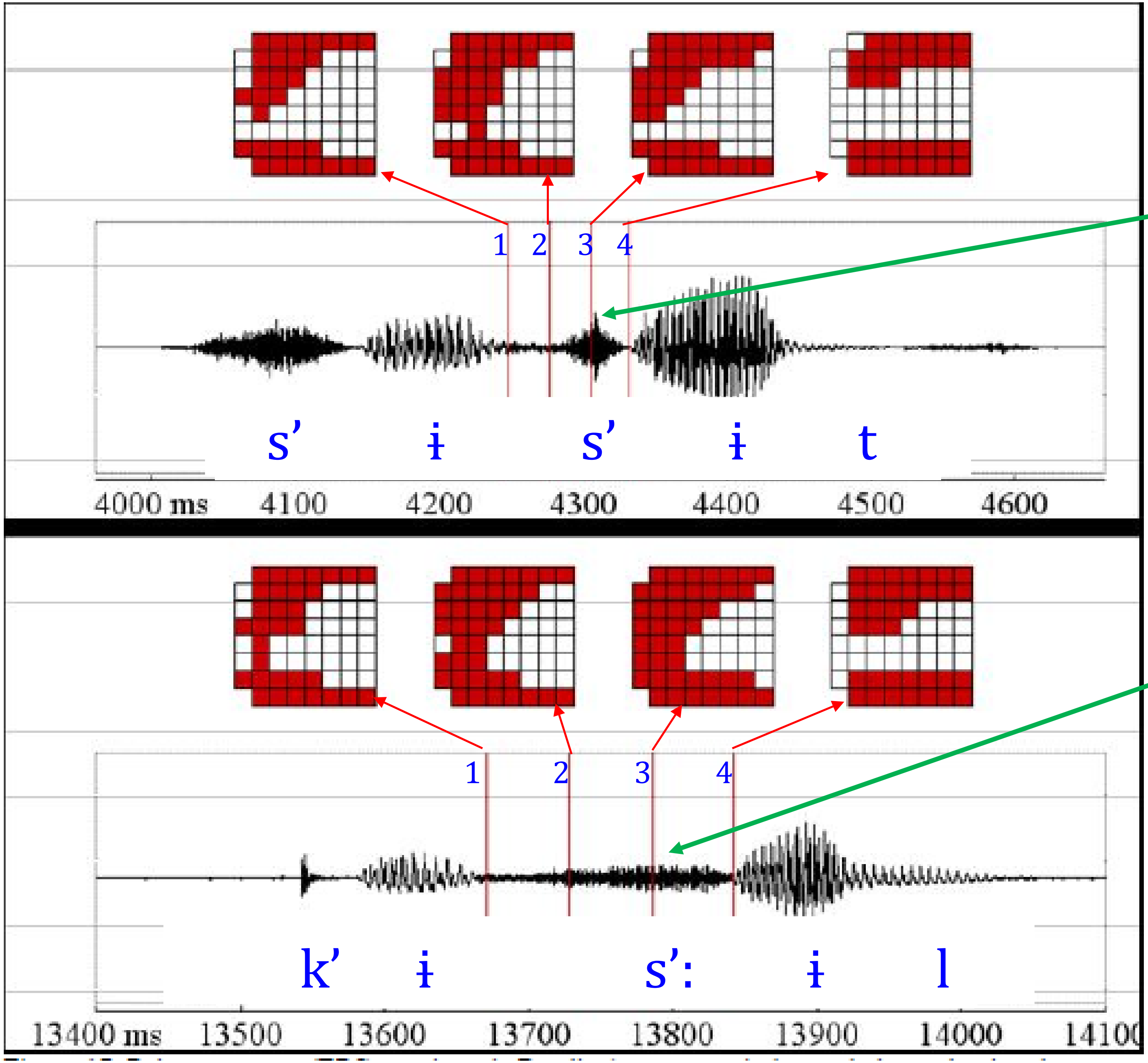


Amharic

s’ – s’:

Phases [s’] 2 and [s’:] 1,2,3 seem to be produced with a closed constriction.

The audio waveforms show that airflow is still passing through a very thigh constriction even if the EPG electrodes are activated.



Fast larynx elevation

Delayed larynx elevation

Control of lip gestures

Iraqw

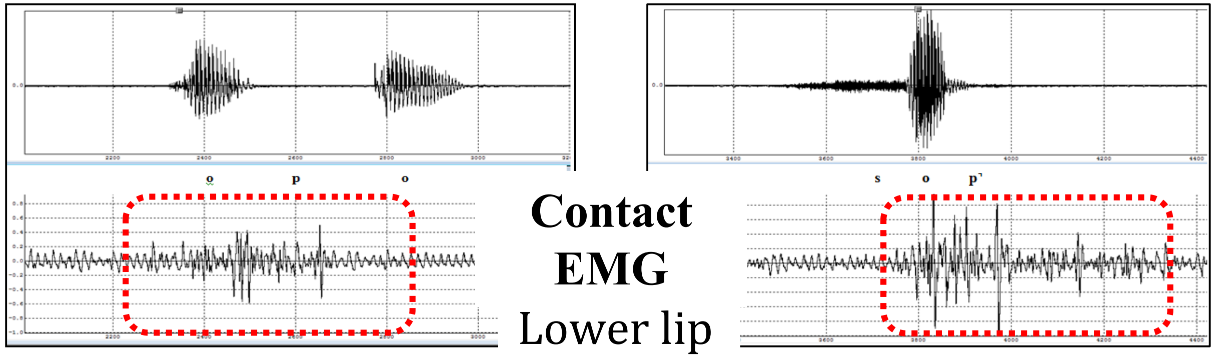
[kʷ] release : no lip rounding

[w] release : lips rounding and protrusion

Mous, Ghio & Demolin(2020)

Karitiana

Allophones as phonological categories?

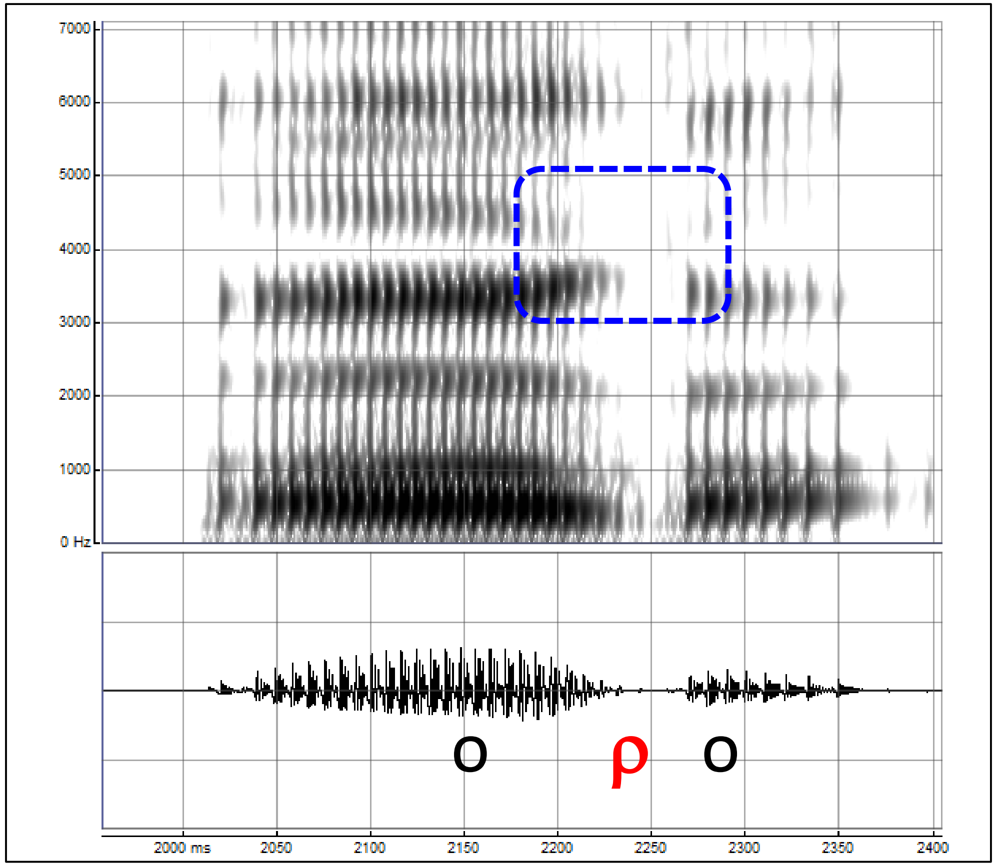


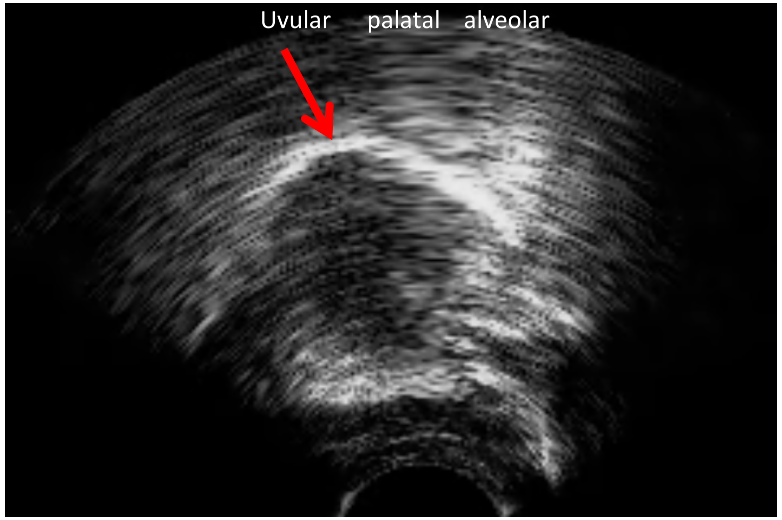
Storto & Demolin (2002)

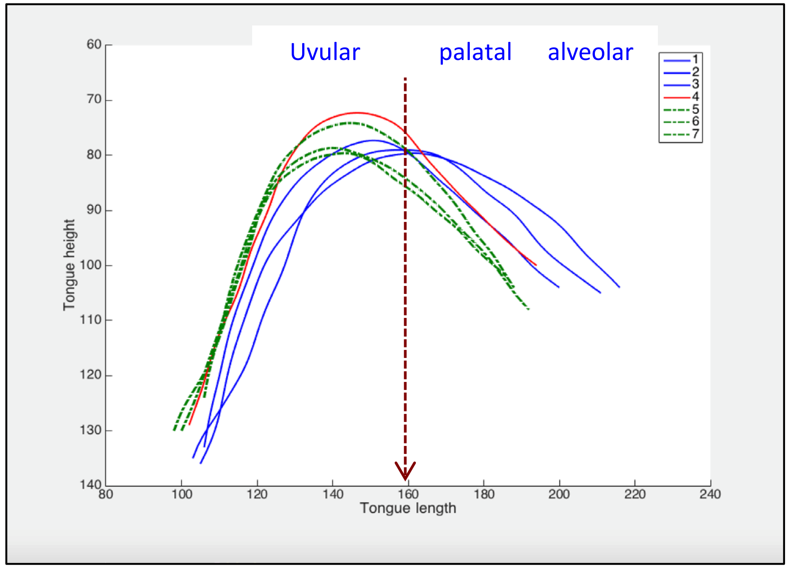
Kalapalo (Carib) uvular tap

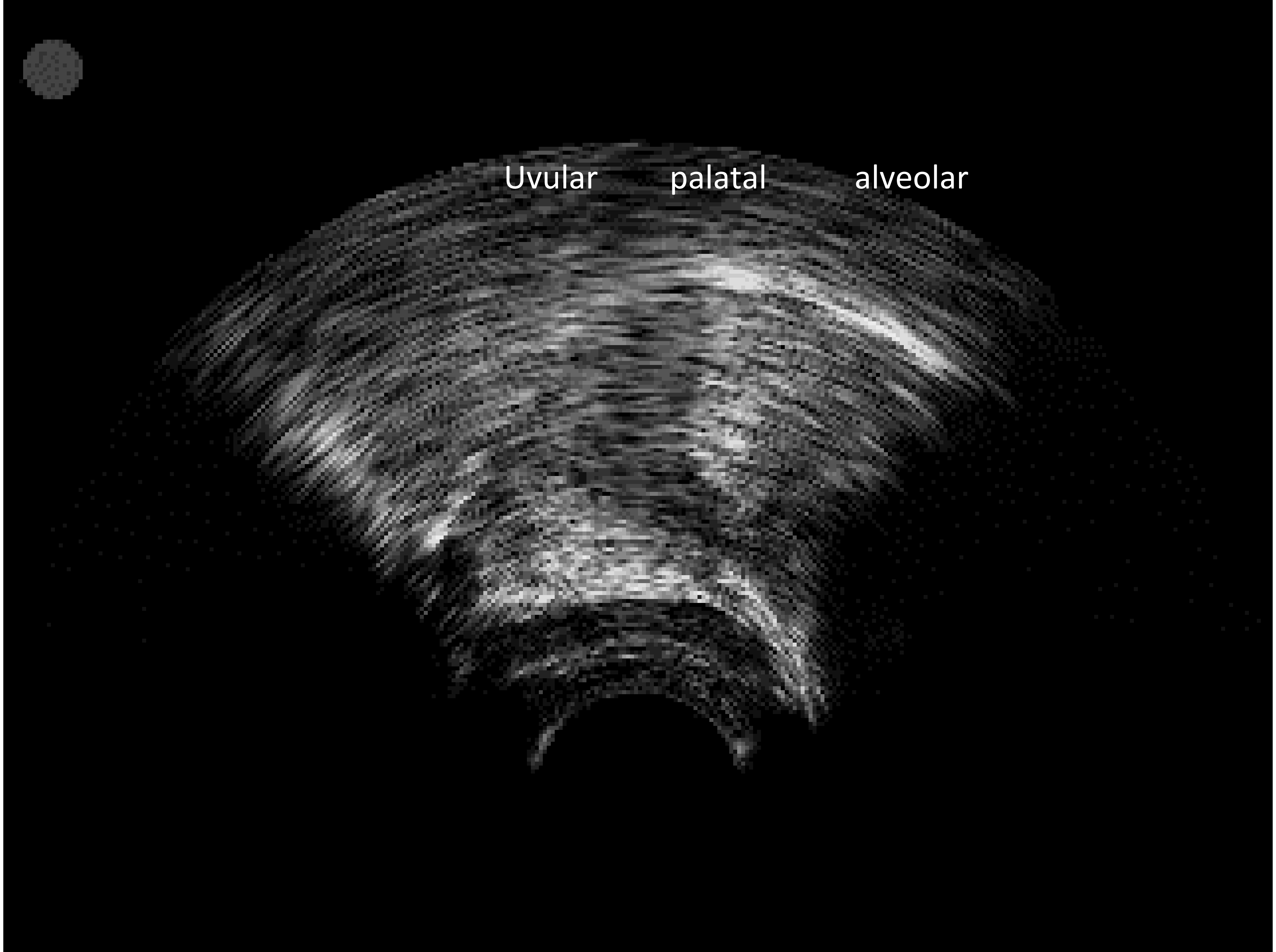
In1887 von den Steinen described a sound, in the Upper Xingu (Amazon): “*complicated to pronounce, between gl and ri, always forming a syllable equivalent to r followed by a reduced vowel*”.

Pole F4/F5 * r/ɾ > l > ρ

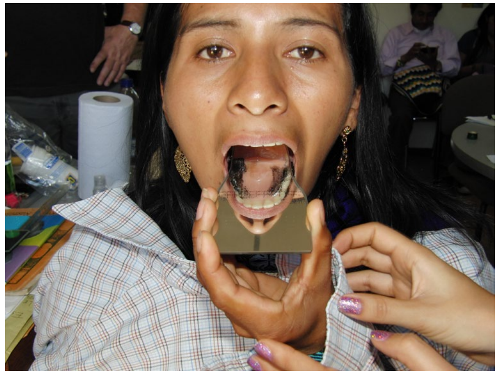







Uvular
palatal
alveolar

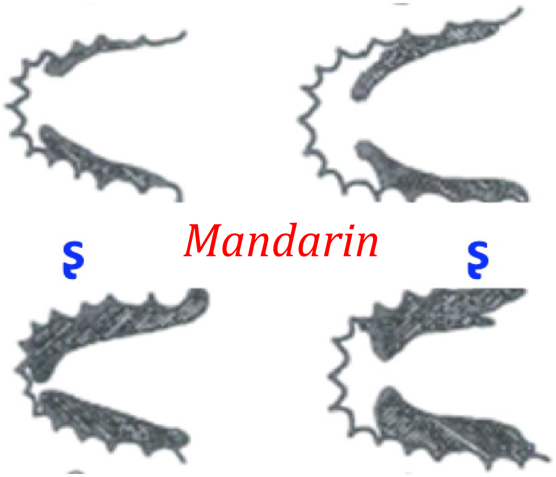
Retroflex place of articulation? *Namtrik* (Colombia)/*Mandarin* retroflex fricatives



Gonzales, Rojas & Demolin (2015)

ʂ *Namtrik* ʂ

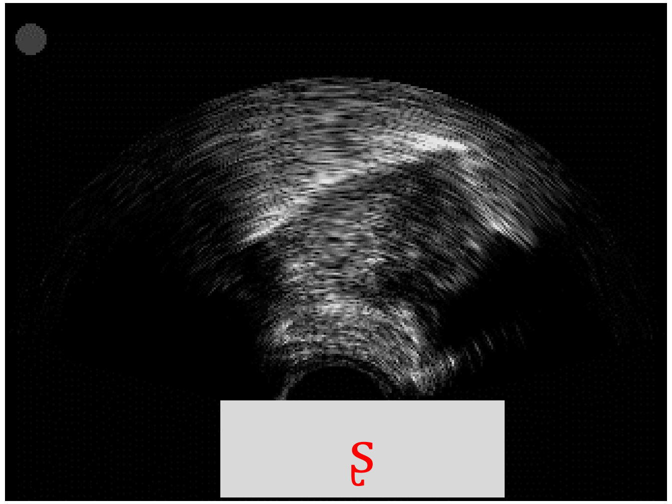
ʃ ʃ



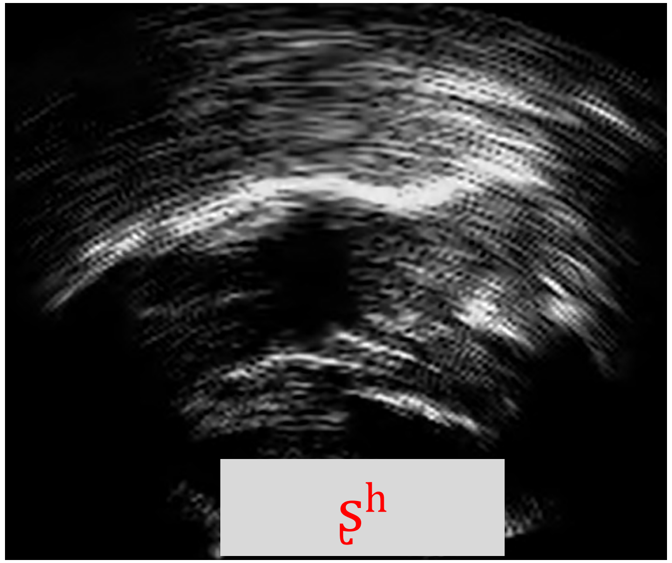
ʂ *Mandarin* ʂ

ɕ ɕ

Ladefoged & Wu Zongji (1982)



[uʂə] ‘*polvo*’



师表 [shi 'biao 3]

Sublingual cavity and low F3

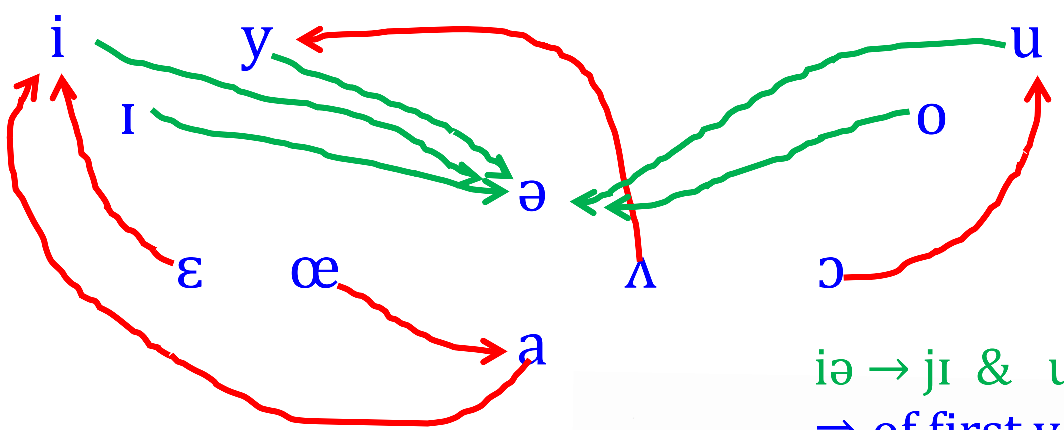
Ming & Demolin (2014)

Frisian vowels, diphtongs & dispersion theory

Short vowels [i, ɪ, ɛ, œ, ə, ɑ, ɔ, o, u]
Long vowels [i:, e:, ɛ:, y:, ø:, a:, ɔ:, o:, u:]
Diphtongs [iə, ɪə, ɛi, œa, ai, ʌy, yə, ɔu, oi, oə, uə] de Graaf (1984)

Dispersion

Attractor?



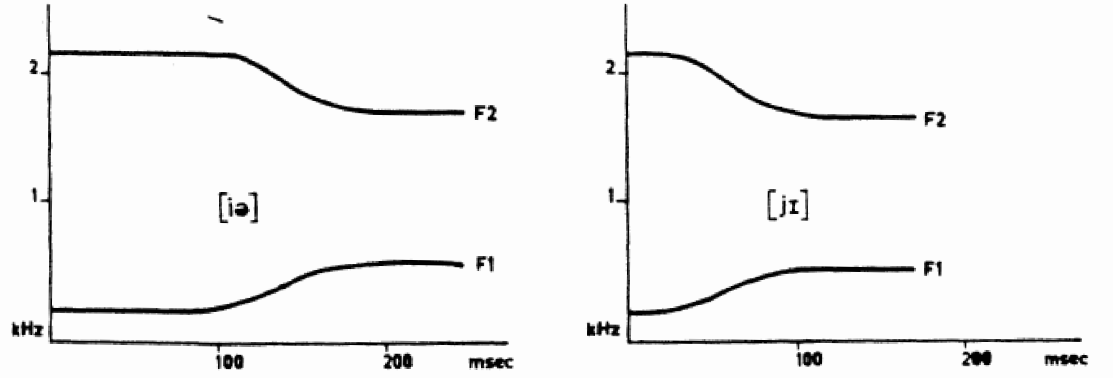
iə → jɪ & uə → wo
⇒ of first vowel shortening

Breaking diphtongs

iə	→	jɪ
ɪə	→	jɛ
oə	→	wo
uə	→	wɑ

fliər → fljɪren 'floor' sg. - pl.
fuət → fwotsjə 'foot' 'small foot'

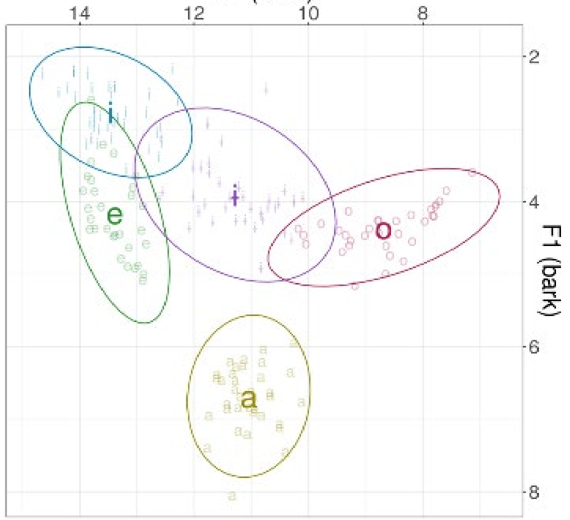
Peripheral vowels are more energetic, they mask the timbre of the 2nd vowel

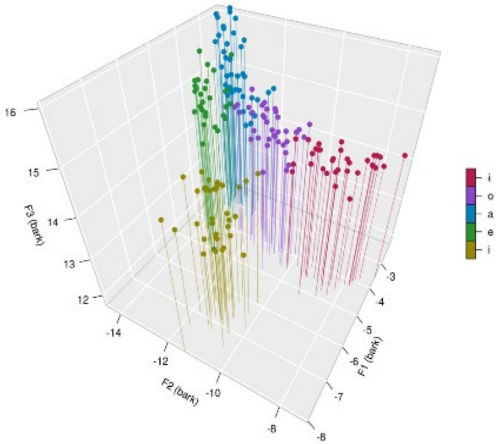


Data and models

Karitiana 5 vowel system [i, e, a, o, ɨ]

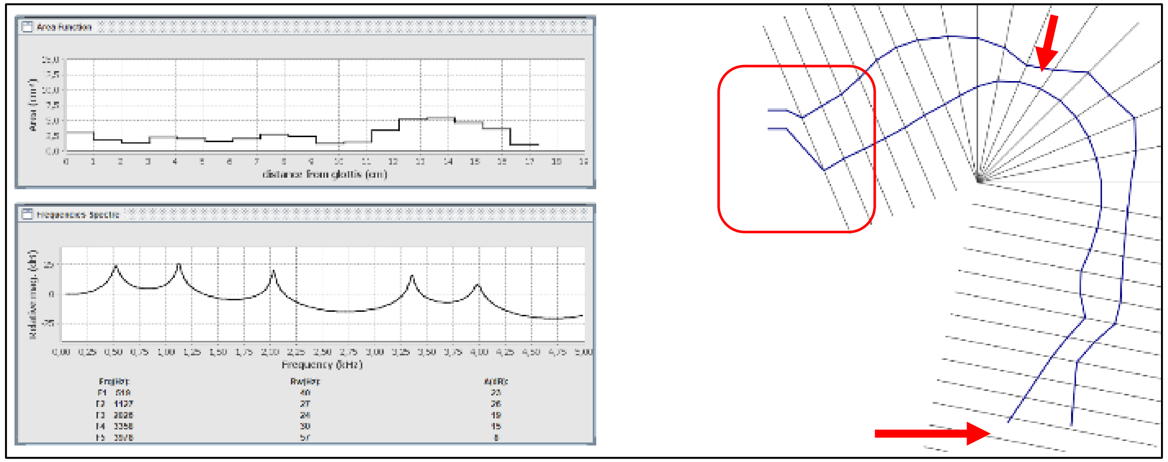
Storto (1999)





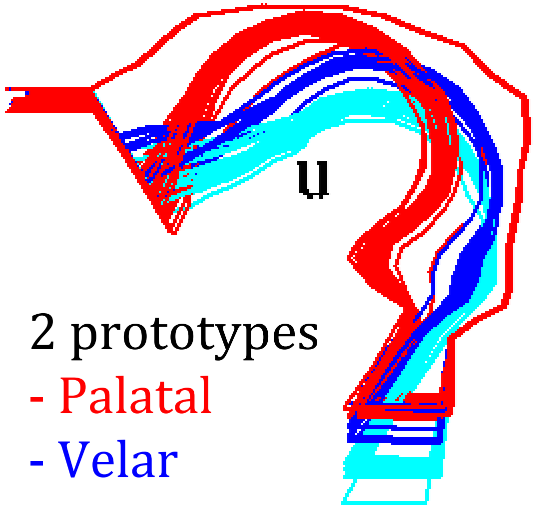
[o] Karitiana

Vaissière & Maeda



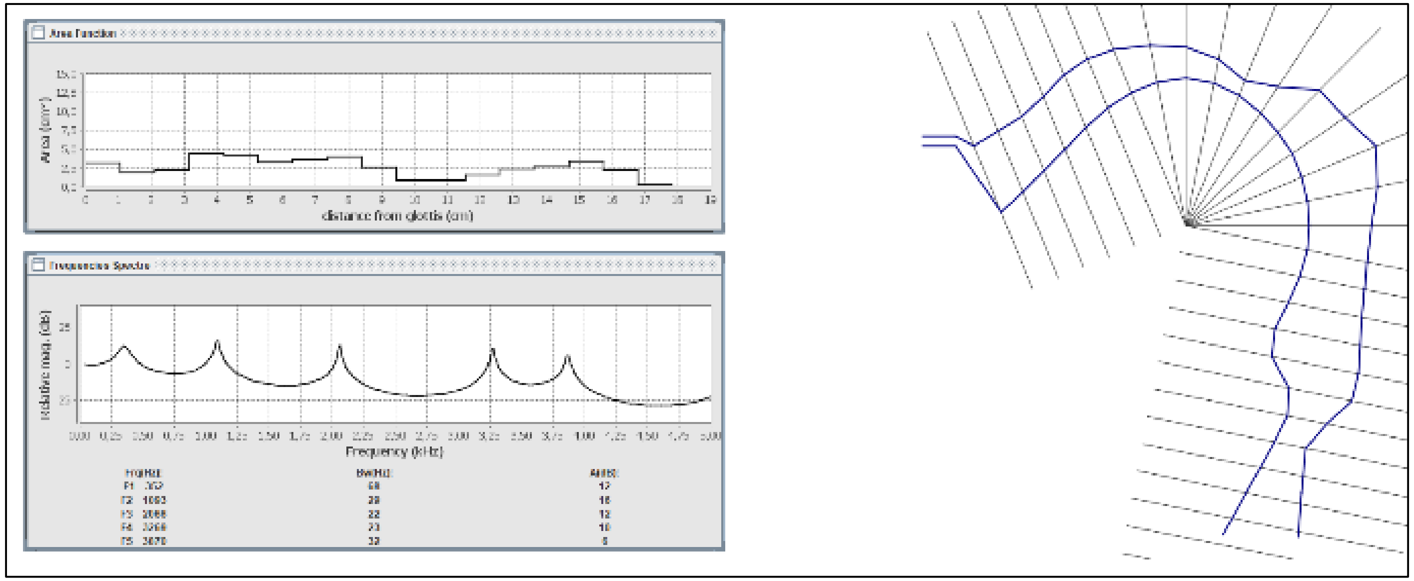
[o] Brazilian Portuguese

[ɨ] Karitiana ?

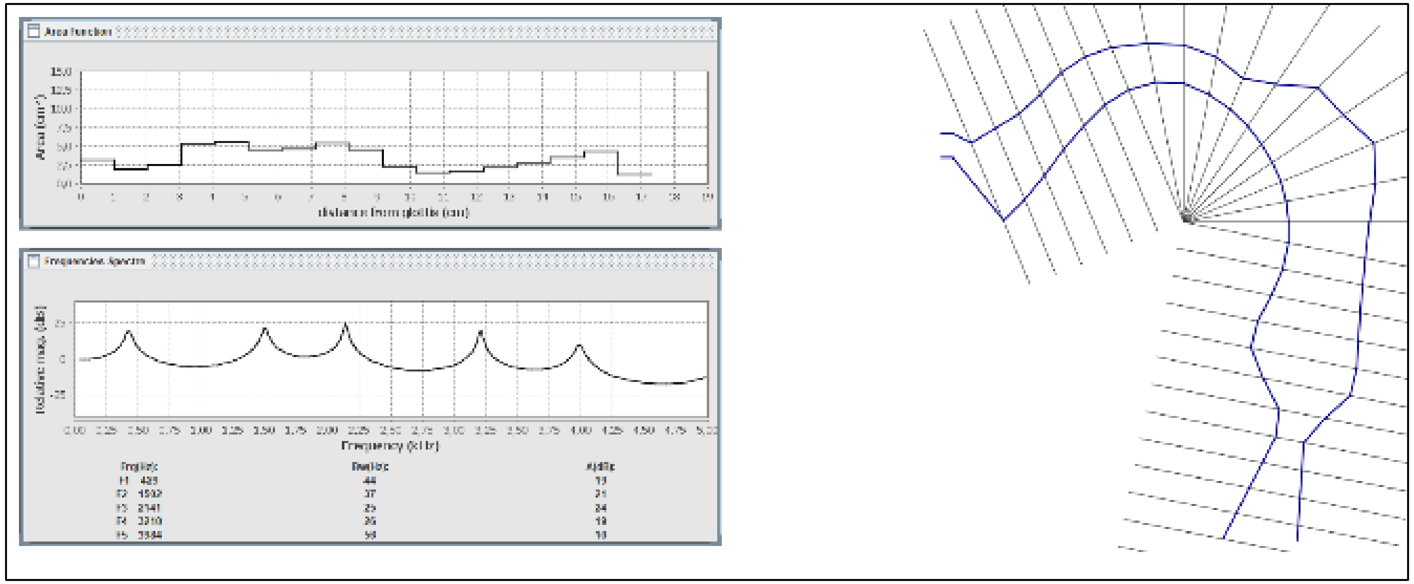


2 prototypes
- Palatal
- Velar

Ménart (2002)



[u] Brazilian Portuguese



[ɨ] Karitiana

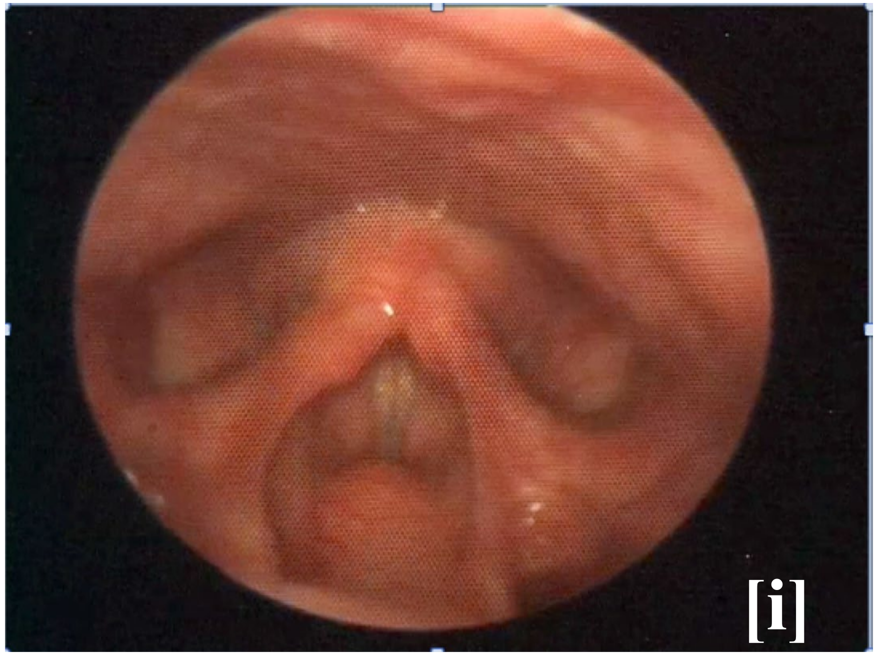
Vaissière & Maeda

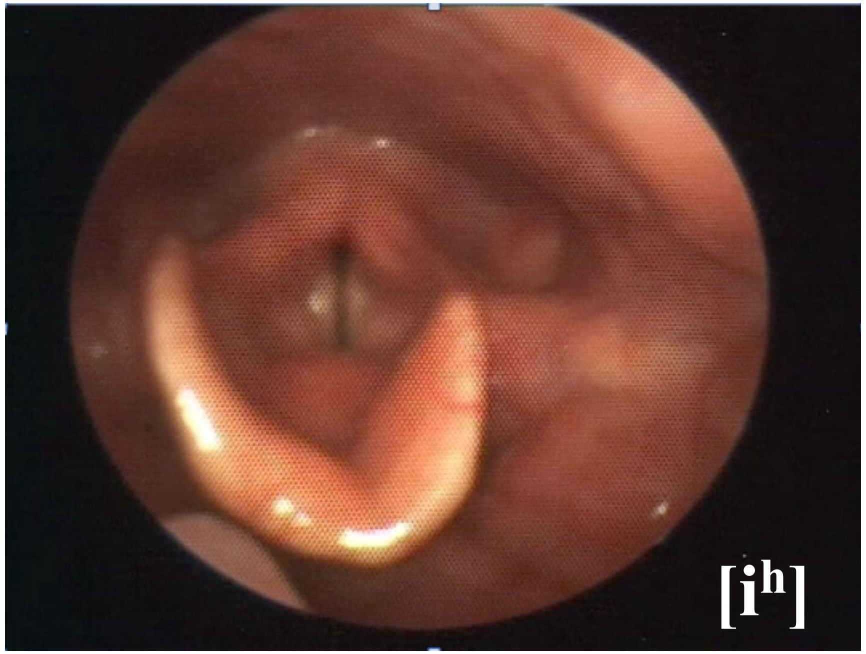
Complexification

Nasa Yuwe (Paez) Colombia

4 vowels timbre [i e a u] - 32 phonological vowels
(Rojas 1998, Diaz 2019)

Oral vowels				Nasal vowels			
i	e	a	u	ĩ	ẽ	ã	ũ
iʰ	eʰ	aʰ	uʰ	ĩʰ	ẽʰ	ãʰ	ũʰ
iˀ	eˀ	aˀ	uˀ	ĩˀ	ẽˀ	ãˀ	ũˀ
iː	eː	aː	uː	ĩː	ẽː	ãː	ũː



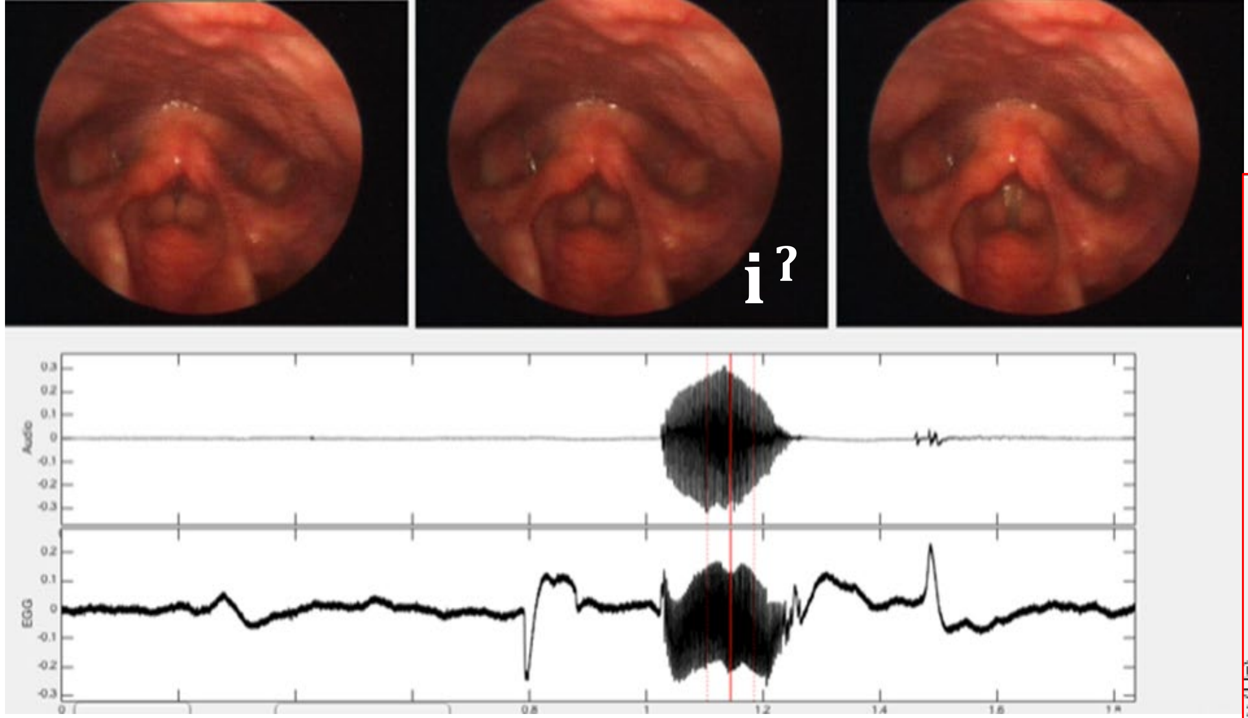


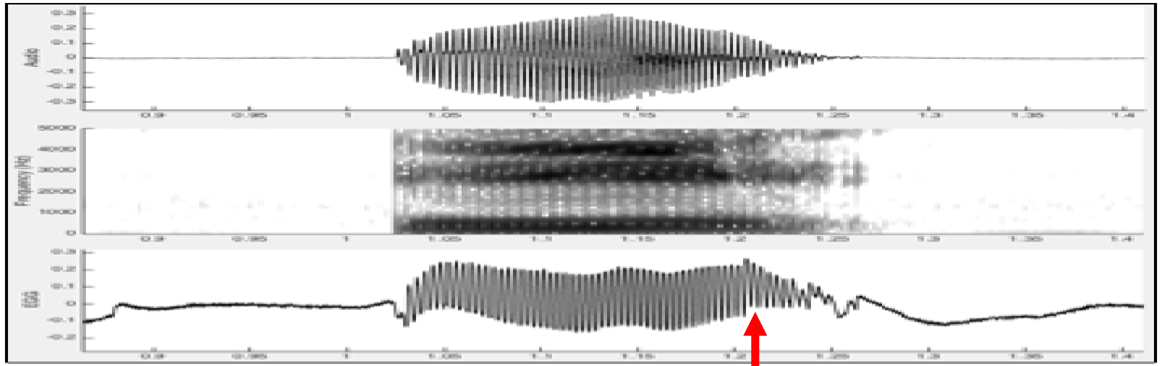
[i^{ʔ}] Glottal closure with a contribution of the ventricular bands.

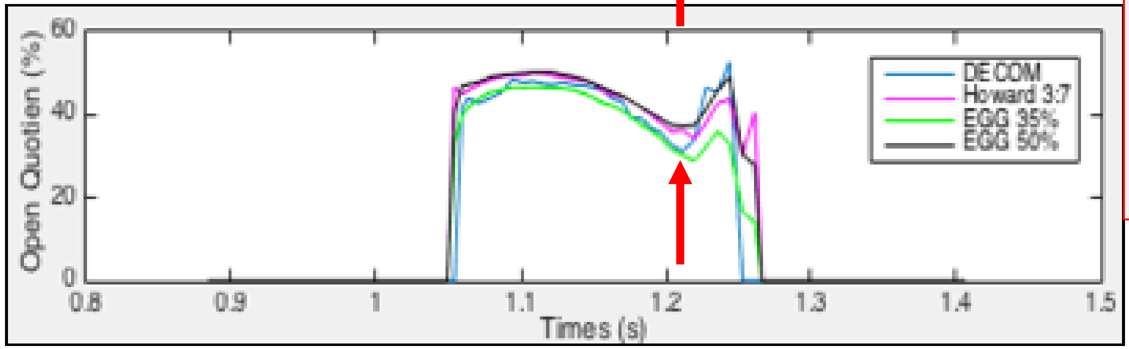
There is no closure of the epilaryngeal tube.

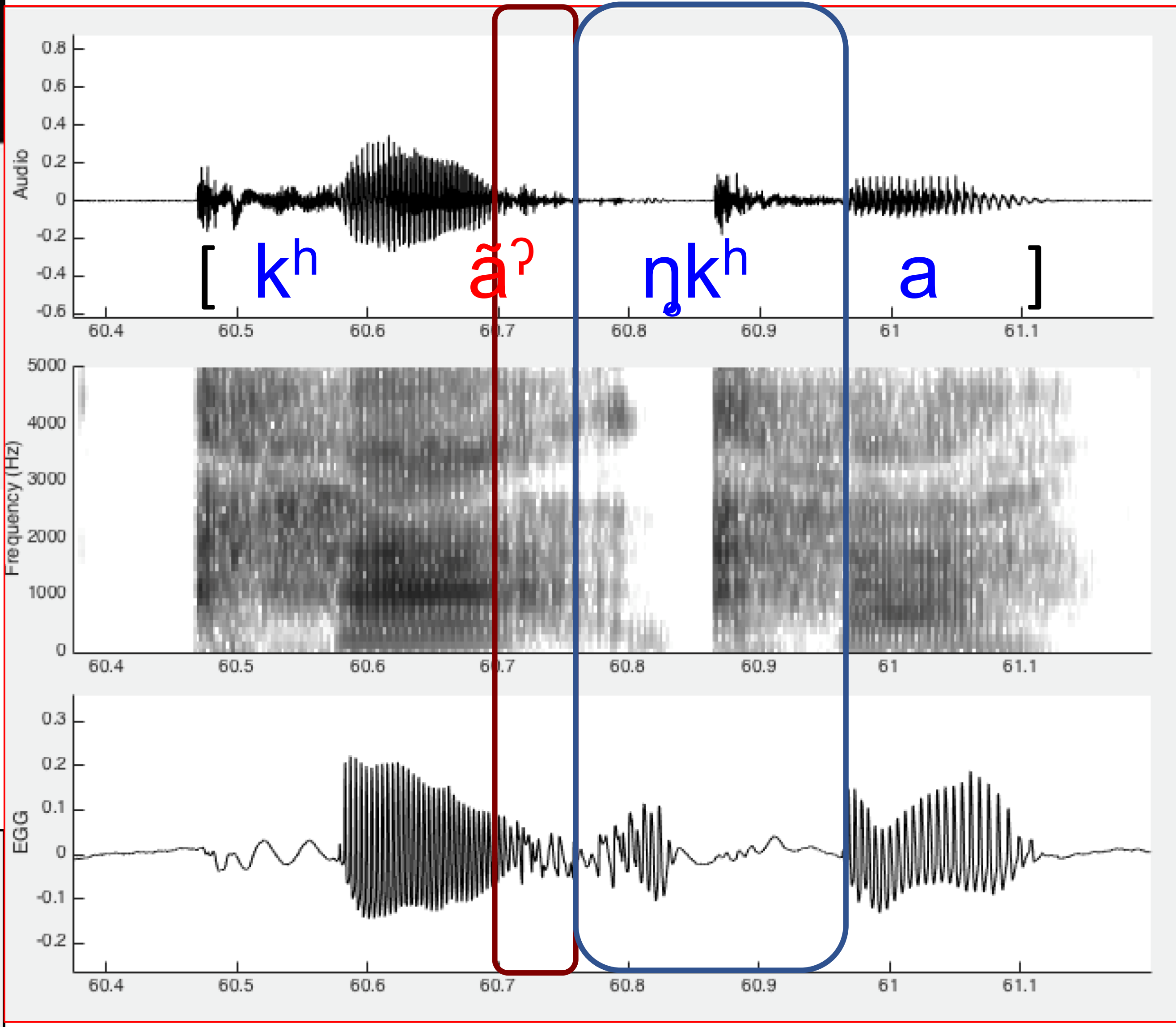
3 Valves of the LAM & the ± CET feature

Glottal vowels, the epilaryngeal tube & LAM

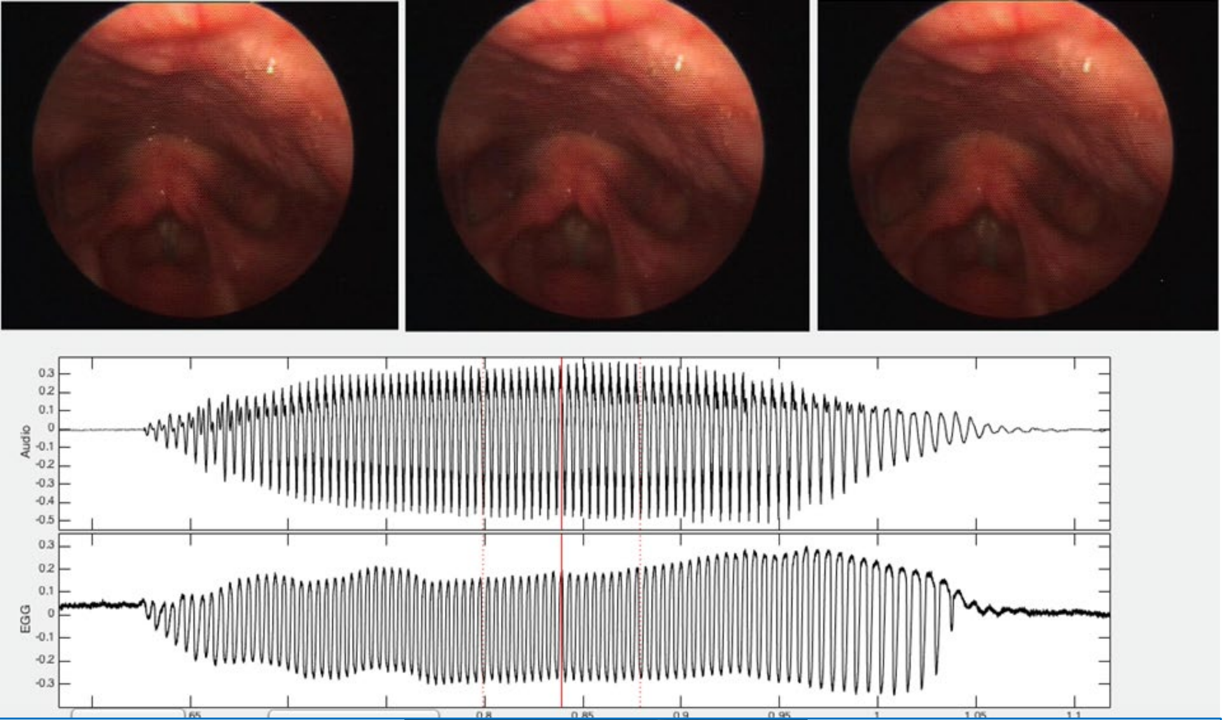


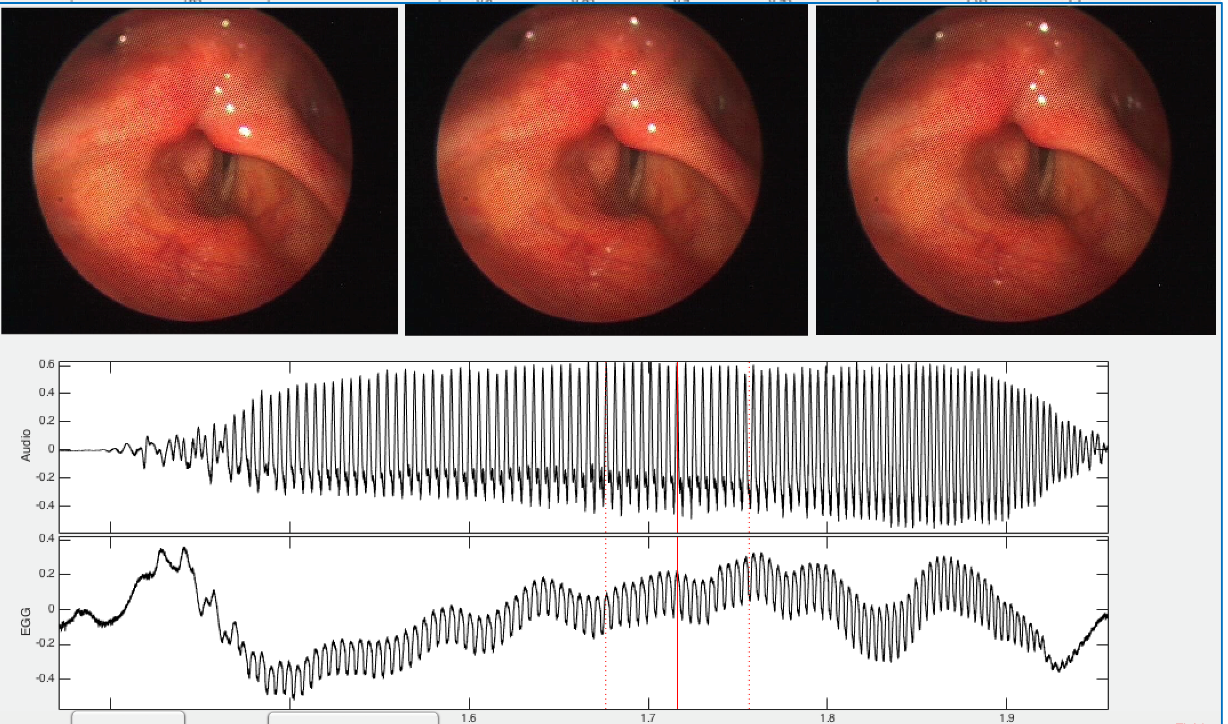






Rojas, Diaz & Demolin (2015)



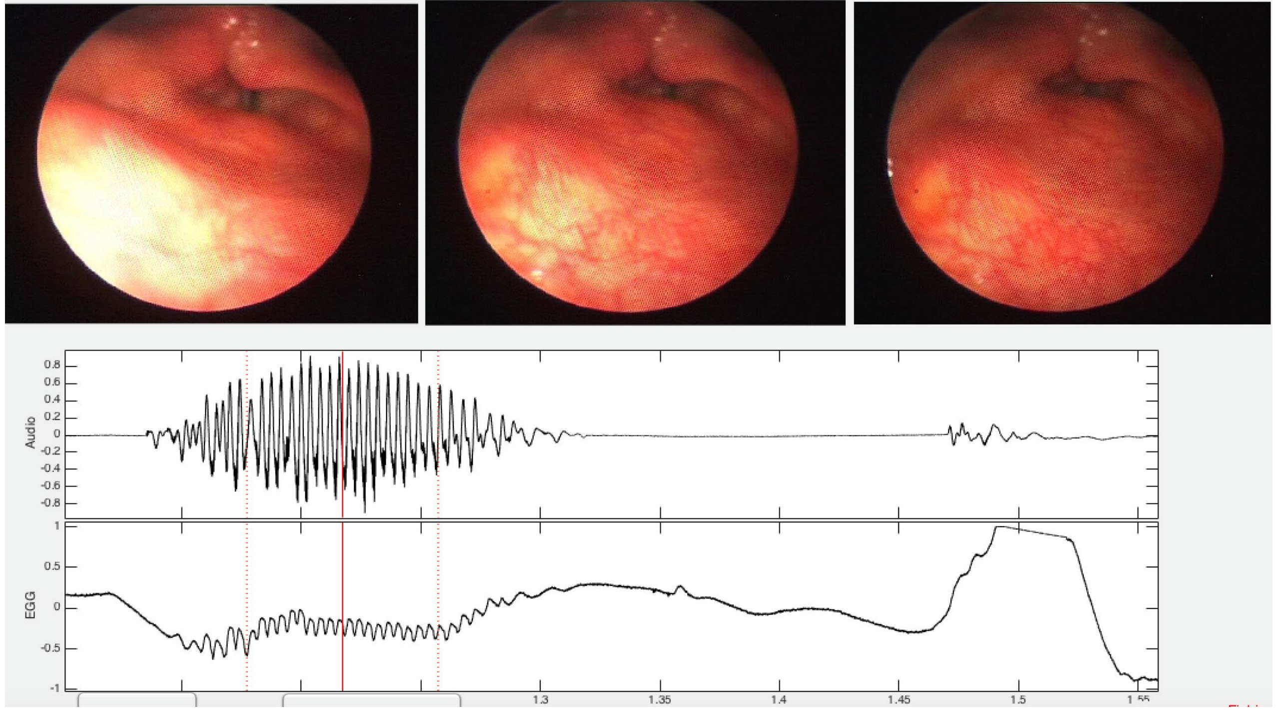


Two *nasa yuwe* speakers.

One of the two subjects had long-standing left recurrent palsy with forward tilting of the left arytenoid above the posterior glottic plane, and had the left vocal fold immobile in the paramedian position.

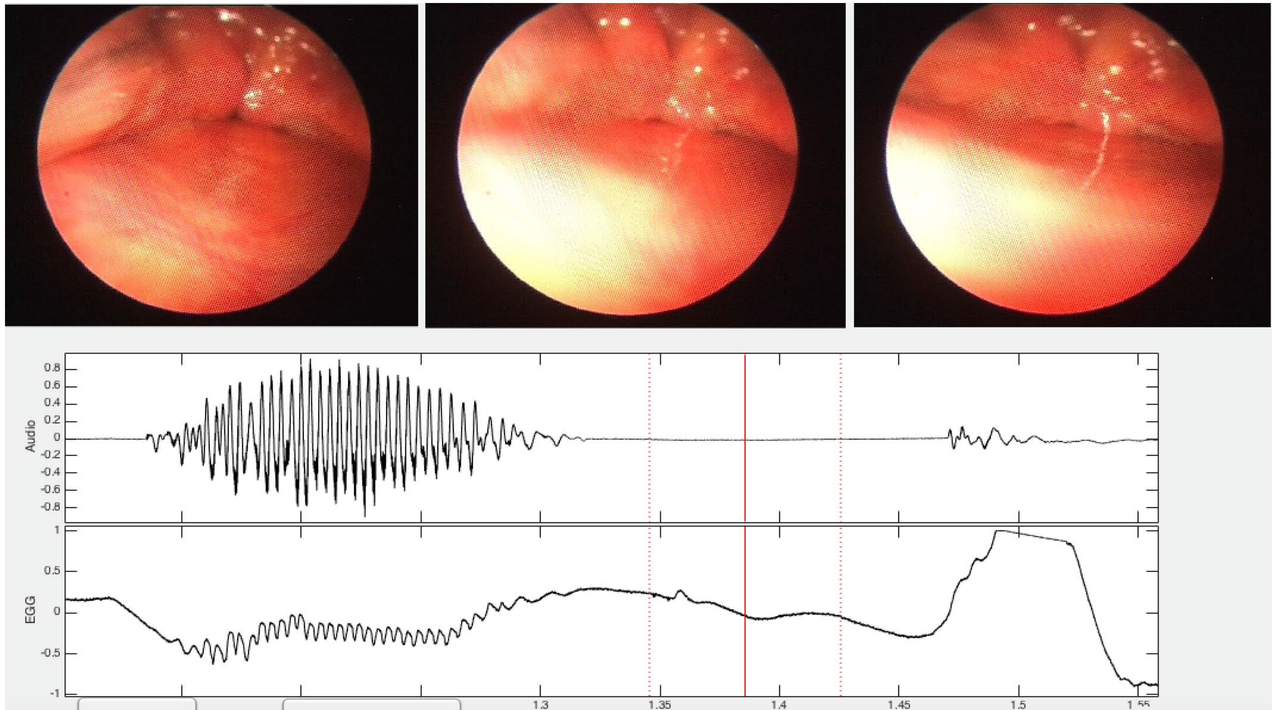
How does the subject with recursive paralysis compensate or adapt to produce the complex phonatory types of *nasa yuwe*?

Demolin, Amelot & Crevier Buchman (2019)



Compensated glottal vowel

Valve 1 (VF) > Valve1
Valve 2 (VB) > Valve 3 (ArF)



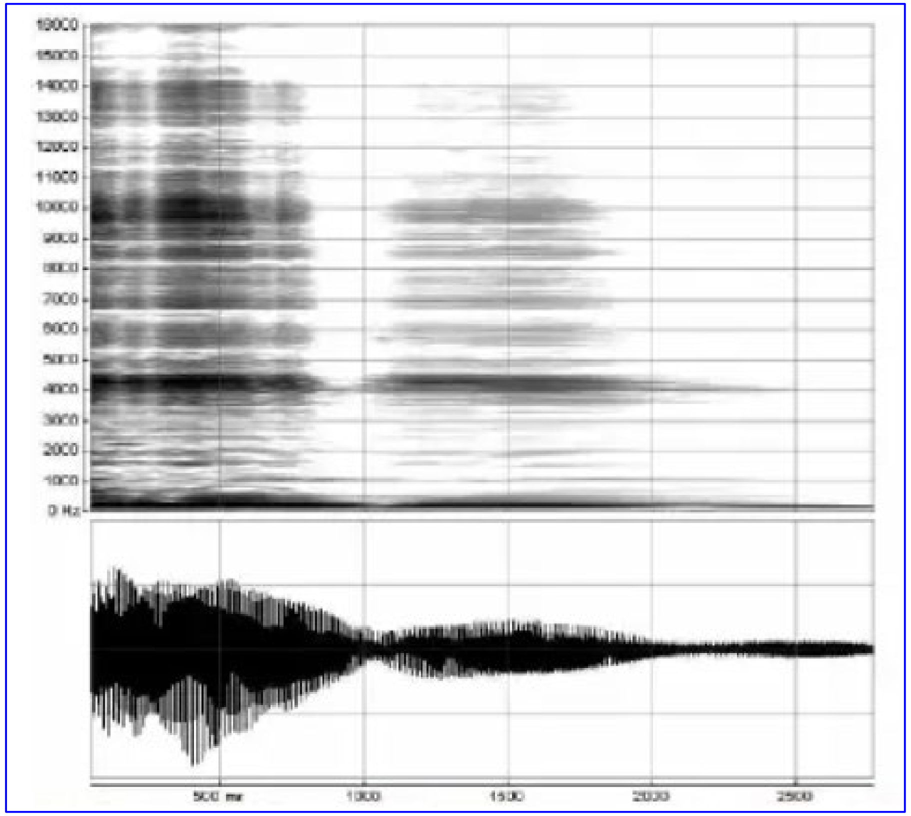
The Ethiopian Lyre Bagana : an instrument for emotion

Stéphanie Weisser

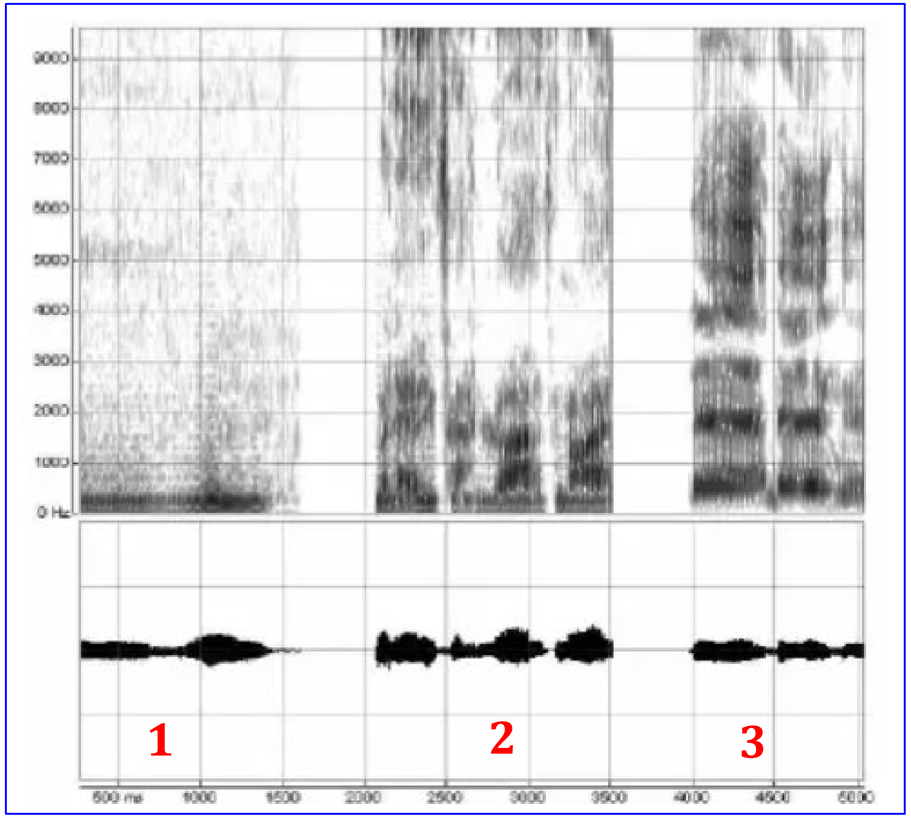
1. Emotion and traditional music.

2. Perception and musical characteristics

3. Culturally-coded timbres.



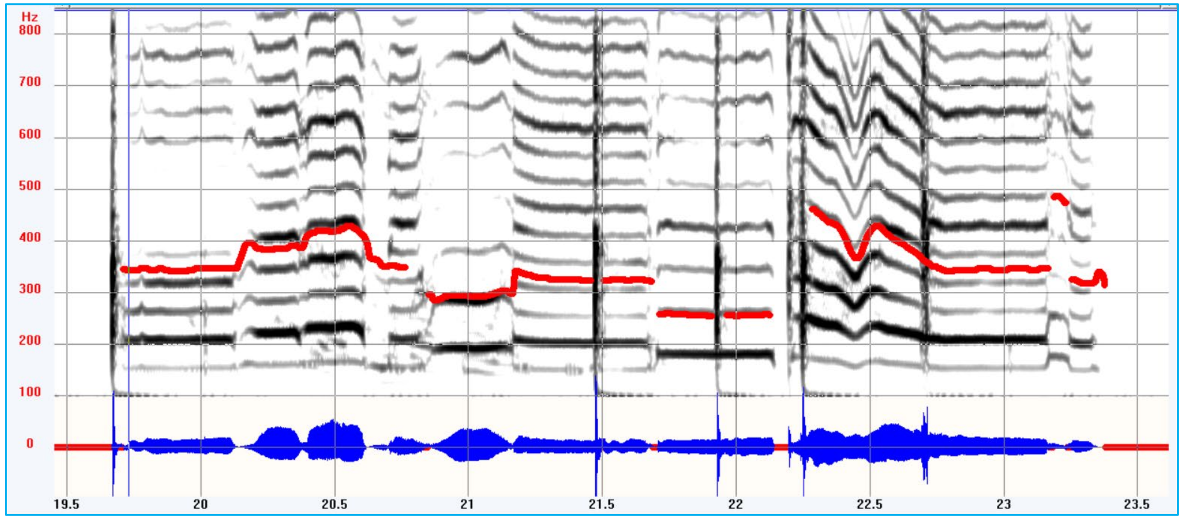
Narrow-band spectrogram of the buzzing sound of the bagana. Upper limit of the analysis : 15 kHz.
Recorded in Addis Ababa, 2003

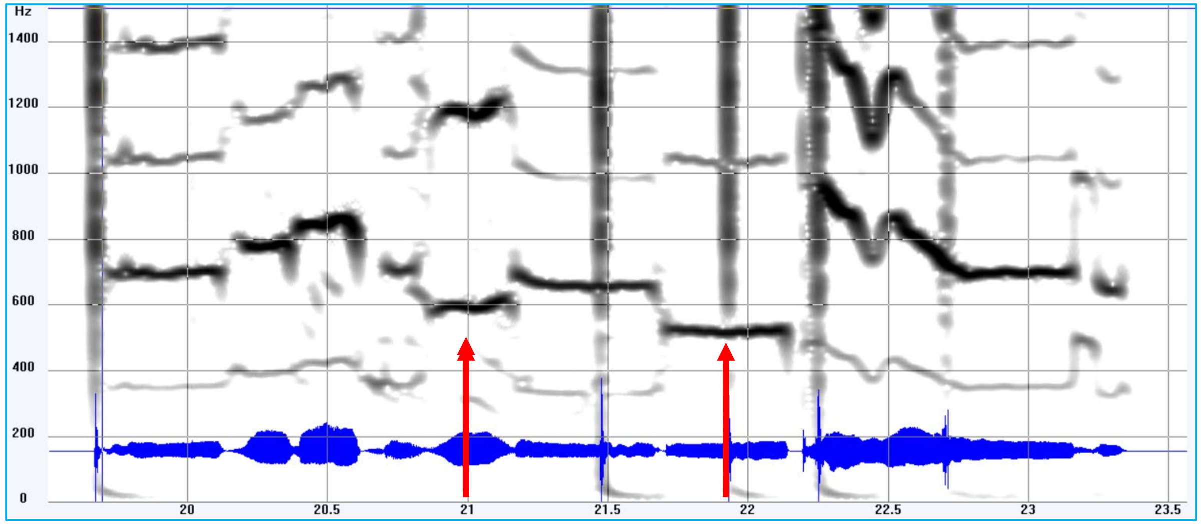


Wide-band spectrogram of: 1: Breathy voice by a bagana player 2: Harsh voice by a bagana player 3: Singing voice used with the Amhara other chordophones.
Recorded in Addis Ababa, 2004 and 2005

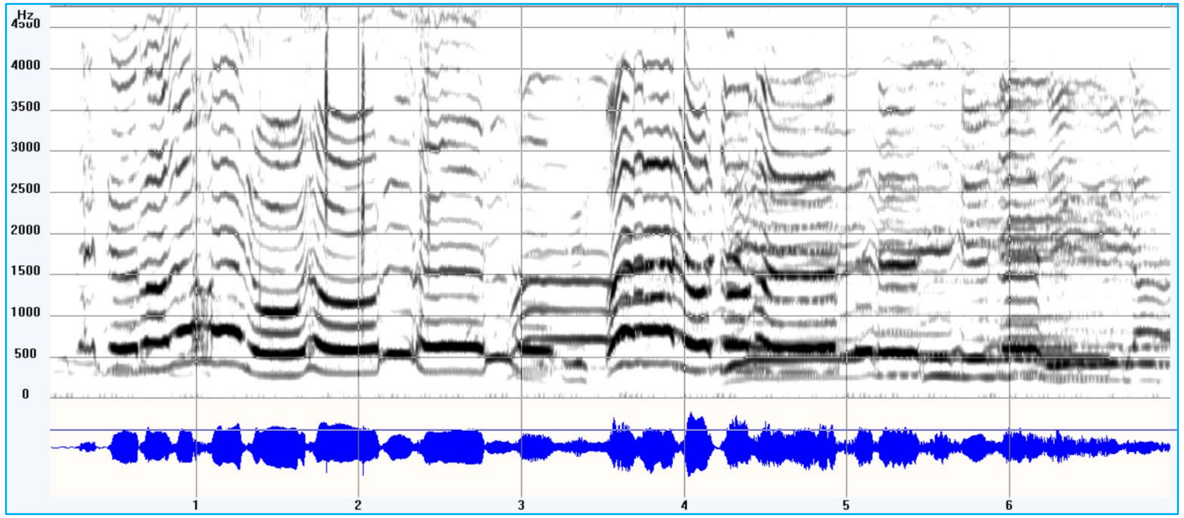
Yodel Djofe Pygmies

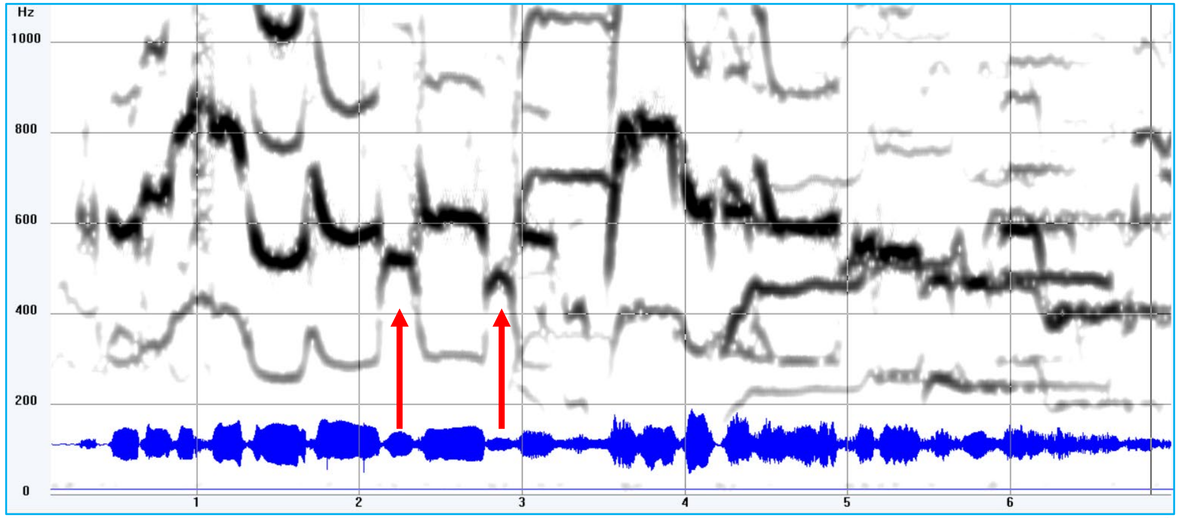
Women





Men





Relevant speech sounds in languages

(90% of the world's languages are unwritten)

Maasai interjections and ideophones

Conative animal interjections

ɣɣ	sound produced to incite bull to fight' (no API symbol)		
ḿḿ	bad smell	↓m̀m̀	good smell
↓p	calling come	↓s	go

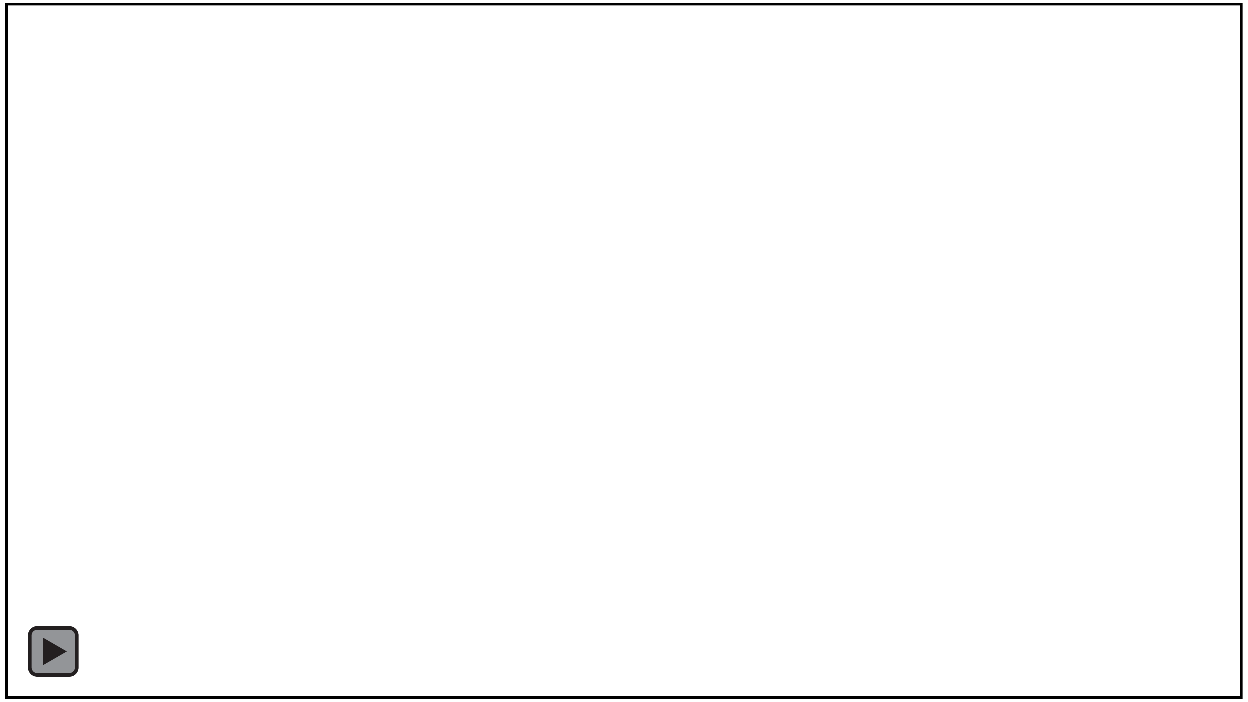
Ideophones and sound symbolism

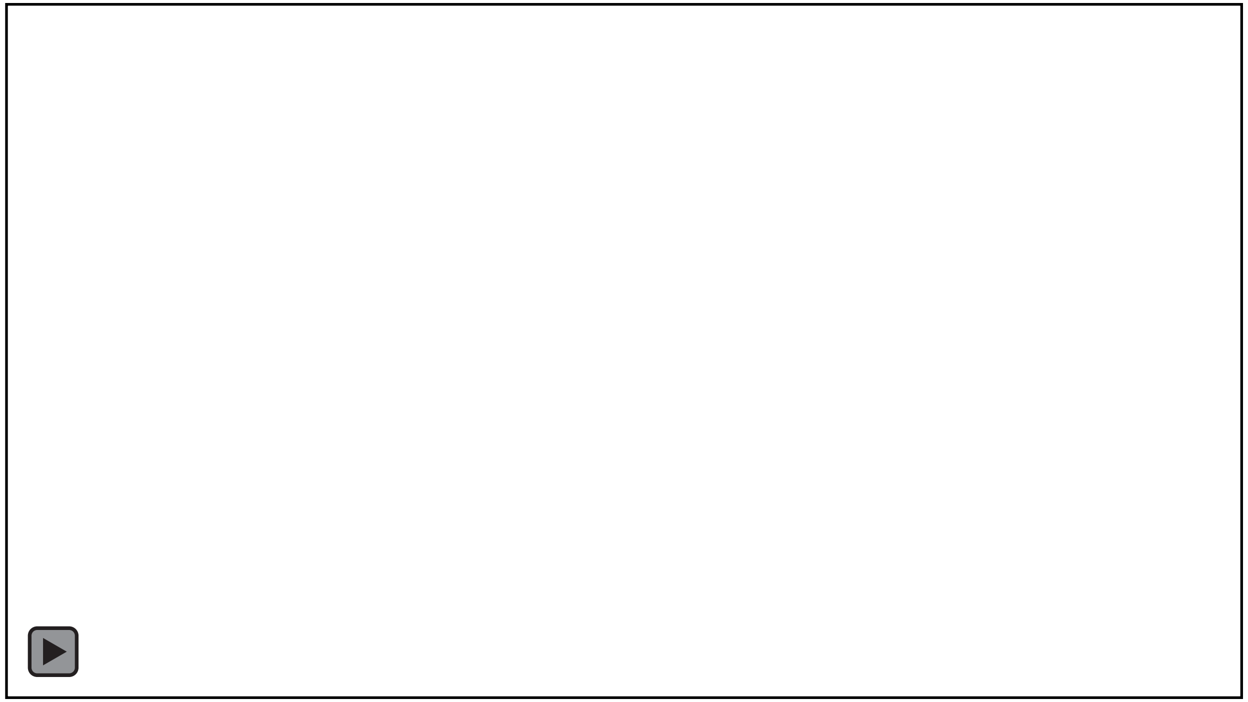
píal	brightly - of white	kúlukúlu	sound of pouring
píd	quickly - of jumping	kúm	hard bang sound
pío	depiction of bad smell	kúmúkúm	sound of walking
tír:ya	quietly	kúr:kúr:	sound of thunder
tíak	sound of dropping something	ŋóɓ	sound of gulping

Conative animal interjections: ɣɣ

‘sound produced to incite bull to fight’ (no API symbol)

Interjections: ḿḿ ‘bad smell’; ↓m̀m̀ ‘good smell’.





Karani, Ghio & Demolin (2022)







2. Sound system structures and their complexity

Basic > complex > elaborate (Lindblom & Maddieson 1988)

Complexification of structures and speech sound production

The complexification of labial consonants follows a pattern similar to the acquisition of speech in children → m, p, b, w > m, p, b, f, v, w….

This follows the development of motor skills. The jaw and lips form a simple control unit in the early stages of babbling.

The mandible and lower lip are highly correlated.

Then there is a differentiation between the lower and upper lips towards actions, but independent in relation to lip gestures.

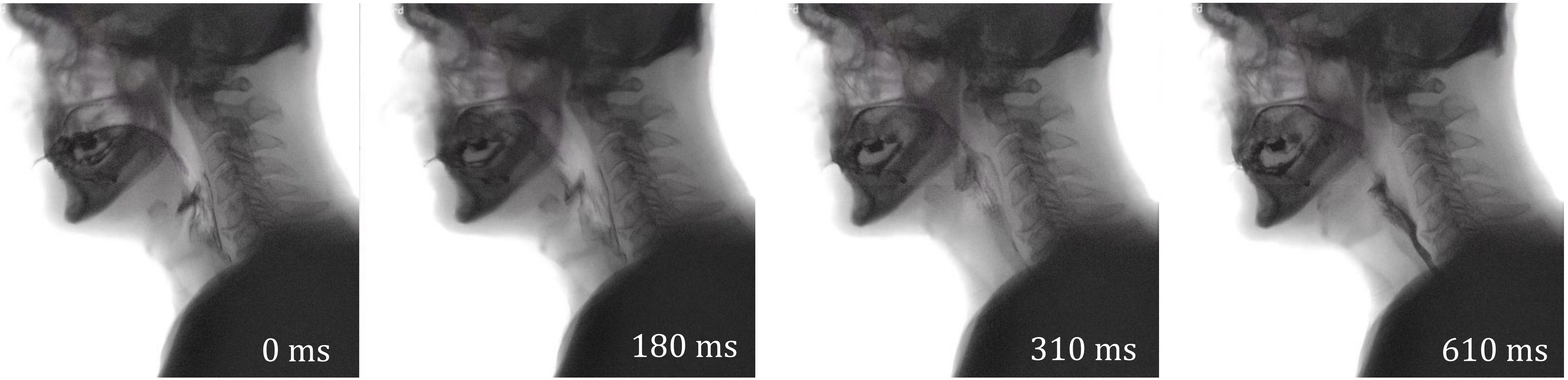
Differentiation > Integration > Expansion (Refinement)

Characteristics of Open Systems

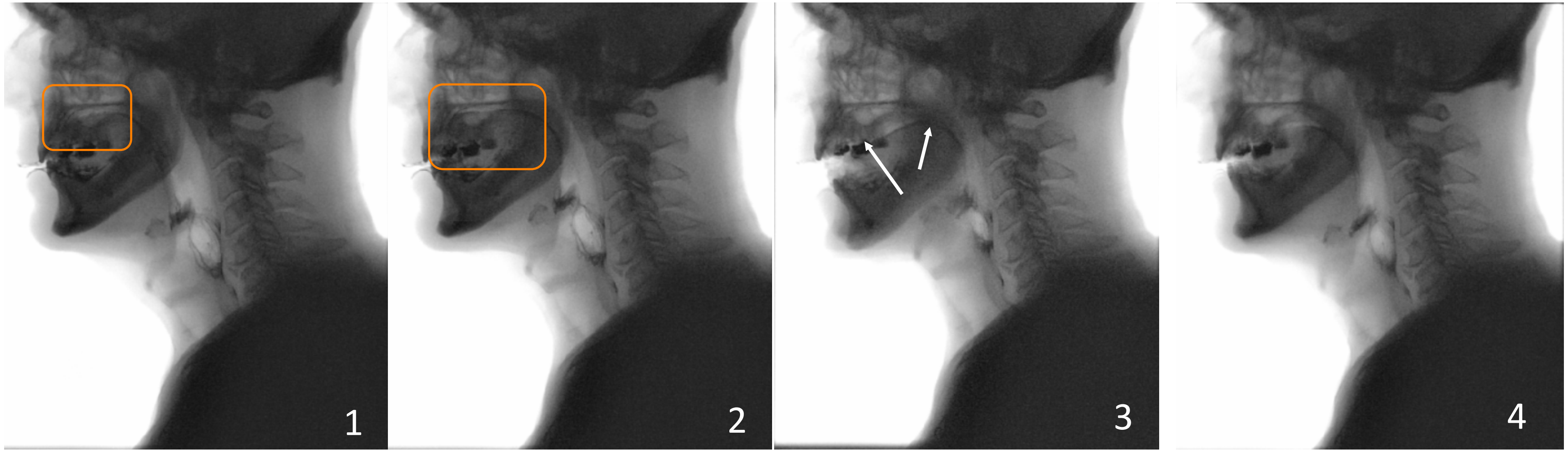
m	p	b	f	v	w	ʙ	ⱱ	ʘ	
mp	ɓ̥	ɓ	ɸ	β	β̞	ʙ̥		ʘ^{h}	kp
mb	p^{h}	b^{ɦ}	f^{w}			mʙ		ʘq^{h}	gb
bmb	p^{w}	b^{w}						kʘ^{ʔ}	kʙ̥
bm	p^{x}							ŋʘ	gɓ
m^{ʔ}	pʼ							ʘqʼ	qɓ
m^{w}									ɢɓ

Speech embodiement : swallowing and clicks

Phases of swallowing saliva



Palatal click [ǂ] mechanism phases



Demolin
Kingston
& Circiu
2023

Are clicks relics of early sounds produced by humans?

They occur in languages spoken by people whose common genetic ancestor lived more than 35,000 years ago, perhaps as long as 55,000 years ago (Knight et al., 2003; Tishko et al., 2007).

Clicks are no more likely to be retained than other speech sounds (Sands & Güldemann (2009), Traill (1986), Traill & Vossen (1997)

They might have been innovated independently in the languages where they now belong to their consonant inventories.

Clicks emerged in Khoesan languages within a linguistic area where there is already a great diversity of sounds and sound systems, an environment that would dynamically encourage further increases in complexity.

Sound change

Epigenetic or differentiative regulation reflecting changes of state.

⇒ shows the quantal nature of speech.

Laws for the dynamics of regulatory networks: feedback loops.

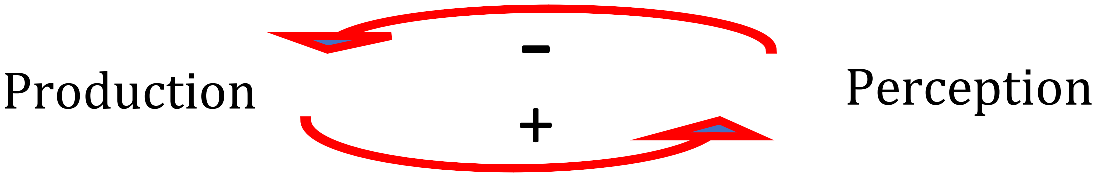
(Thomas 1998)

Propagation > principles of dynamic and self-organized systems.

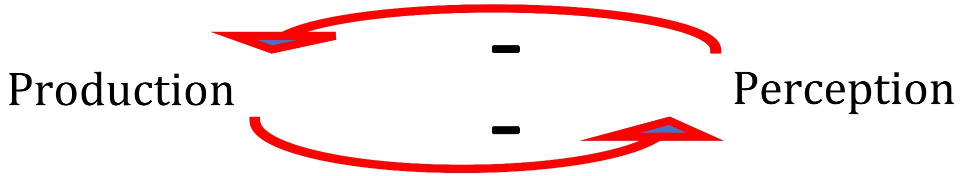
⇒ Population dynamics & models, logistic equation.

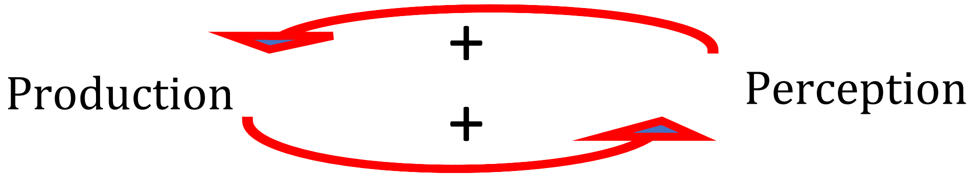
r-K continuum (intrinsic rate of natural increase and carrying capacity)

Speech regulation: homeostatic regulation



Sound change: differentiative or epigenetic regulation

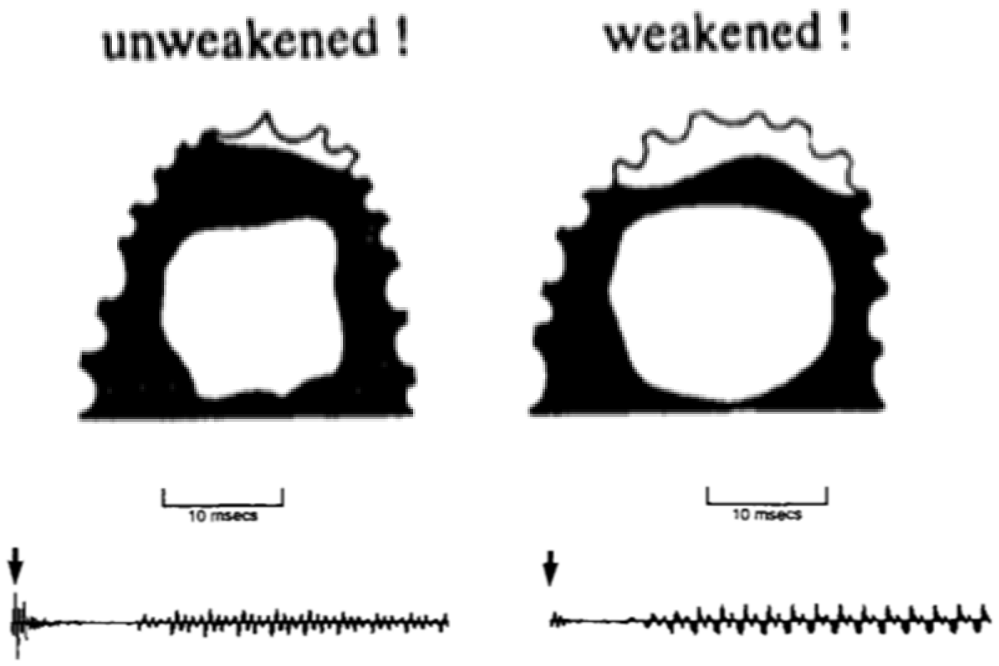




A feature is lost and another already present in the signal is amplified: tonogenesis * ba > pà * pa > pá

An ambiguous sound is misinterpreted > changes from front to back rothic consonants (r > ʀ: Dutch, Portuguese, Swedish, Carib languages)

Loss of clicks: ! > k



! > g, k, ŋ

!go	> go	ant
!xan	> kxan	saw
!nu	> ŋu	land

An articulation is weakened and this has an acoustic cost.

! > k are both grave and abrupt.

k has a weaker intensity.

The alveolar articulation is not encoded at all in the acoustic filtering of the impulse produced there.

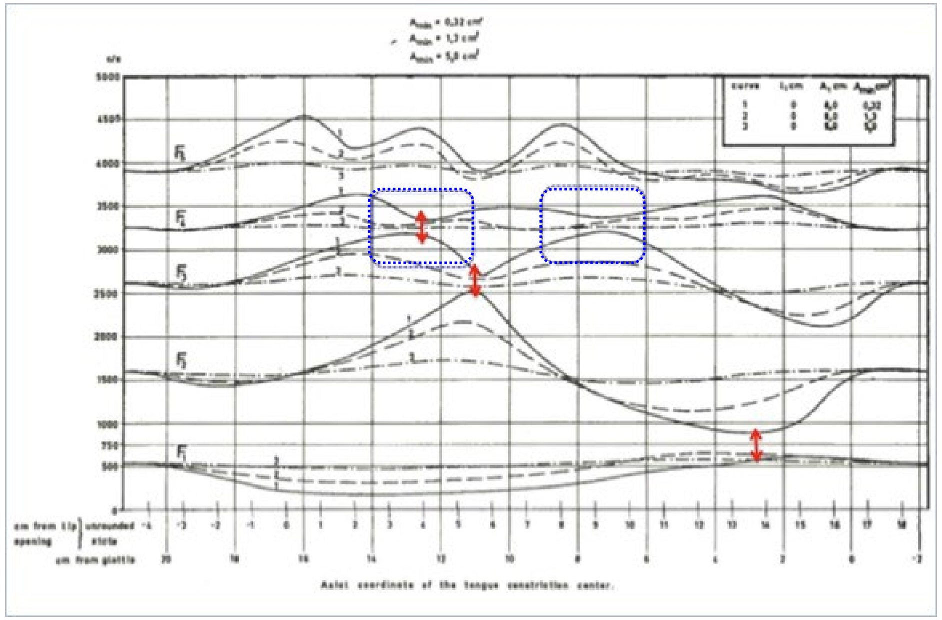
It is the cavity delimited by the back closure of the click which determines the grave spectrum.

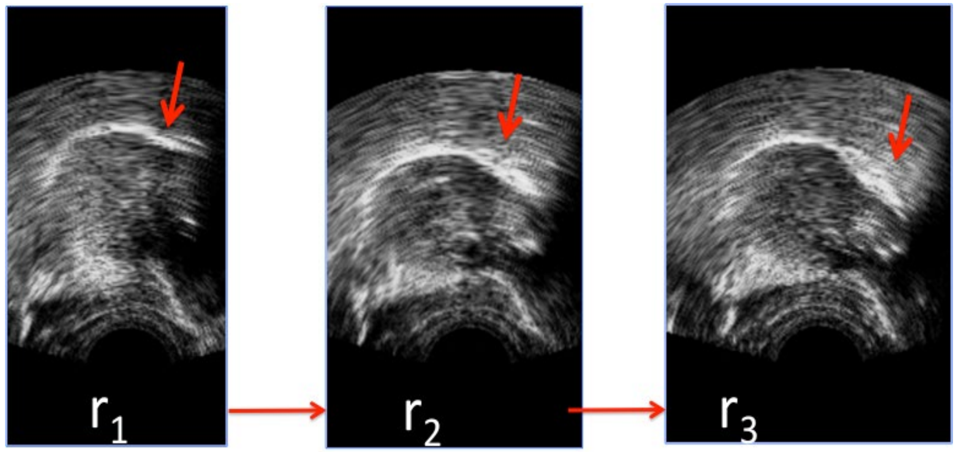
Traill p.c. clicks are (acoustically) just consonants like others but more intense!

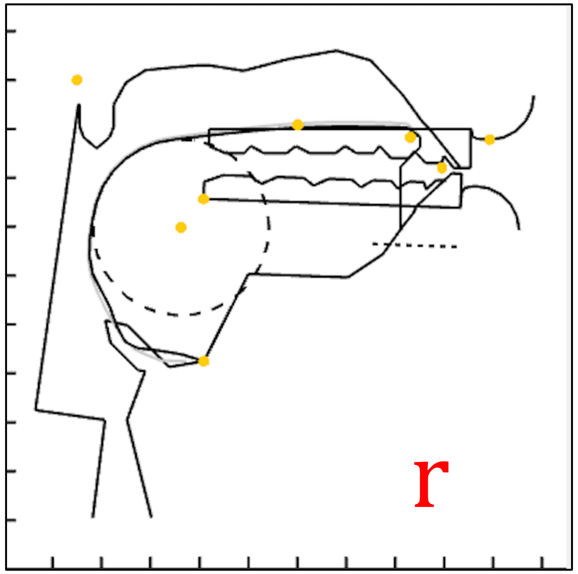
Traill & Vossen (1987)

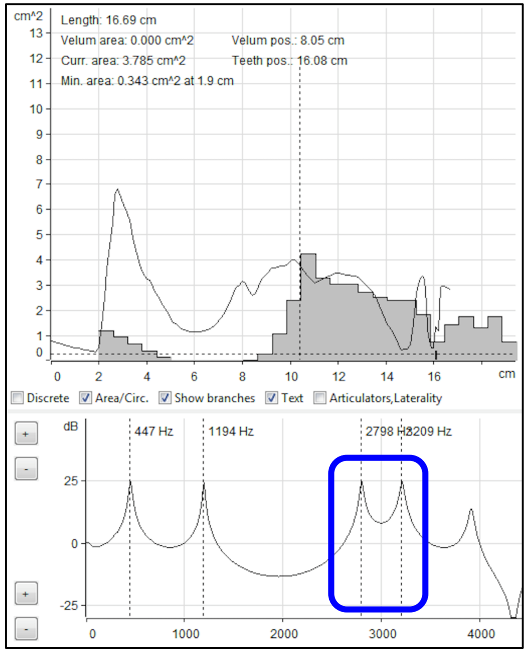
Quantal change

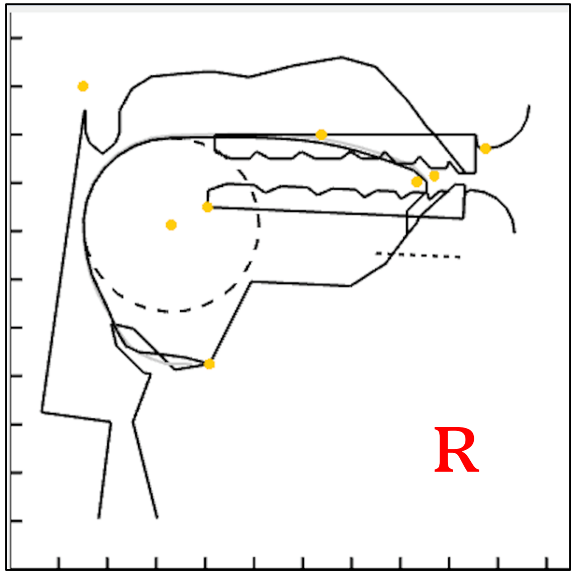
r > R

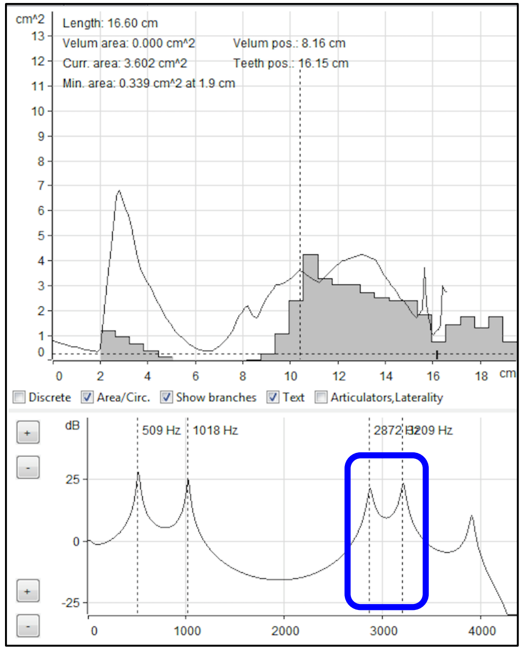










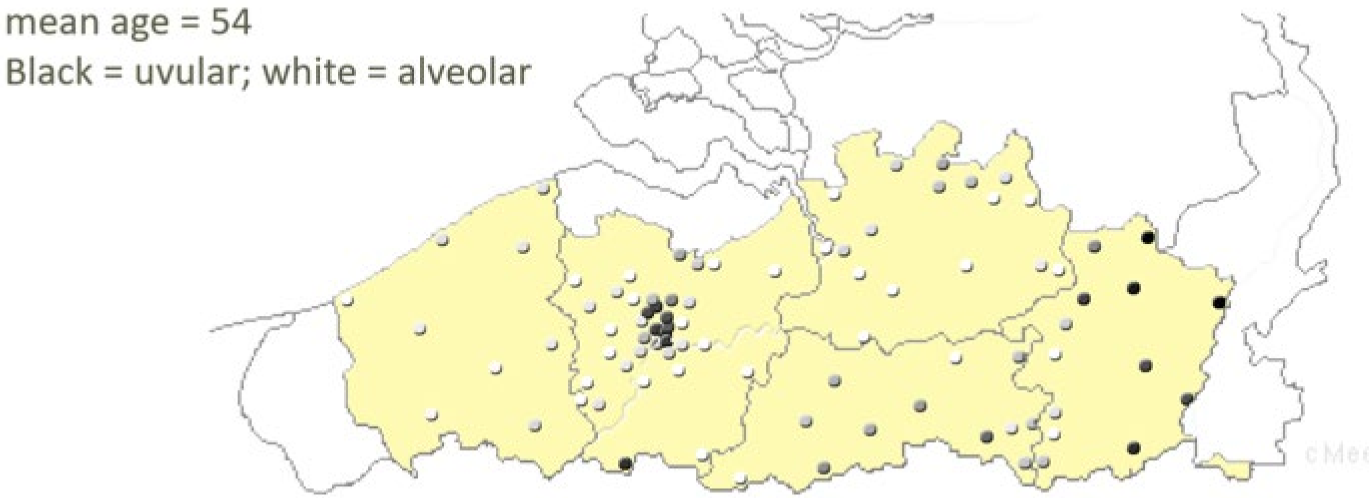


Demolin
Van de Velde
2022

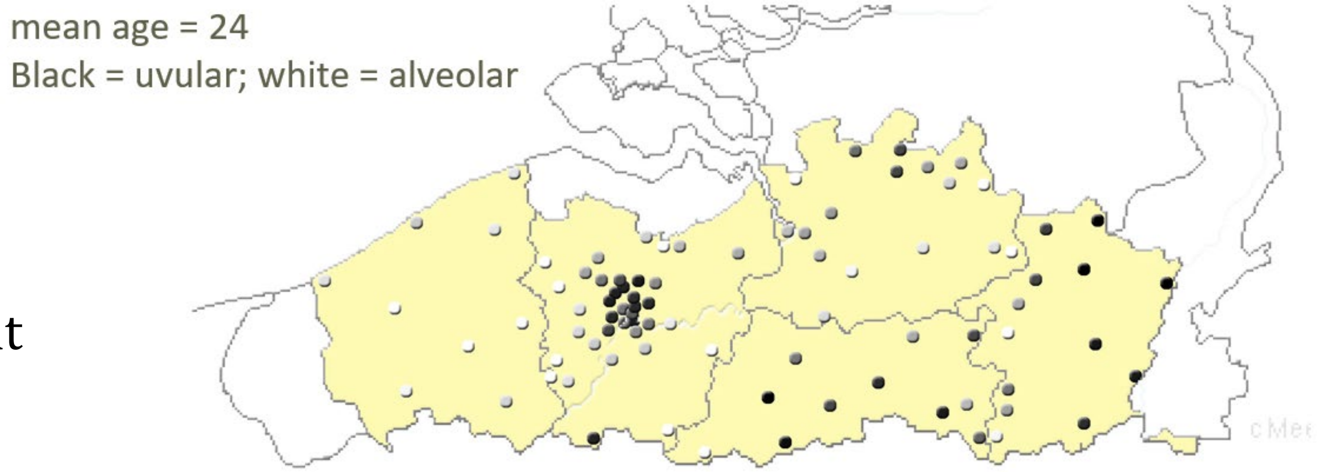
Sound change Propagation

r > ʀ in Flanders

Front-back scores old speakers



Front-back scores young speakers



Tops (2009)
Van de Velde, Tops & Van Hout (2013)

Logistic equation

$$\frac{dN}{dt} = rN \, \frac{(K-N)}{K}$$

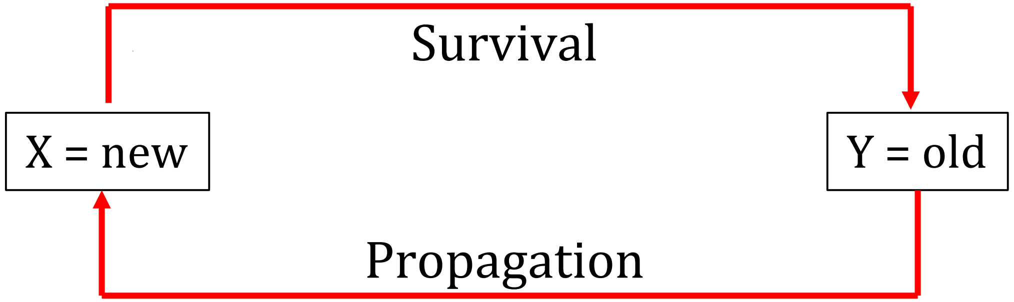
r-K continuum → *r* intrinsic rate of natural increase and *K* carrying capacity

→ Speech : sound variability and perceptual discrimination

Attribute	r selection	K selection
Temporariness	Variable unpredictable	More constant predictable
Population size	Variable below carrying capacity	Constant close to carrying capacity
Intra-interspecific comp.	Variable, often weak	Usually strong
Selection favors	Rapid development	Slow development
Length of life	Usually shorter	Usually longer
Leads to	High productivity	High efficiency

Stage-structured matrix of sound variation growth

Projection map for a population $$N_{t+1} = \lambda N_t$$



$$X_{t+1} = mY_t$$
$$Y_{t+1} = pX_t$$

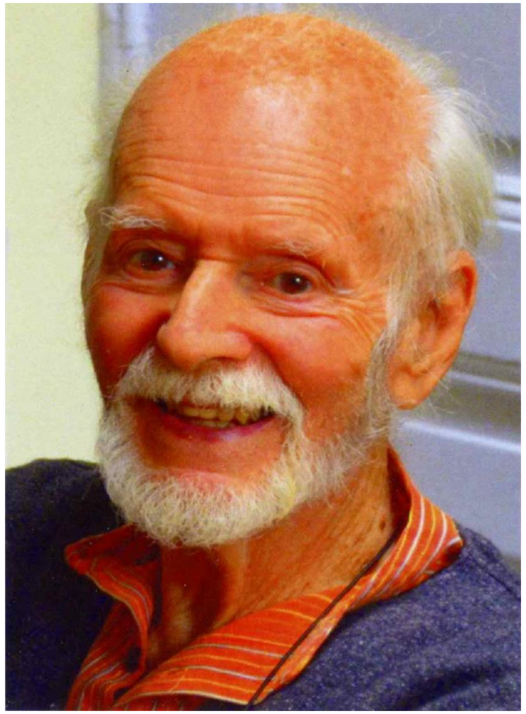
m = number of variants produced in an age group

p = probability that a variant will survive to become an old variant

$$\begin{vmatrix} X_{t+1} \\ Y_{t+1} \end{vmatrix} = \begin{vmatrix} 0 & m \\ p & 0 \end{vmatrix} \begin{vmatrix} X_t \\ Y_t \end{vmatrix}$$

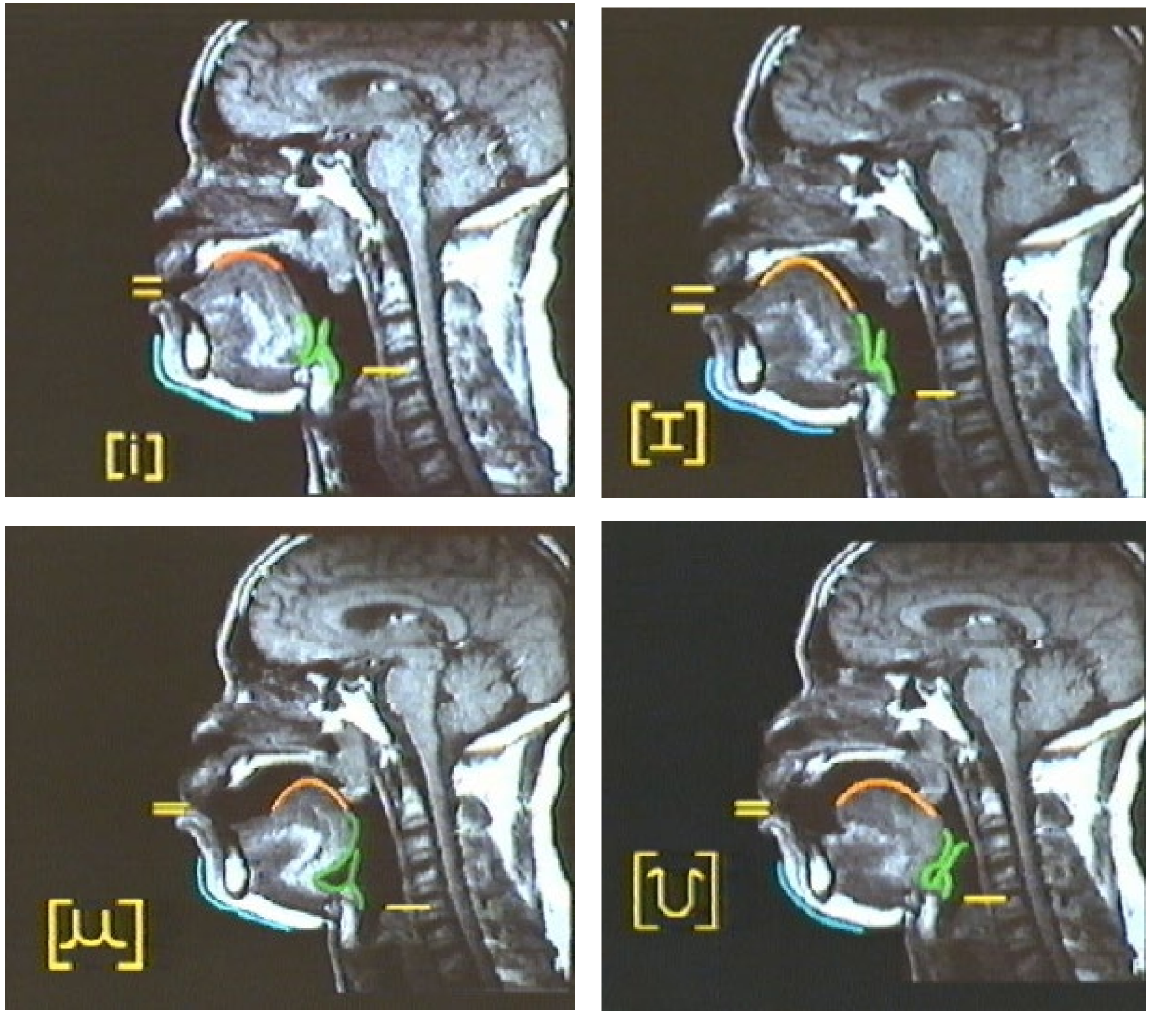
$$\boldsymbol{N_{t+1} = PN_t}$$




3. Biological, physiological and physical mechanisms of speech ⇒ development of experimental methods

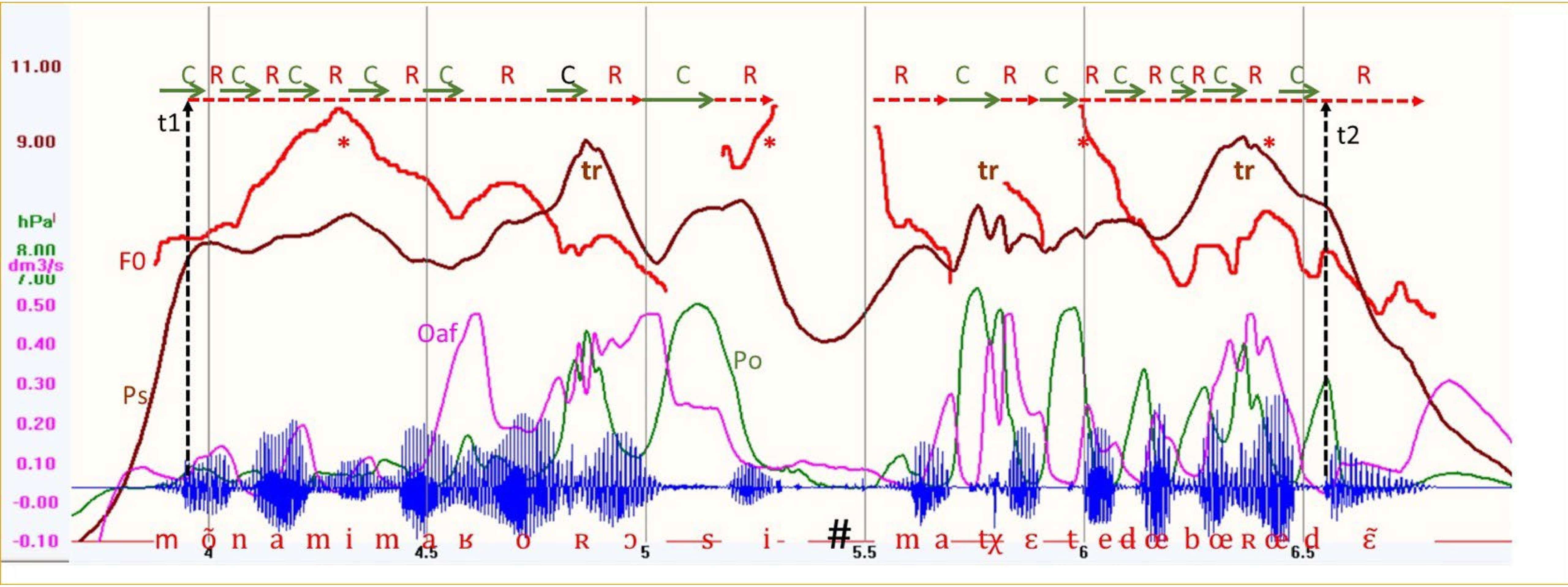
+/- ATR vowels in *Mangbetu*



Nande high/back super closed vowels

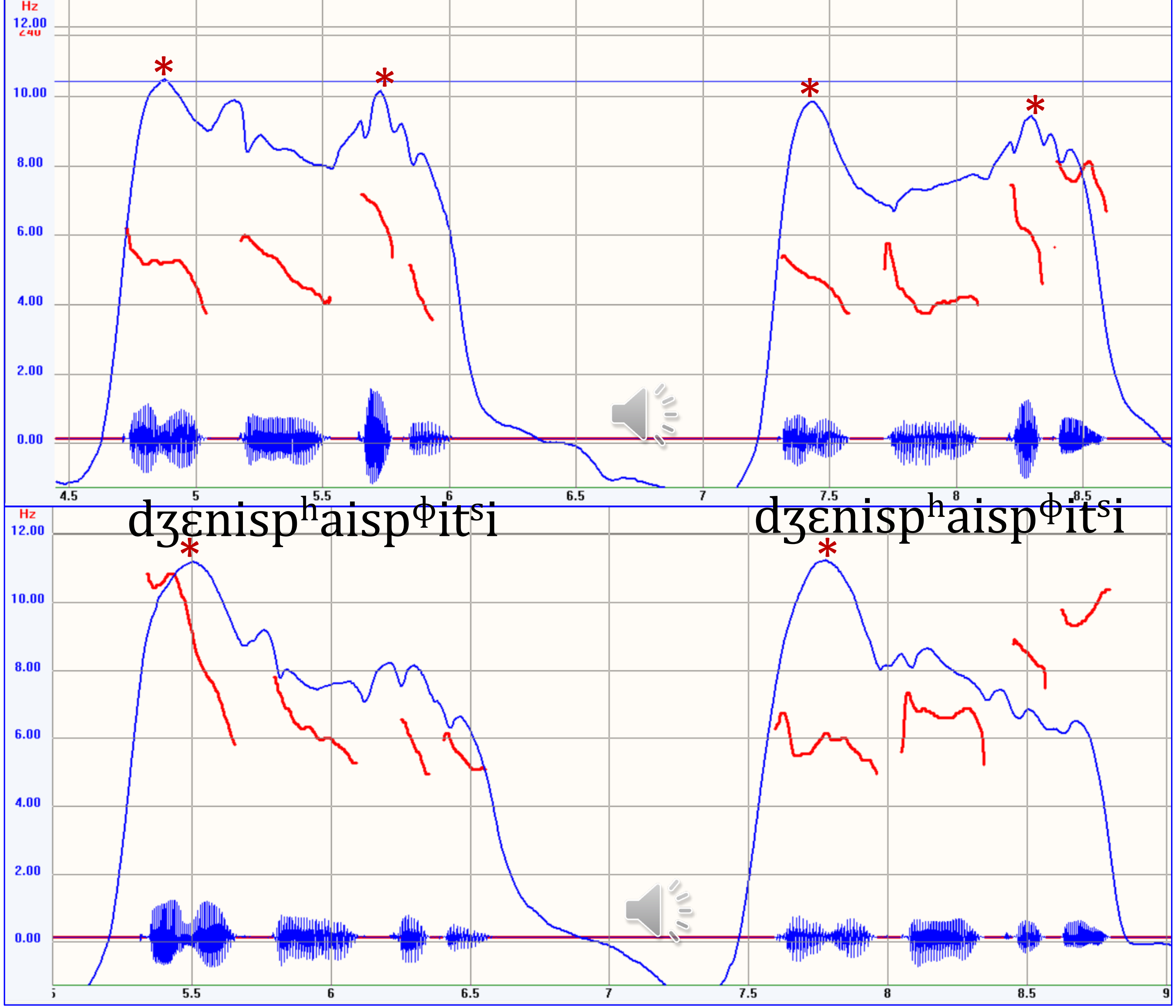
An old and not solved problem requiring subtle answers

The relation between subglottal pressure (Ps) and fundamental frequency (F0)
F0 declinaison? Stress? Intonation? How are Ps and F0 controlled?



repeat photo twice

F0 Ps

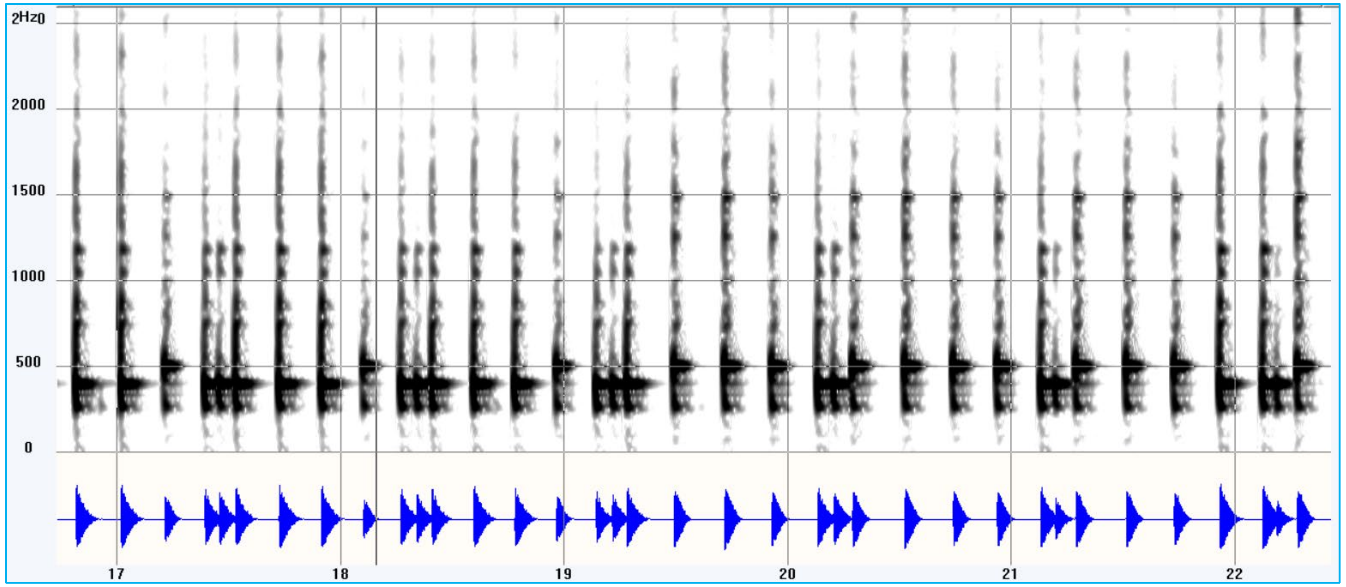


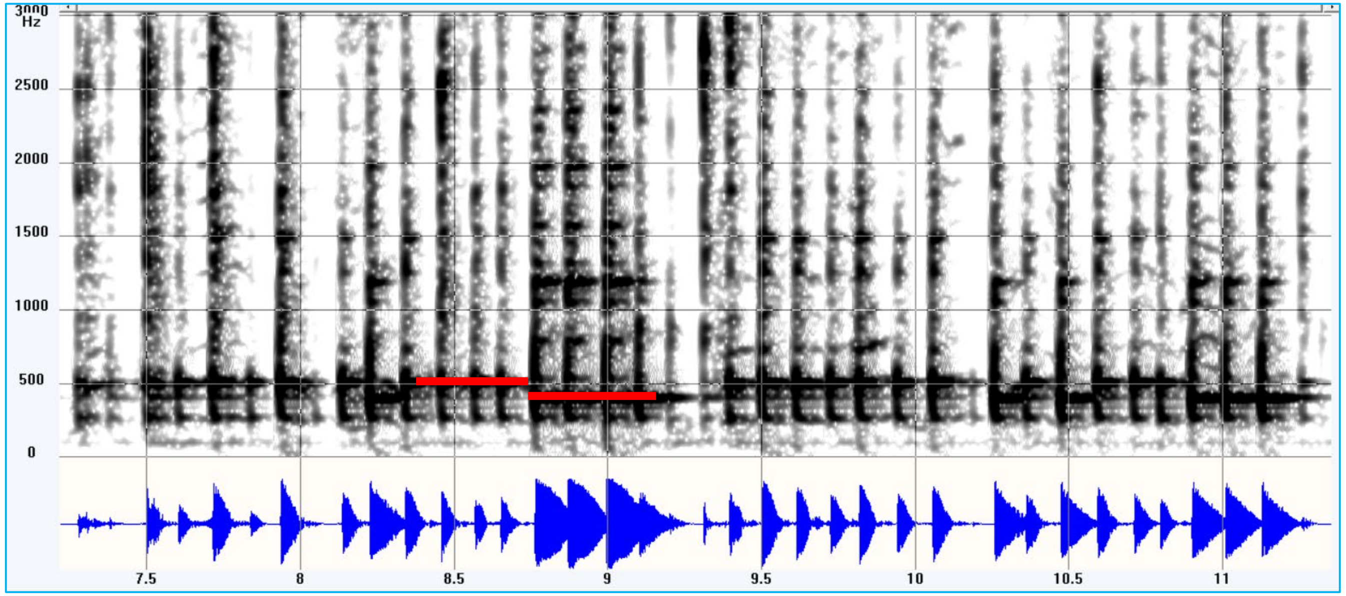
Statement

Interrogative

Ethnomusicology

Mangbetu

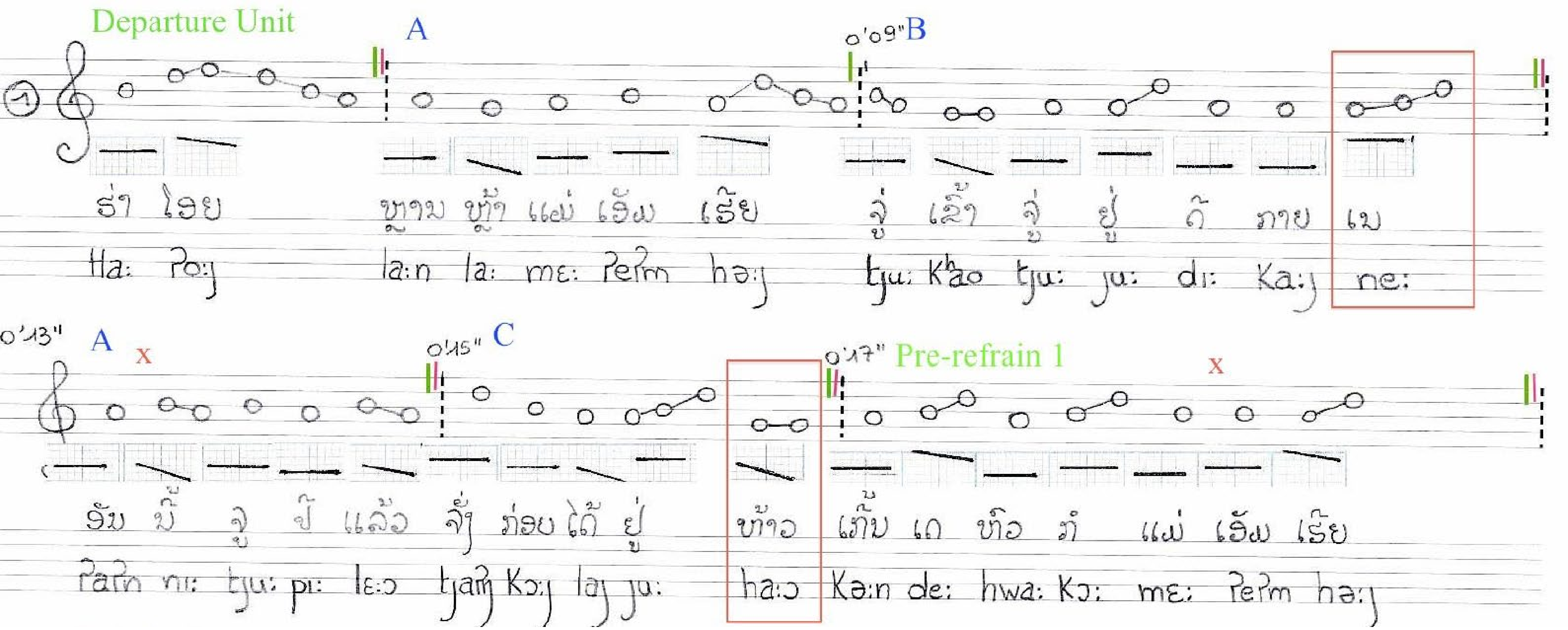




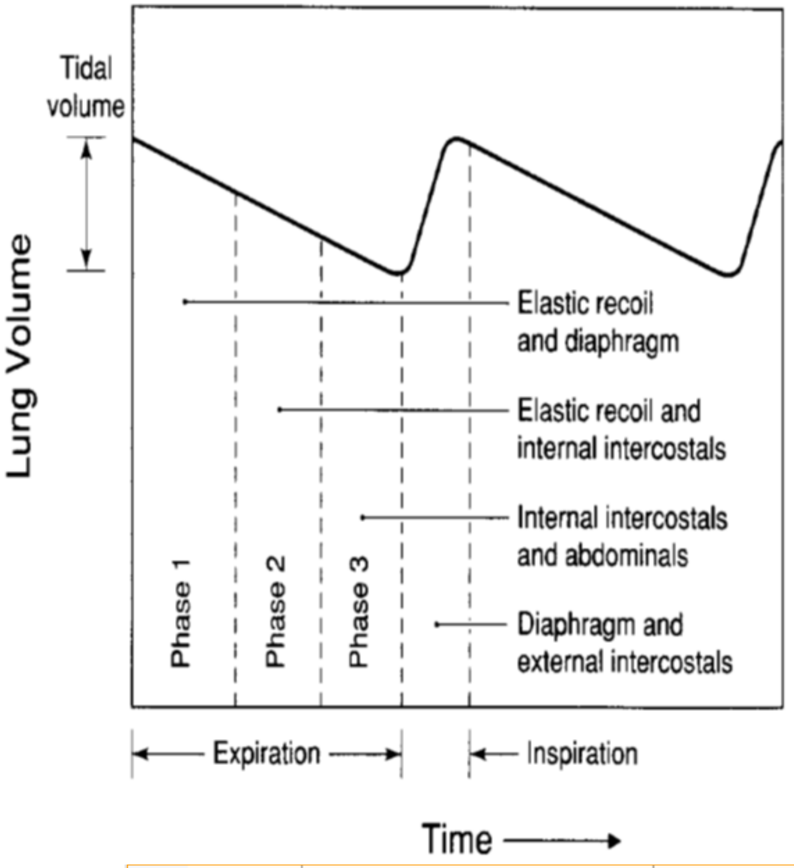
The ratio between tonal melody in speech and song

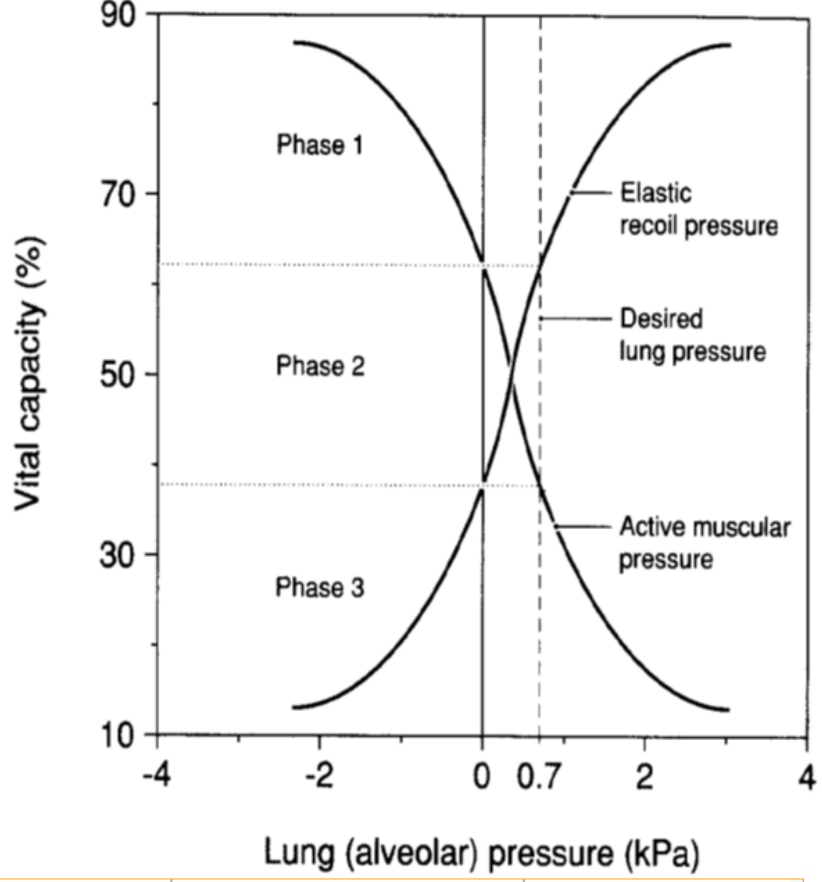
Enregistré par Marie-Pierre Lissoir
Na Kay (Hua Phan, Laos), 13-02-2011
Interprété par Lek
Références: audio STE-016.wav
vidéo M2U00187.mpg

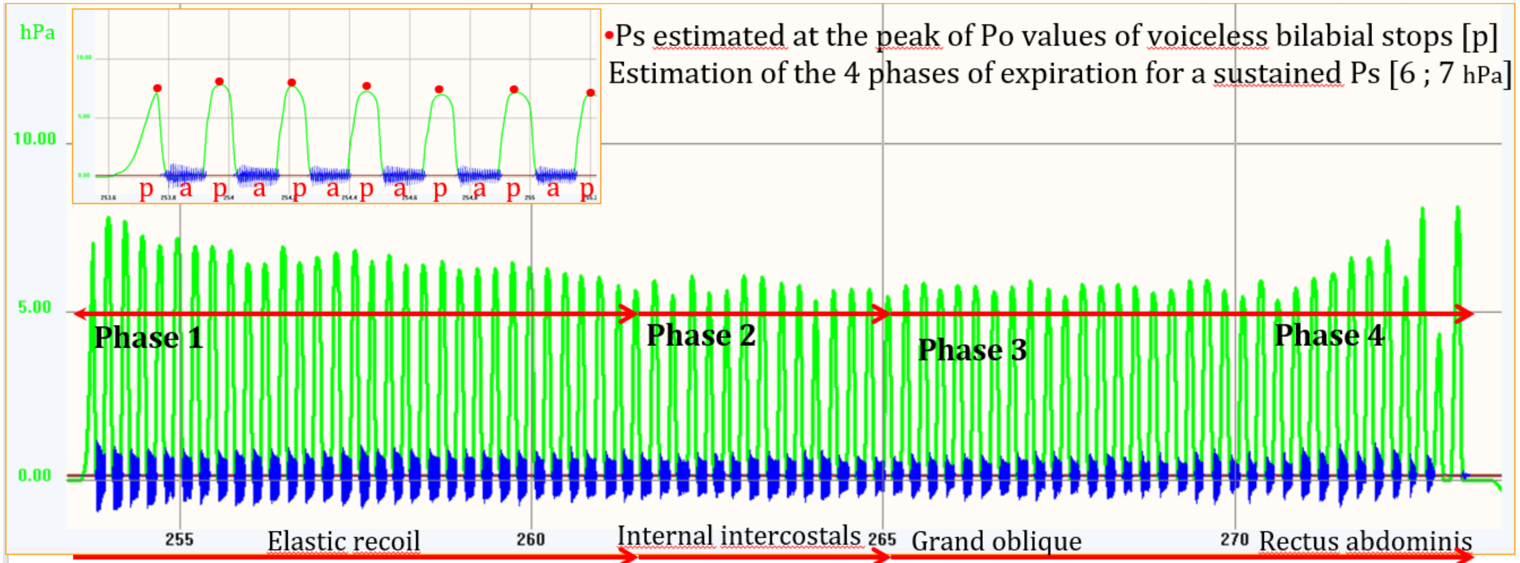
Khap 18
Khap khuan, par Lèk de Na Kay



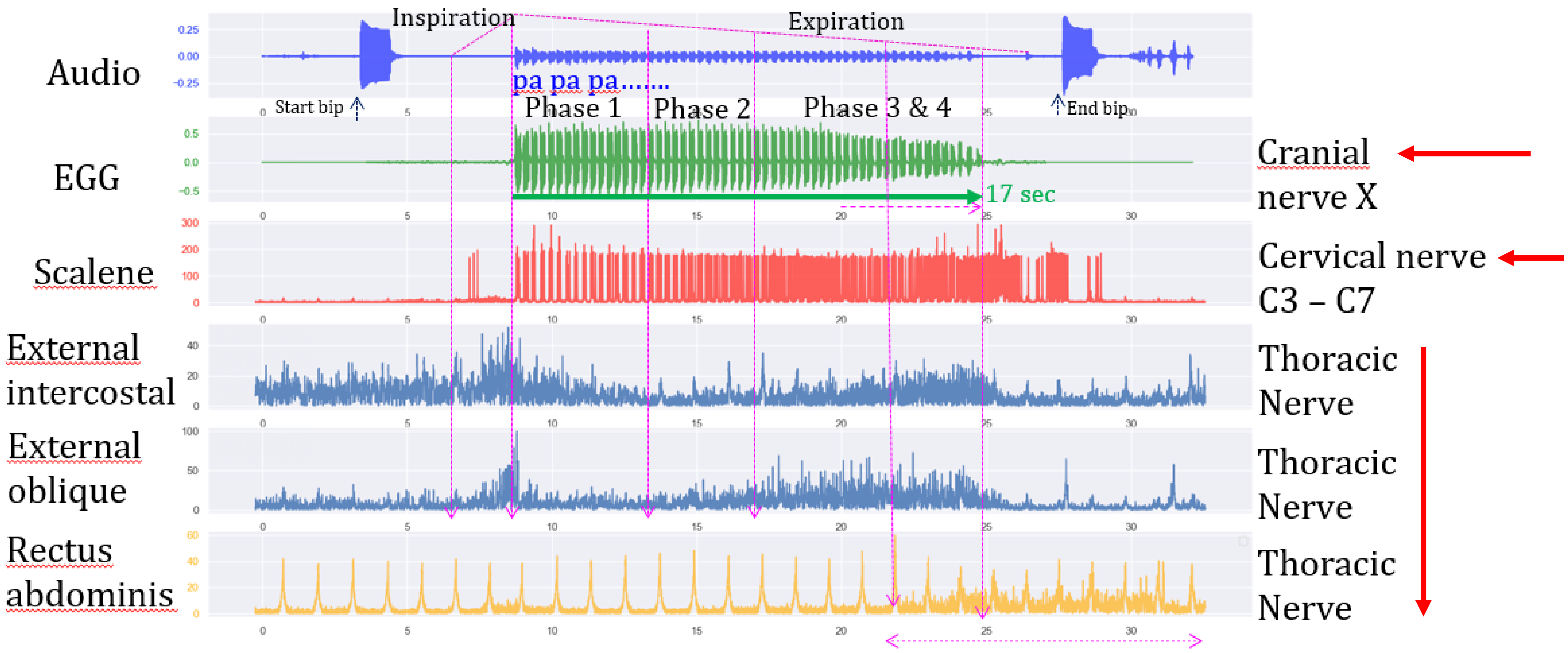
Control and regulation of Ps during respiration







Control and regulation of Ps

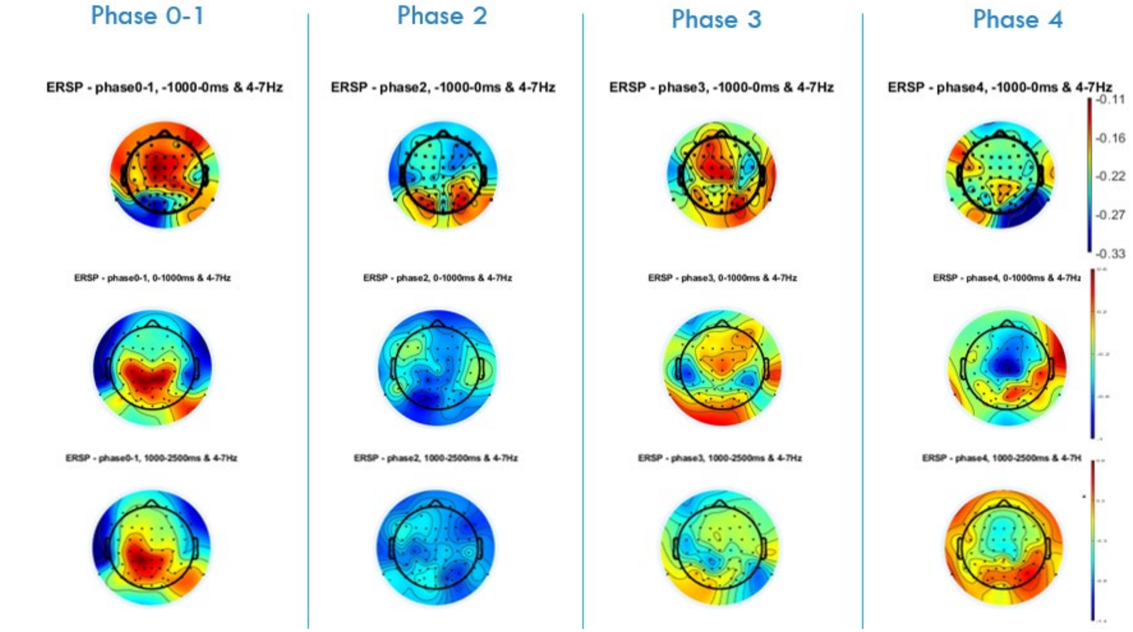


Phase 1 : Elastic recoil ; Phase 2 + internal intercostals ; Phase 3 External oblique; Phase 4 Rectus abdominis

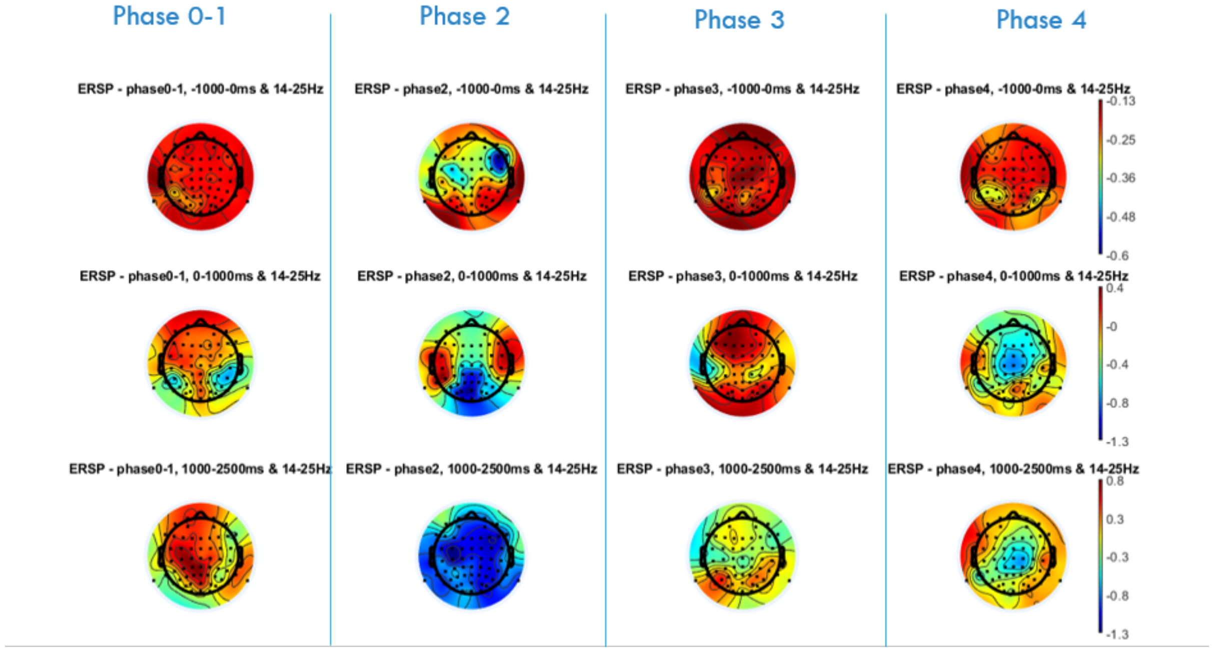
Brain activity? Role of chemo-receptors (O_2 cusumption)? Role mechanoreceptors?

Hashemi, Cheron, Demolin & Cebola

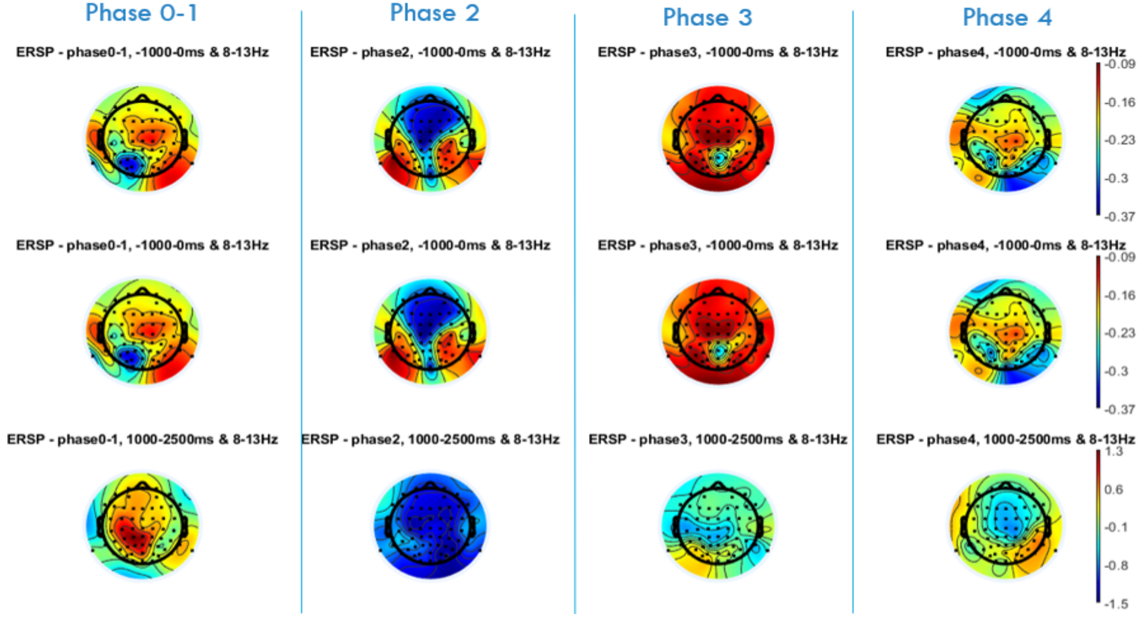
THETA TOPOGRAPHY



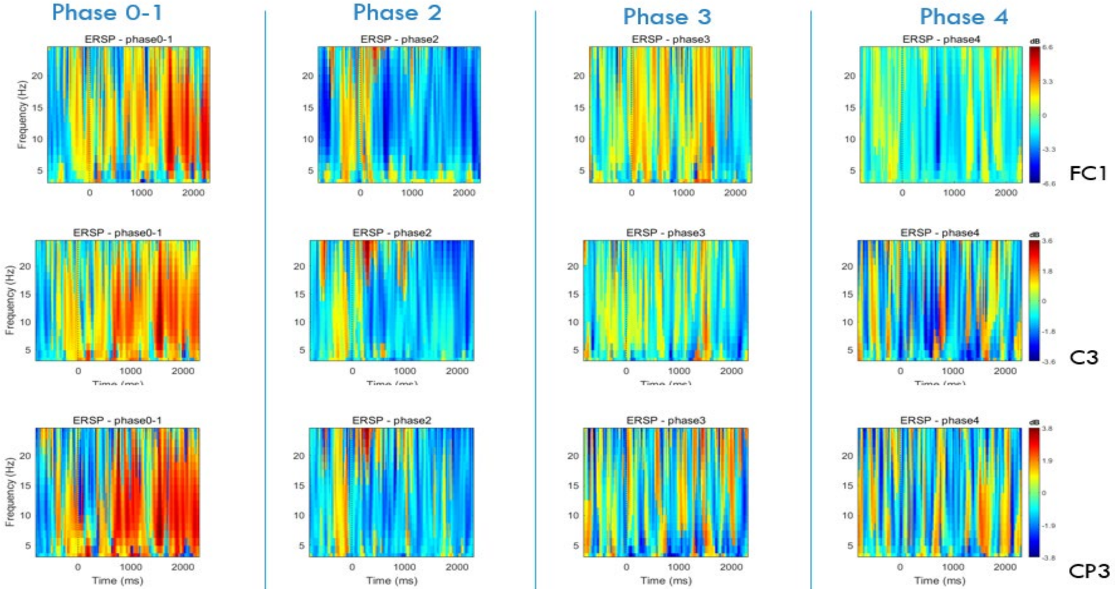
BETA TOPOGRAPHY



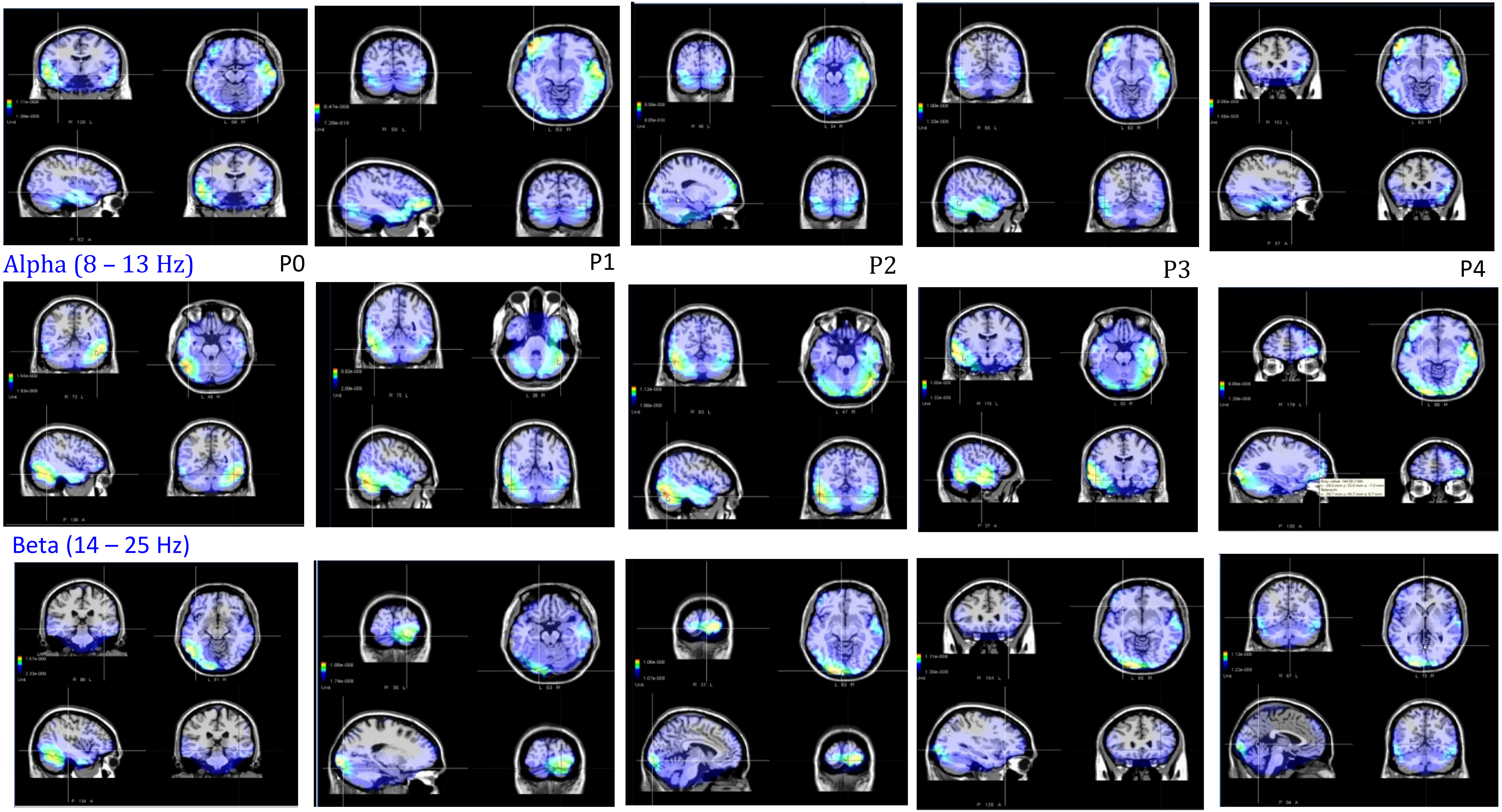
ALPHA TOPOGRAPHY



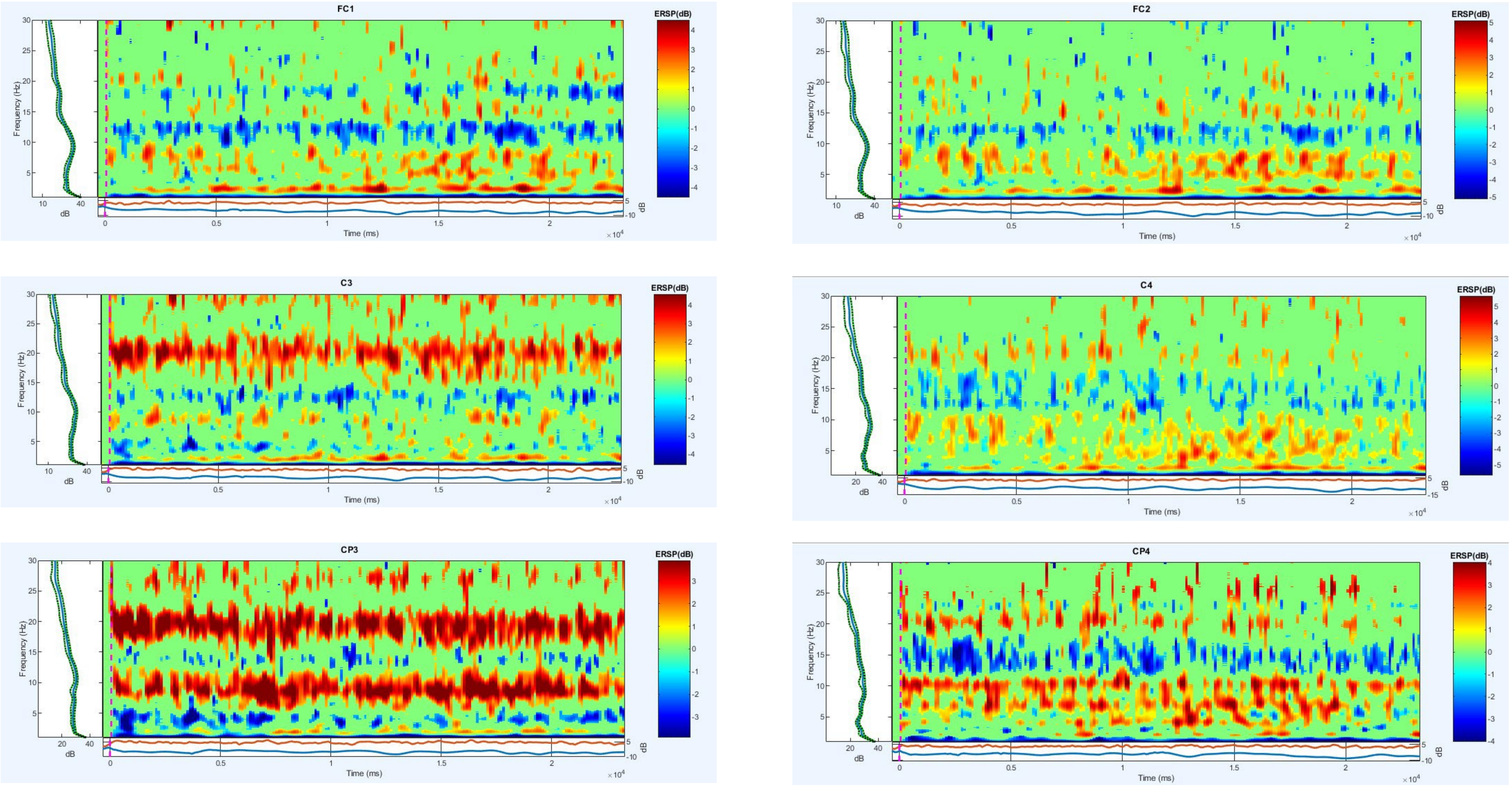
TIME FREQUENCY ANALYSIS BY PHASE ('PA' CONDITION)



Brain localization. Theta (4 – 7 Hz) localization of EEG activity over a period of 500ms before and after each event



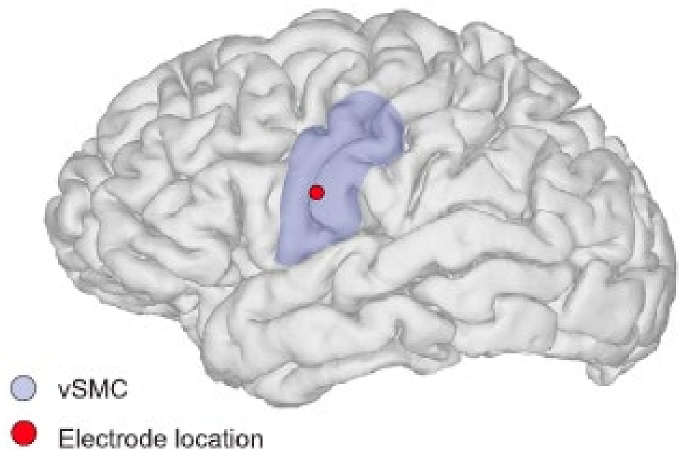
20 sequences of [pa]: unobstructed ear condition, eyes closed
Baseline: repetition of the syllable [pa] during normal breathing

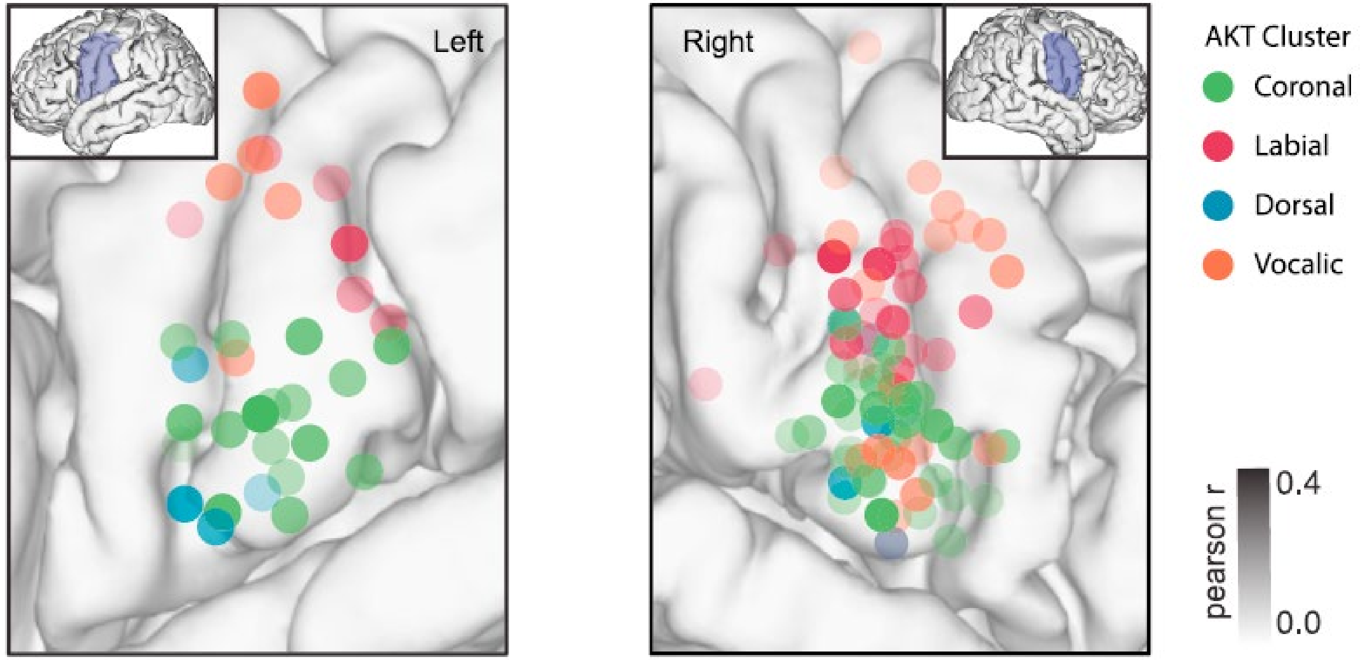


Control and timing of articulatory gestures.

Chartier et al. (2019): Encoding of Articulatory Kinematic Trajectories in Human Speech Sensorimotor Cortex.

Chartier et al. (2019) used deep neural networks to infer speakers' articulator movements from produced speech acoustics.

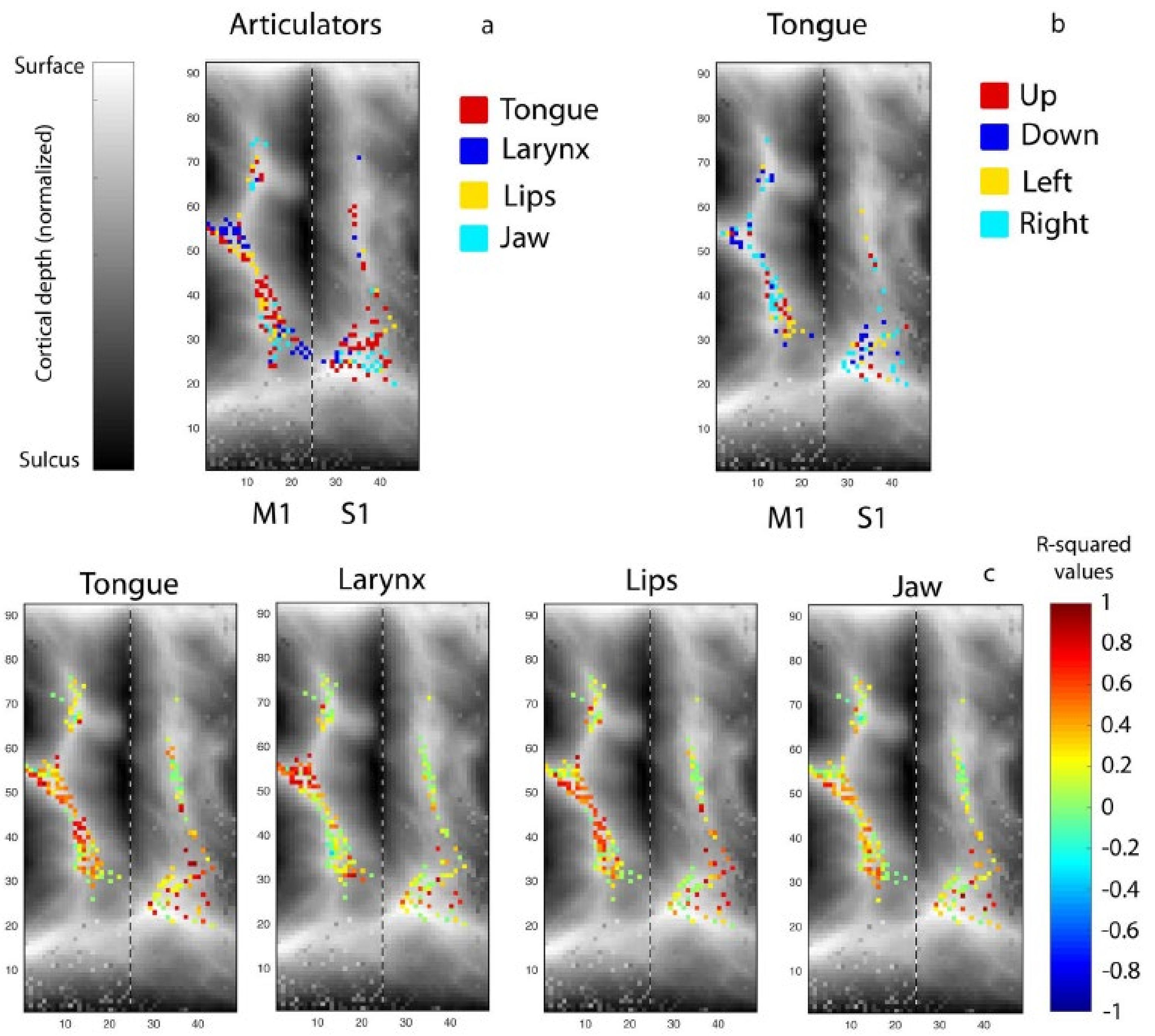




Spatial Organization of Vocal Tract Gestures

Salari et al. (2019) show that recordings from small parts of the sensorimotor cortex (from epileptic patients) contain information about different articulator movements.

Accurate classification was obtained, on average 92% for different articulators and 85% for different tongue directions.



Neuronal oscillations

For the majority of neurons in the nervous system, there are intrinsic processes, located at the level of the membrane of the neurons, capable of producing oscillations of electrical activity without the presence of influx from the environment.

The essential function of neural oscillations is represented by movement.

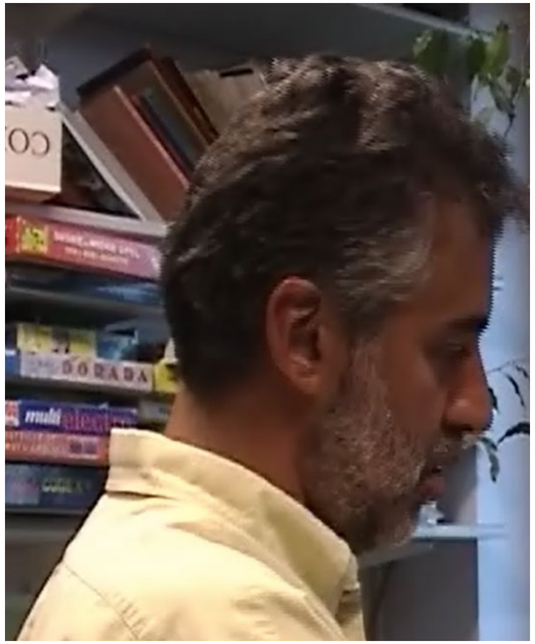
The alpha rhythm is produced by oscillations of the synchronized membrane potentials of all of the different types of excitatory or inhibitory neurons.

The performance of a gesture is dependent on the alpha oscillation phase that precedes its execution.

Neuronal integrator for audition?



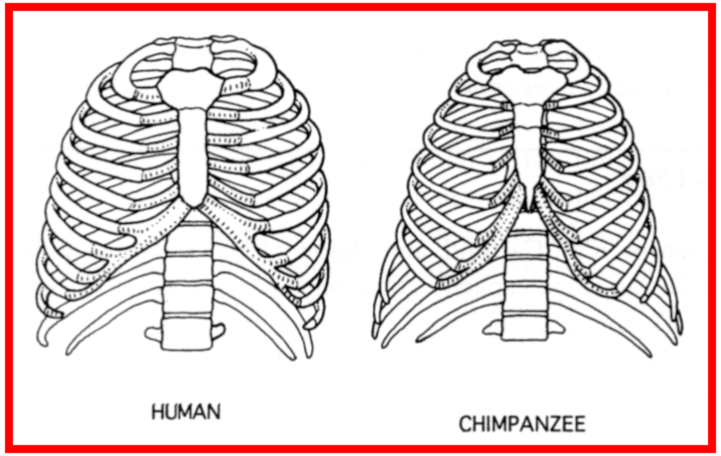



DORADA


4. Origin and evolution of sound production and perception in primates

Change in the shape and musculature of the respiratory system in great apes & hominids.

⇒ greater respiratory control in vocalizations and speech (Mac Larnon 1999).

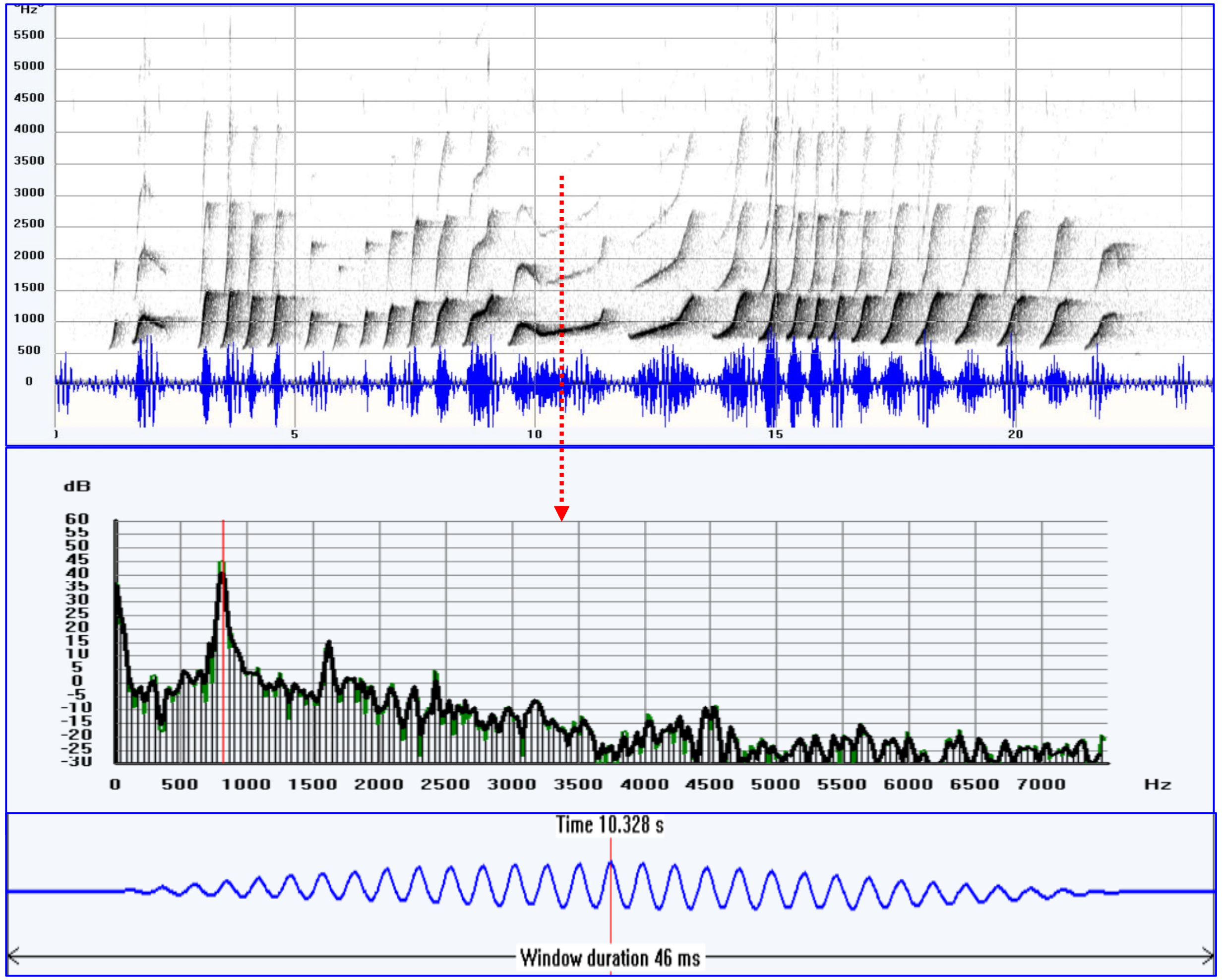


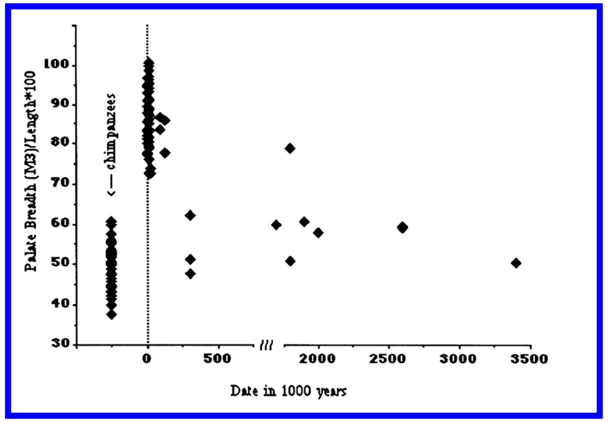
Control of the PCA during vocalizations contributing to produce the difference between voiced/voiceless sounds (ad-abduction of the VF).
There are no traces of this in non-human primates.

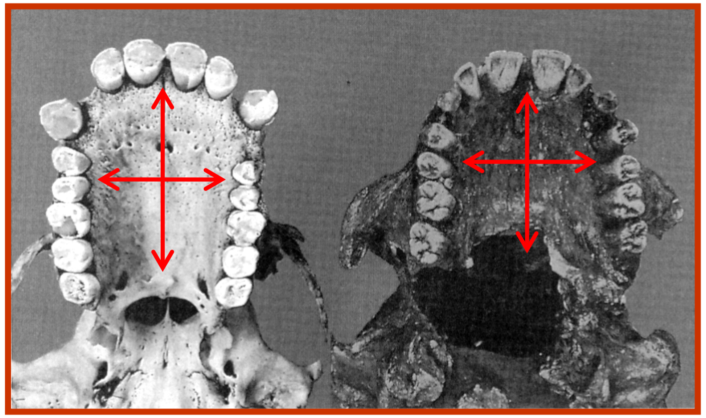
Control of constrictions and types of proprioception
in the VT.

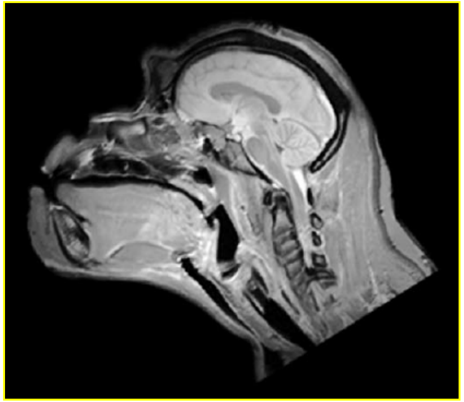
Whistled (and noisy) source with a high F0 > 700 Hz and harmonic spectrum towards a source with a lower F0 (100 to 300 Hz) ⇒ formants.

Change in the size of the vocal tract. Palate length/width 2/1 → 1/1 ⇒ changes in the geometry of the VT (vowels) & constrictions at different places (consonants).

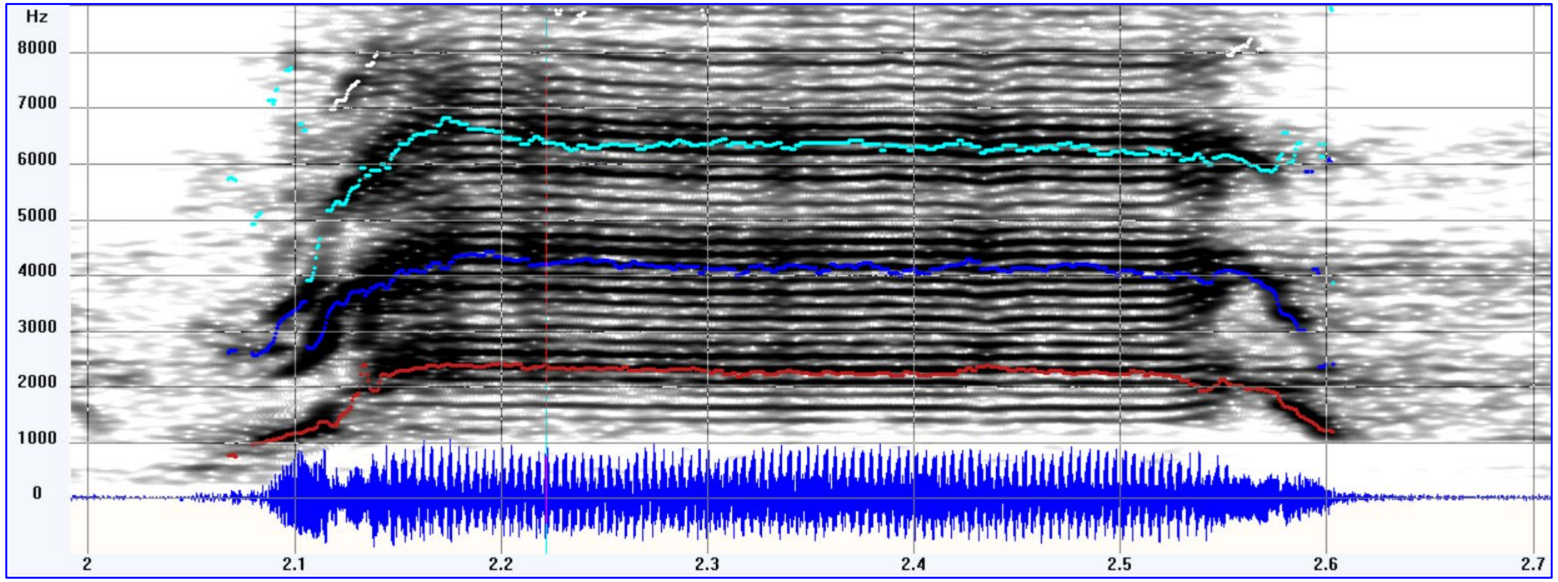


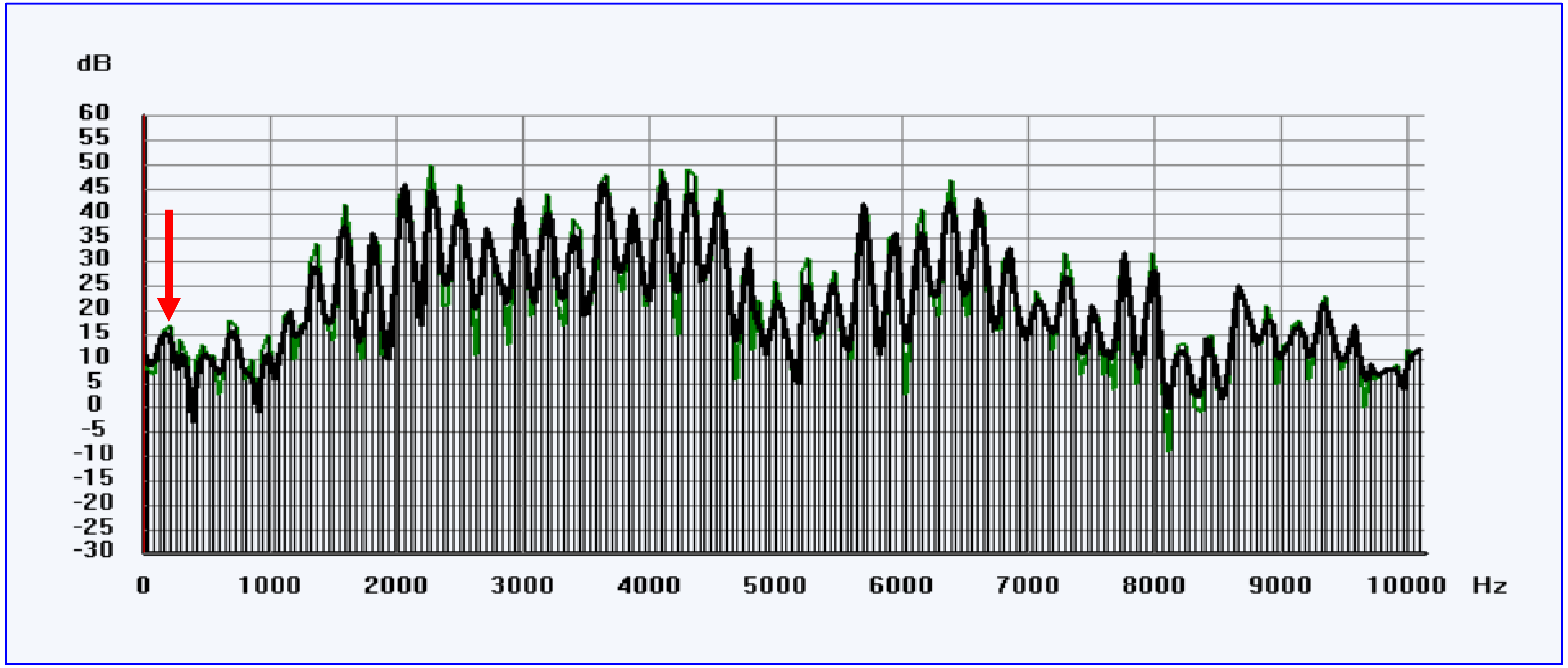


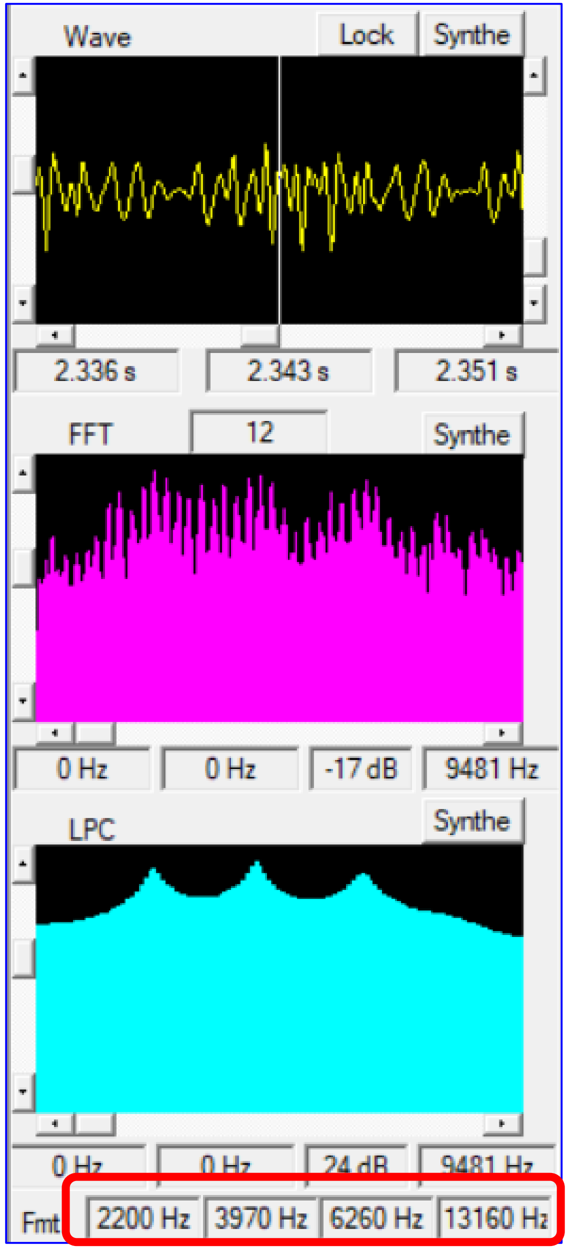


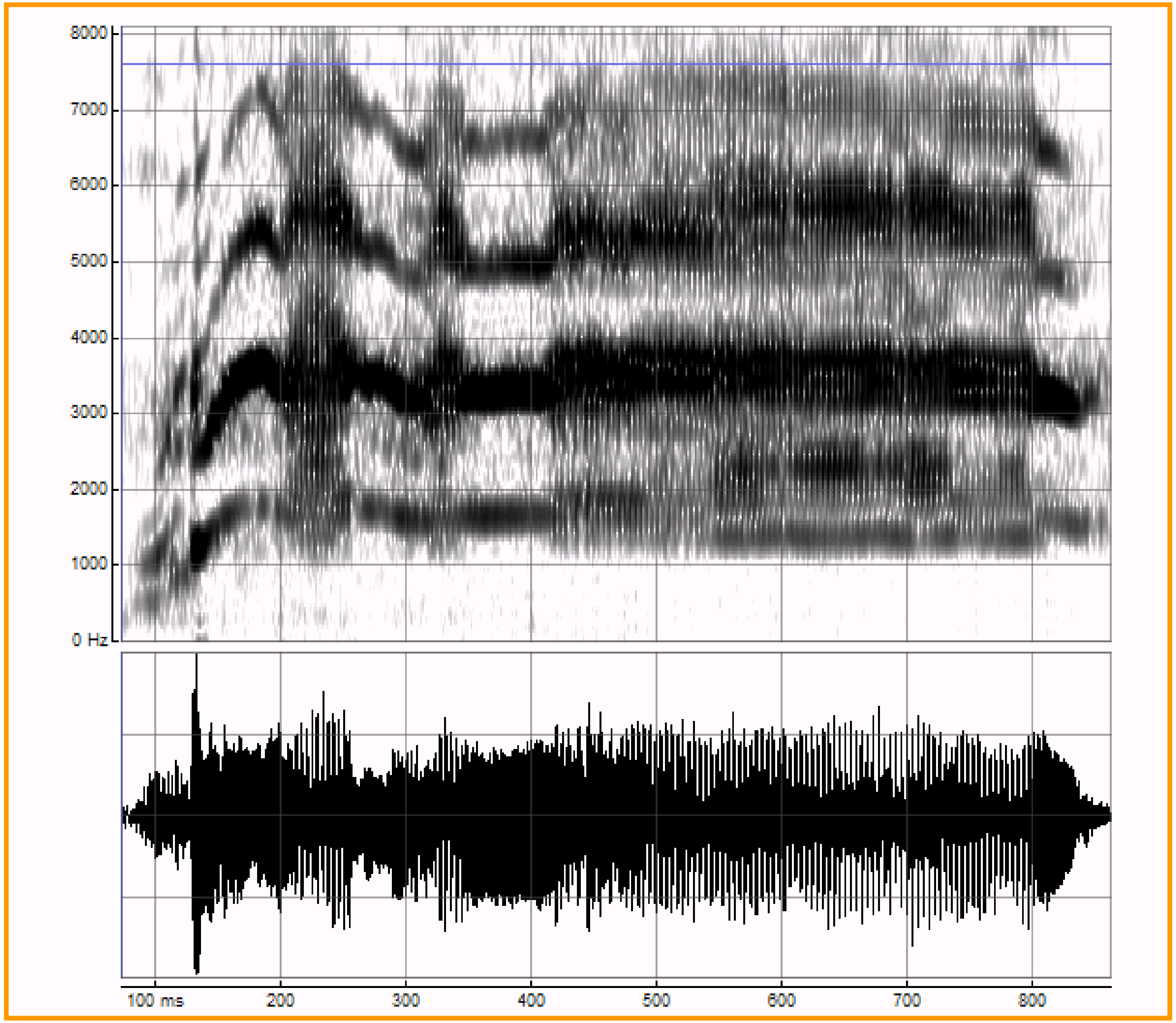


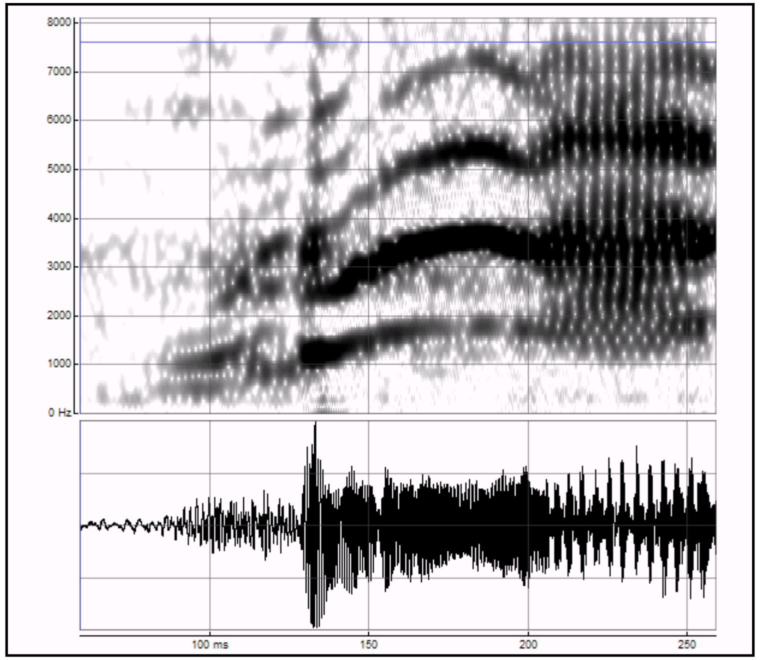
Bonobo double source ⇒ glottis (cartilagenous & membranous dimensions)

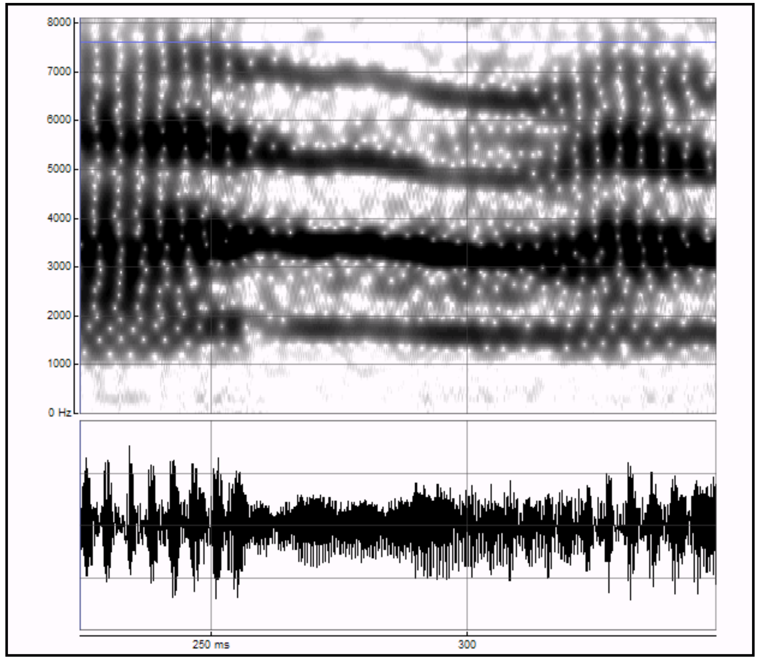


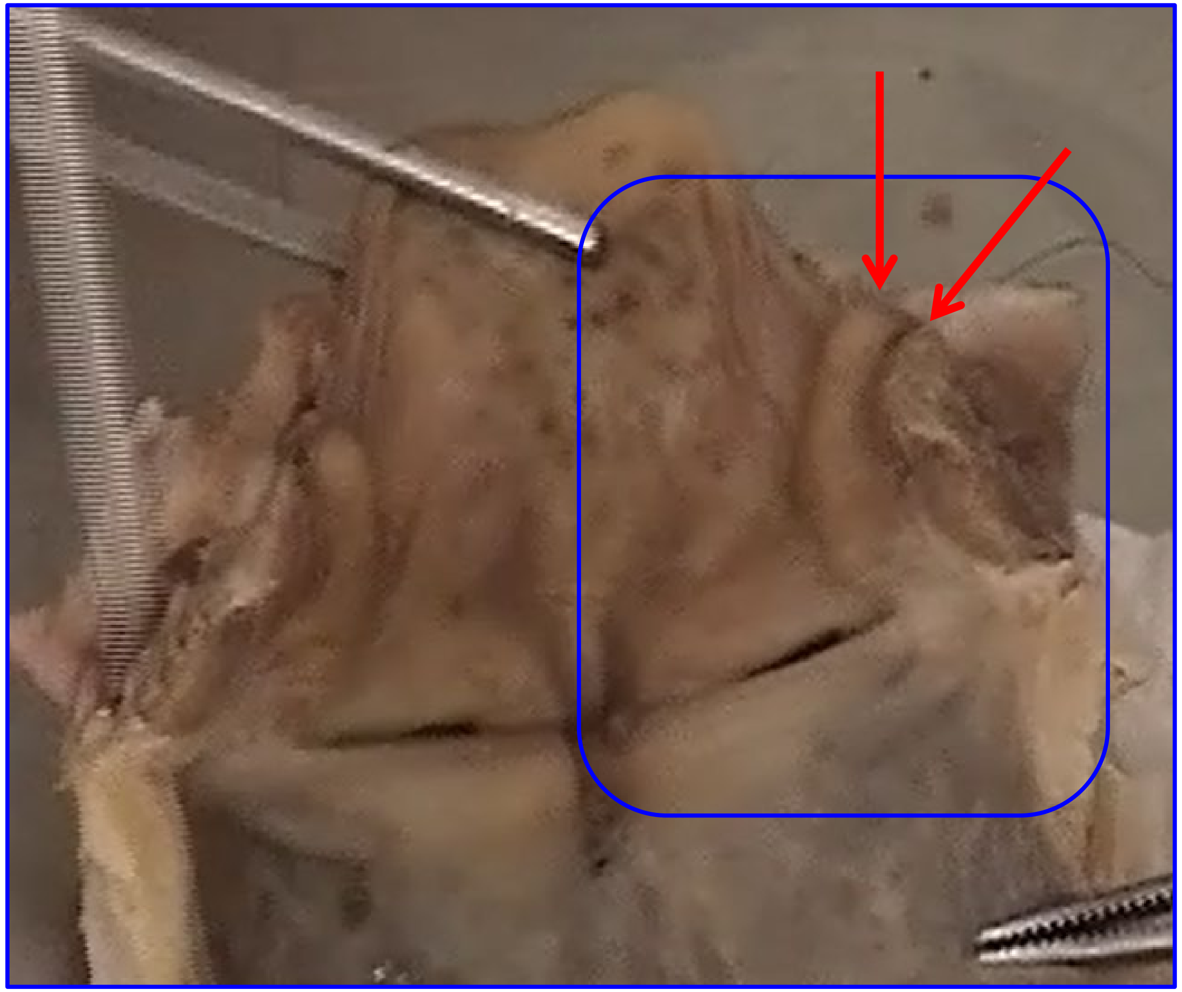


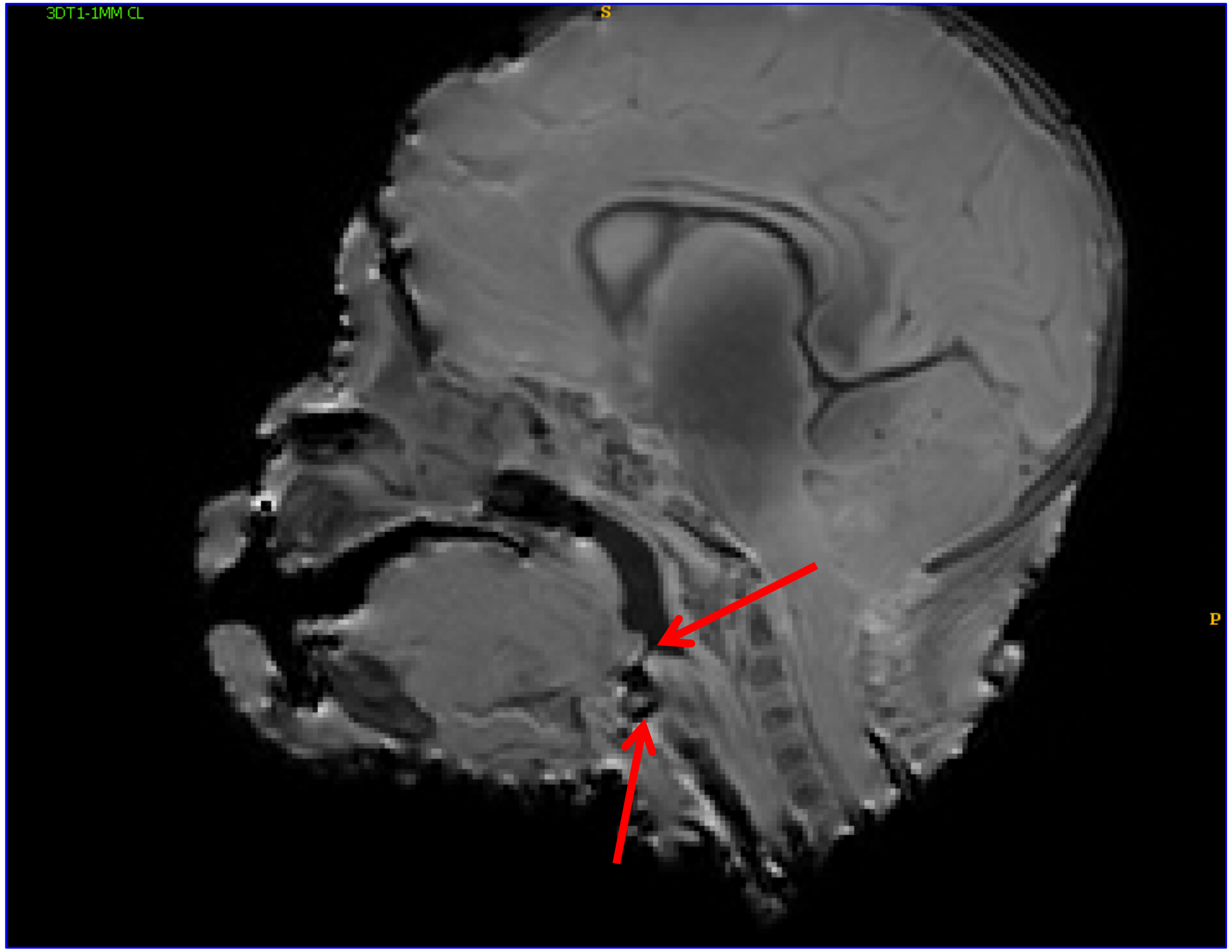


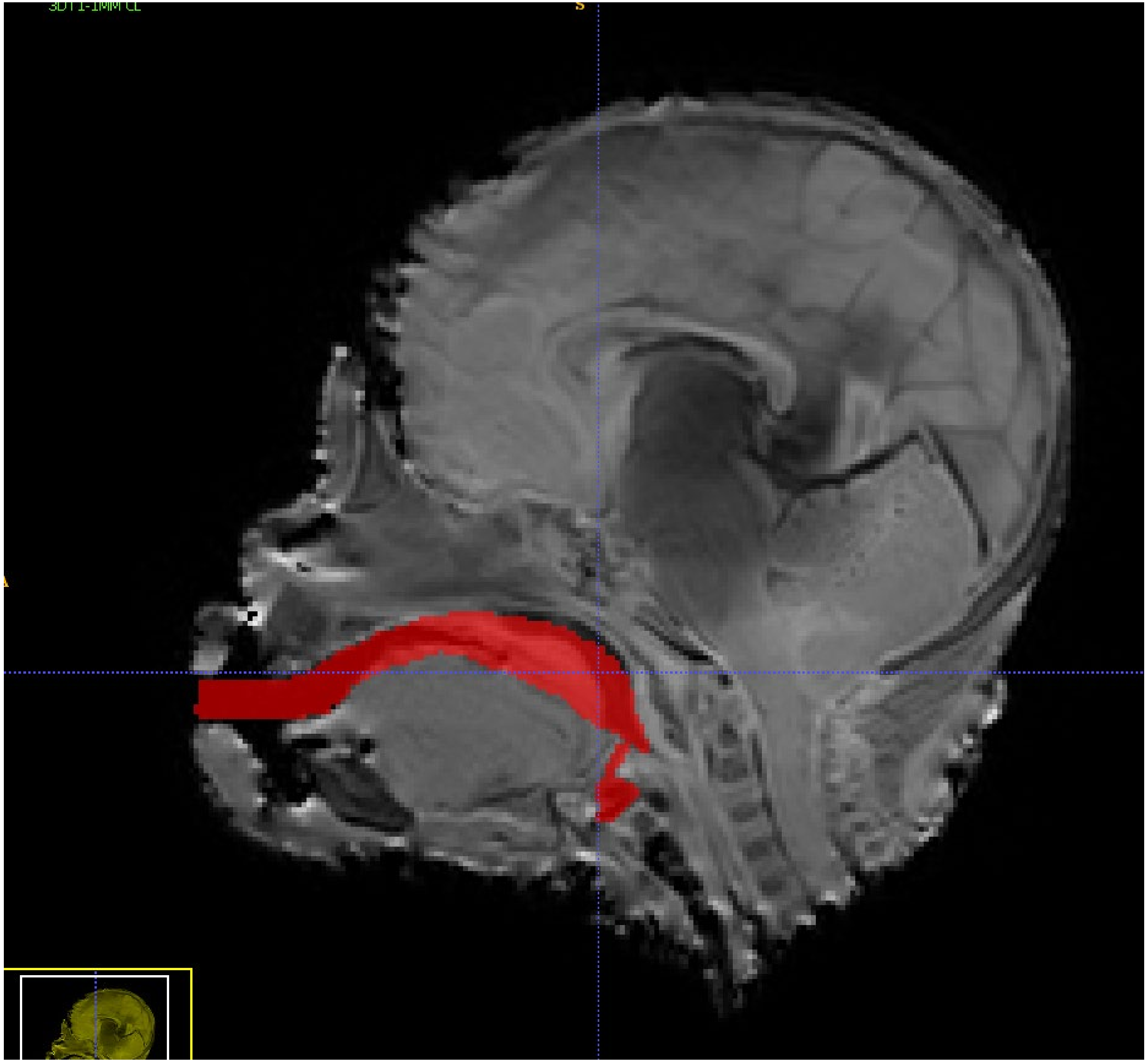

8000
7000
6000
5000
4000
3000
2000
1000
0 Hz
100 ms
200
300
400
500
600
700
800


8000
7000
6000
5000
4000
3000
2000
1000
0 Hz
100 ms
150
200
250


8000
7000
6000
5000
4000
3000
2000
1000
0 Hz
250 ms
300

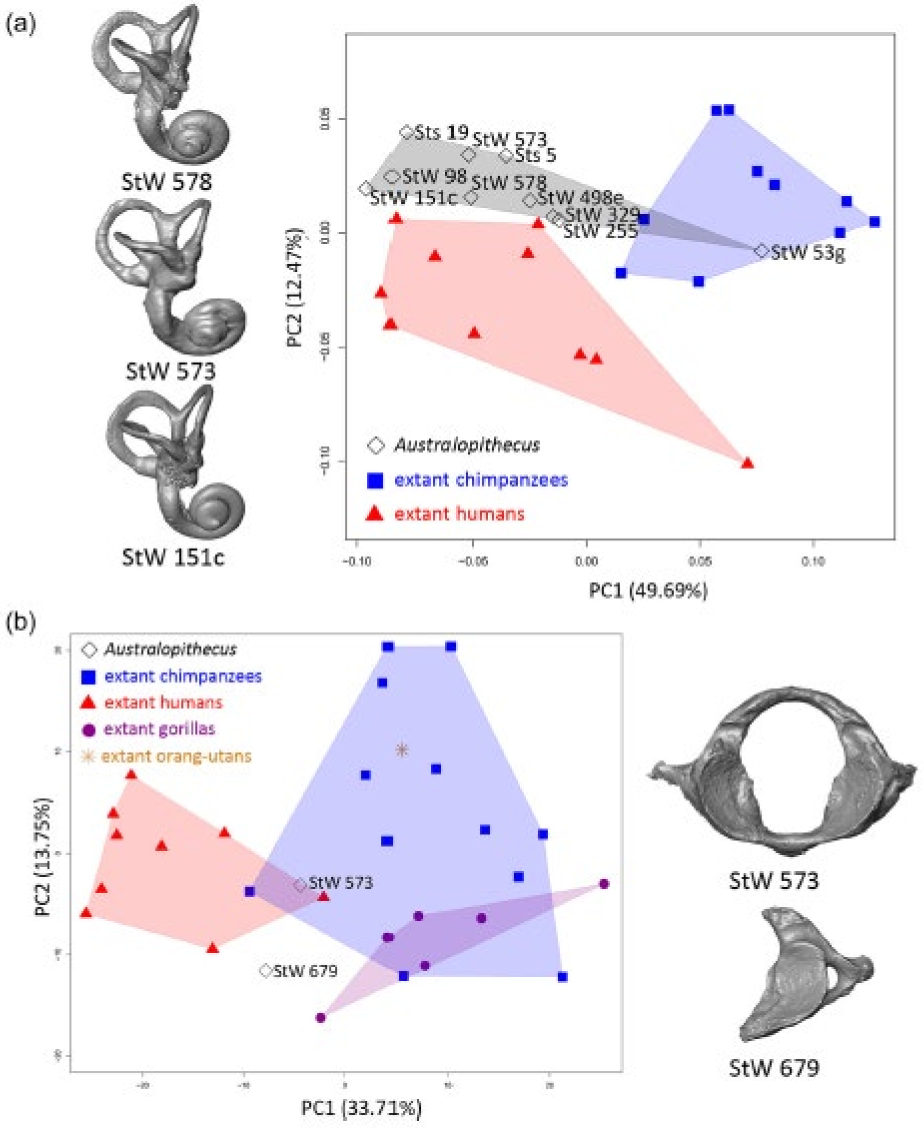



3DT1-1MM CL
S
P

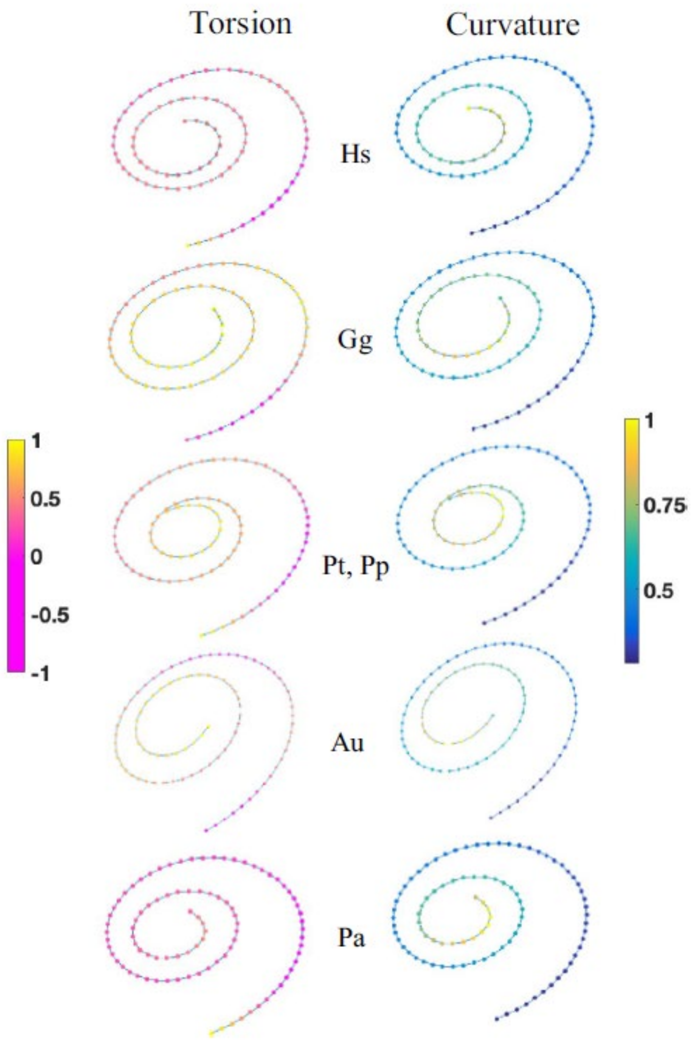

3DT1-1MM CL
S

Auditory constraints on sound perception and vocalizations

Sound perception shaped distinct ecological adaptations among African early hominins.



Beaudet 2023



Braga et al. 2021

Audition structured the formation of phonetic categories.

Increase in the cochlea size.

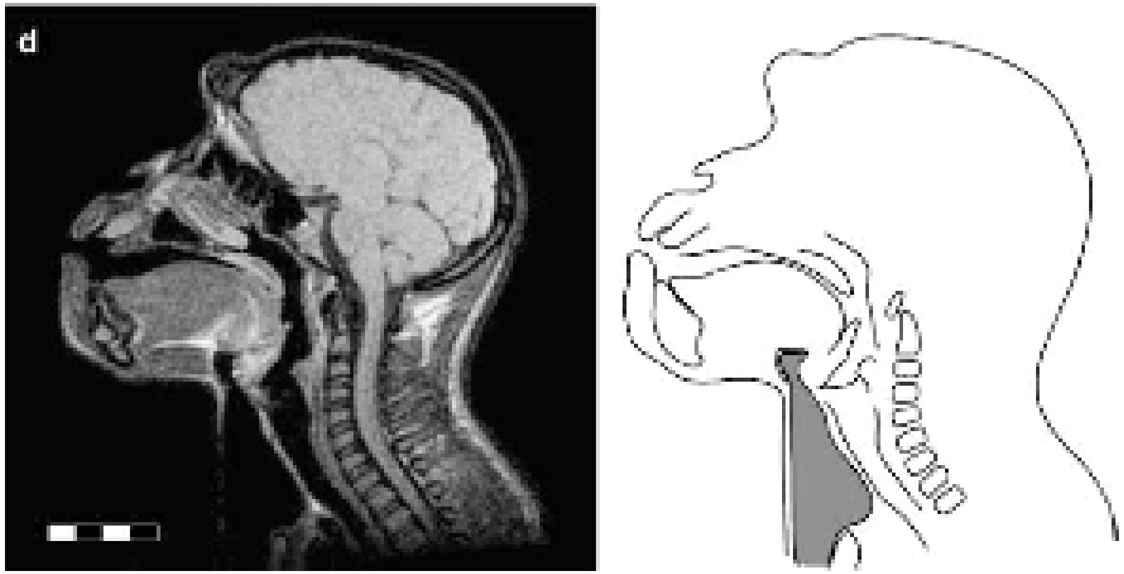
The form and structure of eardrums evolved to respond to large frequency bands.

Hearing evolved for changing sounds.

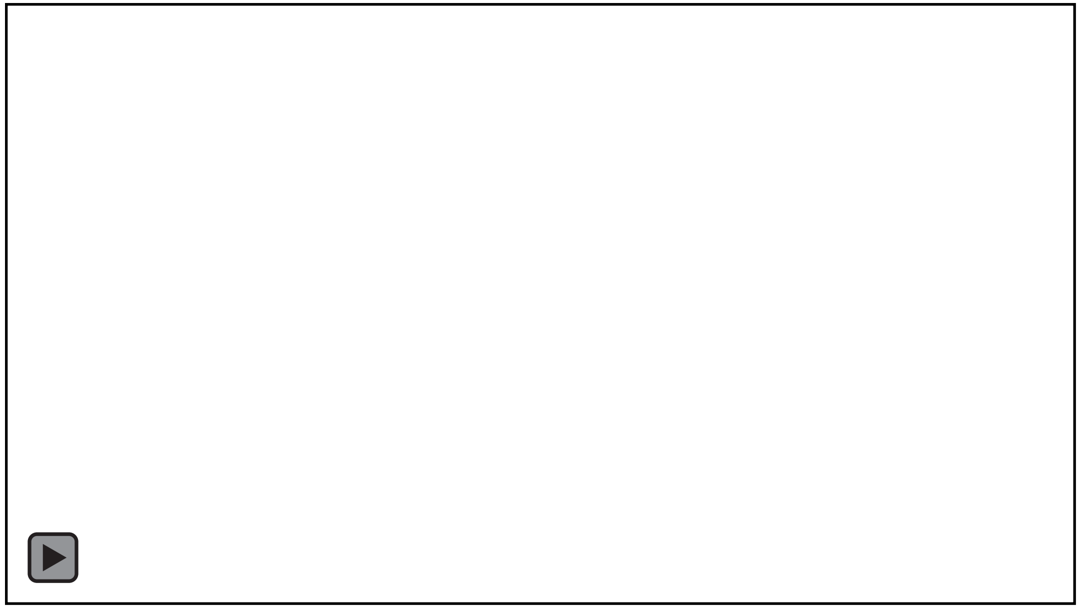
Hearing developed as an alert system ⇒ attention to instantaneous and unusual sounds.

→ Attack transients ⇒ consonants.

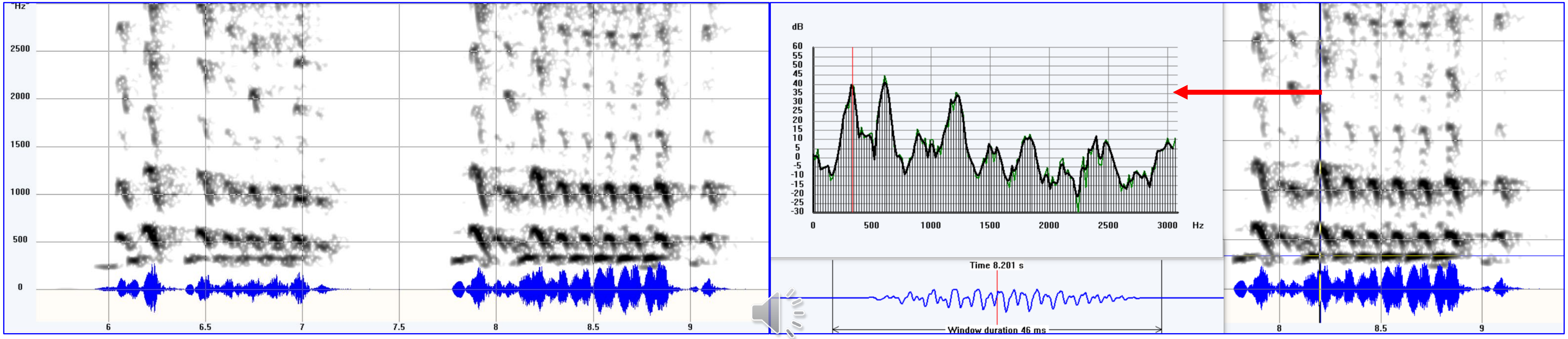
Laryngeal air sacks? Function? Sexual selection?



Nishimura 2006



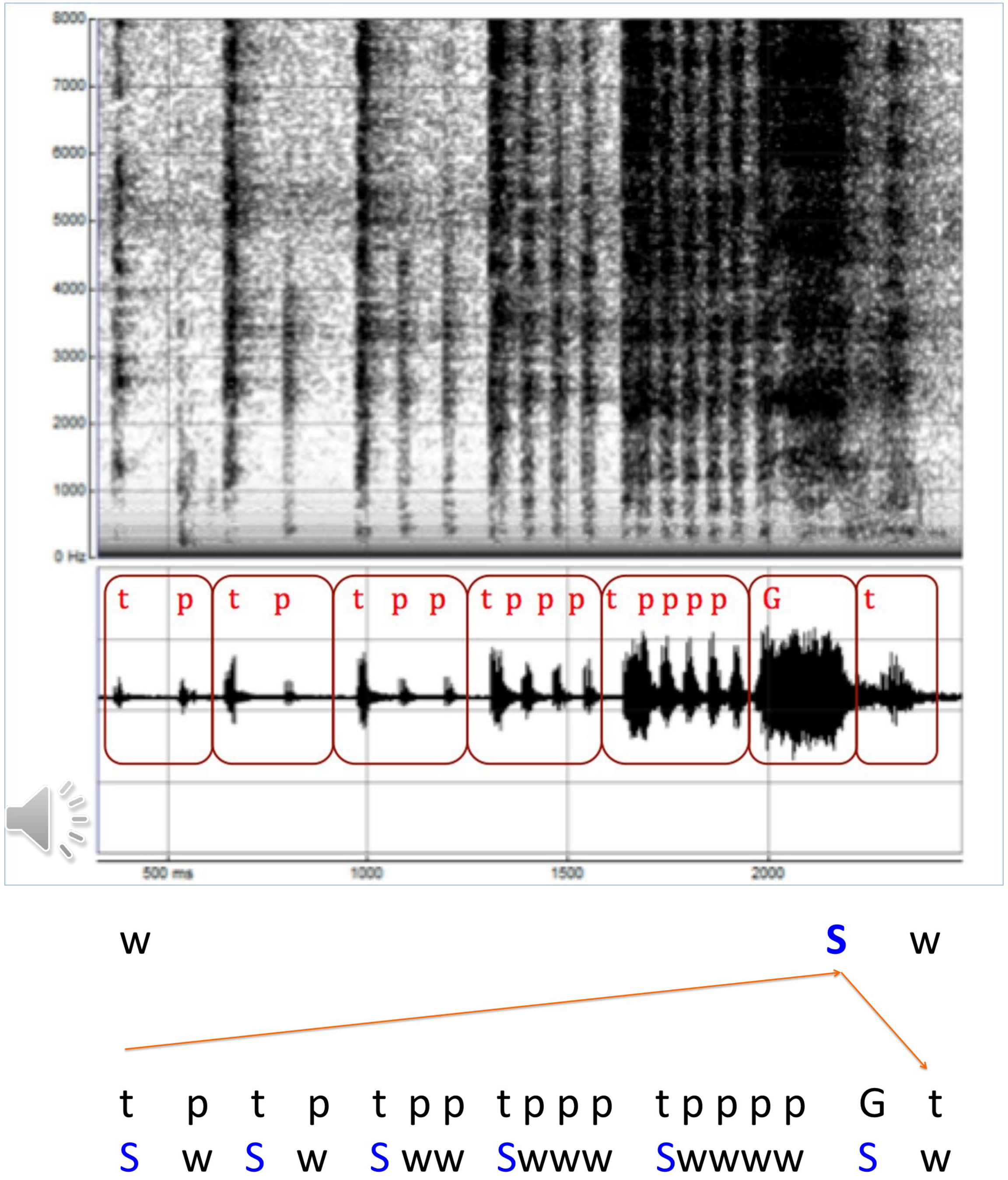
Contribute to the radiation of the sound source at species-specific frequencies. To mark presence/location



Gibbons

Geissmann

Muriqui the hippie monkey : syntax and recursion



Muriqui have recursion and grammar

A sort of **context sensitive grammar** that will generate the pattern is:

$S \rightarrow tpt$

$[Xtp^{n}t]s \rightarrow Sp^{n+1}t$

The first rule is obvious.

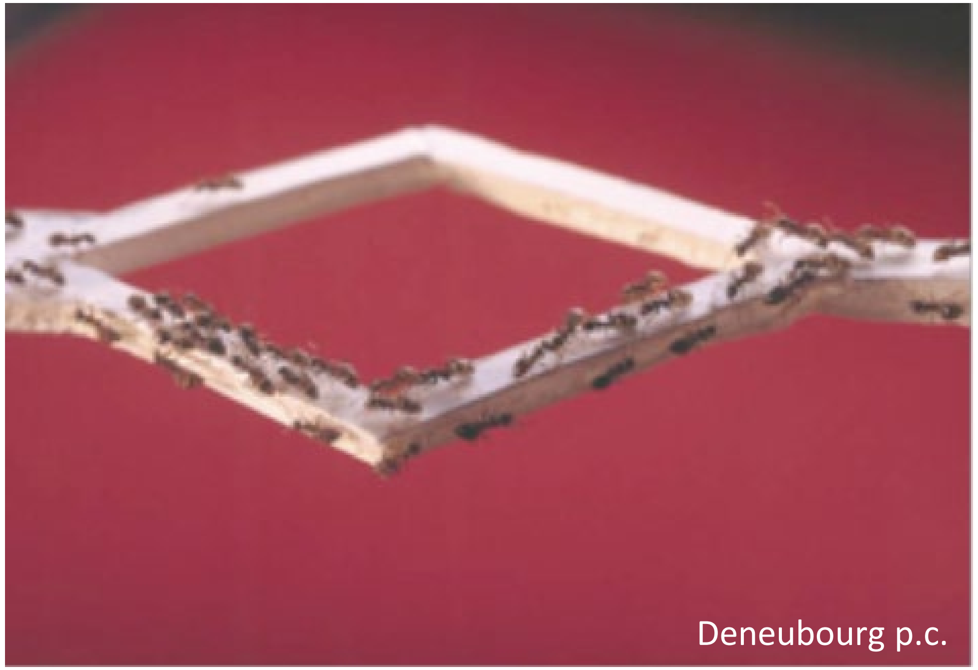
The second rule says, given an S with the structure: anything (X) followed by a t followed by n p's followed by t can be rewritten as the original S followed by n+1 p's followed by a t.

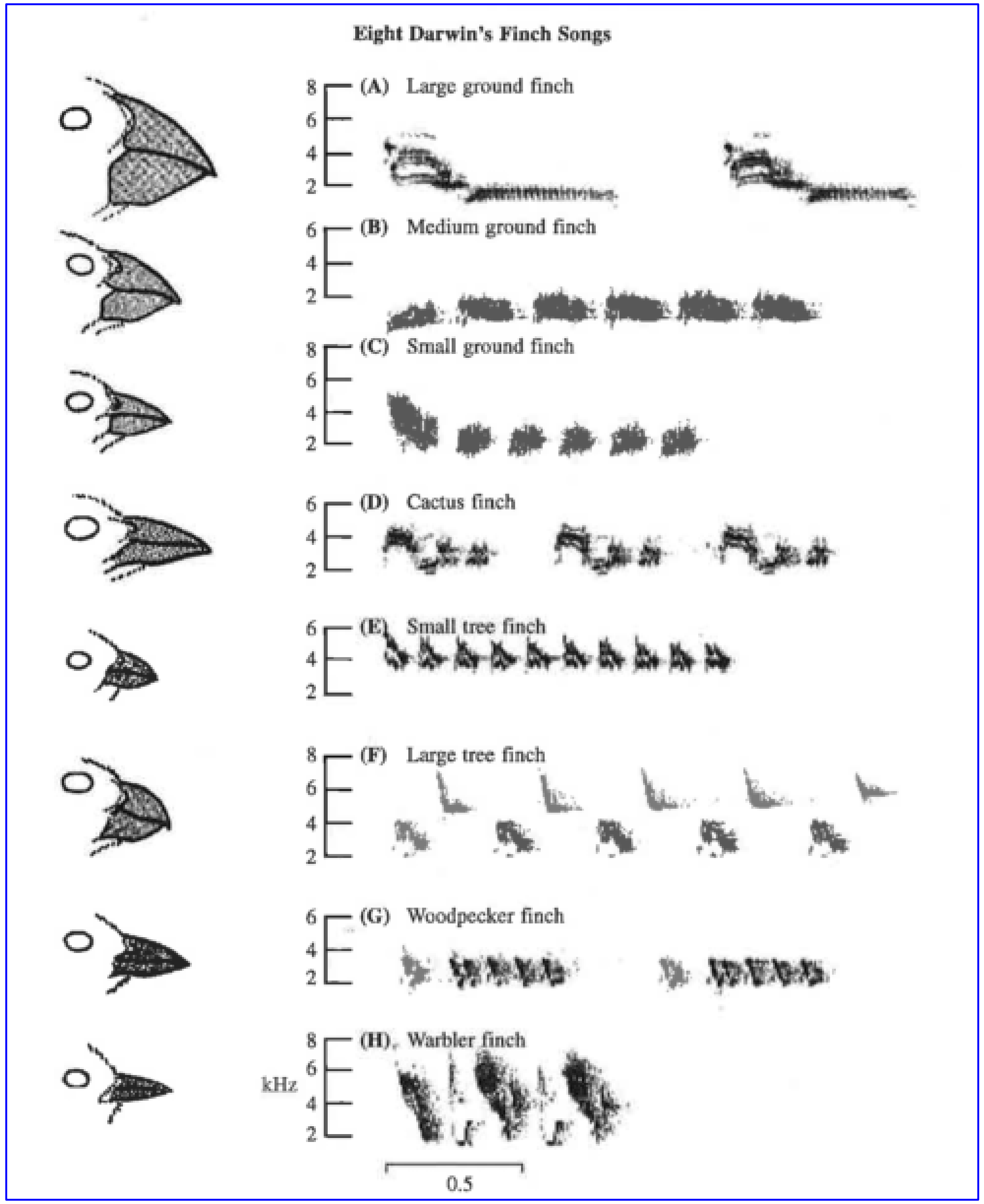
Demolin, Ades & Mendes (2010)

Contribution of animal communication models

Darwin's Finch beak morphology and songs variability

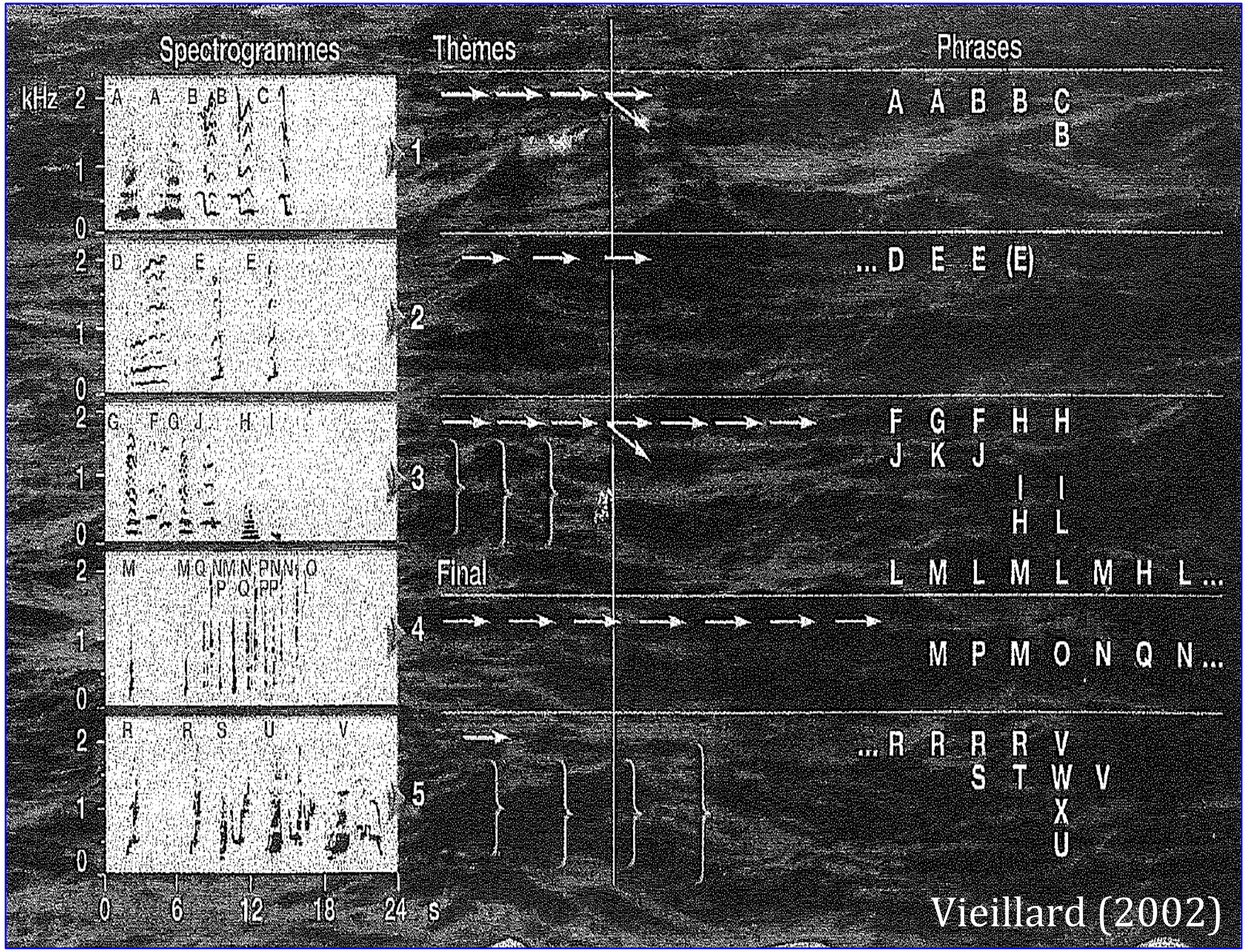
Collective intelligence patterns: Ants, termites





Podos (2001)

Dialectology : birds and whales.



Vieillard (2002)

Humpback whale

AABBC
AABBC

AABBC
XYYZZ

AABBC
AABBC

AABBZ
ABBZZ

AABBZ
AABBZ

ABBZZ
ABBZZ

ABBZZ
ABBZZ

ABBZZ
ABBZZ

‘*Lastly, if this theory of indefinite modifiability be sound, what meaning can be attached to the term language, and what definition can be given of it so as to distinguish a language from a dialect?*’ Lyell (1863)






How are speech sounds made and controlled?

Rousselots' views on phonetics are still true

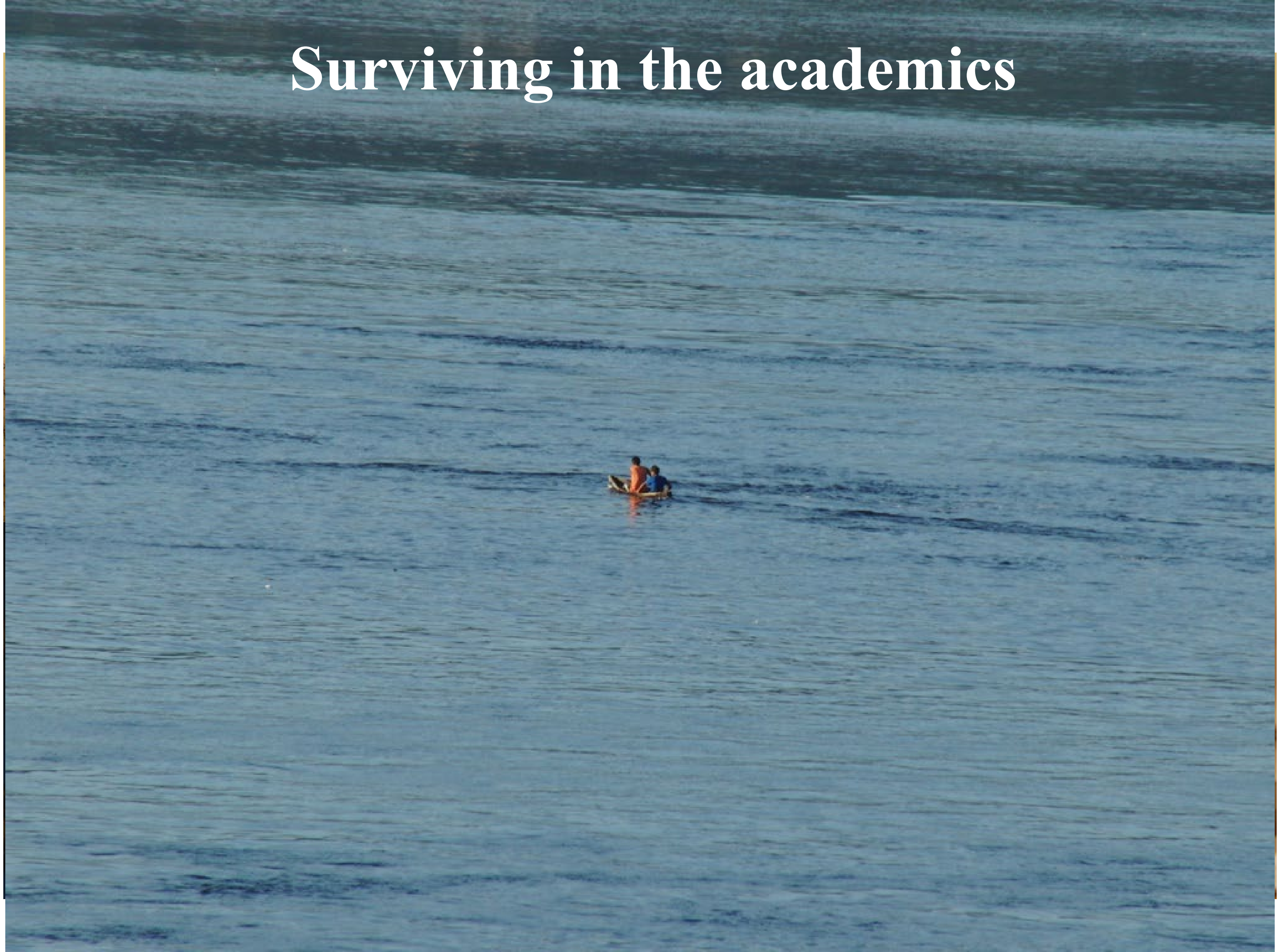
They have to be complemented by the understanding of the brain electrochemical and neurological processes that are linked to language and speech.

There lies the key for future fundamental explanations.

This will likely necessitate the use of deep learning and AI techniques.

What about doing ear tracking?

Surviving in the academics



Measuring clouds

Strier 1992

m ɛ χ s i