FROM SENSITIVITY FUNCTIONS...
...TO MACRO-VARIATIONS

L.-J. Boë, P. Badin and P. Perrier
Institut de la Communication Parlée, Grenoble, France

ABSTRACT

This paper presents an approach to study and to predict the effects of articulatory gestures on formants for vowels. The classical sensitivity functions have been extended to macro-variations: large variations have been applied to the seven control parameters of an articulatory model for ten French vocalic prototypes. Non linearities have been analysed and the importance of the different articulators has been expressed in terms of audibility.

FROM SENSITIVITY FUNCTIONS...

In the tradition of dividing the vocal tract into a set of n-tubes of varying lengths and sections, sensitivity functions have been considered as a fairly efficient tools to study articulatory-to-acoustic relationships. In fact the expression *sensitivity function* was first coined in 1974 by Fant & Pauli [1], who then developed this notion in the frame of speech production theory. Sensitivity functions enable an evaluation of the consequences of small variations in area function ($\Delta A_i / A_i$ or $\Delta l_i / l_i$) of an undamped vocal tract on the corresponding resonance frequencies: transversal and longitudinal sensitivities.

Recently a variational formulation of the resonance modes in the vocal tract has been presented by Jospa [2]. This formulation provides and explicits (non linear) link between the formant frequencies and the area function: analytical expressions for the sensititivity functions that take into account wall admittance and lip radiation effects.

If $[l_A]$ and $[l_x]$ are units of area and length, the sensitivity function defined by Jospa, for a resonance mode n is:

$$S_n(x) = [l_A]^{-1}\, \delta f_n / f_n$$

It expresses the ratio of variation due to a small variation of the area function at the position x and corresponding to a Dirac distribution:

$$\delta A(x) = [l_A]\, \delta(x' - x)$$

Sensitivity functions can be useful to study the effects of small changes in area function, particularly at the constriction. However, it is not possible to extrapolate sensitivity functions to large variations. This is due to the fact that the relationship between area function and the corresponding formants can be highly non-linear (sometimes non-monotonous). This important property is at the origin of Stevens' Quantal Theory [3].

...TO MACRO-VARIATIONS

In fact, the approach which considers the vocal tract as a set of n-tubes whose dimensions (area and length) can be manipulated independently, without referring to an underlying articulatory model, even with few articulatory constraints, seems intrinsically very limited. The area function may not be the right control parameter for the vocal tract. This explains why the acoustic properties attributed to a set of n-tubes are of little use for the study of speech production if they do not correspond to realistic operations at the articulatory level. We thus propose to deal with the relations between vocal tract and acoustic by means of an articulatory model. Moreover, we have introduced the notion of *macro-variation* [4] corresponding to large variations (and not differential) of an *articulatory command* for a given articulation.

Non linearities

Using Maeda's model [5] implemented in an adapted environment [6], we have generated macro-variations for the prototypes of French vowels [i e ɛ a y ø œ u o ɔ] elaborated by Boë et al. [7, 8] for variations of the seven command parameters (Lip Height and Protrusion; Jaw Height; Tongue Dorsum, Body and Apex; Larynx Height), within a range of ± 1σ, and a step of 0.25σ (except for the larynx: range of ± 3σ, and step of 0.1σ). Configurations with constrictions under a threshold of 0.3 cm^2 were not considered as vowels and thus discarded. A dictionary of 1,421

configurations has thus been generated.

To evaluate the non linearity of the articulatory-acoustic relationship, the first four formants have been expressed as first degree polynomial combinations of the seven articulatory parameters:

$$F(P) = w_o + \sum_{i=1}^{7} w_i\, P_i$$

where the w_i coefficients have been determined by optimisation [9]. The relative mean quadratic errors (root mean square error / standard-deviation) are respectively, 43%, 18%, 46% and 54% for the first four formants. For 40% of the items in the dictionary, the explained percentage of the variance is less than 90% with a linear model. Two forms of non linearity can be observed: saturation and non-monotony (Figure 1).

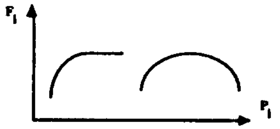

Figure 1: Forms of non linearity between articulatory parameters P_i and formants F_i: saturation and non-monotony.

Vowels presenting non linear relationships are listed in Table 1. The relation between LipH and F_1 for [i, y] is a typical example of the saturation mentioned by Stevens [3]. Dorsum presents non-monotonous relationships with F_1 and F_2.

	F_1	F_2	F_3	F_4
LipH	i y	i y	i y	i e y ɔ
Jaw		ɔ		
Body	y		u y ɔ	i e ɛ a ø œ u o
Dorsum	i y ø œ o	e ɛ a y œ ɔ		
Apex	œ		a	

Table 1: Vowels presenting non-linear relationships between aticulatory parameters and formants (explained variance < 50% with a linear relation).

From the point of view of the articulatory-acoustic relationship, independence can also be considered as a non linearity. This is the case of labial protrusion for all vowels and all formants. Independence and saturation result in *non-audible gestures*, i.e. gestures without acoustic effects, as defined by Abry and Schwartz [10].

Multinomial fits

Each formant frequency F_i in the dictionary was modelled by a separate multinomial function (degree up to 3) of the seven articulatory parameters [9].

$$F(P) = w_o + \sum_{i} w_i\, P_i + \sum_{i} \sum_{j \geq i} w_{ij}\, P_i P_j + \sum_{i} \sum_{j \geq i} \sum_{k \geq j} w_{ijk}\, P_i P_j P_k$$

Table 2 shows the RMSE (Root Mean Square Error / Standard Deviation) as a function of the polynom degree. A fit of degree 3 leads, for the main cardinal vowels [i a u], to relative errors less than 6% for the first four formants, except for a 10% error on F_1 of [u].

	deg. 1	deg. 2	deg. 3
F_1	43.1	13.8	4.5
F_2	18.1	8.8	3.8
F_3	45.7	19.1	8.4
F_4	53.8	26.6	12.6

Table 2: RMSE in % and polynom degree

Influence of gestures on formants

The influence of an articulatory parameter on a formant can be expressed by the relative variation of the formant for a given variation of the parameter. Thus, the resulting variations of the formants for variations of ±1σ of each of the seven parameters, and for each vowel prototypes, have been compared to perceptual thresholds, in order to determine their audibility. The thresholds were 5% for F_1 and 10% for F_2 F_3, approximately following Flanagan [11].

Non audible gestures

Tables 3 and 4 indicate the cases where the variation of a given parameter (within the range of ± 1σ, whenever possible, but avoiding constrictions under 0.3 cm^2) is not audible. Note that the larynx has no effect at all. As well, protrusion is an articulatory parameter with low acoustic influence.

	LipH	LipP	Jaw	Body	Drsm	Apex
i	•	•	•		•	•
e	•	•				
ɛ		•				•
a		•	•		•	•
y		•				
ø		•				•
œ		•				•
u		•	•			
o				•	•	
ɔ		•	•			

Table 3: Non audible gestures for an increase of lip height and protrusion; a raising of jaw, dorsum and apex; a backward displacement of tongue.

	LipH	LipP	Jaw	Body	Drsm	Apex
i	•	•		•	•	
e		•			•	•
ɛ		•				•
a		•	•		•	•
y	•	•		•		•
ø		•			•	•
œ		•			•	•
u	•					
o	•		•		•	
ɔ		•	•			•

Table 4: Non audible gestures for a decrease of lip height and protrusion; a lowering of jaw, dorsum and apex; a forward displacement of tongue.

Audible gestures

Table 5 presents the relative variations of F_1 and F_2 that are larger than 10% for F_1, and 20% for F_2. Note that variations for F_3 did never reach 20%. The body is the most influencing for all vowels, mainly for F_2 [a ø œ u o ɔ], but also for F_1 [i e ɛ y ø œ]. LipH is mainly influent on F_1 of all vowels except [i]; Jaw has influence on F_1 only, for [i e ɛ y ø]. At last, dorsum influences only F_1 of [ɔ].

An example of combination of audible and non-audible gestures

Taking into account the sensivity functions of [i], it is impossible to infer the gesture of the [iy] transition. At the acoustic level: stability of F_1, displacement of the focalisation F_3-F_4 towards a focalisation F_2-F_3 with lowering of F_4, F_3 and to a lesser extent of F_2 [12]. At the articulatory level, this is achieved by an important decrease of lip height, an important increase of protrusion [13], a slight backward displacement of tongue body [14] and a slight larynx lowering [15]. The observation of macro-variations corresponding to these gestures allow to predict these acoustic effects. Protrusion, which has no acoustic effect, could be interpreted as a gesture facilitating the accurate realisation of a small area at the lips.

	F_1	F_2
i	+Body +41 -Jaw +24 -Apex +15	
e	-Body -25 -LipH -24 +Body +19 +Jaw -15 -Jaw +14	
ɛ	-LipH -22 -Body -20 +Body +13 +Jaw -10	
a	-LipH -24	-Body +22 +Body -21
y	+LipH +17 -Jaw +15 +Jaw -13 +Body +13	
ø	-LipH -32 +LipH +19 -Body -17 +Jaw -11	-Body +21 +Body -20
œ	-LipH -29 -Body -17	-Body +24 +Body -21
u	+LipH +73	-Body +26
o	+LipH +40	-Body +37 +LipH +26
ɔ	-LipH -27 +Drsm -12	-Body +30 +LipH +29

Table 5: Percentages of variation of the audible gestures corresponding to ± 1 σ variations of articulatory parameters.

CONCLUSION

With an articulatory model which integrates morphological constraints, it becomes possible to relate geometric variations and associated formant changes to real behaviours in speech production systems.

We have presented an approach that exploits the notion of macro-variations, instead of using sensitivity functions. Thus, the macro-sensitivity functions can be interpreted in terms of non-audible and audible gestures of speech production and motor control commands.

Non-linearities have been quantified by means of a polynomial fit, and it has been show that a multinomial fit of at least the third degree is needed.

The influence of the different articulators on formants for the French vowels has been analysed. Particularly, non audible gestures have been enlighted: larynx height and lip protrusion variations, observed during speech production, seem to have small effects on formants. Protrusion may be interpreted as a gesture facilitating the control of a parameter such as lip area. Oppositely, it has been verified and quantified that the forward/backward displacement of tongue body is one of the most imortant control parameters for F_1 and F_2, with lip height and jaw opening for F_1.

ACKNOWLEDGEMENT

This work has been partially funded by the EC ESPRIT/BR project *Speech Maps*. We are indebted to Andrew Morris of the ICP who has developped and processed the fitting of the articulatori-acoustic relationship.

REFERENCES

[1] Fant G.& Pauli S. (1974) Spatial characteristics of vocal tract resonance modes. *Speech Communication Seminar* 2, 121-126.

[2] Jospa P. (1994) Formulation variationnelle du lien acoustico-articulatoire. *20^{e} Journées d'Étude sur la Parole, GFCP SFA*, 113-118.

[3] Stevens K.N. (1972) *The Quantal nature of speech: Evidence from the articulatory-acoustic data.* In *Human Communication: A Unified View*, 51-66, E.E. David Jr. & P.B. Denes (eds.), Mac-Graw Hill, New-York.

[4] Majid R., Abry C., Boë L.-J. & Perrier P. (1987) Contribution à la classification articulatori-acoustique des voyelles. Étude des macrosensibilités à l'aide d'un modèle articulatoire. *XIth Int. Congr. of Phonetic Sciences*, 2, 348-351.

[5] Maeda S. (1989) *Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model.* In *Speech production and modelling*, 131-149. W.J. Hardcastle & A. Marchal (Eds.), Academic Pub., Kluwer.

[6] Boë L.-J. (1993) *Speech Maps Interactive Plant "SMIP"*.WP2 From speech signal to vocal tract geometry. Sound-to-Gesture Inversion in Speech. ESPRIT/BR n° 6975.

[7] Boë L.-J., Perrier P. & Morris A. (1992) Une prédiction de l'"audibilité" des gestes de la parole à partir d'une modélisation articulatoire. *19^{e} Journées d'Étude sur la Parole, GFCP SFA*, 151-157.

[8] Boë L.-J., Vallée N., Perrier P., Payan Y. & Savariaux C. (1994) *Codebook and sound prototypes with the final state plant.* WP2.2, RP2, Deliverable 17, Sound-to-Gesture Inversion in Speech, SPEECH MAPS (ESPRIT/BR n° 6975).

[9] Morris A. (1992) Least-Squares Fit to Maeda model dictionary. Summary of procedures used and results to date. *Technical Report, Institut de la Communication Parlée, Grenoble.* 8p.

[10] Abry C. & Schwartz J.L. (1988/89) in "La perception visuelle de la parole" Cathiard M.A., *Bull. Institut de Phonétique de Grenoble* 17-18, 110-114.

[11] Flanagan J.L. (1955) A difference limen for vowel frequency. *J. Acoust. Soc. Am.* 27, 613-617.

[12] Mantakas M. (1989) *Application du second formant effectif F'2 à l'étude de l'opposition d'arrondissement des voyelles antérieures du français.* Doctorat INP, Grenoble.

[13] Abry C. & Boë L.-J. (1986) "Laws" for lips. *Speech Communication* 5, 97-104.

[14] Wood S. (1989) Vowel gestures and spectra: from raw data to simulation and application. *12th Int. Congr. of Phonetic Sci.*, 1, 215-219.

[15] Riordan A. (1977) Control of vocal tract length in speech. *J. Acoust. Soc. Am.* 62, 998-1002.