SYNTHESIS-BY-RULE FOR FRENCH

G. Bailly and M. Guerti

Institut de la Communication Parlée, UA CNRS n°368. INPG/ENSERG
46, av. Félix Viallet, 38031 Grenoble, France.
e-mail: bailly@icp.imag.fr

ABSTRACT

Theoretical fundations for a new representation of formantic trajectories are presented. This representation assumes that explicit control of acoustic patterns is done on the underlying resonances of the vocal tract. An application of this approach within a synthesis-by-rule system using an extended version of the Klatt synthesizer is described. Current implementation of this application was done using the COMPOST developpement system.

1. INTRODUCTION

Most synthesis-by-rule systems aim to produce stylized trajectories of a "readable" parametric representation of the speech signal. The choice of the representation is a key-problem to write so-called coarticulation rules: the parameters have to be easely related to articulatory movements which are subject to undershoot, anticipatory behevior and degrees-of-freedom in excess.

In light of the revision by Badin & Boe [4] of Fant's explanation [6] for the explanations for the "affiliations" between formants and vocal tract cavities, we propose to model formant trajectories in terms of the underlying resonance trajectories.

2. DEFINITION OF THE VOCAL TRACT RESONANCES

The vocal tract resonances are defined as the union between the individual theoretical resonances of each cavity of the vocal tract: formant trajectories may be deduced from this set of resonances by applying inter-cavity coupling and vocal tract losses.

The figures 1a and 1b sum up the resonance concept thanks revised Fant's nomograms: we used a 4-tube model with the following configurations for the 4 sections:

Pharynx & mouth cavity : 8 cm2, l varying
Tongue constriction: 0.65 cm2, l=4 cm
Lips: 0.16 cm2 (closed tube) and 4 cm2 (open tube), l=1 cm

The constriction center Xc can vary from 2.1 cm to 11.9 cm while keeping vocal tract length constant and equal to 15 cm.

In all captions, the formants detected by an algorithm described in [5] are plotted in ordinate against the abscissa Xc. In the center and right captions the resonances of respectively the front and the back cavity are superposed in dark to the resonance tracts of the entire vocal tract. All these tracts are then superposed in the left caption.

The proximity of a formant of the whole system with a resonance of certain cavity demontrates clearly the affiliation. Resonance crossings (occuring when a resonance of the back cavity is close to one of the front cavity) produce affiliation change of the corresponding formants (supposed labeled in ascending order). Large coupling between front and back cavity may mask affiliation change (see the cases of the middle constriction (/u/) in fig. 1a and the back constriction (/a/) in fig. 1b.More interesting is the formant convergence between F2-F3 in case of a front constriction (/i/) in fig.1b: the half wavelength of the back cavity cross the lower resonance of the front cavity (intermediate between a helmholtz resonance and a half wavelength).

We adopt the the following notations for designation of resonances : R1, R3 are the two first low resonances of the back cavity and R2, R4 are the first two low resonances of the front cavity.

3. RESONANCE TRIANGLE

A carefull examination of vocalic transitions VV uttered by two male speakers for couples of the 10 French oral vowels in light of interpretation of vocalic nomograms in terms of resonances as developped above show an interesting design of the representation space for vowels. Figure 3 presents the R1-R2 plane deduced from the database for one speaker by interpolations on a grid of formant candidates obtained by closed-phase LPC analysis. Figure 2 shows an example of such an interpolation on a /a-i/ vocalic transitions.

A brief comparison between the nomograms of the Fig.1 and the resonance triangle shows the relative adequacy between the theoretical predictions and the natural data: the large dispersion for the F3 values of the back vowels is accounted by the affiliation of this formant with R4 and thus correlated with the lip aperture. On the contrary the F2 of the front open vowels and the F3 of the front closed vowels are affiliated with R3 and thus are not sensible to the lip aperture. It accounts for the non-linearity of the coarticulation characteristics we already published in [7]: /i/ seemed to have the most influence on the successive vowel while bearing no undershoot. In fact target variability for front-open vowels like /i/ and /e/ in French is mostly accounted by F3 (thus R2) and not F2 (thus R3).

4. MODELLING FORMANT TRANSITIONS VIA RESONANCE TRAJECTORIES

We used this resonance representation for modelling formant trajectories: the synthesis-by-rule system we will detail in the following generates resonance trajectories using the COMPOST rule compiler [3]. These resonance trajectories are then converted in formant trajectories using simple equations which capture the essential characteristics of the inter-resonance coupling.

4.1. The COMPOST language

The COMPOST system is a rule-based system for transducing trees: basic atoms on which COMPOST is working are instances of user-defined generic objects. These instances thus inherit of the properties of the class their generic object belongs to: set of features and numerical values. For example, a declaration of word, syllable and phoneme classes will include for French :

```
class Word
/*Dt,Pp stand for determinant & pers. pronoun*/
  object(Noun, Verb, Dt, Pp...) feature(Content)
  Noun, Verb are Content; Dt, Pp are -Content;
endclass
class Syllable
  object(Syl) feature(acc) Syl are -acc;
endclass
class Phoneme
  object(a,i,u,e,y,b,d,g...)  feature(voc,nas,liq,..)
  cue(duration) a,i,u,e,y,b,d,g, have duration=90;
endclass
```

COMPOST then consists of manipulating a complex structure (n-ary tree) whose leaves are instances thanks to an extension of the well-known rewriting rules:

SubTF -> SubTT / SubTL+SubTR;

SubTF is the focus subtree, SubTT is the transformed subtree while SubTL and SubTR are the left and right context subtrees.

The powerfull COMPOST subtree matching is labelled with special instructions for local operations :

-memorization of focus instances and/or subtrees in SubTF and their replication in SubTT (thus enabling tree manipulation)
- numerical capabilities: memorization of numerical attributes in SubTF and affectation of complex numerical expression to numerical attributes in SubTT.

For example, a rule for syllabic parsing for French will include:

```
/* create a father node Syllable for any non-liquid
consonant followed by a vowel */
regS:  [-voc,-liq] -> )Syl(#1 / + [voc] ;
```

A COMPOST sketch consists of a set of grammars (each containing a set of ordered rules working on the internal COMPOST tree structure) and external calls. A library of standard routines may be augmented by the user using the COMPOST C-toolkit.

4.2. Modelling trajectories

COMPOST library includes the routine Gentraj which produce frames of parameters according to instructions present in its actual internal tree. It scans for any instance of the class Phoneme and generates the absolute time reference axis according to the actual values of Duration

cue (expressed in ms). The subtree of each Phoneme is then scanned for any instance of the class Target. The first cue of each Target object gives the delay of this target in ms according to the beginning or the end of the father Phoneme (according to the feature ±Final). The following cues precise the entire set of parameters used by the synthesizer. The targets will be then connected together according to the name of the generic object (splines, straight lines, step functions...).
Like in object-oriented design all cues of a certain generic object are allocated on instanciation. To avoid obligatory synchronization of all parametric trajectories, the target is validated only if the parameter's value differs from a default value specific to each parameter.For example, the following rule generates the two targets (oral and naso-pharyngeal parts) for a French nasal vowel /ɑ̃/:

```
-------------------
class Target
  object(X0,X1,X3) feature(Final) cue(t,F1,F2,F3)
endclass
an: ɑ̃ -> #1(<X3, t=20, F1=515, F2=1350, F3=
2100><X0, t=50, F1=530, F2=1000, F3=2200>);
-------------------
```

The set of F1-F4 targets for each phoneme is then taken for R1-R4 targets as developped in the preceeding chapters. Once the entire resonance and residual parametric trajectories have been computed by Gentraj, two other functions of the library are then called to produce sound: Coupling and Klatt. Coupling computes new ordered values for F1-F4 using the following algorithm:

```
-------------------------
- sort F1-F4 in ascending order
- for all Fi do:
    for all j<>i do
      d=Σ 200.sgn(i-j).exp(-|Fi-Fj|/200)
      Fi = Fi+d
    enddo
enddo
-------------------------
```

4.3. Vocalic target resonances

Examples of resonance target values for the 10 French oral vowels and some consonants are given below:

```
-------------------------
```

//	R1	R2	R3	R4	B1	B2	B3	B4
a	620	1240	3330	2500	75	60	130	150
ɔ	460	1000	3000	2600	80	65	130	50
o	370	800	2260	3000	55	65	110	100
u	250	750	2100	3000	60	60	95	110
œ	480	1420	3200	2300	70	70	90	95
ɸ	360	1600	2150	3000	60	70	80	75
y	250	1720	2060	3000	75	95	125	80
ɛ	590	1900	3300	2300	60	100	110	120
e	320	2700	2000	3300	55	85	80	100
i	250	3000	2000	3400	55	60	100	100
j	250	2500	2100	3200	85	80	100	110
w	250	750	2100	3000	60	90	90	110
ɥ	250	1600	2200	3200	75	95	120	80
m	250	1300	2300	3000	80	130	140	140
n	250	1450	2600	3300	90	50	150	130
r	250	1100	2300	3400	90	90	140	160
l	250	1650	2100	3400	90	120	145	190

```
-------------------------
```

5. CONCLUSIONS

Spectrogram reading in light of resonance structures enables a clear acoustic-to-articulatory inversion [1] and thus enables an easy way of modeling formant trajectories. Consonant loci have to be revised to take account of this new vocalic structure. We think that this acoustic representation of the speech signal may suggest new investigation for a number of key problems such as effective formant calculation [8]: F'2 could be the result of tracking of R2 instead of a large-scale integration.

Acknowledgements

We are very thankful to P. Badin for his suggestions on resonance structures.

REFERENCES

[1] ATAL, B.S., CHANG, J.J., MATHEWS, M.V. & TUKEY, J.W. (1978) "Inversion of Articulatory-to-Acoustic Transformation in the vocal tract by a computer sorting technique", *J. Acoust. Soc. Am.*, 63, 1535-1555.

[2] BAILLY, G., MURILLO, G., AL DAKKAK, O. & GUERIN B. (1988à "A text-to-speech synthesis system for French by formant synthesis", *7th FASE Symposium*, 255-260.

[3] BAILLY, G. & TRAN A. (1989) "COMPOST: a rule-compiler for speech synthesis", Eurospeech, 136-139.

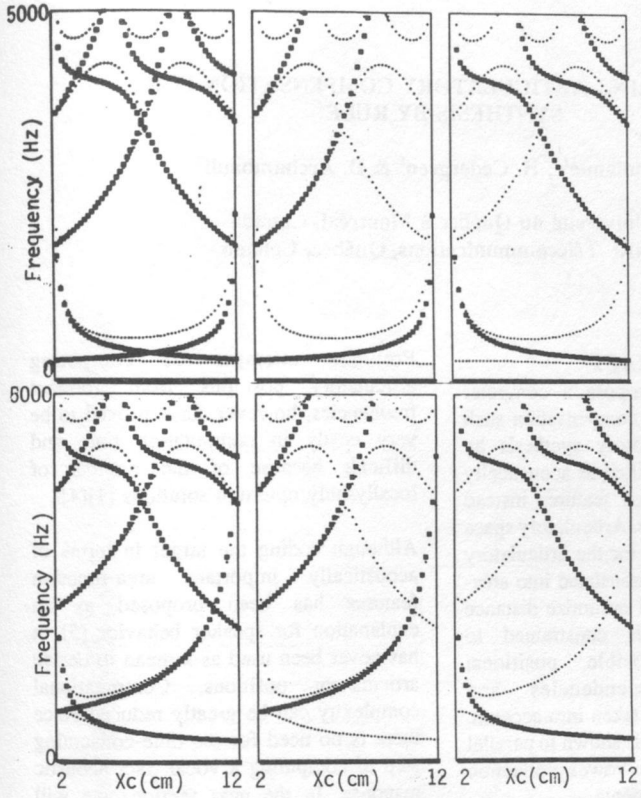
[4] BADIN, P. & BOE, L.J. (1987) "Vocal tract nomograms: acoustic considerations", *Proc. of XIth Int. Cong. of Phon. Sci.*, 352-355.

[5] BADIN, P. & FANT, G. (1984) "Notes on vocal tract computation", STL-QPSR, 2-3, 53-108.

[6] FANT G. (1960), "*Acoustic Theory of Speech Production*", The Hague: Mouton & Co.

[7] GUERTI, M. & BAILLY G. (1990) "Anticipation et retention dans les mouvements vocaliques du français", XIII° Journées d'Etudes sur la parole, 292-295.

[8] SCHWARTZ , J.L. & ESCUDIER, P. (1987) "Does the human auditory system include large scale spectral integration ?", *The psychophysics of speech perception* (Schouten, M., Editor), Nato Asi series, Dortrecht, 284-292.



Figs.1a & b: Vocalic nomo-grams: on the nomogram for the 4-tube model are superposed from right to left: the nomograms of a) the back cavity alone, b) the front cavity and c) both cavities considered as independent. Top captions are for the closed case (Al=0.16 cm2) and bottom for the open case (Al=4.0 cm2).

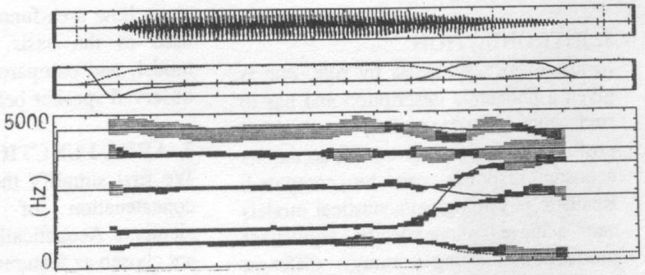


Fig.2: Resonance tracks for an /a-i/ sequence: from bottom to top are presented a) the formant and resonance tracks, b) the interpolation functions issued by a Temporal Decomposition technique and c) the time signal.

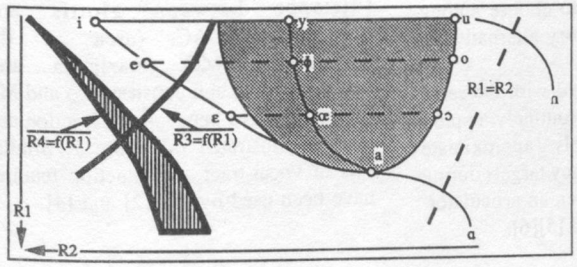


Fig. 3: Vocalic triangle in the R1-R2 plane: the classic F1-F2 vocalic triangle is figured in dark. Realization space for R3 and R4 are figured in heavy dark. Resonance targets for extreme back vowels /ɲ/ and /ɑ/ are suggested.