A FEEDFORWARD CONTROL STRATEGY CAN SUFFICE FOR ARTICULATORY COMPENSATIONS

Shinji Maeda

Département SIGNAL, CNRS URA-820
Ecole Nationale Supérieure des Télécommunications
46, rue Barrault, 75634 Paris Cedex 13, France

ABSTRACT

An articulatory model was used to analyze cineradiographic and labiofilm data. The variation in "target" values of two model parameters, the jaw and tongue-dorsum positions, during the production of the vowels, /i/ and /a/, was examined. The "target" values of these two parameters for the same vowel vary much more than the corresponding acoustic ones. The scattergram of each vowel exhibited a linear relationships which can be regarded as an indication of the coordination between the jaw and tongue. When the coordination effects are subtracted, the articulatory variability becomes comparable to that of the acoustic (F1/F2) one. Calculations with the model indicated that the coordination is used by speakers to achieve an acoustic compensation. These findings suggest that vowel production is compensatory and that compensation can be modelled effectively by a feedforward strategy.

1. INTRODUCTION

Bite-block vowel experiments have demonstrated a speaker's ability to compensate for the effects of blocked jaw position by readjusting the other articulators to produce specified vowels. Observing a speaker's ability to compensate immediately, Lindblom, Lubker and Gay have suggested that normal speech production itself is compensatory [3]. If this is the case, we should observe in normal speech a high degree of variability in the individual articulatory positions and a lower degree in the corresponding acoustic patterns, for example, in the formant patterns. Moreover, if compensation occurs in an arbitrary manner, it is not effective to specify vowel targets in terms of articulatory parameters. This appears to be one of reasons why the targets are often described by the vocal tract area function [2]. If compensation occurs in a lawful manner however, the vowel targets can be specified directly by the individual parameters with some calculations reflecting the laws. We shall investigate these questions by analyzing X-ray and labiofilm data with an articulatory model.

2. ARTICULATORY DATA AND MODEL

The data consist of more than 1000 digitized tracings of vocal tract shapes corresponding to 10 French sentences uttered by two female speakers, PB and DF [1]. Each of the data frames describing the vocal tract profiles from the glottis to the lip opening and the frontal lip shapes was obtained by manually tracing radiofilms and labiofilms shot simultaneously at a rate of 50 frames per second. The digitized version of the data has been kindly provided by the Phonetic Institute of Strasbourg, France.

The measured vocal tract shapes were analyzed statistically. A factor analysis has resulted in a linear articulatory model with seven parameters. In this study, we shall focus our attention on two parameters, the jaw and tongue-dorsum positions for two reasons: these two parameters are most important for specifying the tongue profiles and they can acoustically compensate for each other, specifically in the production of unrounded vowels, such as /i/, /e/, and /a/ [4].

3. ARTICULATORY VARIABILITY

With the linear model, the value of each parameter is calculated directly from the measured vocal tract shape. The articulation along a sentence can be described,

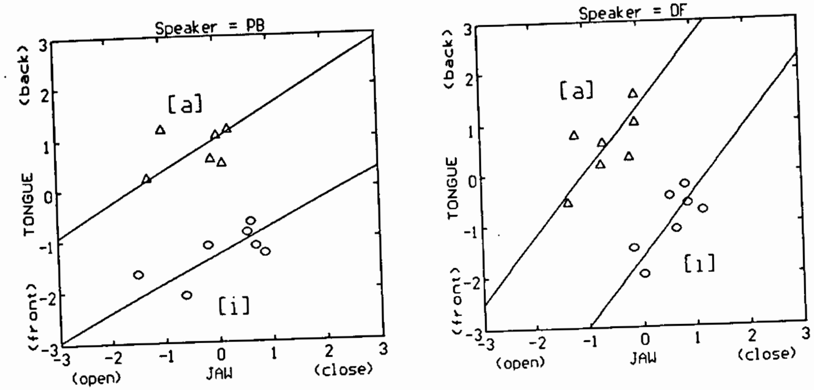

Fig.1 Scattergrams of jaw and tongue-dorsum parameters at the "articulatory targets" for the two vowels /i/ (indicated by the circles) and /a/ (by the triangles). The ordinate and abscissa have standardized units. Zero corresponds to the arithmetic mean calculated for all the utterances by each speaker. 1 (-1), 2 (-2), and 3 (-3) represent 1, 2, and 3 standard deviations, respectively, from the mean. Data for the two speakers, PB and DF are shown.

therefore, by the frame-by-frame variation of the calculated articulatory parameter values. The resultant data have indicated a considerable articulatory variability for the same vowel from different phonetic contexts. In order to assess the range of variability, trajectories of the two parameters, jaw and tongue dorsum, had been plotted on the jaw-tongue articulatory space. Then an articulatory "target" position was determined as the turning point on each trajectory. The result is shown in Fig.1.

The straight lines plotted on Fig.1 were determined by means of a principal component analysis of the scattergrams associated with each of the two vowels: they correspond to the first principal axis. Although the scattergrams exhibit a great degree of variations, the data points for /i/ and /a/ are distributed without overlap. Furthermore, each cluster is distributed roughly along the straight line. These straight lines can be regarded as linear approximations of the inter-articulatory coordination between jaw and tongue-dorsum. The observed variability, therefore, can be separated into a controlled context-determined variation and an unexplained residual, say, "true" variability. Since the proportion of the variance extracted by the first principal component varies between 65% (in the case of [a] uttered by speaker PB) and 88% ([i] by speaker DF), the true articulatory variability for jaw and tongue ranges from 35% to as small as 12% of the observed variance.

4. ACOUSTIC VARIABILITIES

The articulatory variability can be examined more meaningfully, if it is compared with the corresponding acoustic variability. In this study, the first (F1) and second formant (F2) frequencies, as the acoustic characteristics of the two vowels, were calculated using the articulatory model. The F1-F2 calculations were done only for speaker PB, since the data for DF lacks the lip section and thus F1 and F2 cannot be calculated. All seven parameter values were derived from the corresponding data frame. The area function and then formant frequencies were computed from model specified vocal tract shapes. The resultant F1/F2 plots are shown in Fig.2. The data points for the vowel /u/ are added to indicate the vowel space of speaker PB.

Comparing the articulatory target scattergrams in Fig.1 (for speaker PB) and the corresponding acoustic ones in Fig.2, it appears that the acoustic scattergram points are distributed more tightly than

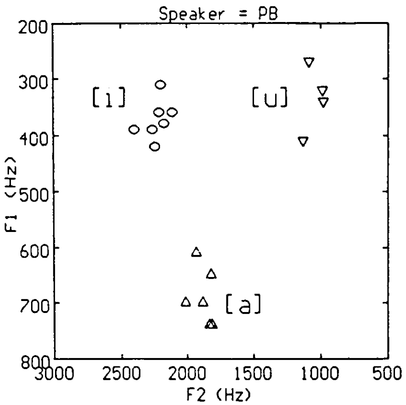

Fig.2 The first (F1) and second (F2) formant scattergrams corresponding to the articulatory target scattergrams shown in Fig.1 (for speaker PB). The scattergram for the vowel /u/ is also plotted to indicate the speaker's vowel space.

articulatory ones, *i.e.*, acoustic variability seems to be less than articulatory one. For a quantitative comparison, let us propose a variability index, υ (an averaged normalized variance), for two articulatory or two acoustic variables as follows:

$$\upsilon = 100 \times \sqrt{\frac{1}{2}\left(\frac{\sigma_1^2}{\sigma_{1_{max}}^2} + \frac{\sigma_2^2}{\sigma_{2_{max}}^2}\right)} \qquad (\%)$$

where σ_i^2 is the variance of variable i (= 1 or 2 in our case), and $\sigma_{i_{max}}^2$ is the possible maximum variance of variable *i*. Since a sufficient amount of data to determine the possible maximum variance is not available, we have assumed, as a gross approximation, that $\sigma_{i_{max}}$, of articulatory and acoustic data can be substituted by the values of half of the range of the individual variables. In the calculation of the articulatory variability index, $\sigma_{i_{max}} = 3$ is used for both jaw and tongue-dorsum data, corresponding to half the range, since parameter values rarely exceed the range from -3.0 to 3.0. The acoustic variability index is computed assuming that $\sigma_{1_{max}}$ (for F1) equals to 300 Hz, and $\sigma_{2_{max}}$ (for F2) to 1250 Hz. The calculated index values are listed in Table 1.

Table 1 Articulatory and acoustic variability indices (in %) for the two speakers

	PB		DF	
	/i/	/a/	/i/	/a/
Jaw/Tongue				
υ	21.7	16.2	17.0	18.4
$\upsilon_{residual}$	7.5	8.1	3.6	4.6
F1/F2				
υ	7.9	8.9	---	---

The index values span around 20% for the articulation and less than 10% for the acoustics. The residual articulatory variability indices are listed at the rows marked "$\upsilon_{residual}$" in Table 1, which are calculated from the proportion of variance corresponding to the residual. These index values are less than 10%, a value which is less than half of the corresponding total raw variability, and which compares well with the index calculated for the F1/F2 scattergrams of PB shown in Fig.2. For speaker DF, the true articulatory variability is four times less than the observed raw variability. The calculation have indicated that although the variability of the individual articulators is relatively great, if the coordination term is subtracted, the articulatory variability compares well with the acoustic one.

5. COMPENSATORY ARTICULATION

What mechanism lies behind this significant reduction of the variability from articulatory to acoustic by means of coordination? In our previous studies [4], we have already shown that in case of unrounded vowels such as /i/ and /a/, jaw and tongue-dorsum positions can acoustically compensate for each other, as mentioned earlier. The compensation means that a deviation in the position of one articulator can be compensated by a readjustment of other articulator(s) to keep the deviation in the acoustic pattern to a minimum. It is reasonable, then, to hypothesize that the inter-articulatory coordination, in fact, results in the acoustic compensation of the type just described above. If this is the case, the principal axis representing the coordination in Fig.1, is also an acoustical "equi-line", *i.e.*, changes in the values of the two parameters along these lines result in relatively invariant acoustic patterns that depend only on the vowel identity.

In order to demonstrate the acoustic equivalence for the two vowels, F1 and F2 values were calculated at different jaw positions from -2.0 (low) to 1.0 (high) for /i/ and from -3.0 to 0.0 for /a/, with 1.0 step size. The corresponding tongue positions were determined by their linear relationships. Note that a change in jaw position influences not only the tongue shape, but also the lip aperture and, to some extent, the larynx position. The values of the remaining five parameters were kept fixed at those originally determined from the corresponding vocal tract data frame. The results are listed in Table 2. The index related to the equi-line of /a/ is 3.2%, which is much smaller than observed acoustic variability. As far as the vowel /i/ is concerned, the index becomes extremely small, about 1%, indicating that the equi-line produces an almost invariant F1-F2 pattern.

Table 2 F1/F2 variability indices calculated along the equi-lines of PB in Fig.1.

	range	index
/i/	-2.0 ⇔ 1.0	1.1 %
/a/	-3.0 ⇔ 0.0	3.2 %

Although the acoustic compensation along the equi-lines is not perfect, it is safe to state that articulatory manoeuvres along an equi-line tend to result in fairly invariant acoustic patterns around the target vowel. It should be emphasized here that the equi-lines are derived from the observation of data. It is tempting to speculate then that the speakers have integrated these equi-lines in their mental process and exploit them to place individual articulator positions differently but appropriately for particular phonetic contexts, yet producing relatively invariant acoustic targets.

It may be noteworthy to mention that the coordination does not necessarily always means compensation. In the case of /u/ for example, the raw variabilities of the jaw and lip parameters (height and protrusion) were relatively small, less than 10%. In detail however, scattergrams indicated that closing the jaw, and narrowing and protruding of the lip opening occur concomitantly, enhancing together a narrow and long lip tube. The acoustic consequences of this kind of coordination would be exactly in the opposite of compensation.

6. CONCLUDING REMARKS

It has become clear that the apparently large variability of the individual articulator positions during the same vowel but from different contexts can be explained, at least in part, by the inter-articulator coordination. Moreover, the coordination is such as to achieve an acoustic compensation which results in the realization of a relatively invariant acoustic target, thus supporting the idea of speech production as a compensatory process [3]. Surprisingly, the coordination and thus the compensation can be specified directly in terms of articulatory parameters. The implication of this is important. If the relationship is well defined in such a simple fashion, it is not unreasonable to assume that speakers know exactly how to coordinate in advance. Then a feedforward control mechanism can be assumed for the compensatory articulation, without resorting to acoustic or to sensory feedback.

7. REFERENCES

[1] Bothorel, A., Simon, P., Wioland, F. and Zerling, J-P. (1986). *Cinéradiographie des Voyelles et Consonnes du Français*. Travaux de l'Institut de Phonétique de Strasbourg.

[2] Gay, T., Lindblom, B. and Lubker J. (1981). "Production of bite-block vowels: Acoustic equivalence by selective compensation", J. Acoust. Soc. Am., 69(3), 802-810.

[3] Lindblom, B., Lubker, J. and Gay, T. (1979). "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation", J. Phonetics, 7, 147-161.

[4] Maeda, S. (1990). "Compensatory articulation during speech: evidence from the analysis and synthesis of vocal tract shapes using an articulatory model". In *Speech Production and Speech Modeling* (W.J. Hardcastle & A. Marchal, editors), pp. 131-149. Kluwer Academic Publishers.