

Aerodynamic speech database : some new questions

Didier Demolin

Gipsa-Lab, Grenoble

Edinburgh, December 2(th, 2010

Respiratory activity in speech

Intensity is correlated with subglottal pressure (P_s).
 P_s can raise if there is an increase in glottal resistance.

Glottal resistance :

$$R = \frac{P}{U}$$

(pressure between the glottis)
(flow volume between the glottis)

⇒ Fundamental frequency control

⇒ Declination (downdrift)

Air flow

$$U = A \cdot (P_1 - P_2)^a \cdot c$$

U = Volume velocity in cm³/s

A = section's area.

P₁ = pressure

P₂ = pressure

c = constant

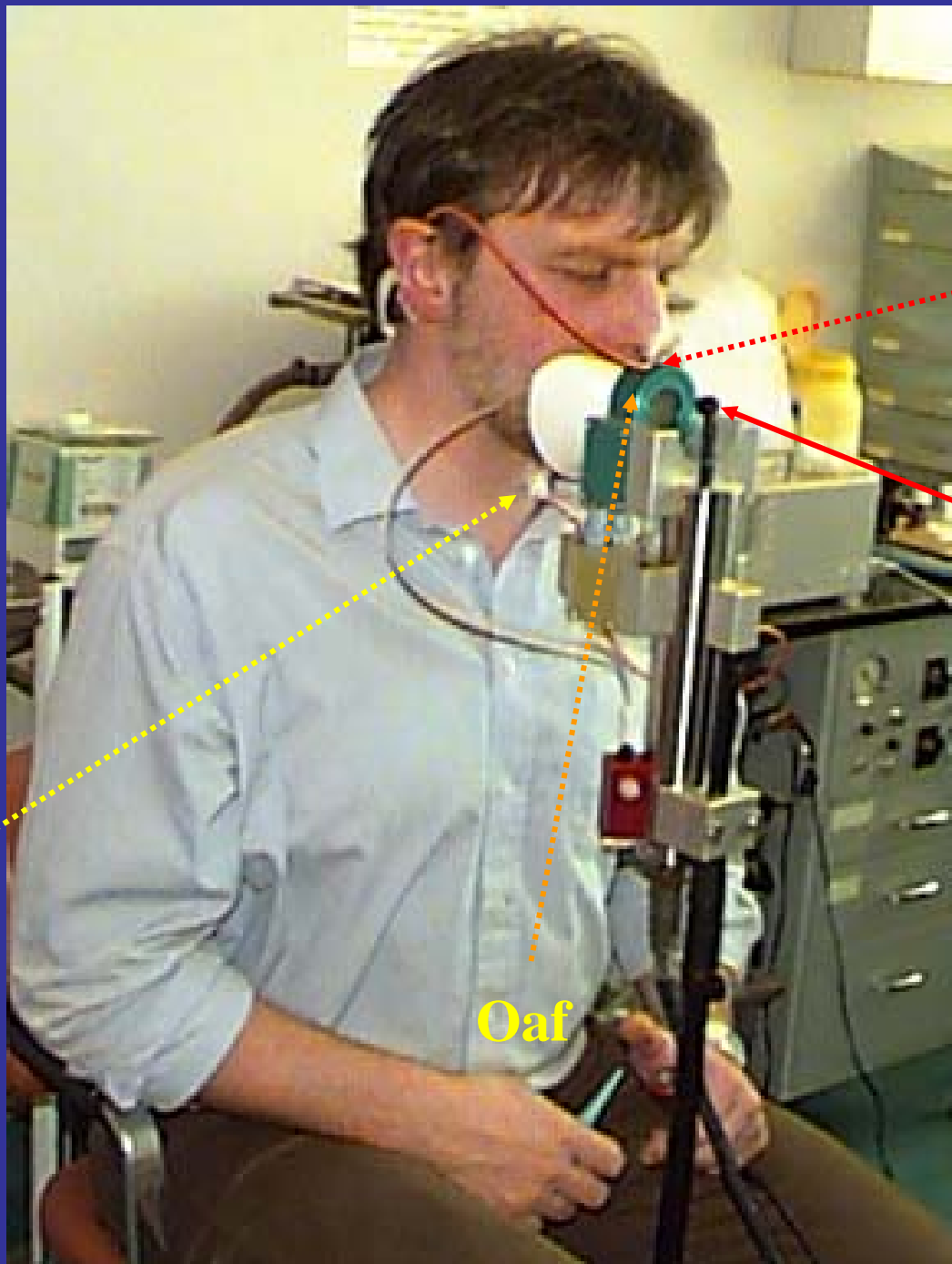
^a from 0.5 to 1 according the nature of the flow.

1 → laminar

0.5 → turbulent



Ps



Oaf

Po

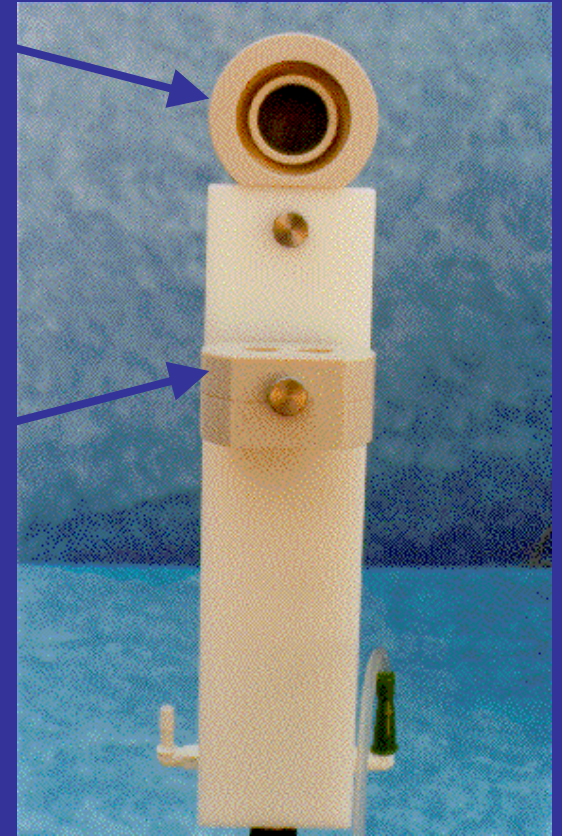
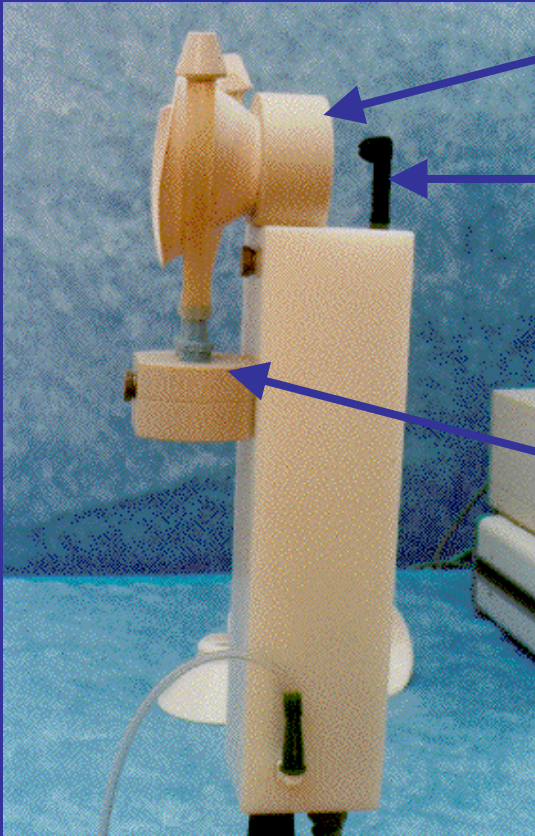
Audio

Eva2

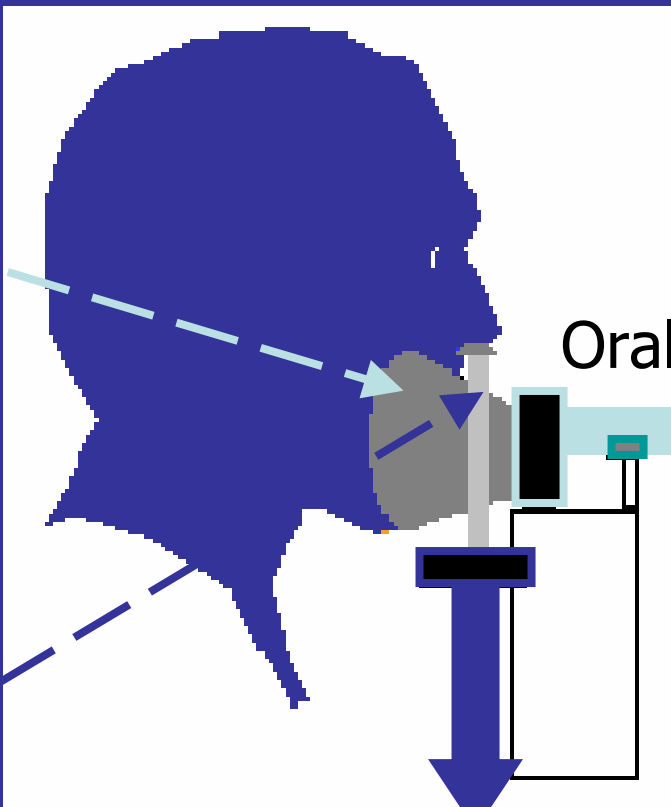
Oral airflow

Microphone

Nasal
airflow



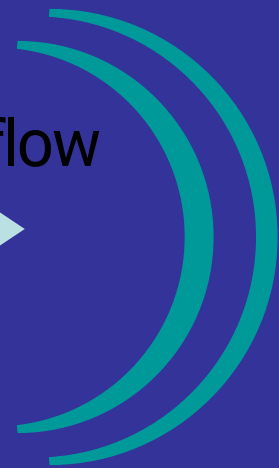
Mask



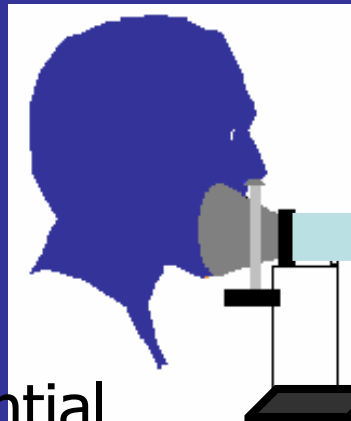
Oral airflow



Sound

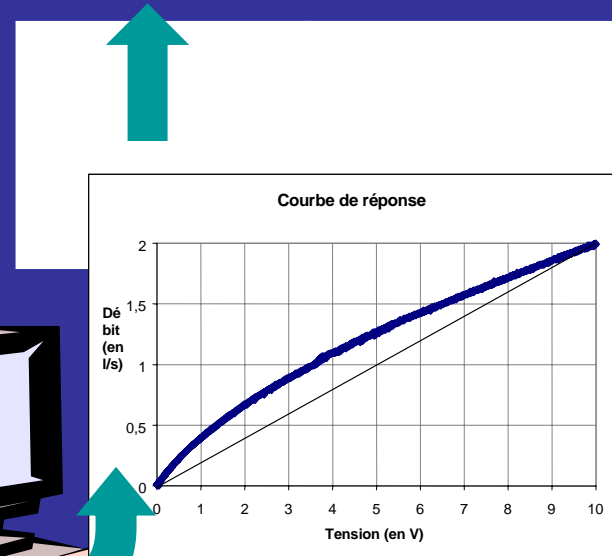
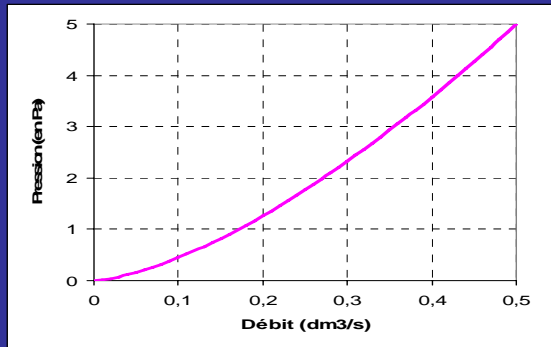
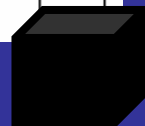


Flow measurement linearization



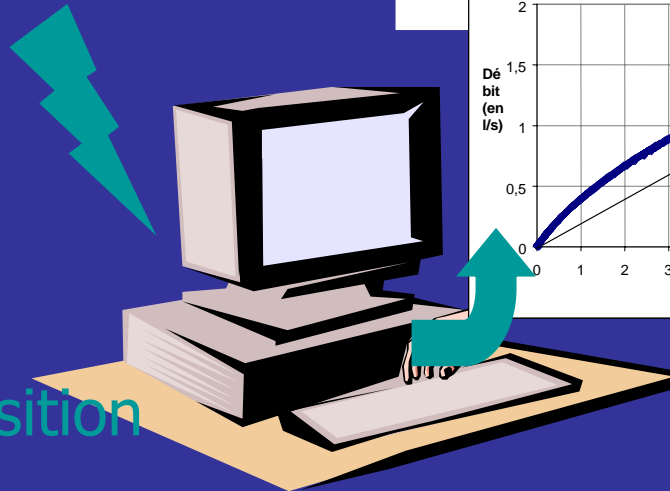
Oral flow

Differential pressure captor







Acquisition

Computerized linearization



Procedure

Six Belgian speakers of French, three males and three females, were recorded while saying nonsense words in phrases such as *Sisi, dis sisi encore* ('Sisi, say sisi again'). The nonsense words used in the present study combined each of the consonants /f,     C*/ with the vowels /i, a, u/ to form two syllable words with the first syllable being the same as the second.

Recordings were made of the oral airflow, the pressure of the air in the pharynx (P_o) and the pressure of the air in the trachea (P_s) just below the vocal folds.

The pharyngeal pressure (P_o) was recorded via a catheter (i.d. 2 mm) passed through the nose so that its open end could be seen in the pharynx below the uvula.

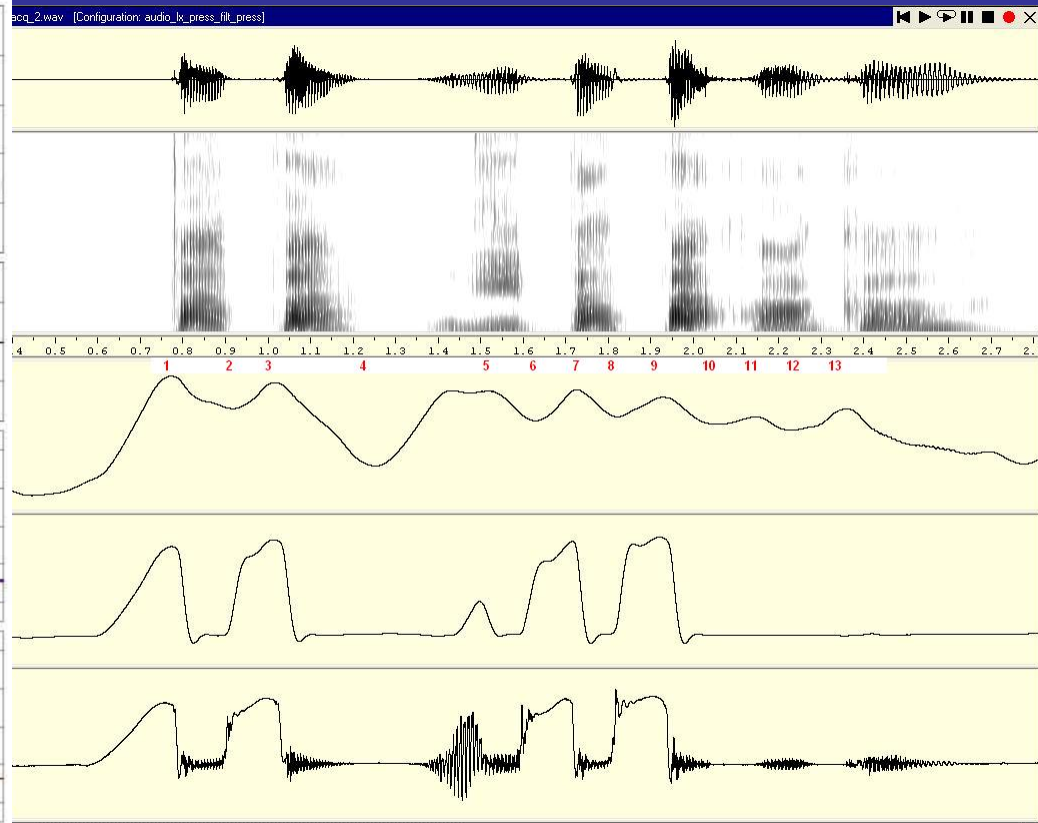
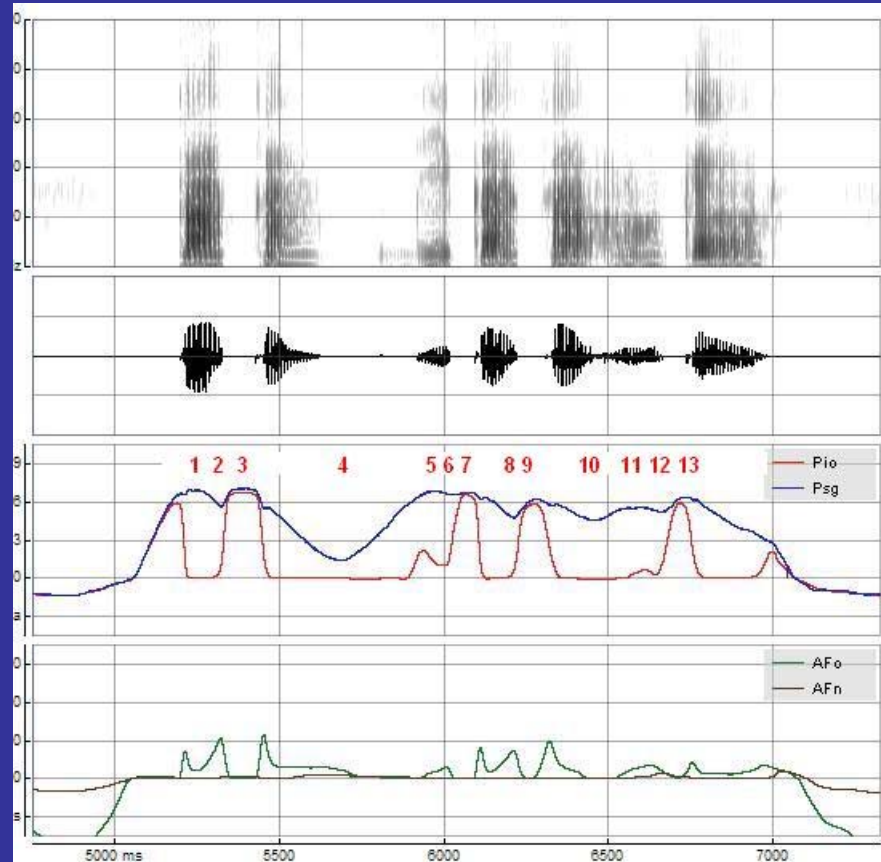
(P_s) was measured via a tracheal puncture between the first and second rings of the trachea or between the cricoid cartilage and the first tracheal ring. (The best location for most speakers was between the tracheal rings).

A local anesthetic was applied both externally and inside the trachea by means of a fine needle. A larger needle with an internal diameter of 2 mm was then inserted and connected to the pressure transducer. Both pressure signals were transduced by the Eva 2 and Physiologia systems described by Teston and Galindo (1990, 1998).

The audio signal was recorded in a synchronous way with the aerodynamic parameters.

The audio signal was digitized at 16000 Hz and the physiological data at 2000 Hz for transfer onto a computer.

Direct tracheal puncture vs plethysmograph

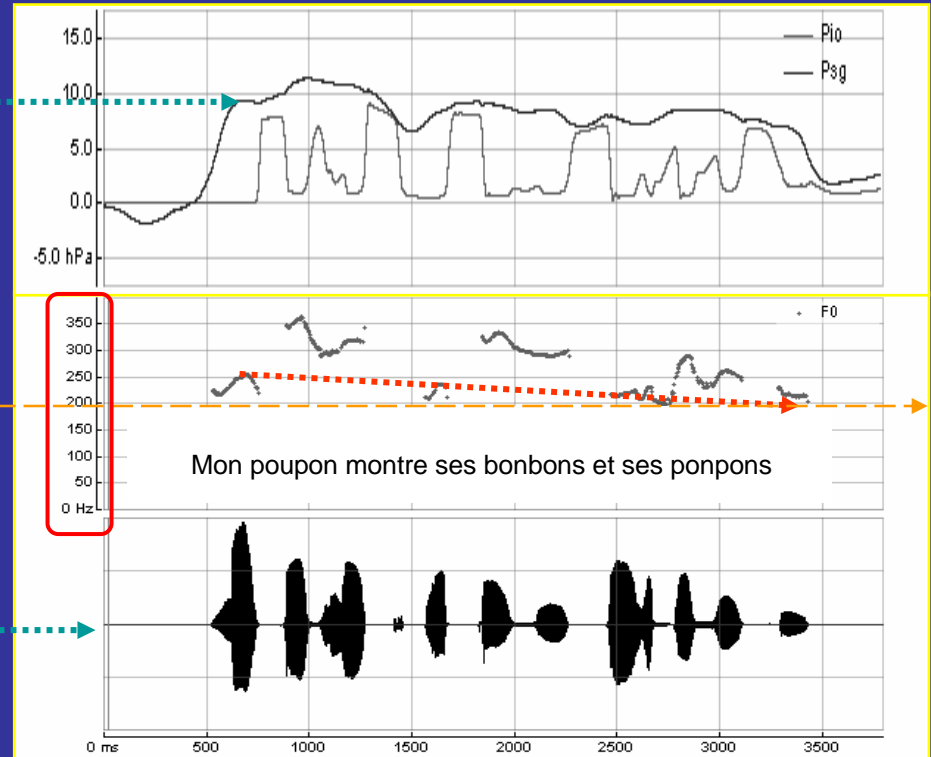
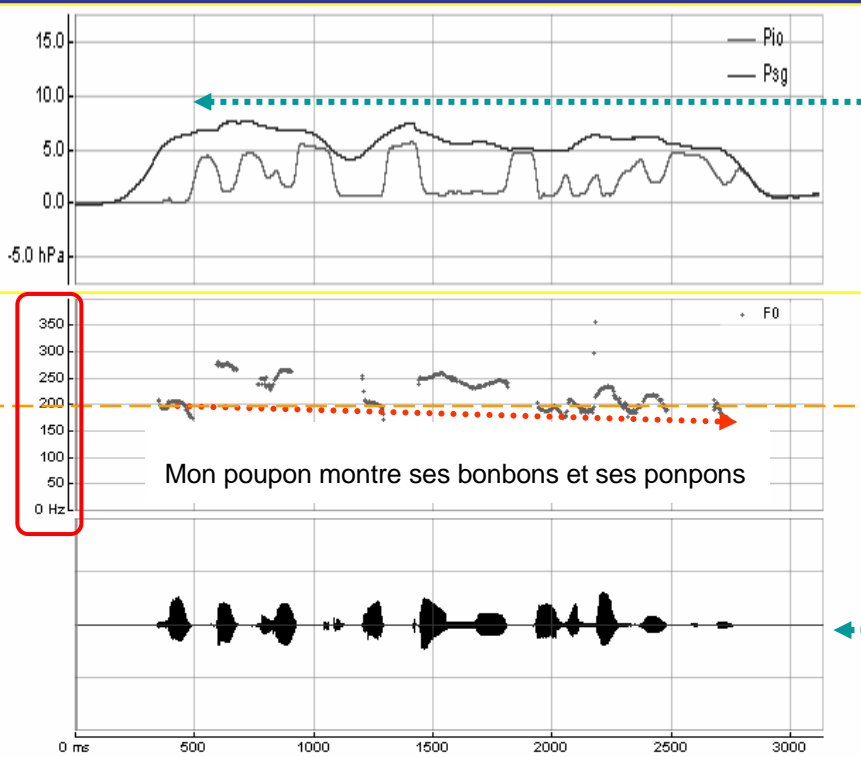


‘Papa, dit papa encore’

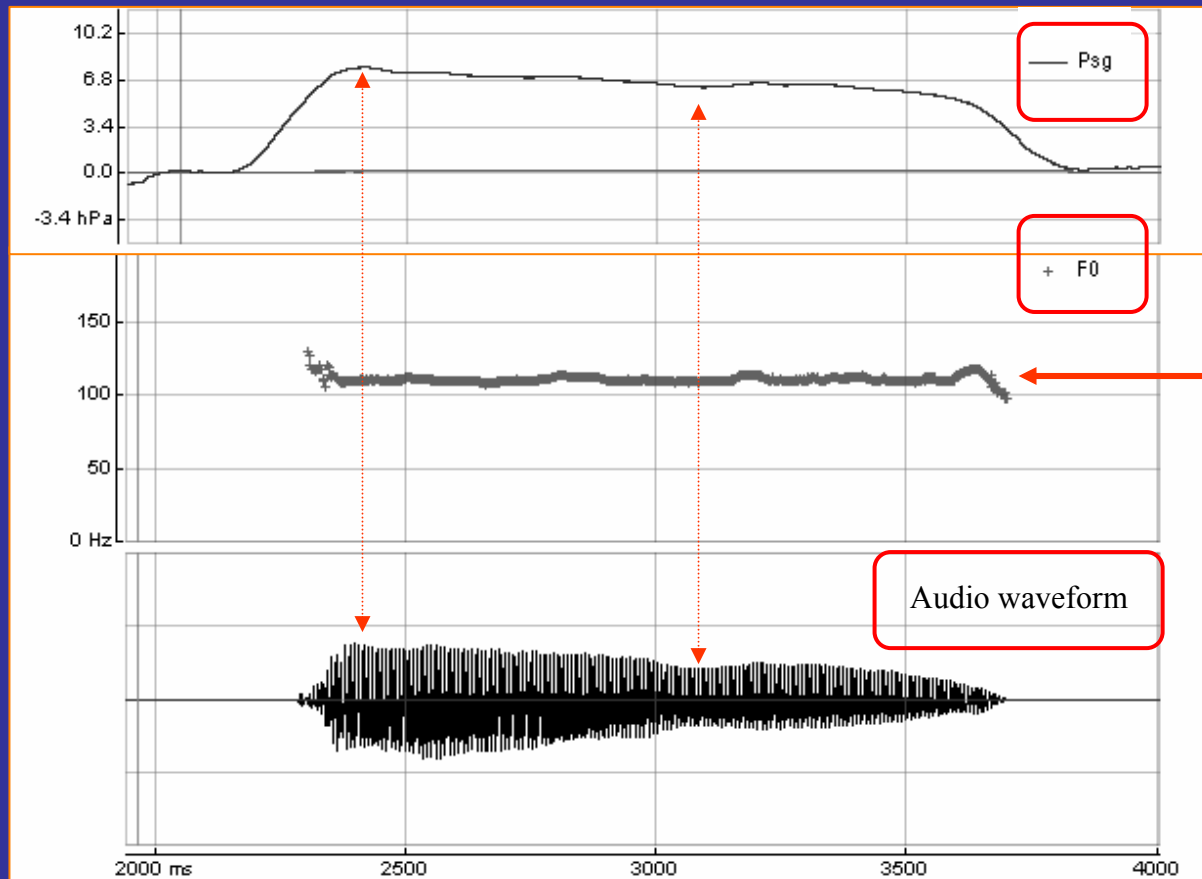
‘Papa, dit papa encore’

- Humans can modulate f_0 independently of intensity and P_s .

The main effect of P_s on f_0 is that it elevates the f_0 baseline.

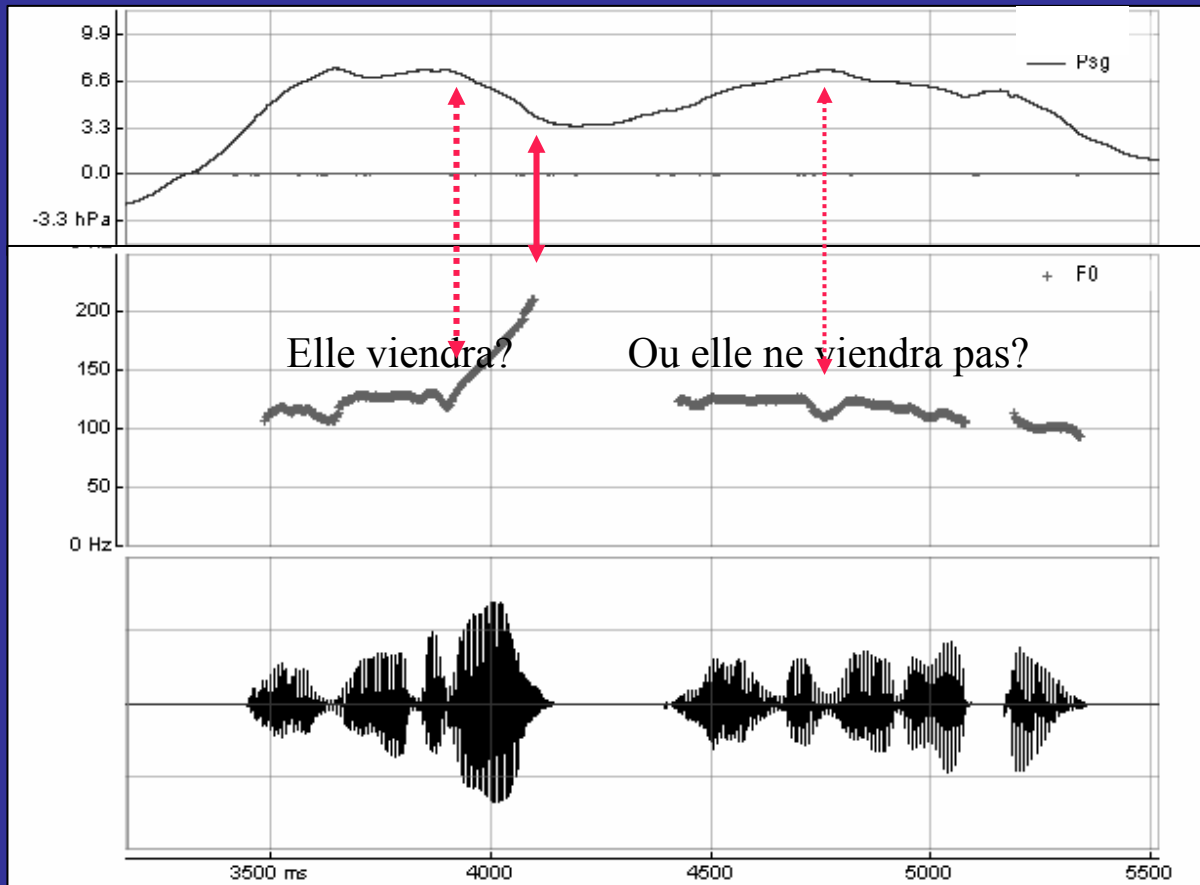


Independence of fo from Ps and intensity in humans



▪ Otherwise f_0 and intensity are largely independent.

⇒ This implies a specific mechanism of control.



**Subglottal and intra-oral pressure
in the production of fricatives**

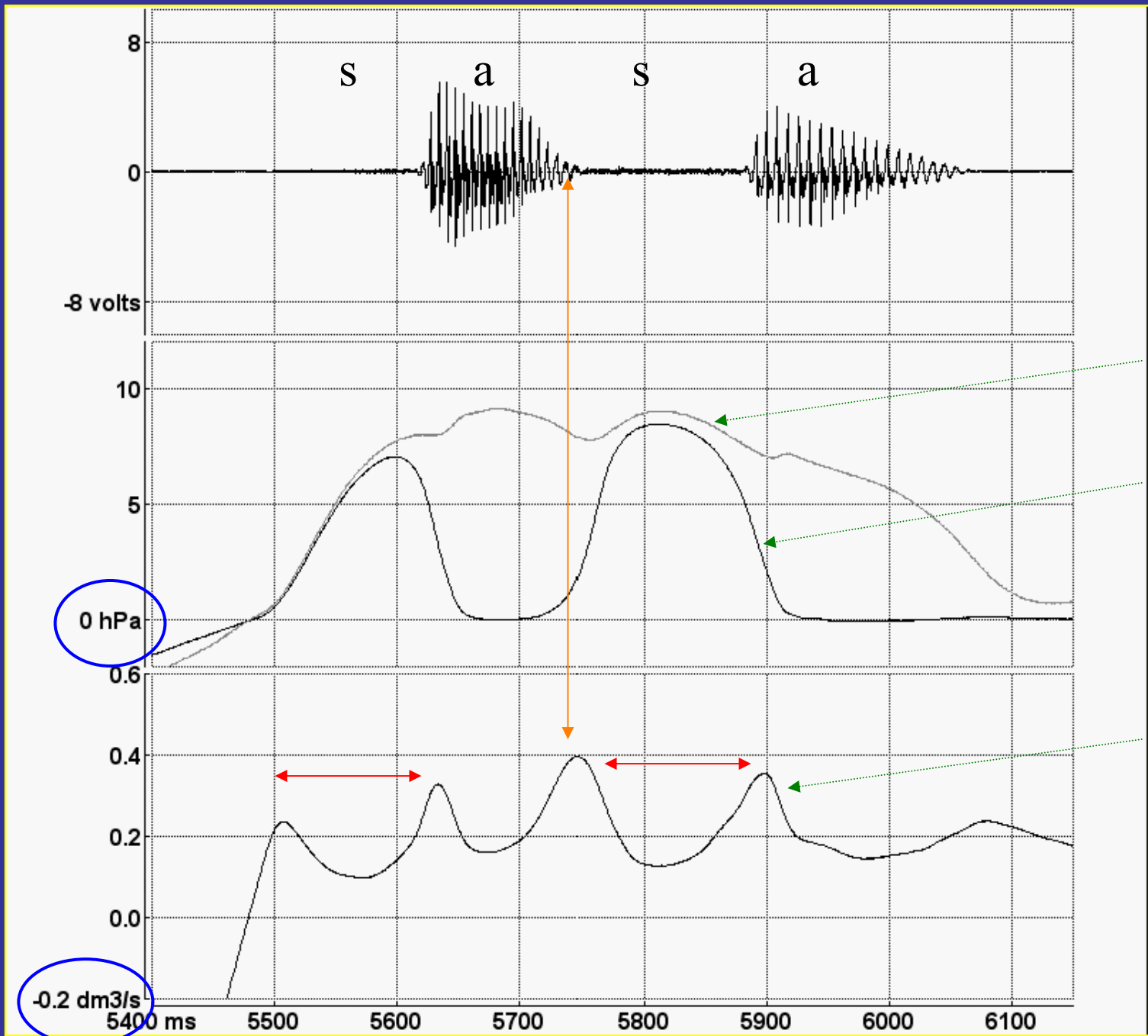
Can we predict the point where frication will start?

What is the dimension of the constriction and what is the necessary flow velocity to generate a turbulency?

What is the critical Reynolds (Re) number for a given geometry of vocal tract?

Voiceless fricatives

$$\Delta P_{\text{oral}} = (P_o - P_{\text{at}}) = \max$$



P_s

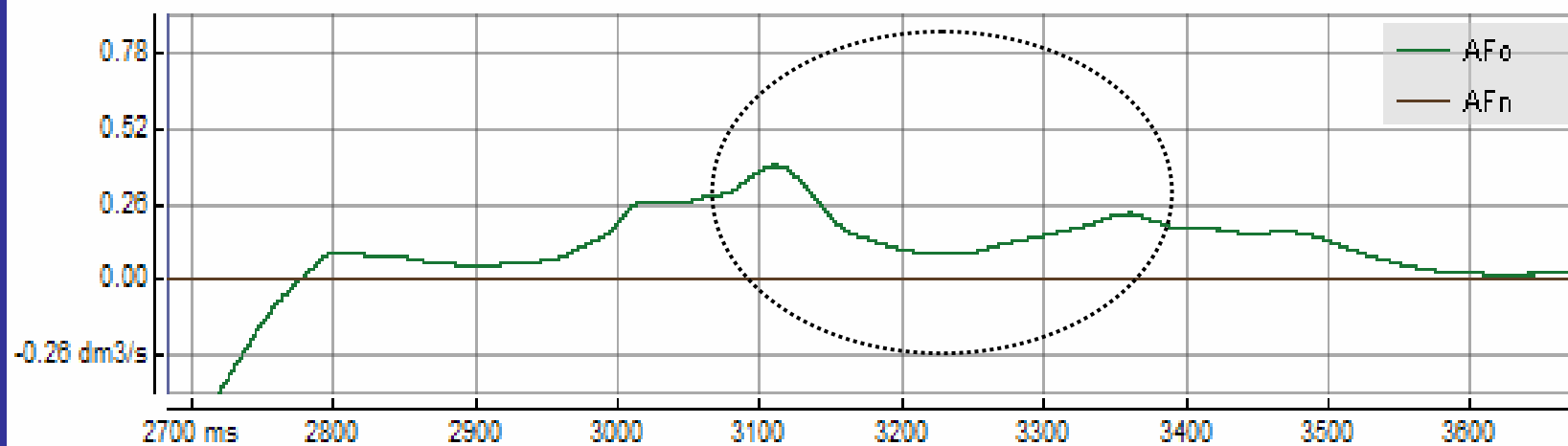
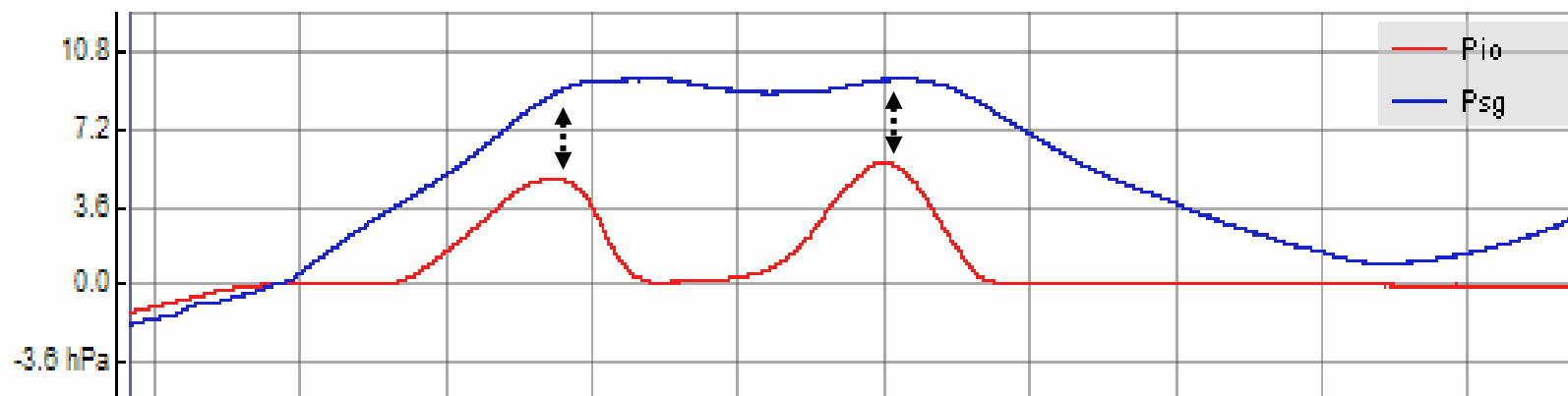
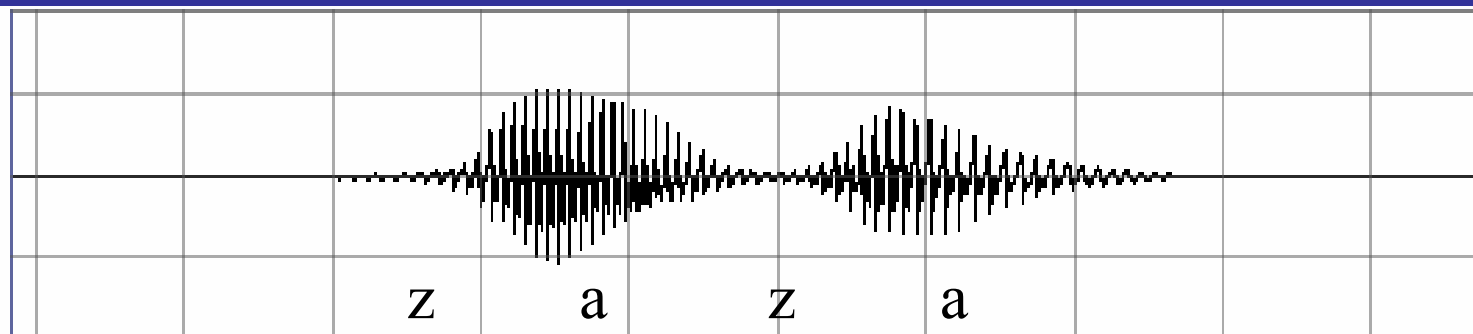
P_o

O_{af}

Voiced fricatives

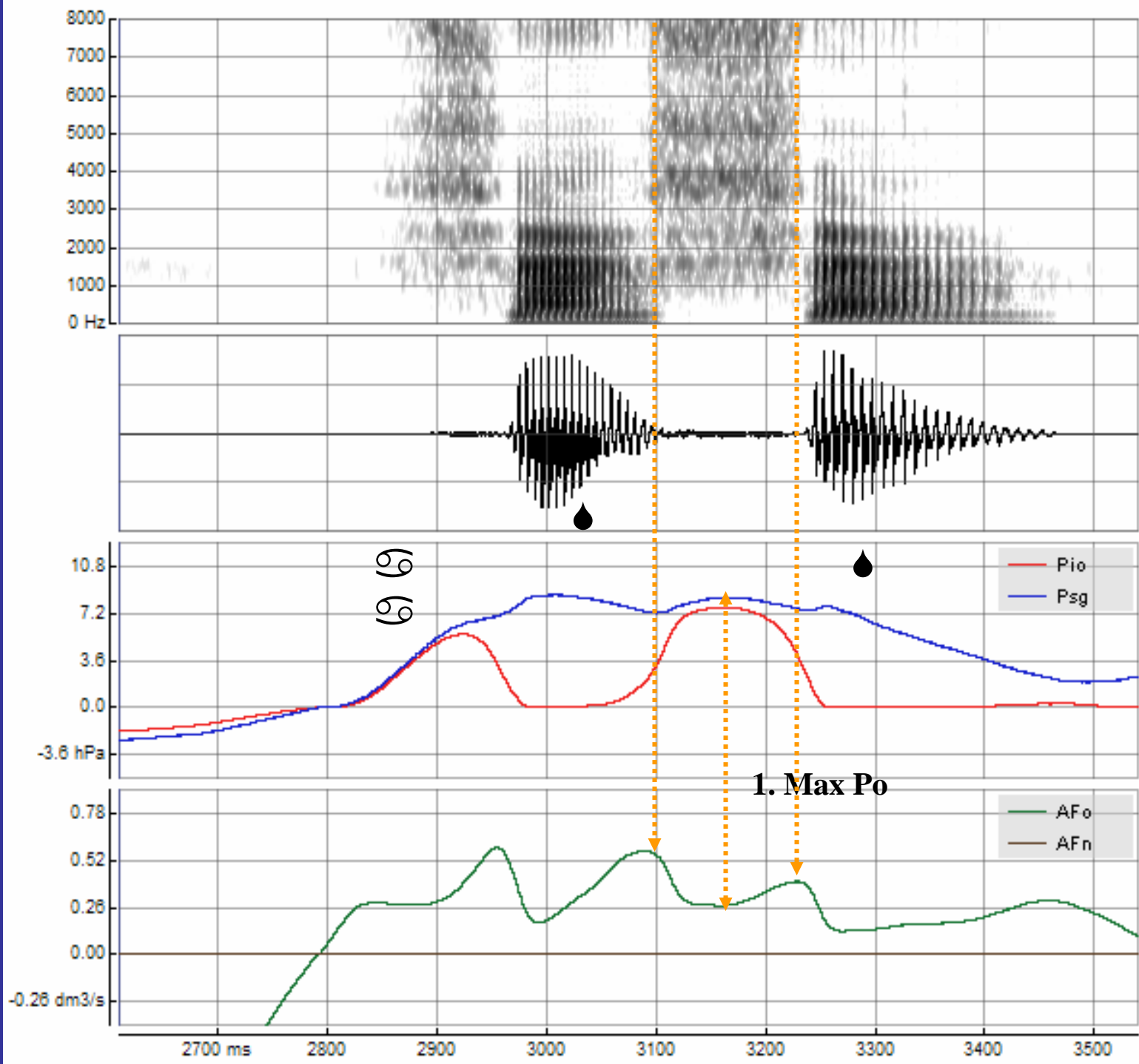
$$\Delta P_{\text{glot}} = (P_s - P_o) = \text{max}$$

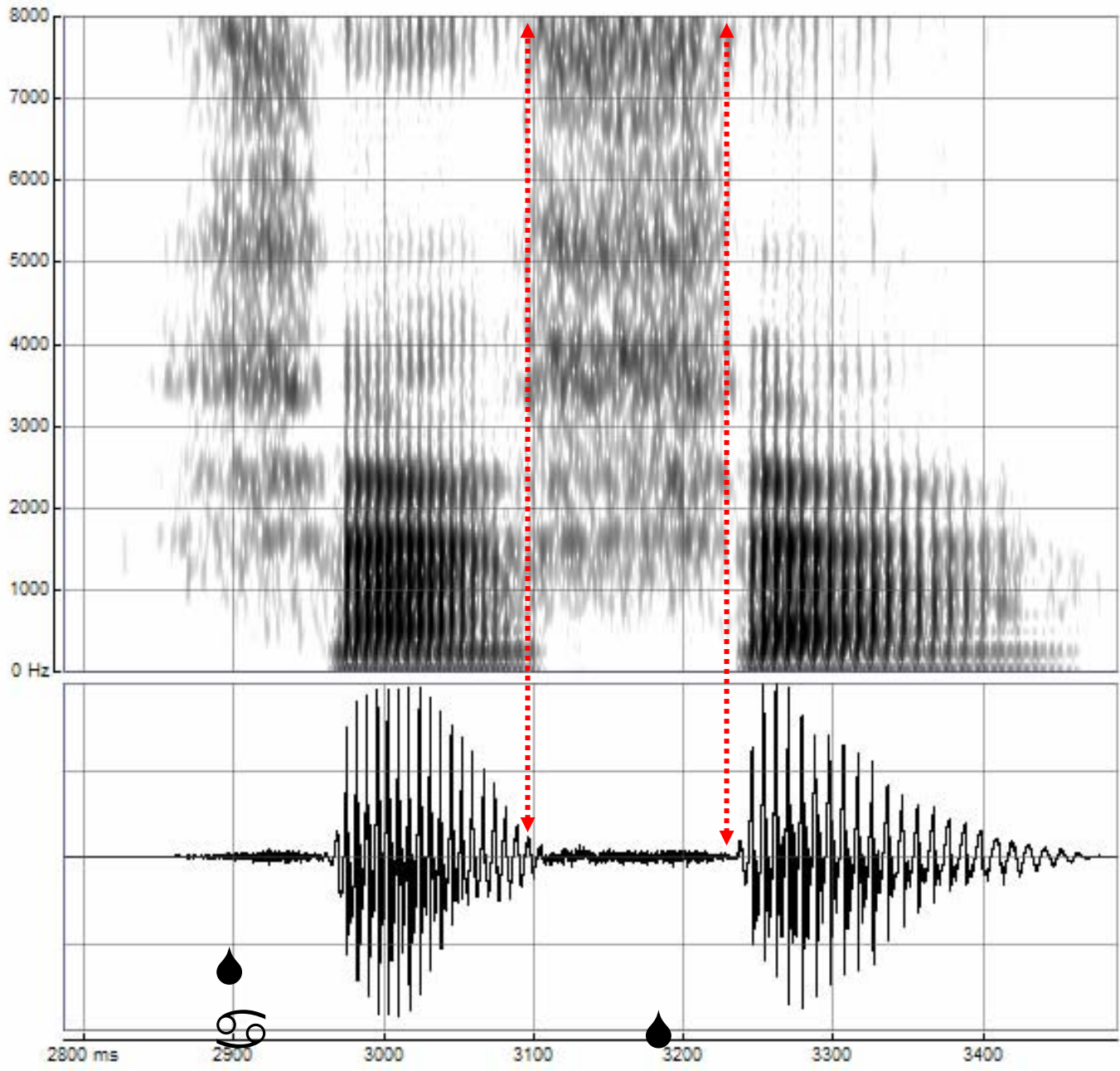
$$\Delta P_{\text{oral}} = (P_o - P_{\text{at}}) = \text{max}$$



Mean peak subglottal pressure (P_s), intraoral (P_o) and oral airflow measurements at three different points:

1. At the level of maximum P_o
2. At the beginning of the fricative (i.e. when frication noise can be detected on the audio waveform).
3. At the end of the fricative when frication noise disappears.



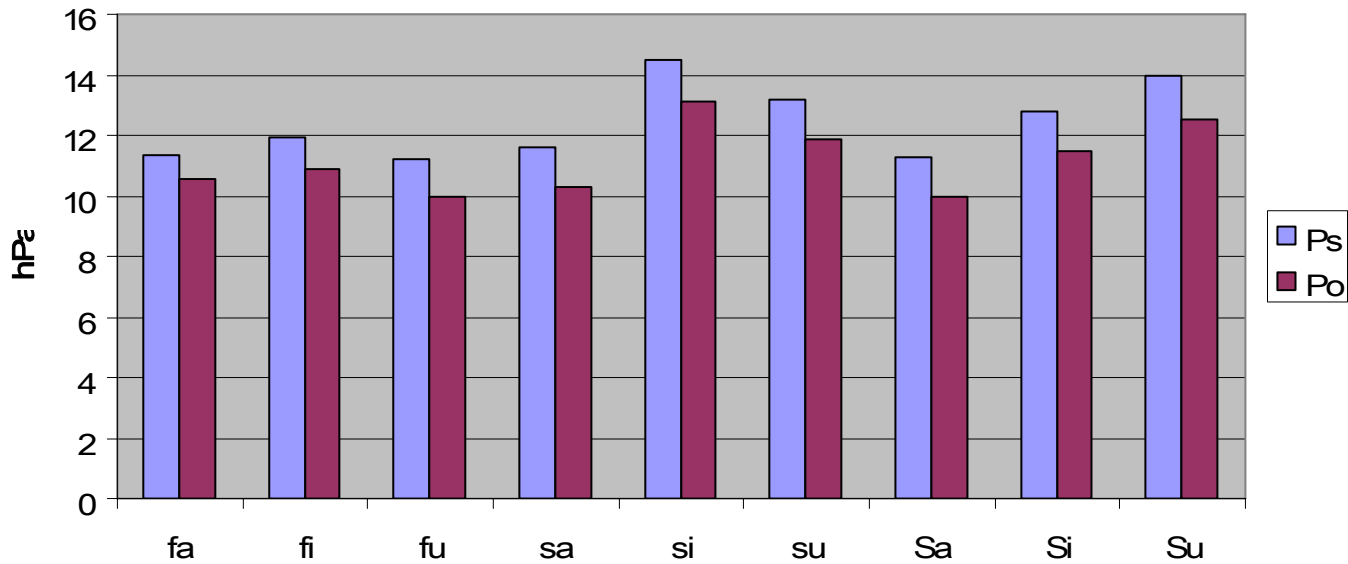


99

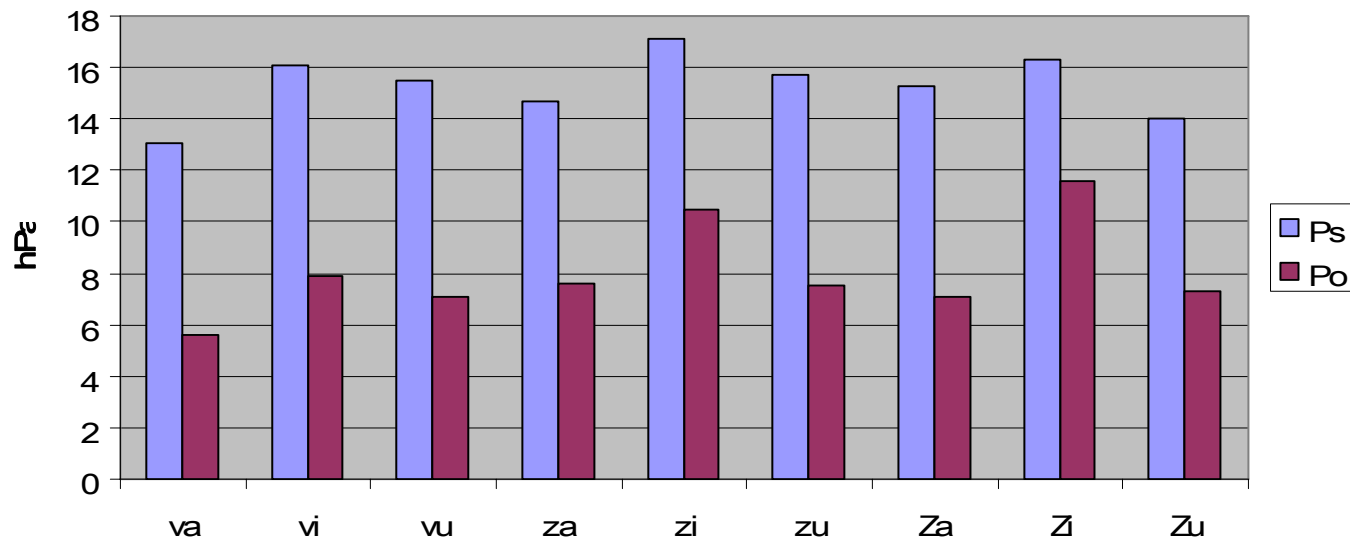
$\Delta P_s - P_o$ (in hPa) (1CmH₂O = 10.2 hPa) n=120

	♈	♠	♾
P_o	4.7	4.6	8.7
P_s	5.8	5.3	9.3
ΔP	1.1	0.6	0.6
	♠	♋	♁
P_o	2.9	2.8	5.3
P_s	6.8	6.6	9.4
ΔP	3.9	3.8	4.1

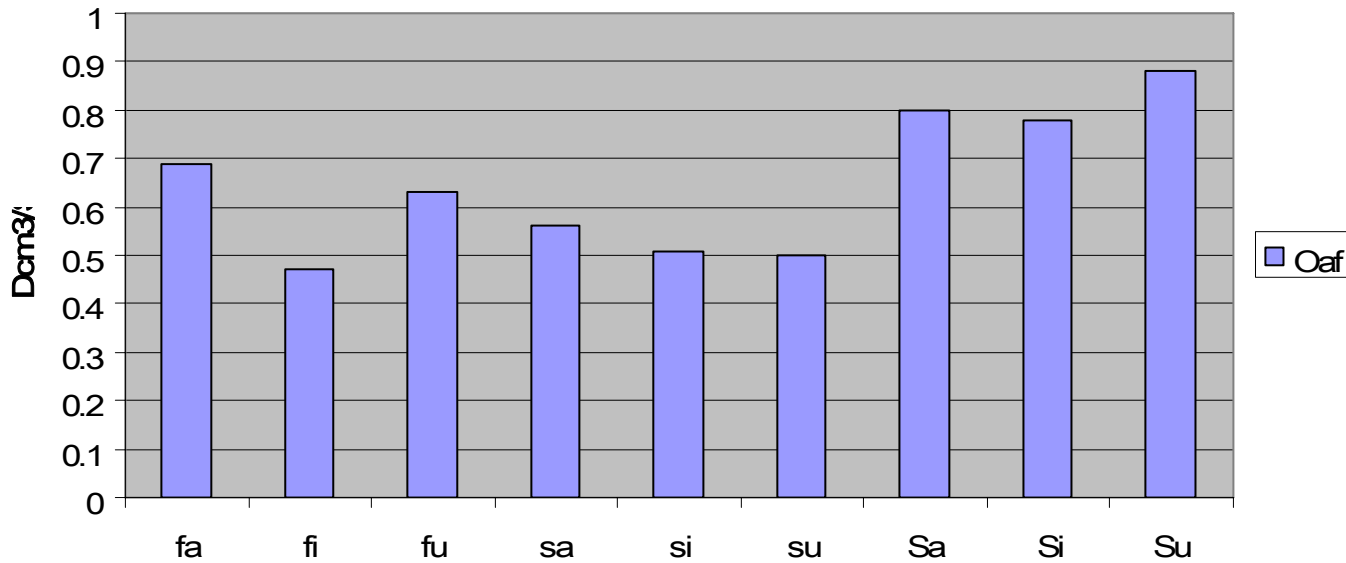
Ps and Po for Voiceless fricatives



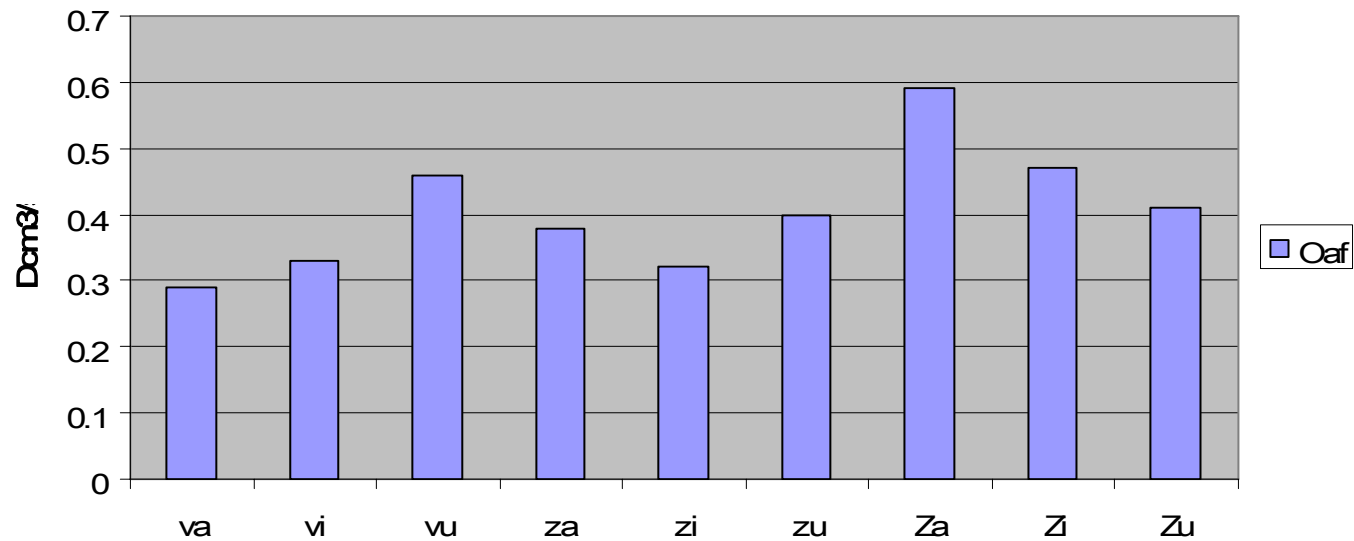
Ps and Po for voiced fricatives



Mean Oaf for voiceless fricatives



Mean Oaf for Voiced fricatives



Fricatives

$\Delta P_s - P_o >$ is greater for voiced fricatives (as expected)

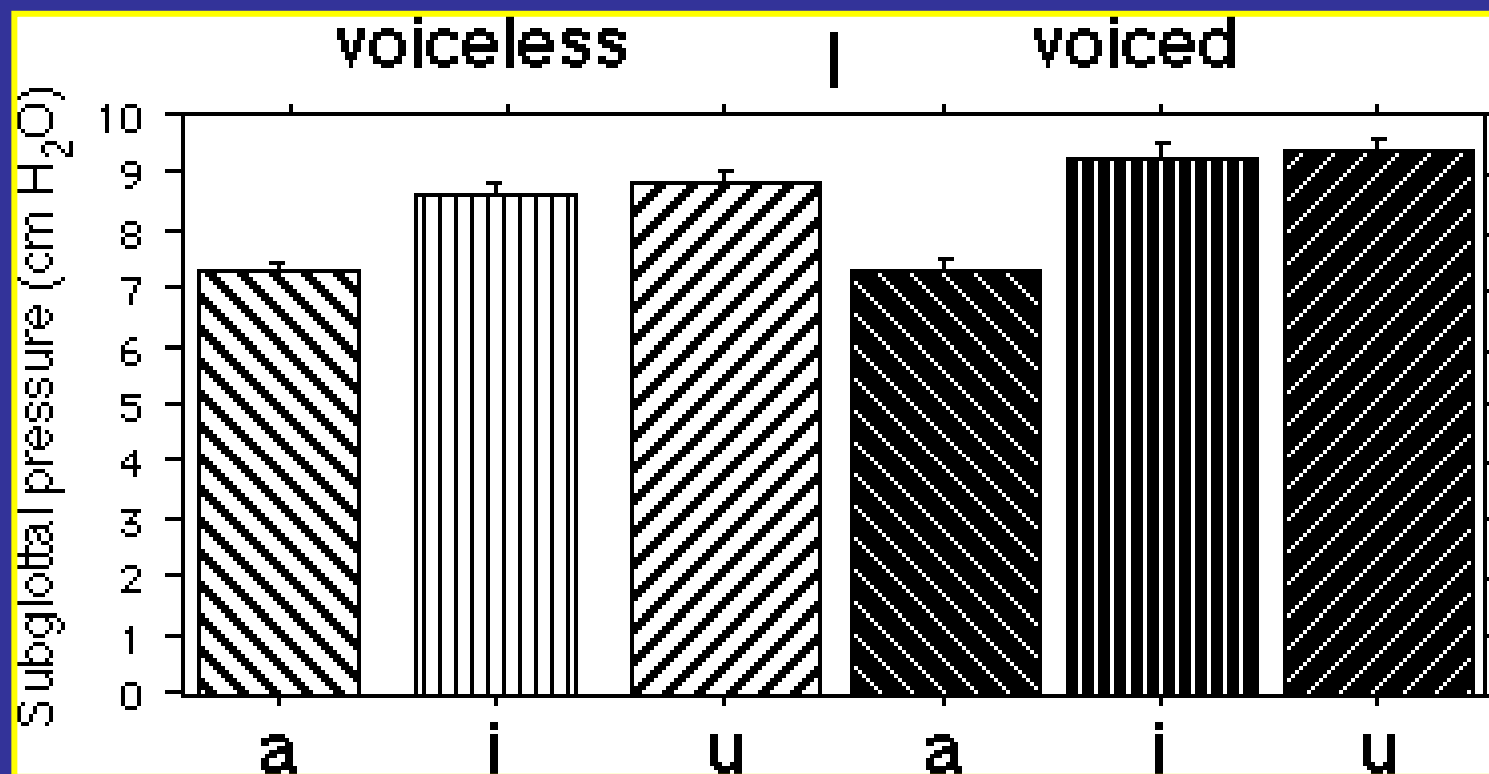
The ΔP observed for the voiceless fricatives is likely achieved by an effect of **R**

P_s is higher for voiced fricatives

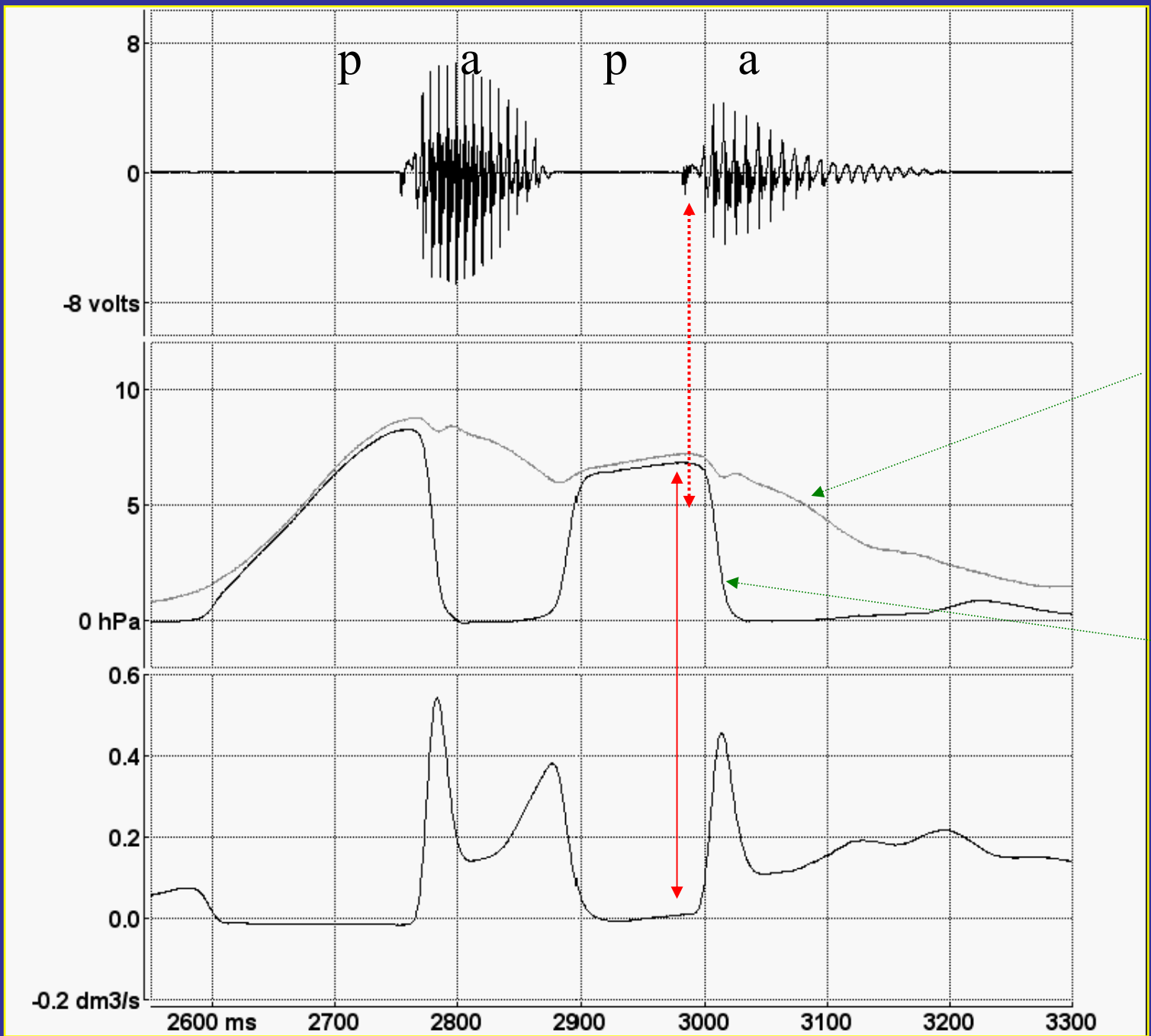
P_s and P_o are much higher for the palatal fricatives compared to the labio-dentals and the alveolars

This could be due to a larger constriction and a higher aerodynamic impedance

(Demolin, Hassid, Ladefoged & Soquet), found that there is a considerable difference in the **subglottal** pressure of voiced and voiceless stops (*measured before the burst*) when they are associated with the vowels [i, a, u].



Differences in the mean subglottal pressure for voiceless as opposed to voiced stop consonants immediately before the vowels /i, a, u/.



P_s

P_o

Across all other variables, the difference in subglottal pressure between /a/ and /i/ is 1.65 cm H₂O and between /a/ and /u/ is 1.76 cm H₂O.

Both these differences are highly significant ($p < 0.0001$) in a one way ANOVA. The difference between /i/ and /u/ is 0.11 cm H₂O, and is not significant ($p = .303$).

Some speakers contributed more to these differences than others. The next Figure shows the results of an analysis of the voiced stop data of four of the five speakers, (One speaker was omitted because less than half the desired measurements were made, most of the missing measurements being for the voiced consonants.)

All the speakers were significantly different from each other ($p < .0001$) except for 'c' and 's' ($p = .7316$).

Two further points can be made with respect to this analysis.

Firstly, although for each of the speakers /a/ has the lowest mean subglottal pressure, the difference between /u/ and /i/ varies among these speakers. Three of them have a larger subglottal pressure for /u/ than for /i/, but one of them, speaker 'd', had a significantly higher pressure for /i/ than for /u/ ($p < .0001$).

Secondly, the difference between /a/ and /i/ for speaker 'd' is 2.973 cm H₂O, a remarkable 30% increase. There are no significant differences due to sex. Speakers 'c' and 'd' are male, and speakers 's' and 'v' are female.

The differences are slightly more marked for voiced consonants in comparison with voiceless consonants. The mean difference between /a/ and /i/ is 1.97 cm H₂O, and between /a/ and /u/ is 2.08 cm H₂O.

Voiced consonants have a higher subglottal pressure than the voiceless consonants.

The subglottal pressure differences could be due a modification of the tract shape made in order to articulate the vowel, or to a greater stiffness in the tract walls. It is difficult to explain why these facts would account for the higher subglottal pressure found for voiced consonants which are produced with adducted vocal folds.

Turning to the impedance of the glottis or to the position of the larynx, it is also difficult to find an explanation.

With a constant driving force produced by a constant rate of decrease in lung volume there would be a higher, not a lower, subglottal pressure for /a/.

Since the respiratory system is generally regarded as producing voluntary variations in intensity but not (at least in most well known languages) producing voluntary increases in pressure for particular sounds, this finding is puzzling.

We could conclude that French speakers may be using voluntary increases in subglottal pressure to distinguish some sounds or particular segments but this is hard to accept given what we know about the regulation of respiration in speech.

Does this account for a language specific fact, or does this account for some universal aspect of the mechanism of speech production?

Prominence

Benguerel (1970) :

“Accent” in French, or “stress”, in English, is somehow a misnomer for a phenomenon which apparently has very little to do with intensity, fundamental frequency, sub-glottal pressure, airflow, or muscular effort of some kind.

*‘In fact several linguists have thought for some time that one or several of these parameters were relevant to unemphatic stress, but this proves not to be the case. Thus “**prominence**” would have been a better choice, but at this point, it would be futile to attempt such a change of name’.*

Benguerel (1970)

Unemphatic stress and emphatic stress

Unemphatic stress best correlates to the longer duration of the last syllable of each rhythmic group.

Emphatic stress correlates highly with **sub-glottal pressure**.

However, we found that in all cases, there is a rise of sub-glottal pressure on or before every emphatically stressed syllable.

This rise is affected at any instant by two factors: **the action of the lungs and the respiratory system**, and **the resistance offered to the airstream by the glottis** and the articulators.

‘In our experiments, and within the limits of accuracy imposed by the method employed to measure sub-glottal pressure, the state of the glottis appears to be one of the two significant parameters to affect sub-glottal pressure. The other one is laryngeal tension’.

In the case of unemphatic stress, sub-glottal pressure is kept fairly constant and fundamental frequency changes are likely produced by laryngeal adjustment.

In the case of emphatic stress, sub-glottal pressure is presumably, in french, the main cause of fundamental frequency changes, although in some cases, it may be overridden by laryngeal adjustment.

⇒ Need for a study of stress examining the action of the vocal folds.

Ladefoged (1967)

Experiments about the relation between subglottal pressure and stress.

‘These results reinforce the view stated earlier that stress is best described in physiological rather than in acoustic terms. Because of the interaction of vowel quality and intensity, and the trading relationships between intensity, frequency and duration, there is no single simple acoustic event that always occur in all stressed syllables in English. But it is apparent that every stress is accompanied by an increase of sub-glottal pressure’.

⇒ Two new physiological parameters: **pressure drop between the vocal folds and vocal fold’s tension.**

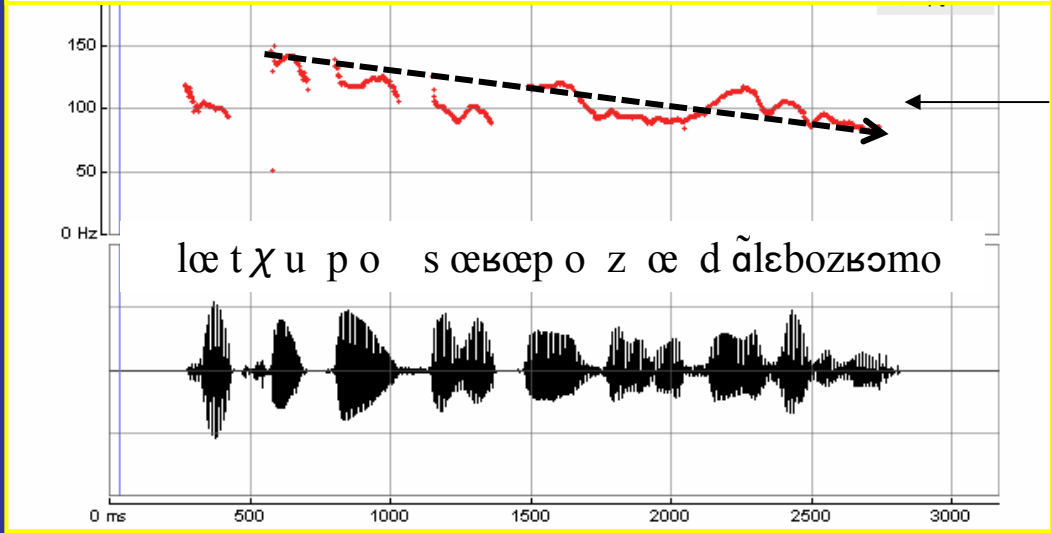
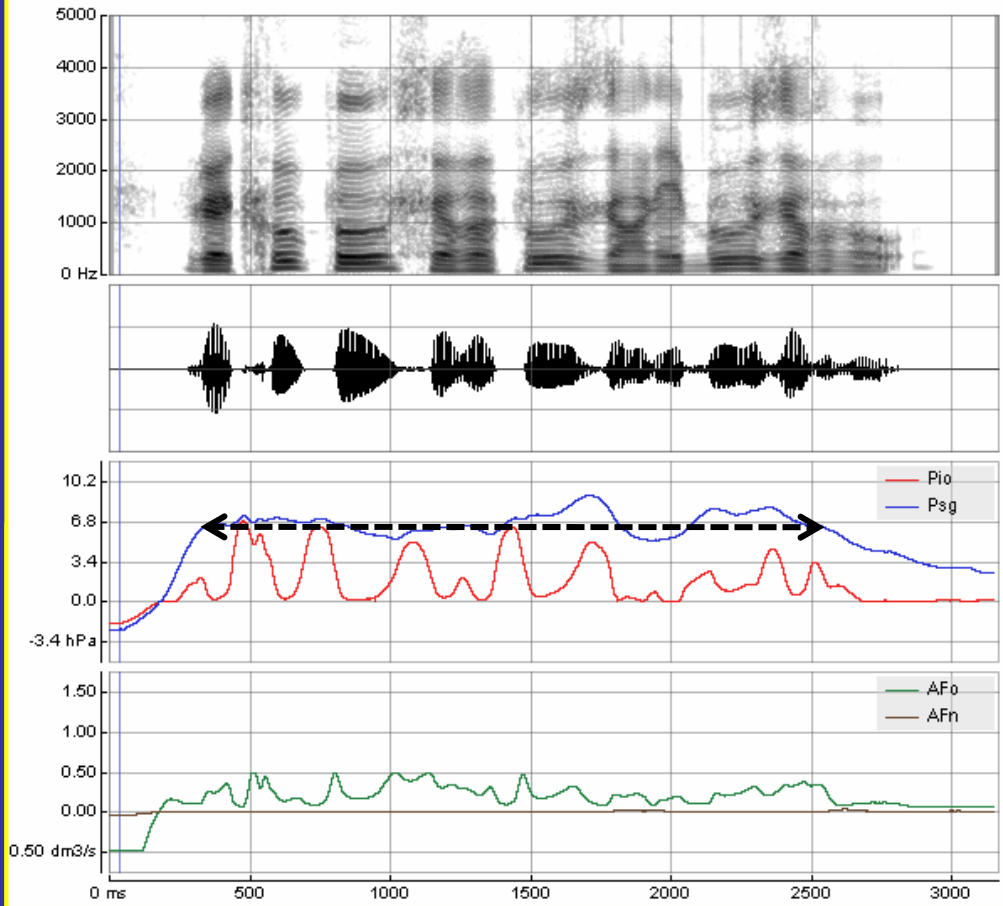
Fant, Kruckenberg, Liljencrants & Hertegård (2000) :

‘It was found that assessments of word prominence closely followed those of the maximally stressed syllable within the word.’

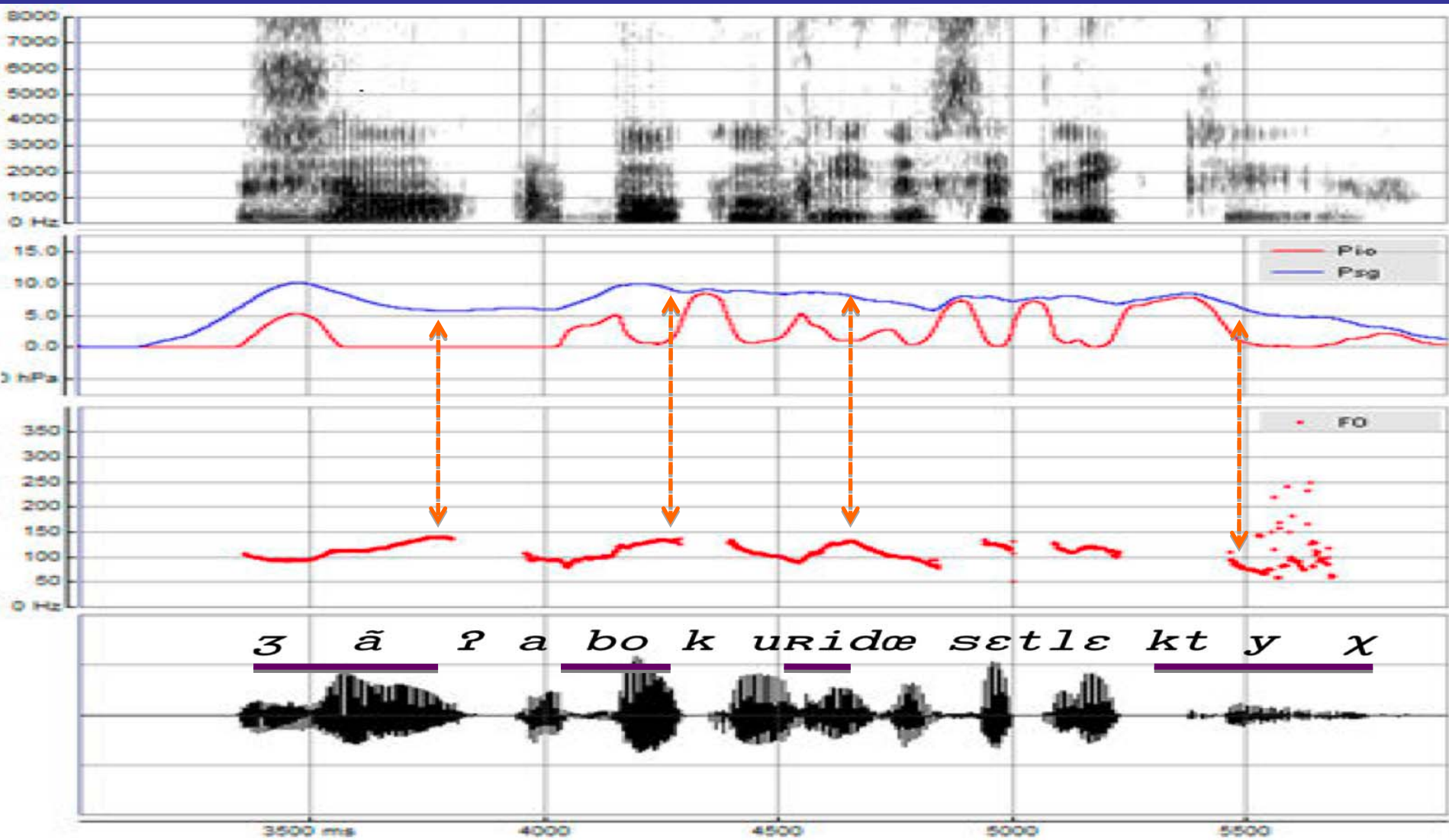
How sub-glottal pressure promotes perceived prominence?

Correlates of prominence :

Subglottal pressure; duration; intensity; ‘spectral tilt’; f₀ pattern.



F0



Data show that speakers might have a greater control on fine details of speech production than what is generally assumed.

The understanding of how aerodynamic parameters are controlled in speech production requires the collaboration of several disciplines and new models of the aerodynamics of speech production.