STATIC AND DYNAMIC RELATIONS BETWEEN VOCAL TRACT CONFIGURATIONS AND ACOUSTICS

M. Mrayati* and R. Carré**

*)Scientific Studies and Research Center, POBox 4470, Damascus
**)Telecom Paris, Unité Associée au CNRS, 46 rue Barrault, Paris

ABSTRACT

The relation between vocal tract configurations (VTC) and the acoustic signal represents a complicated non linear process. This phenomenon is inherently dynamic. In this article, aspects of this relation will be presented using a new speech production model, based on the Distinctive Regions and Mode (DRM) [8]. The parameters of the DRM model and its command strategy simplify and formalize the relation between the VTC and the acoustic parameters.

The command inventory of the model is a set of dynamicaly defined gestures. In the acoustic domain formant transitions with time are formalized into a primitive lexicon which is related to the primitive gestures of the DRM model. Formant trajectories in the F1-F2-F3 space are also stylized and related to model gestures or combinations of gestures.

1. LINKS IN VTC-ACOUSTICS

Relations between variations in the vocal tract configurations A(x,t) and the output acoustic signal can deal with different parameters of this signal. Examples of acoustic parameters of the signal are: (1) formant frequencies Fn, (2) formant transitions in the time domain (FT), (3) formant bandwidth (Bn), (4) formant trajectories in the F1-F2-F3 space, (5) noise type, frequency, intensity and duration. One can trace two main schools relating to the consideration of the variations of the VTCs. The first emphasises the articulatory aspect involved in the process, and consequently, concentrates on articulatory models and parameters and then relates the acoustic output to them. The second school emphasises the acoustic tube depicting the VTC, and as such concentrates on acoustic models and parameters taking into account physiological constraints. The DRM model is a model of the second school.

It is not the object of this article to review literature on VTC-acoustic relations, nevertheless, it is helpful to mention some exemples of it. Chiba and kajiyama [4] related the increase or decrease in a resonant frequency of an acoustic tube to constricting the tube near the maximum point of the pressure standing wave, or near the maximum point of the volume velocoty respectivly.

Perturbation theory has been successfully used to study the relation between a small area-function variation and a crresponding acoustic parameter variation [10]. Starting in 1967 Fant used this theory to develop the concept of Sensitivity Function. Sensitivity functions of an arbitrary area function A(x) relate, for a given formant, small local spatial perturbations to formant frequency or bandwidth. Sensitivity functions for length perturbations are also introduced by FANT and served as a measure of "formant-cavity" affiliations [6].

Sensitivity functions can also help to define formant stability as a function of local perturbations. The DRM model is based on the concept of sensitivity function.

Several models of the vocal tract have been proposed to study the relation between VTC and formant frequencies. These models are capable of providing insight into some area-function-acoustic relations or articulatory-acoustic relations. Fant [7] elaborated a model composed of four cavities representing the vocal tract. He provided nomograms relating the five formant frequencies to the dimensions of these four cavities. These nomograms reflect the variations in the formant frequencies due to variations in place and area of the constriction.

As early as 1955 Stevens and House [11] proposed a three-parameter model for vowel production. They presented several nomograms relating acoustic parameters (F1, F2, F3) to articulatory parameters (place of articulation, degree of constriction, and lip parameter).

This article presents the use of a new model of the vocal tract to study the VTC-acoustic relations. Results on simple and formalized relations between model gestures and formants are presented.

We argue that in the research for relations between speech production and acoustics, the adopted model should not overlook important considerations such as: First, the model should be part of, and coherent with, an underlying unified concept linking the different phases involved in the speech communication process; namely: the central representation and motor control, the articulatory, the acoustic and the phonologic levels. Second, the acoustic signal is actually the output of an apparatus for the conversion of muscular energy into acoustic energy. This conversion is hypothesized to be effecient. The modeling of this apparatus and the choice of its parameters and its command strategy are crucial for the right detection and explanation of the relations between the different levels of the speech communication process. Third, speech production is inherently a dynamic process. We are forwarding the hypothesis that the DRM model incorporates these considerations [9].

2. THE DRM MODEL, GESTURES AND ACOUSTICS

For an acoustic tube, closed at one end open at the other, there exist distinctive spatial regions (R) having specific Formant Transition Behavior (FTB). These FTBs are monotonic as long as the variations of the cross sectional area (S) of the different regions (R) are within specific limits defining a mode denoted One Tract Mode (OTM) (approximately between 1 and 16 cm^2, if the neutral position is 4 cm^2). Two other modes can also be defined depending on region cross sectional areas. These are the Transition Mode (TM) corresponding to narrow S(R) (approximately between 0.05 and 1 cm^2); and the Two-Tract Mode (TTM) corresponding to practically closed S(R). These regions and modes are deduced from sign changes of sensitivity functions of the uniform tube (Figure 1). If one is interested in the first three formants, eight regions, can be distinguished.

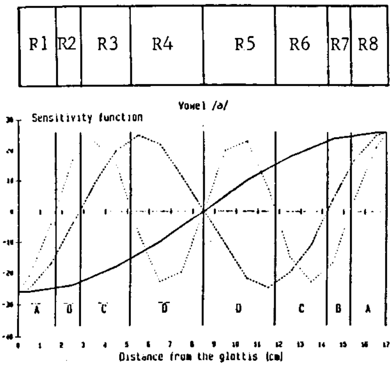

Fig1 - Eight spacial regions from the sensitivity functions for the three formants obtained from a uniform tube and the DRM model.

These regions have the following lengths respectively: L/10, L/15, 2L/15, L/5, L/5, 2L/15, L/15, L/10 (L being the total effective length of the tube). It is clear from the sensitivity functions that, for each of the regions, an increase or a decrease of its cross sectional area leads to a well defined variation sign of the three formants (as far as the OTM is concerned). This aspect distinguishes the eight possible variation signs (three formants with 2 possible variation signs for each result in 8 combinations. The four front regions are antisymetric with the four back ones. As a consequence, an increase in one of the front regions can acoustically compensate a similar increase in the corresponding back region (compensation effect) and area changes in the opposite direction involves a maxmial formant change (synergetic effect). If one considers one resonant mode only, there exists two regions. When only the first two formant are considered, four regions have to be taken into account.

It can be shown that region area changes around the neutral result in efficient simultaneous three formant modulations (efficient transition). When the cross sectional area of a region varies substantially away from the neutral, the change in formant frequency saturates. Such cases represent quantal acoustic targets (Tac) [9]. The definition of fixed spatial regions, delimited for the neutral, permits the existence, in a same model, of efficient transitions of formants on the one hand and stable targets on the other hand. The transversal command strategy for controlling the DRM model is simple, and can take advantage of the inherent synergy principle.

The comand of the model is achieved by means of what we call region-gestures S(R,t). There are eight classes of primitive region-gestures, one for each region (R1 to R8). The parameter of each gesture are the span S(R) and the tempo S(R,t). These gestures are defined around the neutral. In simple or primitive uterances, such as(∂C∂) one region-gesture is involved. In natural uterances serveral region-gestures are combined. Actually, the complete inventory of gestures involves other types of gestures such as the velic and the glottal ones.

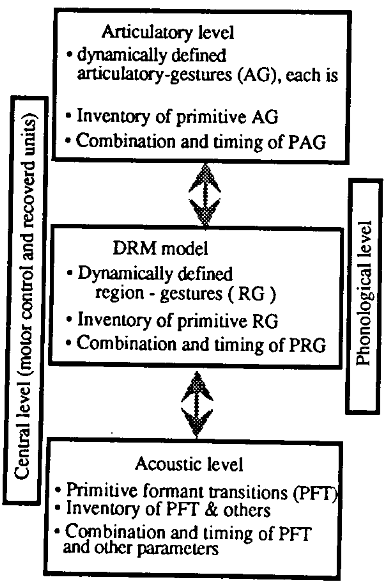

Fig 2 - The DRM model as a gestural acoustic model and its relation to acoustic and articulatory levels.

In the articulatory domain, it is postulated that to realize a region-gesture, one or a group of articulator is controlled in a coordinated manner. The units of the articulatoy gestures are dynamically defined. The articulatory gestures producing the primitive region-gestures could form an inventory of articulatory primitive gestures. These gestures are combined to form natural uterances.

In the acoustic domain, the units corresponding to the primitive region-gestures are specific patterns of formant transitions in addition to other units such as noise patterns. Figure 2, shows a schematic representation of this concept, the central representation level and the phonological level could be added, but it is out of the scope of this study.

It is supposed that in each of these domains, the set of units form a primitive inventory. These primitive units are combined or organized in the temporal domain, taking into acount anatomical, acoustical, phonological and perceptual constraints.

In the following paragraphs we shall concentrate on the relation between the region- gestures and the formant statics and dynamics in the acoustic domain.

3. STATIC RELATIONS

In this paragraph, we point out certain static VTC-acoustics relations dèmonstrated by the DRM model. It can be shown that any area-function A(x) representing a VTC can be modeled using the DRM model and consequently mapped into a point in the vocalic space F1-F2-F3. It can also be demonstrated that the model can produce a vocalic space larger than any other known model of comparable limits on A(x). This VTC-acoustic property of the DRM model is due to the pseudo-orthogonality of its region-gestures [8] [2]. Thanks to the symetrical aspeats of the model, it can incorporate the influance of all parts of the vocal tract. One consequence of this is that what is important in some vowel production is not only the dégree of constriction, but also the opening of the cavity which is at distance from the mid-point equal to that of the constriction.

It has been shown that the model incorporates inherently quantal acoustic tragets Tac [9]. For exemple it is shown, that when the VTC defined by the model becomes close to that of a cardinal vowel, one formant at least becomes quasi-stable, i.e. $dFn/dS(R) = 0$ (see figure 1 in [9]). This interesting VTC-acoustic relation deserves extensive investigation.

4. DYNAMIC RELATIONS

The speech production process is inherenty dynamic and non linear. In this paragraph, we shall treat briefly two aspects of the OTM mode of the DRM model. The first one is the relation between the primitive region-gestures (PRG) and the primitive formant transition behavior (PFT). These gestures are combined in natural utterances to produce all possiple formant patterns. The second one is the relation between region-gestures and corresponding trajectories is the formant space F1-F2-F3. Particularly, trajectories of vowel-vowel transitions (V1-V2) are formalized and related to region-gestures.

4.1 Relations between PRG and PFT

Starting from a neutral tube as a reference (schwa or /ə/), we define a simple PRG to be the complete or partial closure of a region. We have, eight such PRGs, one for each region. Knowing that the regions of the DRM model are favorite places of articulation of consonants [8], therefore the eight PRGs represent sounds of the type /əC/. For example closing region eight (R8: the lips) will lower the three formants, i.e. produces the corresponding PFT. Inversing a PRG, i.e. opening a region, will produce the inverse of the PFT (i.e. raising the three formants in our example). Combining the PRG with its inverse will simulate sequences of the type /əCə/. Figure 3 schematizes the eight PFTs corresponding to the eight PRGs. These region-gestures and their acoustic counterparts are confirmed by well known data on natural speech. Figure 4 gives examples that can be found in the literature and mainly by Delattre [5]. We added five cases for five arabic consonants having their place of articulation in the back half of the vocal tract, (complete results on particular arabic consonants are being prepared for publishing). The following remarks are worth mentioning:

(1) the vowel /ɛ/ is presented for natural speech because data on /ə/ is not available; (2) all the PFTs given for the front half of the vocal tract are those reported by Delattre except for /ʃ/ and /ʒ/ where we presented them non labialized, where the F3 transition is inverted to keep the gestures primitive; (3) other PRGs are defined for the DRM model and the combination of serval gestures can produce any formant pattern [2].

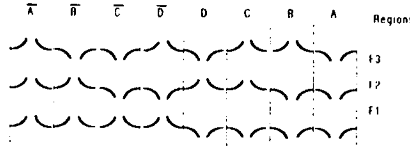

Fig 3 - The schematized eight PFTs corresponding to the eight PRGs.

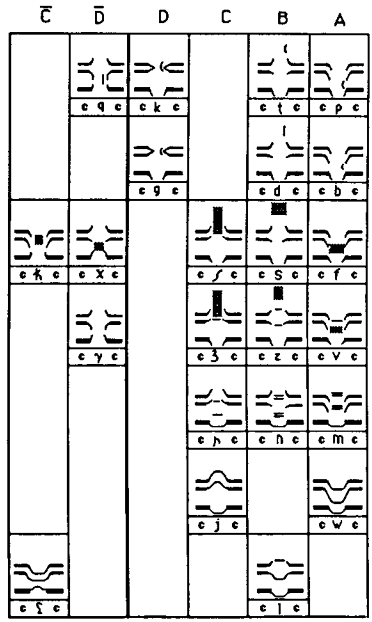

Fig 4 - /ɛCɛ/ transitions deduced from the natural speech, with corresponding regions marked.

4.2 Region-gestures and trajectories in the formant space

The mapping of Region Gestures into the vocalic space F1-F2-F3 has certain pseudo-orthogonal properties. For clarity purposes these trajectories are calculated by computer simulation for a four region model and F1-F2 only.

The model trajectories, their main axis, their spans, and the parameters translating them in the vocalic space, are analysed

[3]. Comparing such trajectories with those obtained for natural V1-V2 utterances, one can investigate and understand the rôle played by the regions of the vocal tract in producing V1-V2 sounds.

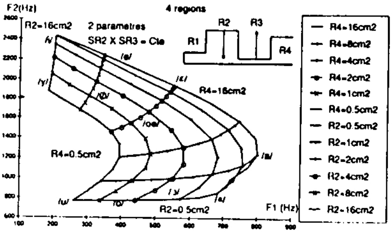

Fig 5 - Trajectories obtained in the F1-F2 plane. Four regions are taken into account. R1 is equal to 1.4 cm^2. Regin areas are varied logarithmically from 0.5 to 16 cm^2. The product of R2 area by R3 area is constant (synergetic area command).

Figure 5 shows an example of trajectories in the F1-F2 plane, corresponding to the combination of two gestures simultaneously (synergetic mouvement). Analysis of vocalic trajectory representing a natural utterance can be achieved by projecting it on such a plane. For exemple, the trajectory of a natural /ai / transition is easily compared with that produced by the model as in figure 5 and for R4 =16 cm^2.

The DRM model is capable of producing any vocalic traget and realistic trajectories by means of simple region-gestures.

5. CONCLUSION

The new DRM model, has been used to show new insight into the relation between VTCs and formant dynamics. Uterances of the form /əCə/ were analysed. The consonants /C/ were Arabic pharyngial ones. Results confirmed the formant-transition patterns predicted by the DRM model for regions R3 and R4 in the back half of the vocal tract ($\bar{C}$ and $\bar{D}$).

The model with its parameters and command strategy seems to be appropriate for the gestural task dynamic approach of the speech communication process [1]. Finally, the model has an explanatory power in relating VTCs to acoustics.

A nasal tract could be added to the model using the same DRM concepts, while noice production in the TM mode is to be investigated.

REFERENCES

[1] BROWMAN, C.P., and GOLDSTEIN, L. (1987), "Tiers in articulatory phonology, with some implications for causal speech", Haskins Laboratory, Status Report on Speech Research, SR 92, 1-30.

[2] CARRE, R., MRAYATI, M. (1990), "Articulatory-acoustic-phonetic relations and modeling, Regions and Modes", in *Speech production and speech modeling*, (W.J. Hardcastle and A.Marchal, editors), Kluwer Academic Publishers.

[3] CARRE, R., and MRAYATI, M., (1990), "Vowel-vowel trajectories and Region modeling", 2nd Seminar on Production: Models and Data, Leeds 13 - 15 May. To appear in the "Journal of Phonetics".

[4] CHIBA, T. and KAJIYAMA, M., (1941). "*The vowel. Its nature and structure*", Tokyo

[5] DELATTRE, P. (1967). "Des indices acoustiques aux traits pertinents", Proc. of the 6th ICPHS, 35-46.

[6] FANT, G., (1980), "The relations between area functions and the acoustic signal", phonetica, 73, 55-86.

[7] FANT, G., (1960). "*Acoustic theory of speech production*", Mouton, the Hague.

[8] MRAYATI, M. CARRE,R. and GUERIN,B. (1988), "Distinctive regions and modes : a new theory of speech production", Speech Communication, Vol 7, 257-286.

[9] MRAYATI, M. CARRE, R. and GUERIN, B. (1990), "Distinctive Regions and Modes: articulatory-acoustic-phonetic aspects", Speech Communication, 9, 231-238.

[10] SCHROEDER, M. (1967), "Determination of the geometry of the human vocal tract by acoustic measurments", J. Acoust. Soc. Am. 41, 1002-1010.

[11] STEVENS , K.N. , and HOUSE, A.S. (1955), "Development of a quantitative description of vowel articulation", J. Acoust. Soc. Am., 27, 484 - 493.