

On articulatory and acoustic variabilities

Shinji Maeda

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

Received 25th July 1990 and in revised form 29th October 1990

In our previous studies, an articulatory model has been established by applying a factor analysis to cineradiographic and labiofilm data. Seven parameters are sufficient to specify the entire vocal tract shape with reasonable accuracy. With this model, the temporal variations are described in terms of the successive frame-by-frame samples of these seven articulatory parameters. In this paper, temporal patterns of two of the seven parameters, corresponding to the jaw and tongue-dorsum positions during the production of the two unrounded vowels /i/ and /a/ were examined in detail. Articulatory “target” values of these two parameters for each of the two vowels embedded in different phonetic contexts vary significantly more than the corresponding acoustic targets expressed in terms of first (F_1) and second (F_2) formant frequency values. An index is proposed to express quantitatively the degree of variability for both articulatory and acoustic targets. Using this index, articulatory target variability is found to be about twice as large as acoustic target variability. For each vowel, the distribution of the targets in the jaw/tongue-dorsum space exhibits a fairly linear relationship between the two dimensions, which can be regarded as an indication of the inter-articulatory coordination between the jaw and the tongue-dorsum. When the articulatory variability index is recalculated by subtracting the predicted effects of coordination, its value becomes comparable to that of the acoustic (F_1/F_2) variability. Further calculations with the model suggest that the coordination is used by the speaker to achieve acoustic compensation. These findings confirm that vowel production is indeed compensatory and suggest that compensation can be accurately modeled without resorting to acoustic or other sensory feedback mechanism.

1. Introduction

This paper attempts to provide some direct evidence for compensatory articulation in normal speech production by comparing articulatory variability and the corresponding acoustic variability. In the past, 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. Although bite-block vowel production is a special case of speech production, it is important because of its implication for the theory of speech motor programming (Lindblom & Sundberg, 1971; Lindblom, Lubker & Gay, 1979; Lubker, 1979; Fowler & Turvey, 1980; Gay, Lindblom & Lubker, 1981). Observing a speaker’s ability to compensate

immediately, Lindblom *et al.* (1979) have suggested that normal speech production itself is compensatory. 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.

A salient feature of our study is that simultaneous cineradiographic and labiofilm data are analyzed statistically with a linear component model (Harshman, Lageföged & Goldstein, 1977; Maeda, 1978; Jackson, 1988). Special care was taken in the factor analysis procedure to ensure that the resultant components could be interpreted in articulatory terms; thus the model can be considered as an articulatory one (Shirai & Honda, 1976; Maeda, 1979). The time-varying vocal tract configurations are described by frame-by-frame variations of the model parameters, such as jaw position, tongue-dorsum position, lip height parameter, and so on. Since entire vocal tract shapes from the glottis to the lips are specified by the model, their acoustic characteristics can be calculated. By systematically manipulating the model parameters, articulatory-acoustic relationships can be investigated (Majid, Boë & Perrier, 1986). Our recent studies (Maeda, 1989, 1990) have indicated that jaw position and tongue-dorsum position can acoustically compensate for each other for unrounded vowels, such as /i/, /e/, and /a/. For the rounded vowels, such as /u/ and /o/, compensation is possible between jaw and lip height (lip aperture) parameters. When jaw and tongue positions for the same vowel embedded in different phonetic contexts were plotted, these two parameters exhibited linear relationships, suggesting that speakers actually exploit the compensation during speech production. The purpose of the present paper is to add a new dimension to this earlier investigation by comparing the articulatory and the acoustic variabilities.

2. Articulatory data and linear component 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. 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. Detailed information, such as the list of sentences, selected X-ray tracings and the corresponding spectrographic data, can be found in Bothorel, Simon, Wioland & Zerling (1986).

Tongue shapes are measured using a semipolar coordinate system whose position is fixed with respect to the rigid maxillary structure. The variations of the tongue shapes are therefore measured relative to the maxilla. The vertical distance between the upper and lower front incisors is considered as a measure of the jaw position. The lip opening tube is approximated by a uniform tube having an elliptic cross-section. The lip tube length is identified as the distance between the upper incisors and the point where the minimum separation of the upper and lower lip occurs. The minimum separation is defined as the height of the ellipse. The width of the lip opening ellipse is determined from the lip frontal trace.

An articulatory model is derived as the result of a factor analysis on the measured vocal tract shapes. This does not necessarily ensure that the resulting model is an articulatory one. It is necessary to have an idea about what an articulatory model should be, providing a basis for the judgement. In this regard, we followed a jaw

based model proposed by Lindblom *et al.* (1971), where tract shapes were determined as a function of parameters such as jaw, tongue-body, and tongue-tip positions; lip height and protrusion; and larynx height. The underlying hypothesis is that during speech production the complex muscular activities of the vocal tract organs are organized into a small number of independently controllable but interlinked functional blocks. Let us call these blocks "elementary articulators" and their action "elementary gestures". In our study, these elementary articulators are extracted from the measured data by means of a statistical analysis (Maeda, 1979). The activity of each articulator is assumed to be specified by a single parameter, such as jaw position, tongue-dorsum position, and so on. The analysis has resulted in an articulatory model with seven parameters. The model consists of three time varying parts: the lip opening tube, the tongue-body profile and the larynx. The region of the hard and soft palates, the velum and the rear pharyngeal walls are assumed to be fixed. Jaw position affects the states of the three parts. In addition to the jaw, the lateral tongue shapes are determined by three parameters: dorsal position, dorsal shape and apex position. There are two intrinsic parameters for the lips: height (or aperture) and protrusion. The larynx tube is specified by its intrinsic larynx height parameter and by the influence of jaw position.

In this study, we shall focus our attention on two parameters, jaw and tongue-dorsum, since articulatory-acoustic relationships have indicated that these two parameters can acoustically compensate for each other, specifically in unrounded vowels, such as /i/, /e/, and /a/ (Maeda, 1990), as discussed above.

3. Temporal patterns of articulatory parameters and their variabilities

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, therefore, by the frame-by-frame variation of the calculated articulatory parameter values. For illustration, the calculated temporal patterns of jaw and of tongue-dorsum are shown in Fig. 1. They correspond to the phrase *Une pâte à choux* uttered by subject PB. The ordinate for both jaw and tongue-dorsum is in a standardized unit, normalized by the mean and standard deviation calculated for all the utterances by that speaker. The zero corresponds to the mean value; nonzero values indicate the number of standard deviations from the mean.

Let us compare these two temporal patterns for the vowel labeled as [a] in *pâte* and in the function word *à*. These two vowels can be phonologically distinguished by the contrast back *vs.* front, /a/ for *pâte* *vs.* /a/ for *à*. In modern French however, the degree of the distinction varies from dialect to dialect and from speaker to speaker. An inspection of the corresponding spectrogram indicated that the formant values at the center of the two vowels are quite similar. We judge, therefore, that speaker PB did not distinguish between these two vowels. A large difference in the jaw position is noticed however: the jaw is clearly open during the first vowel [a] (−1.4 at the local minimum which occurs at frame number 31) while during the second vowel [a] it is at about the average position (0.0 at the frame number 40). The temporal pattern of the tongue-dorsum position is different also: the first vowel exhibits a falling pattern (i.e., a fronting gesture), while the second vowel indicates a peak (i.e., backing followed by fronting). At the center of each vowel, the dorsal position of the first vowel is somewhat less back than that of the second. The second speaker, DF, showed similar articulatory and acoustic patterns, except the

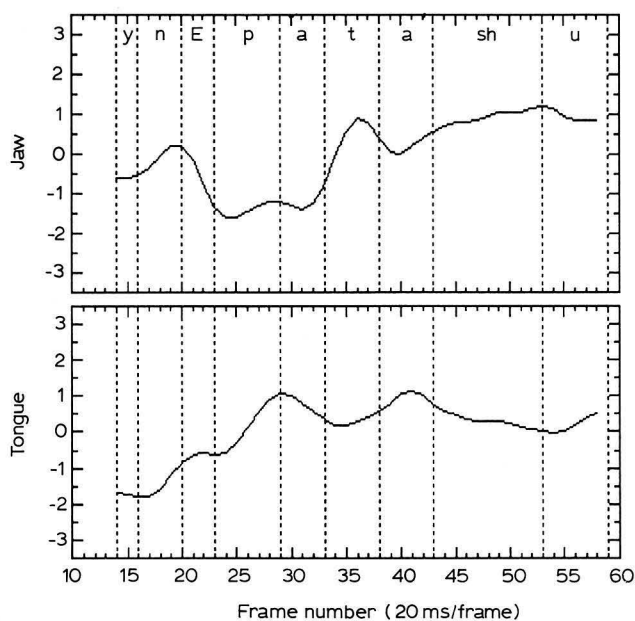


Figure 1. Temporal patterns of the two articulatory parameters, jaw (top) and tongue-dorsum (bottom), during the phrase *Une pâte à choux* uttered by subject PB. The ordinate in Figs 1, 2 and 3 has standardized units. Zero corresponds to the arithmetic mean calculated for all the utterances by each speaker. 1 (-1), 2 (-2) and 3 (-3) represent 1, 2 and 3 standard deviations, respectively, from the mean.

difference in the temporal patterns of these two articulatory parameters during the two tokens of [a] was less salient than for PB.

The data have indicated that there exists a considerable articulatory variability for the same vowel. In order to assess the range of variability, patterns corresponding to the same vowel are superimposed by aligning the curves at the vowel onset, as shown in Fig. 2 for the two speakers, PB and DF. The onset and offset of each vowel token were determined by inspecting the corresponding spectrogram. Temporal patterns of only two vowels, /i/ and /a/, are shown, for the reason stated above, and also because there were too few samples for the other vowels.

There are two remarks to make: first, the variability of individual articulators is about one third of the maximum range of the parameter observed for all utterances combined (which is almost always within plus and minus 3.0, i.e., three standard deviations either side of the mean); second, there is no clear contrast between the two extreme vowels /i/ and /a/, in particular, for the jaw position.

How is such a great magnitude of variability possible? Why is a clear contrast between the two extreme vowels /i/ and /a/ lacking? The most obvious answer is that the temporal patterns exhibited by the two elementary articulators are not independent, but rather reflect coordination between tongue and jaw. In order to make this point clearer, the corresponding trajectories of the two parameters were plotted on the jaw-tongue articulatory space. Then an “articulatory target position” is detected from each trajectory as the turning point on the trajectory. The detected

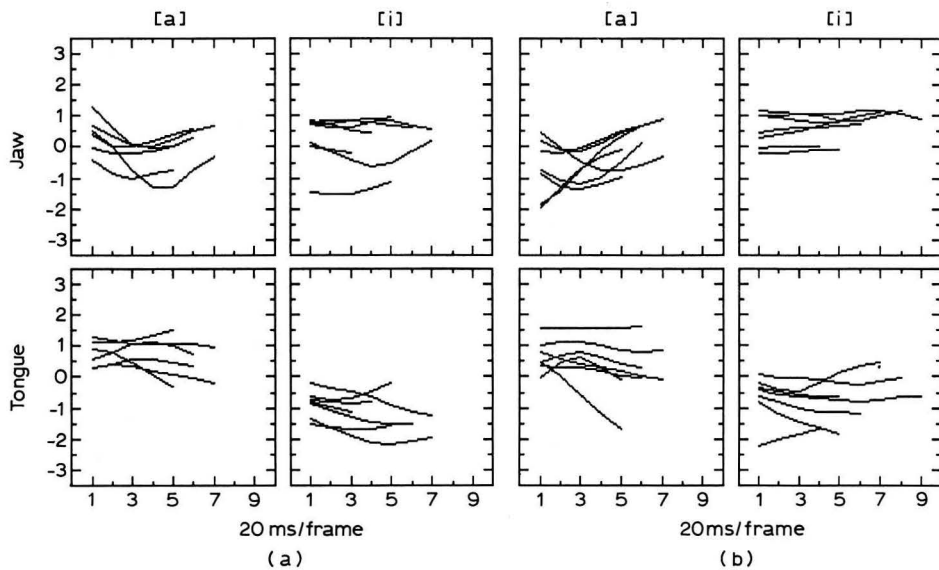


Figure 2. Superimposed temporal patterns of jaw and tongue-dorsum parameters, for the two extreme French vowels /a/ and /i/ uttered by the two speakers (a) PB and (b) DF. The patterns are extracted from different phonetic contexts and aligned at the acoustic vowel onset which corresponds to the frame number 1 in each figure.

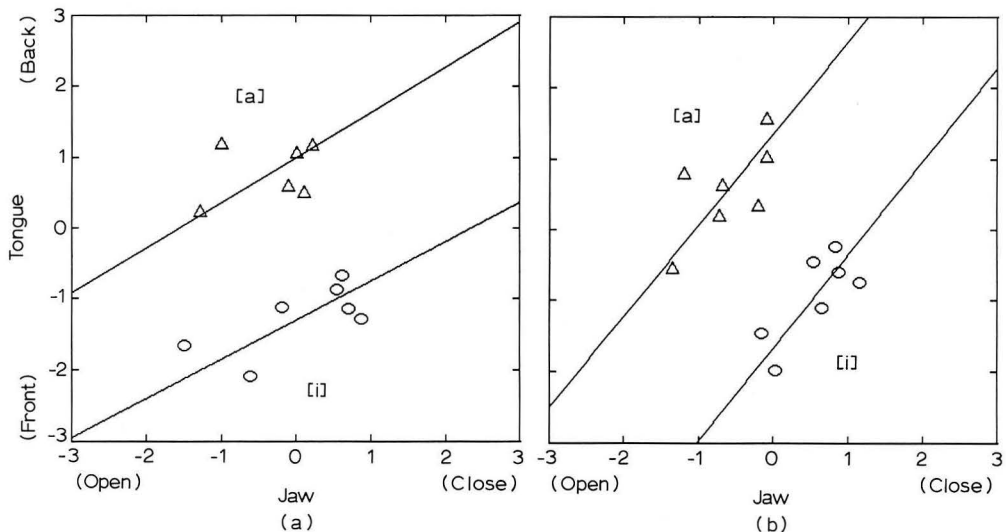


Figure 3. Scattergrams of jaw and tongue-dorsum parameters at the "articulatory targets" for /i/ (circles) and /a/ (triangles), determined from the patterns shown in Fig. 2. The straight lines indicate the first principal axis of the corresponding scattergram of each vowel. Data are for speakers (a) PB and (b) DF.

targets for /i/ and for /a/ are indicated, respectively, by the circles and the triangles in Fig. 3, for PB and for DF. In the case of speaker PB, F_1/F_2 trajectories are also used similarly to detect acoustic targets.

The straight lines plotted on Fig. 3 are determined by means of a principal component analysis (e.g., Morrison, 1976) of the scattergrams associated with each of the two vowels: they correspond to the first principal axis. Note that the data points for /i/ and /a/ are scattered without overlap. Furthermore, each cluster is distributed roughly along the straight line. These straight lines can, therefore, be regarded as linear approximations of the inter-articulatory coordination between jaw and tongue-dorsum. In other words, the observed variability can be separated into a controlled context-determined variation and an unexplained, 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 corresponds to only 12 to 35% of the observed variance.

4. Articulatory and acoustic variabilities

The articulatory variability can be examined more meaningfully if it is compared with the corresponding acoustic variability. In this study, the first and second formant frequencies were calculated using the articulatory model. We did not attempt to determine F_1 and F_2 values from the corresponding speech signal, since only spectrograms are available to us, and it is difficult to determine accurately F_1 and F_2 from the spectrogram, especially for female speakers. Even if the signals had been available, the processing would have been difficult, since the signals are contaminated by the X-ray camera noise. One other, perhaps more essential, reason why F_1 and F_2 are calculated with the model will be discussed later.

The F_1 – F_2 calculations were done only for speaker PB. Unfortunately, the lip section of the articulatory data for speaker DF is not complete. Consequently, the model for DF lacks the coefficient values specifying the lip opening tube, and thus formant frequencies cannot be calculated. As mentioned before, the model specifies vocal tract shapes with the seven parameters including jaw and tongue-dorsum positions. All seven parameter values were calculated from the corresponding data frame. The area function and then formant frequencies were computed from model-specified vocal tract shapes. The resultant F_1 – F_2 plots are shown in Fig. 4. The data points for the vowel /u/ are added to indicate the speaker's vowel space.

Comparing the articulatory target scattergrams in Fig. 3 (for speaker PB) and the corresponding acoustic ones in Fig. 4, it appears that the acoustic scattergram points are distributed more tightly than articulatory ones; i.e., acoustic variability seems to be less than articulatory one. However it is obvious that we cannot compare meaningfully these two sets of data having different physical dimensions, a dimensionless normalized unit for articulatory position and frequency for F_1 and F_2 . It is necessary to have a dimensionless common measure of variability for comparison. We propose therefore a somewhat *ad hoc* variability index, v (an averaged normalized variance), for two articulatory or two acoustic variables as follows:

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

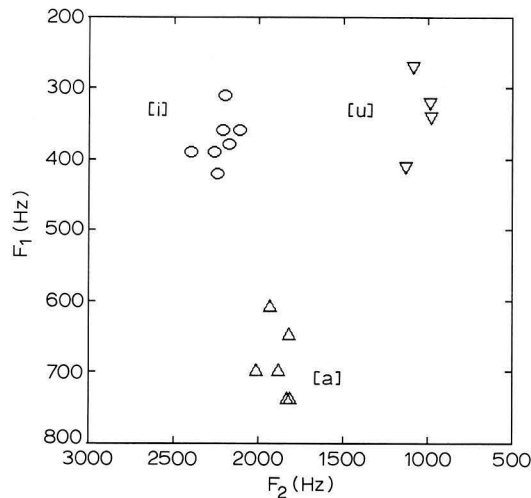


Figure 4. The first (F_1) and second (F_2) formant scattergrams corresponding to the articulatory target scattergrams shown in Fig. 3 (for speaker PB). The scattergram for the vowel /u/ is also plotted to indicate the speaker's vowel space. F_1 and F_2 values are calculated from the vocal tract configurations specified by the model with the seven parameters including jaw and tongue-dorsum positions.

where σ_i^2 is the variance of variable i ($i=1$ or 2 in our case), and $\sigma_{i_{\max}}^2$ is the maximum possible variance of variable i . (I thank Louis-Jean Boë for suggesting the normalization of each individual variance instead of normalizing the sum of variances.) Strictly speaking, the maximum articulatory variances should be calculated from the distribution of all possible articulatory target positions for that speaker. The maximum acoustic variances then could be computed from the corresponding F_1 – F_2 scattergram derived by using the model, as described just above. Since a sufficient amount of data to determine the articulatory target distribution is not available, we have assumed that the maximum standard deviations, $\sigma_{i_{\max}}$, of articulatory and acoustic data can be replaced 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 as the maximum standard deviation of 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 F_1) equals 300 Hz, and $\sigma_{2_{\max}}$ (for F_2) equals 1250 Hz. Admittedly the substitution for the maximum possible standard deviations of values of half the range is a gross approximation, but it appears to be adequate for our present purpose. In the actual calculation of the acoustic variability index, formant frequencies in hertz were converted into the Bark scale (Traunmüller, 1990), in order to enhance the perceptual importance of the variability in the first formant frequency.

The calculated index values for the articulatory (in Fig. 3) and acoustic (in Fig. 4) scattergrams for PB are shown in Table I. The index values span around 20% for the articulation and less than 10% for the acoustics. As mentioned before, the observed articulatory variability contains a component which results from the

TABLE I. Calculated articulatory (jaw/tongue-dorsum) and acoustic (F_1 – F_2) variability indices for the two speakers PB and DF. v_{total} indicates the raw variability of the articulatory target data shown in Fig. 3, and of the acoustic target data shown in Fig. 4. v_{residual} is the index after the subtraction of the influence of the coordination between jaw and tongue-dorsum from the raw variability

		PB		DF	
		/i/	/a/	/i/	/a/
Jaw/tongue	v_{total} (%)	21.7	16.2	17.0	18.4
	v_{residual} (%)	7.5	8.1	3.6	4.6
F_1 – F_2	v_{total} (%)	7.9	8.9	—	—

inter-articulatory coordination. The first principal component indicated by the straight lines in Fig. 3 was interpreted as the coordination component. The residual then can be considered the “true” variability, unexplained by coordination. The true articulatory variability indices are listed at the rows marked “ v_{residual} ” in Table I, which are calculated by weighting the variances, σ_1^2 and σ_2^2 in the above equation, with each proportion of variance corresponding to the residual. These index values are less than 10%, a value which is less than half of the corresponding raw variability, and which compares well with the index calculated for the F_1 – F_2 scattergrams of PB shown in Fig. 4. For speaker DF, the true articulatory variability is four times smaller than the observed raw variability. The calculation suggests 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: equi-line

How does the articulatory coordination effect this significant reduction of the variability in going from articulatory space to the acoustic space? In previous studies (Maeda, 1989, 1990), we have already shown that in the case of unrounded vowels such as /i/ and /a/, jaw and tongue-dorsum positions can acoustically compensate for each other. In the case of rounded vowels, lip-rounding compensates for a deviation in jaw position. Compensation means that a deviation in the position of one articulator can be counteracted by a readjustment of other articulator(s) to minimize the deviation in the acoustic pattern. It is reasonable, then, to hypothesize that the inter-articulatory coordination, in fact, results in an acoustic compensation of the type just described. If this is the case, the principal axis representing the coordination in Fig. 3, is also an acoustic “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 /i/ and /a/, F_1 and F_2 values were calculated at seven different jaw positions from -3.0 to 3.0 with a step size of 1.0 . The corresponding tongue positions are determined by the linear relationship. 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

TABLE II. Calculated acoustic (F_1 – F_2) variability indices when the parameter values of jaw and tongue-dorsum positions are systematically varied along the principal axis of each of the two vowels, shown in Fig. 3 for speaker PB. In full range, the jaw parameter value is varied from -3.0 (very low) to 3.0 (very high) with a step size of 1.0 . In sub-range, the jaw value is varied within a more realistic range for each of the two vowels, as indicated in the table

	Full range (-3.0 to 3.0)	Sub-range	
/i/	3.2%	(-2.0 to 1.0)	1.1%
/a/	6.7%	(-3.0 to 0.0)	3.2%

the data frame corresponding to the target for /i/ or for /a/. If index values calculated in this way are significantly smaller than those from the data, listed in Table I, then the principal axis can be considered an equi-line. This comparison is particularly valid, since in both cases the F_1 – F_2 scattergrams and, from them, the indices, are calculated using the same articulatory model. In fact, this is one of the reasons why calculation with the model, instead of measurement on the original speech signals, was used to derive the F_1 – F_2 scattergrams.

The calculated indices of F_1 – F_2 variability of the two vowels are listed in Table II in the column labeled “full range”. For the vowel /i/, the index value (3.2%) is much smaller than the corresponding one (7.9%) in Table I. The index value of the equi-line associated with the vowel /a/ (6.7%), however, is only slightly smaller than the corresponding index of the observed scatter (8.9%). The F_1 – F_2 scattergram of the equi-line exhibits a small dispersion on the F_1 axis, but a relatively large dispersion on the F_2 axis, which has a strong influence on the index value even on the Bark scale.

This unexpected result for /a/ is partly due to the fact that the full range of jaw position is not realistic. For this reason, we have recalculated the index for a more realistic range. The results are listed in the second column, entitled “sub-range”, in Table II. Now the index for the equi-line of /a/ becomes 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 F_1 – F_2 pattern.

Although the acoustic compensation along the equi-lines is not perfect, it is safe to state that articulatory maneuvers 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 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.

6. Concluding remarks

It has become clear that the apparently large variability of the individual articulator positions during the same vowel in different contexts can be explained, at least in

part, by 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 (Lindblom *et al.*, 1979). Surprisingly, the coordination and thus the compensation can be specified directly in terms of articulatory parameters. In this study the coordination during the production of the two vowels is specified by the linear relationships between the two articulatory parameters, jaw and tongue-dorsum. The implication of this is important. If the relationship is well defined in such a simple fashion, then 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.

In many previous studies of compensatory articulation or of speech motor control in general (for example, Gay *et al.*, 1981; Saltzman & Munhall, 1989), goals of articulatory actions were specified in terms of the vocal tract area function, or in terms of the constriction location along the vocal tract and its degree. It is obvious, as shown in our study, that such goals cannot be directly specified in terms of the individual articulatory positions, which vary significantly for the same vowel. It then becomes necessary to introduce a mechanism for controlling individual articulator movements, in order to achieve a target area function. Often feedback mechanisms are proposed to explain the speaker's ability to realize the targets. However, even reflex sensory feedback, not to mention slower auditory feedback, may not be fast enough to account for the observed human performance, as shown in bite-block experiments (Lubker, 1979). In order to explain the immediate readjustment during the production of bite-block vowels, Lindblom *et al.* (1979) have suggested a mental simulation of articulatory gestures to control the real articulators optimally. In this case, the feedback control occurs in the mental simulation.

Our finding here suggests that although fixed targets cannot be specified in articulatory positions, the required articulatory condition for achieving vowel targets can be specified by means of the linear relationships between the two articulators. It is not necessary then to describe the targets in terms of the area function, since the articulatory positions and acoustic pattern can be related in a direct and computable way. Speech production could operate with an open-loop control mode and the articulatory motions could be purely "ballistic" without any feedback. Is that really what is happening in speech production? Unfortunately, there is no definite answer to this question, since our findings are based strictly on observations of vocal tract shapes. Our assertion is a phenomenological one at best. It is safe only to state that the time-varying shape of the vocal tract during speech production can be effectively modeled by an open-loop system. We are not suggesting, here, a model such as coordinative structures among the articulators defined by "equations of constraint", which is built into the peripheral system as a vibratory "spring-mass" system (Fowler, Rubin, Remez & Turvey, 1978). One crucial factor that must be taken into account in formulating a motor command scheme is anticipatory articulation. Motor programming probably requires an intricate calculation involving both anticipation and coordination at the same time. Indeed, this is the reason why we favor the idea that compensatory articulation is calculated at the motor programming level, not built into the peripheral system itself.

Obviously, a larger body of articulatory data is needed to confirm compensatory

articulation during vowels. It is necessary to explore the inter-articulatory coordination of articulators other than the jaw and tongue-dorsum and to extend the investigation to different vowels and consonants. It is quite possible that more than two articulators coordinate. In that case, the coordination and then the compensation should be specified by an equi-surface, instead of the equi-line described here for the two-articulator coordination. Much more work on compensatory articulation as well as on anticipatory articulation is needed in order to formulate a command scheme for transforming a discrete speech string into continuous articulatory gestures.

I would like to thank Péla Simon, André Bothorel, Jean-Pierre Zerling, and François Wioland at l'Institut de Phonétique de Strasbourg for providing us with the digitized X-ray data. I would also like to thank Celia Scully for helpful comments on the manuscript.

References

- Bothorel, A., Simon, P., Wioland, F. & Zerling, J.-P. (1986) *Cinéradiographie des voyelles et consonnes du français*. Travaux de l'Institut de Phonétique de Strasbourg.
- Fowler, C. A. & Turvey, M. (1980) Immediate compensation for bite-block speech, *Phonetica*, **37**, 306–326.
- Fowler, C. A., Rubin, P., Remez, R. E. & Turvey, M. T. (1978) Implications for speech production of a general theory of action. In *Language production* (Butterworth, editor), New York: Academic Press.
- Gay, T., Lindblom, B. & Lubker, J. (1981) Production of bite-block vowels: Acoustic equivalence by selective compensation, *Journal of Acoustical Society of America*, **69**(3), 802–810.
- Harshman, R., Ladefoged, P. & Goldstein, L. (1977) Factor analysis of tongue shapes, *Journal of Acoustical Society of America*, **62**(3), 693–707.
- Jackson, M. T. T. (1988) Analysis of tongue position: Language-specific and cross-linguistic models, *Journal of Acoustical Society of America*, **84**(1), 124–143.
- Lindblom, B. E. F. & Sundberg, J. E. F. (1971) Acoustical consequences of lip, tongue, jaw and larynx movement, *Journal of Acoustical Society of America*, **50**(4, Part 2), 1166–1179.
- Lindblom, B., Lubker, J. & Gay, T. (1979) Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation, *Journal of Phonetics*, **7**, 147–161.
- Lubker, J. (1979) The reorganization times of bite-block vowels, *Phonetica*, **36**, 273–293.
- Maeda, S. (1978) Une analyse statistique sur les positions de la langue: Etude préliminaire sur les voyelles françaises. In *9èmes JEP, GALF*, pp. 191–199.
- Maeda, S. (1979) Un modèle articuloire de la langue avec des composantes linéaires. In *10èmes JEP, GALF*, pp. 152–164.
- Maeda, S. (1989) Articulation compensatoire de voyelles: analyse de données cinéradiographiques avec un modèle linéaire. In *Mélanges de phonétique générale et expérimentale* (A. Bothorel, J. C. Galdin, F. Wioland & J. P. Zerling, editors), Vol. 2, pp. 544–562. Publications de l'Institut de Phonétique de Strasbourg.
- 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.
- Majid, R., Boë, L.-J. & Perrier, P. (1986) Fonctions de sensibilité, modèle articuloire et voyelles du français. In *15èmes JEP, GALF*, pp. 59–63.
- Morrison, D. F. (1976) *Multivariate statistical method*, 2nd edn. New York: McGraw-Hill.
- Saltzman, E. L. & Munhall, K. G. (1989) A dynamical approach to gestural patterning in speech production, *Ecological Psychology*, **1**(4), 333–382. (Also in *Haskins Laboratories Status Report on Speech Research*, (1990) SR-99/100, 38–68.)
- Shirai, K. & Honda, M. (1976) An articulatory model and the estimation of articulatory parameters by nonlinear regression method, *Transactions IECE Japan*, **J59-A**(8), 668–674. (In Japanese.)
- Traunmüller, H. (1990) Analytic expressions for the tonotopic sensory scale, *Journal of the Acoustical Society of America*, **80**(1), 97–100.