Regenerating the spectral shapes of [s] and [ʃ] from a limited set of articulatory parameters

Noël Nguyen[a)]
Laboratoire Parole et Langage, Université de Provence, 29, av. R. Schuman, 13621 Aix-en-Provence, France

Philip Hoole
Institut für Phonetik und Sprachliche Kommunikation, Ludwig Maximilians-Universität, Schellingstrasse 3, 8000 München 40, Germany

Alain Marchal
Laboratoire Parole et Langage, Université de Provence, 29, av. R. Schuman, 13621 Aix-en-Provence, France

(Received 21 December 1992; accepted for publication 16 March 1994)

This work was aimed at exploring articulatory-acoustic relationships in the production of French fricatives. More precisely, an attempt was made to find out whether the spectral shapes of [s] and [ʃ] can be regenerated from the x and y coordinates of three electromagnetic transducers affixed to the tongue in the midsagittal plane. The corpus was composed of the two fricatives /s/ and /ʃ/ combined with the vowels /a/ and /i/ in sequences of the type /VsʃV/ and /VʃsV/, and was read by one male native speaker of French. The spectrum regeneration was based on a statistical procedure which consisted of estimating the factors explaining the main part of the acoustic variance from the position of the transducers, by means of multiple linear regression. The articulatory-acoustic correlations were high and allowed us to regenerate the fricative spectra with a good accuracy. The way in which the acoustic parameters varied as a function of the articulatory ones in the statistical model was in good agreement with data reported in previous works. The results support the idea that the tongue has relatively few degrees of freedom in the production of [s] and [ʃ].

PACS numbers: 43.70.Aj, 43.70.Jt

INTRODUCTION

Numerous articulatory models are based on the assumption that tongue movements in speech production are characterized by a small number of degrees of freedom, and are controlled by a limited number of parameters. For Liljencrants (1971), the sagittal contour of the tongue is comparable to a pseudoperiodic function whose first three components of a Fourier series suffice to obtain a good approximation. Harshman *et al.* (1977) and Maeda (1990) demonstrated that the sagittal shape of the vocal tract can be schematically represented as a linear combination of a small number of factors obtained by carrying out a factor analysis on cineradiographic data. Jackson (1988) also assumed that articulatory movements are the outcome of a linear transformation applied to a limited set of factors. In a microbeam study, Lindau-Webb and Ladefoged (1989) showed that the x and y coordinates of two pellets located along the tongue midline may suffice to reconstruct the entire contour of the tongue by means of a multiple linear regression. This last work demonstrated that the movements of the various parts of the tongue in the sagittal plane are highly interdependent and, very interestingly, that a small number of spatial measures, taken solely in the anterior cavity, may provide information about the entire vocal tract shape.

It might also be possible to indirectly estimate the number of parameters of tongue control in speech, by attempting to regenerate the spectrum of the acoustic signal from a given set of articulatory parameters. By "regeneration" we more specifically mean that the acoustic spectrum is reconstructed on the basis of articulatory-acoustic statistical correlations, through the use of a statistical method such as multiple regression. If such an operation was proven to be feasible, it would suggest that the spectral shape of the acoustic signal is controlled by a number of articulatory variables which is not significantly higher than the number of parameters used in the statistical model. It is a truism to say that statistical estimations of this kind do not automatically lead to conclusions which can be applied generally. Attention must be paid to the fact that a statistical correlation might be no more than a simple coincidence between two variables that are actually independent from each other. Highly accurate estimations have to be considered with care, since they are likely to be explained in part by speaker-dependent characteristics. In some cases, the correlations between the articulatory parameters and the acoustic ones may be artificially strengthened, depending on the way in which the corpus has been constructed. However, once all the necessary precautions are taken, spectral resynthesis via regression would have the advantage of being very easy to implement since it does not require the determination of the sagittal shape of the entire vocal tract by means of an invasive and time-consuming technique like radiography, nor the compu-

a)Present address: Laboratoire de Psycholinguistique, Faculté de Psychologie, Université de Genève, 9 route de Drize, CH-1227 Genève, Switzerland.

 0001-4966/94/96(1)/33/7/$6.00

tation of the vocal tract area and transfer functions. The present study was aimed at finding out whether it is possible to regenerate the spectral shape of the French fricatives [s] and [ʃ] from articulatory data collected by electromagnetometry.

I. MATERIAL AND PRELIMINARY DATA ANALYSIS

The experiment consisted of an audio recording synchronized with a simultaneous electromagnetic (EMA) recording. The electromagnetic system used was an Articulograph AG100, marketed by Carstens Medizinelektronik (Göttingen, Germany). This system offers the possibility of monitoring articulatory movements in speech production by means of small electromagnetic receivers attached to the articulators in the midsagittal plane. We refer the reader to Tuller *et al.* (1990), Honda and Kaburagi (1993), and Hoole (1993) for an assessment of the Articulograph and to Perkell *et al.* (1992) for an extensive overview of EMA. In our experiment, three coils were attached to the midline of the tongue, one coil was attached to the lower central incisors, and one reference coil to the upper central incisors. The frontmost tongue coil was placed approximately 1 cm away from the tip of the tongue, the rearmost coil about 3.5 cm away from the same point and the mid coil in between. The articulatory data were LP filtered (cut-off frequency 40 Hz), A/D converted (sampling rate 250 Hz), and stored on a PC AT 386 for further processing. The audio signal was recorded by means of a B&K microphone on a DAT recorder, digitized on a laboratory computer (sampling rate 16 000 Hz, LP filtered at 7500 Hz), and aligned with the EMA signal by means of a series of synchronization pulses recorded on the second track of the DAT tape.

The corpus was composed of the two fricatives /s/ and /ʃ/ adjacent to each other in all possible sequences of the type /VsʃV/ and /VʃsV/, where V_1 and V_2 were /a/ or /i/. There were eight VCCV sequences, which were embedded in carrier sentences like the followings: "Toi tu tr*iches si* j'veux" ("You can cheat if I want"); "C'est une cl*asse cha*rgée" ("It's a large class"); "Tu te ca*ches sous* l'lit" ("You hide under the bed"); etc. This corpus was designed to minimize differences in the stress patterns and locations of word boundaries across items. Each sentence was repeated five times by one male native speaker of French (28; no known speech or hearing disorders).

Possible measurement errors due to slight translational movements of the head with respect to the helmet of the Articulograph were minimized by subtracting the coordinates of the reference coil from those of the other coils. In addition, the articulatory data were for each repetition rotated within the *x*-*y* plane so as to orient the first principal axis of the jaw movements in a vertical direction. Finally and for each item, the V_1–C_1 and C_2–V_2 boundaries were located on the acoustic signal using a signal editor, and the measured *x* and *y* coordinates of the tongue-back, tongue-mid, and tongue-tip coils over the entire consonant cluster were extracted and stored in a separate file. These coordinates are shown in Fig. 1. A 3-sigma ellipse[1] has been drawn around the points representing each coil. A rough outline of the upper incisors, the alveolar ridge and the hard palate in the midsagittal plane is also presented.[2]

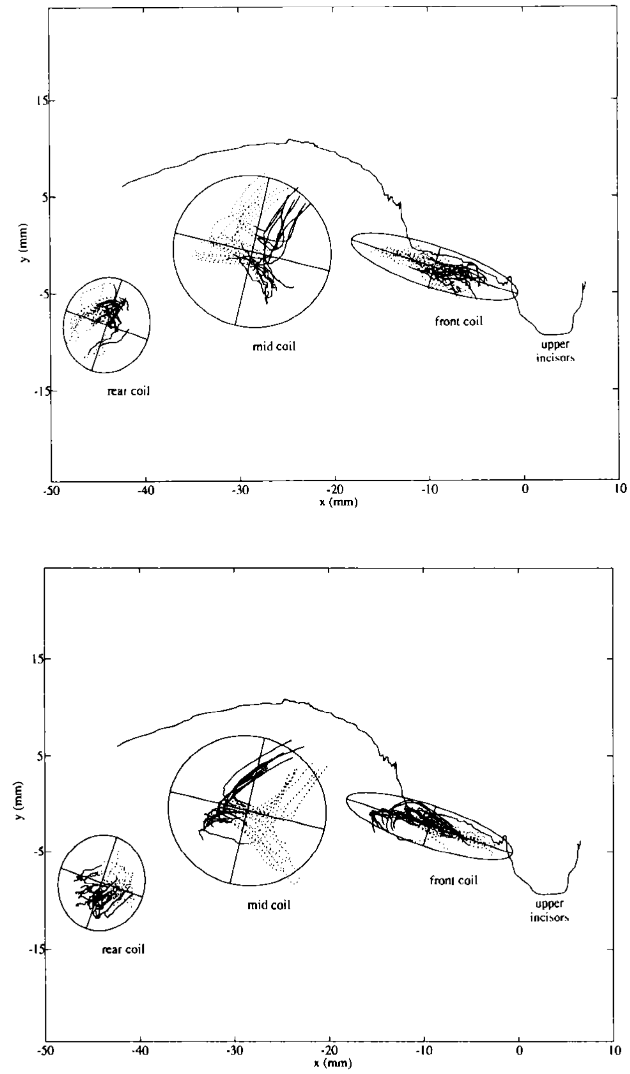


FIG. 1. The *x* and *y* coordinates of the three tongue coils over the entire consonant cluster for each VCCV item and each repetition. A 3-sigma ellipse has been drawn around the points representing each coil. The measurements are given with respect to the reference coil attached to the upper incisors. Anterior is to the right. Upper panel: [sʃ]; lower panel: [ʃs]. The time dimension has been roughly represented by using a dotted line to draw the first half of a coil's trajectory, and a solid line for the second half.

Figure 1 shows that the constriction was probably situated in the vicinity of the tongue-tip receiver coil for both [s] and [ʃ]. Important vowel-dependent variations behind the constriction point also have to be mentioned, especially for the tongue-mid coil which was substantially displaced toward the hard palate in the neighborhood of [i] (see Hoole *et al.*, 1989). Contrary to observations made by others for English (e.g., Perkell *et al.*, 1979), [s]-to-[ʃ] assimilations in [sʃ] clusters did not seem to occur in our corpus.

The acoustic data consisted of a series of 256-point DFT spectra lined up on the time scale with the articulatory measurements, for each item. A 32-ms Hamming window was used for the spectral analysis. The 256 components pertaining to each spectrum were normalized so that their mean was 1, and were next reduced to 21 components by averaging the spectral energy over 1-Bark intervals from 0 to 8 kHz. Figure

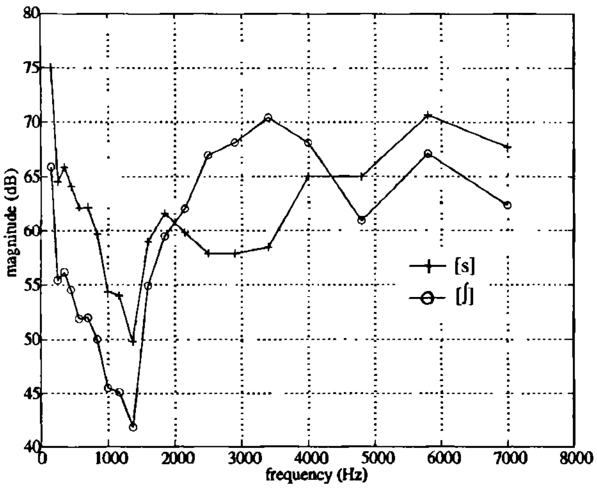


FIG. 2. Average spectral shape of [s] and [ʃ] in the stable part of each of the two fricatives.

2 represents the average spectral shapes of [s] and [ʃ] in the stable part of each of these two fricatives (i.e., 50 ms after the offset of V_1 or before the onset of V_2, depending on the position of the fricative in the C_1C_2 cluster).

The alveolar fricative is characterized by a main spectral peak in the vicinity of 5800 Hz, and by another less important peak below 2000 Hz, which may be related to the second formant of the adjacent vowel (see Soli, 1981). The postalveolar fricative is characterized by the predominance of a centrally located spectral proeminence (around 3400 Hz), while a second peak is also noticeable around 5800 Hz. There is relatively more energy in the lowest part of the spectrum for [s] than for [ʃ]. The data below 8 Bark (center frequency: 840 Hz) were ignored in the following stages, since variations which are irrelevant to our study may occur in that frequency range due to voicing assimilation.

The final stage in the reduction of the acoustic data consisted of a principal-components analysis. The advantage of such an analysis resides in the fact that it provides us with a small number of parameters—the factors—which account for the most important variations in the spectra, and which can be more easily related to articulatory parameters than the spectral components themselves (see Hoole *et al.*, 1989, for a similar approach). The analysis was applied to a matrix composed of 2937 14-dimensional spectra. All together, the first three factors explained a fairly high percentage of the acoustic variance (fact. 1: 41%; fact. 2: 23.9%; fact. 3: 9%; total: 73.9%). The first factor is plotted against the second one in Fig. 3.

Two-sigma ellipses are used to approximately delimit the [s]-region and [ʃ]-region in the factor plane. Each ellipse was constructed on the basis of 40 points representing the values of factor 1 and 2 in the stable part of each fricative, for each VCCV item and each repetition. It clearly appears that the main variations in our acoustic data depend on the consonant category. Figure 4 depicts the first three eigenvectors of the correlation matrix calculated in the factor analysis.

The first eigenvector is characterized by two zero crossings indicating that factor 1 differentiates the spectra which have more energy between 2 and 4 kHz, from those for

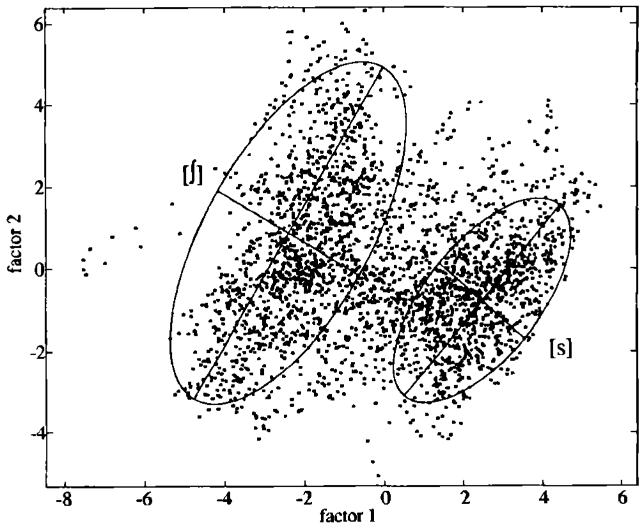


FIG. 3. Acoustic data displayed as points in the plane of the 2 first factors ($n=2937$). Two-sigma ellipses have been drawn around the points representing the values of factors 1 and 2 in the stable part of each fricative, for each VCCV item and each repetition.

which the energy is higher outside this central region. The second eigenvector has one zero crossing which shows that factor 2 is an indication of whether the energy level is on average higher below or above 4 kHz. Thus the first two factors together account for the main differences between the spectral shapes of [s] and [ʃ]. The third eigenvector reflects variations of the energy in more specific parts of the spectrum.

II. SPECTRUM REGENERATION METHOD AND RESULTS

Spectrum regeneration was carried out in two steps. First, the acoustic factors were estimated from the articulatory parameters, using a multiple linear regression (hereafter L_{Reg}). Second, the spectra were themselves regenerated from the fitted values of the factors, by means of another linear transformation.

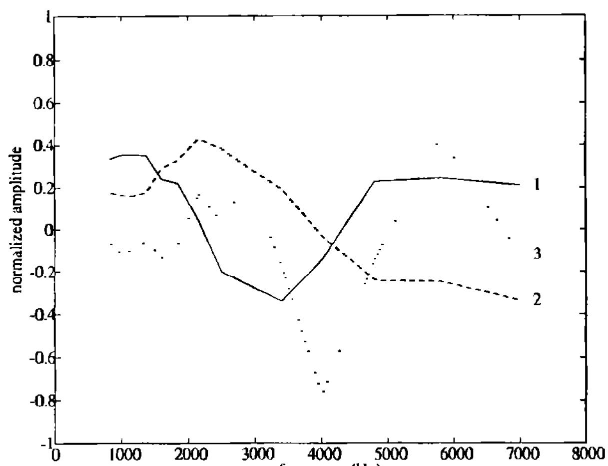


FIG. 4. Eigenvectors 1, 2, and 3 of the correlation matrix calculated in the factor analysis.

TABLE I. Correlations (L_{Reg}: multiple linear regression; E_{Reg}; empirical regression) between measured and fitted values for acoustic factors 1, 2, and 3. For L_{Reg}, all correlations are significant beyond the 0.0001 level.

	n	Factor 1	Factor 2	Factor 3
E_{Reg}	2937	0.982	0.957	0.956
L_{Reg}	2937	0.859	0.386	0.234

In L_{Reg}, the parameters which served as independent variables were the x and y coordinates of the tongue-back, tongue-mid, and tongue-tip coils, while the dependent variables were constituted by the first three acoustic factors. The regression model had the following general form:

$$f'_{ik}=\left(\sum_{j=1}^{3} x_{ij}a_{jk}+\sum_{j=1}^{3} y_{ij}b_{jk}\right)+c_k,$$

$$i=1,...,n,\quad k=1,2,3, \tag{1}$$

where, for each sample i and each factor k, f'_{ik} is the fitted value of the factor, x_{ij} and y_{ij} are the coordinates of the coil j, a_{jk} and b_{jk} the corresponding regression coefficients, and c_k is the intercept. The number of samples n was equal to 2937.

In addition to using L_{Reg}, we determined the results with another regression technique called empirical regression (hereafter E_{Reg}; see Cazes, 1976). In this last case, the estimated value of a given acoustic parameter f for a given tongue "profile" (composed of only three points) was simply defined as the mean f value associated with the profile's k-nearest neighbors in the articulatory space. Such a series of local approximations seems to be suitable for modeling nonlinear relationships between any number of independent variables and the to-be-estimated variable (see discussion below). In this experiment, each articulatory neighborhood was composed of five tongue profiles. A regression summary is presented in Table I.

Table I shows that the accuracy of the estimations based on E_{Reg} was very satisfying for the three factors. The correlation coefficient between the measured and fitted values was higher than 0.95 in all three cases. L_{Reg} also yielded a high r for factor 1. The scores were lower for factors 2 and 3, but still beyond the 0.0001 significance level. Thus there was a close relationship between the EMA parameters and the dimensions of maximal acoustic variance. Measuring the x and y coordinates of three points located on the anterior part of the tongue seems to be sufficient to determine, with good accuracy, the position of the fricative in a factor plane accounting for the main part of the data spread in the original Bark space.

Finally, the fricative spectra were regenerated from the output parameters of L_{Reg}, in the following way (see Zahorian and Rothenberg, 1981):

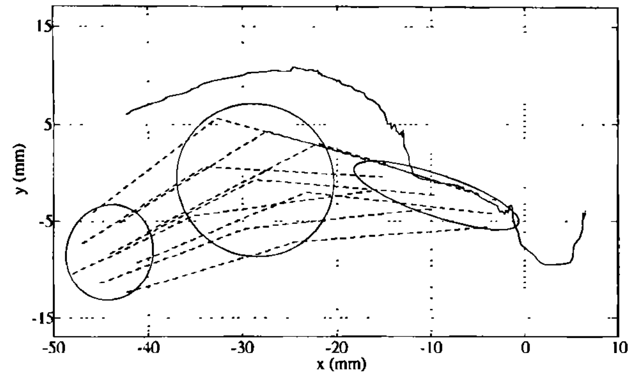
$$S'_{ij}=\left(\sigma_j\sum_{k=1}^{3} f'_{ik}u_{jk}\right)+\bar{S}_j,\quad i=1,...,n,\ j=1,...,p, \tag{2}$$

where, for each sample i, S'_{ij} is the estimated amplitude of the spectral component S_j, f'_{ik} is the estimated value of factor k, u_{jk} is the jth coordinate of the kth eigenvector of the correlation matrix in the factor analysis, $\bar{S}_j$ and σ_j are the mean and the standard deviation of S_j, respectively. The number of spectral components p was equal to 14.

The accuracy of this model was assessed by calculating the correlation coefficient between the measured and fitted spectral components, for each spectrum separately. This coefficient turned out to be on average very high (mean value: 0.906, σ=0.08, n=2937), and stable over the whole length of each C_1C_2 cluster, regardless of the surrounding vowels and of the respective positions of [s] and [ʃ] in the cluster. Thus the regeneration of the fricative spectra can be considered as being quite accurate.

If we now return to the linear regression, it is interesting to examine more closely the nature of the relationships that emerged from the analysis between the coil displacements and the acoustic factors. In theory, the highest regression coefficients should be associated to the articulatory parameters which were the most strongly related to the acoustic changes. But as we know, it is rather difficult to interpret the regression coefficients in such terms when there are high correlations between the independent variables, as was the case in our experiment (since the coils obviously cannot move completely independently of each other). Because of this phenomenon of collinearity, the importance of each independent variable is highly influenced by the presence in the model of the other ones (Berry and Feldman, 1985). Therefore, instead of examining each regression coefficient separately, we simulated a displacement of the three tongue coils at the same time in the midsagittal plane, so as to characterize the way in which the acoustic factors were varying as a function of this displacement in our model. The direction and length of displacement were determined with respect to a coordinate system which was specific to each coil, and was defined by the major and minor axes of the corresponding 3-sigma ellipse.[3] Each coil was moved from -2 to 2 standard deviations in three steps $(-2,0,2)$ along each of these axes. At each step, we calculated the values of factor 1 and 2 corresponding to the tongue "profile" roughly schematized by the three coils. There were nine different tongue profiles, and thus nine calculated values for each factor. The results are presented in Fig. 5.

This figure first shows that a forward displacement of the three coils provokes a significant increase in factor 1 combined with a more limited decrease in factor 2. In other words, this movement is associated with acoustical changes basically consisting in passing from the [ʃ] region to the [s] region in the factor space (see Fig. 3). Such a result is consistent with the acoustical theory of fricative production (Fant, 1960; Heinz and Stevens, 1961), according to which a shortening of the anterior cavity causes a shift of the main spectral peak toward higher frequencies. Figure 5 also indicates that an upward movement of the coils is correlated with a moderate decrease in factor 1. Note that in our simulation, this movement mainly consisted in a raising of the tongue-mid coil, since the dispersion of the EMA data in the vertical direction was significantly higher for this coil than for the



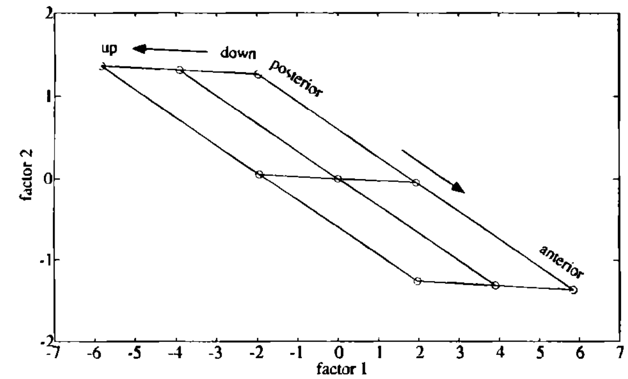


FIG. 5. Variations of the two first acoustic factors induced by a forward and upward displacement of the three tongue coils in the statistical model. Each coil was moved from -2 to 2 standard deviations in three steps $(-2,0,2)$ along each of the two axes of the corresponding 3-sigma ellipse. Upper panel: nine resulting tongue "profiles"; lower panel: values of factors 1 and 2 associated with each profile.

other two. Detailed examination showed that the tongue-mid coil was on average closer to the palate for [ʃ] than for [s] in the stable part of the consonant (possibly because the anterior part of the tongue is ungrooved for the postalveolar, see Stone, 1991), and in the neighborhood of [i]. At the acoustic level, an ANOVA revealed that factor 1 was sensitive to the vocalic context, and more precisely had a lower value for [s] when the preceding vowel was an [i] [interaction consonant $\times$vowel 1: F $(1,64)=8.95$, $p<0.005$]. Thus a raising of the tongue behind the point of constriction, either due to the consonant category or to the influence of the adjacent vowel, is associated with a displacement toward the [ʃ] region in the acoustic factor space. These results confirm previous observations based on electropalatographic data by Hoole *et al.* (1989).

III. DISCUSSION

The accuracy of our statistical model may be partly due to the fact that the speaker in this experiment pronounced [s] and [ʃ] in a rather stable way across repetitions and across contexts. Consequently, the variability exhibited by the spectra could be accounted for by a small number of factors, which in turn made it possible to regenerate the spectra without any major distortions. Nevertheless, the factors themselves proved to be accurately predictable from the EMA data. This means that the articulatory parameters extracted by EMA were closely related to the dimensions of maximal acoustic variance.

Our results are in a good agreement with those that we already have obtained with another subject in a similar experiment (Nguyen *et al.*, 1991). In that previous work however, [s] and [ʃ] were not combined with each other but produced in sequences of the (V)C(V) type. Moreover, articulatory and acoustic data were only gathered at the midpoint of each fricative. As a result, the data divided into two separate subsets (corresponding to the two fricatives) in the articulatory space as well as in the acoustic space. This data-clustering phenomenon may well have artificially strengthened the articulatory-acoustic correlations (Hoole *et al.*, 1989). In the present experiment, an attempt was made to remove this possible artifact by using [sʃ] and [ʃs] sequences. It was assumed that sequences of this kind would constrain the speaker to map out (as far as possible) the entire articulatory region potentially involved in the production of the two fricatives (see Perkell *et al.*, 1979, for a similar approach). A close examination revealed that the data were as expected spread out over a continuum in the articulatory space (see Fig. 1) and in the acoustic factor space. Our results demonstrated that it was again possible using these data to resynthesize the fricative spectra with a high precision, from the x and y coordinates of three tongue coils.

Interestingly, E_{Reg} allowed us to make more precise estimation than L_{Reg}. The disadvantage of E_{Reg} is that it does not offer the possibility of formalizing articulatory-acoustic relationships in the form of algebraic equations. However, E_{Reg} does not involve any presupposition about the *overall* nature of these relationships, while L_{Reg} obviously relies on the postulate that they are linear (to our knowledge, there is still no statistical method for automatically approximating a dependent variable from *many* independent variables by means of a nonlinear equation). Thus the better results obtained by using E_{Reg} may be taken as an indication that the relations between the EMA parameters and the acoustic factors are not really linear. This in turn leads to the possibility that our results might be consistent with the quantal theory of speech production, according to which an [s]-to-[ʃ] (or [ʃ]-to-[s]) gesture is accompanied by an abrupt, "quantal" change in the spectral characteristics of the acoustic output (Perkell *et al.*, 1979; Stevens, 1989). In an [sʃ] sequence for example, the presumably continuous backward movement of the tongue would cause a discontinuous change in the configuration of the anterior cavity (creation of a sublingual space), which would in turn induce an abrupt downward shift of the lowest major spectral peak. In a companion experiment (Nguyen and Hoole, 1993) based on a corpus similar to the one used in the present work, we indeed showed that there was a nonlinear relationship between the position of the constriction point and the lowest peak frequency, when the surrounding vowels were unrounded.

However, attention should be paid to the fact that an acoustic factor can hardly be identified with a measure of the frequency of a single spectral peak. There is as much difference between these two parameters as between a mean[4] and a mode in descriptive statistics for example. Therefore, it is quite possible that a quantal "jump" in the frequency of the lowest main peak was accompanied by a smooth variation of the acoustic factors, just as when a sudden change in the relative height of the two modes of a bimodal distribution is accompanied by a gradual shift of the mean. In this connection, it should be noted that the percentage of acoustic variance (R^2) accounted for by the EMA parameters in L_{Reg} can be considered as being very significant for factor 1 [$F(6,2930)=1370.31$, $p<0.0001$], factor 2 [$F(6,2930)=85.51$, $p<0.0001$], and factor 3 [$F(6,2930)=33.7$, $p<0.0001$]. In addition, it appeared that the estimations based on E_{Reg} degraded rather *smoothly* as the size of the articulatory neighborhood was increased, for all the factors. This leads one to suppose that these factors varied at a relatively slow rate as a function of the displacements of the tongue coils.

It is interesting to note that the fricative spectra were accurately regenerated here in the absence of any information on lip protrusion, a parameter which may play an important role in establishing the distinction between [s] and [ʃ] (e.g., see Faber, 1989). This probably results from a statistical correlation between the degree of lip protrusion and the position of the tongue in our data (since the postalveolar fricative was presumably produced with a rounding of the lips, contrarily to the alveolar for which the lips were unrounded). Because of the existence of this statistical correlation, a lowering of the main spectral prominence due to lip rounding was predictable on the sole basis of the EMA measurements. This hypothesis is all the more plausible since the corpus did not contain any rounded vowels, which might have eliminated the correlation between lips and tongue movements by inducing lip rounding in the production of [s].

In conclusion, the EMA measurements proved to be closely related to the acoustic spectra of [s] and [ʃ] produced by our speaker. This high correlation allowed us to accurately regenerate the fricative spectra from the EMA data by means of a regression. These results are quite encouraging, especially because they concern phonemes whose production requires a fairly precise and complex tongue configuration (see, for example, Stone, 1991). Despite this geometric complexity, our data support the idea that the tongue has relatively few degrees of freedom and is governed by a small number of underlying parameters in the production of [s] and [ʃ]. It must also be mentioned that the acoustic correlates of a fricative mainly depend on the portion of the vocal tract between the point of maximum constriction and the lips (Hughes and Halle, 1956). This was empirically confirmed by Hoole *et al.*'s experiment (1989), in which high correlations were found between [s] and [ʃ] spectra on the one hand, and their electropalatographic patterns on the other hand. From that point of view, it might turn out to be more difficult to reconstruct the spectrum of a vowel from EMA data, since it would possibly require attaching the posterior coil further back on the tongue than in our experiment.

ACKNOWLEDGMENTS

This work was supported by the European ESPRIT-ACCOR project. We are grateful to M. T. T. Jackson, R. N. Ohde, M. A. Svirsky and an anonymous reviewer for helpful comments on an earlier version of the paper.

[1]The use of 3-sigma ellipses instead of more usual 2-sigma ellipses was only aimed at making Fig. 1 clearer.

[2]The experimenter traced the subject's palate contour by moving once, along the palate midline, a coil fastened to the end of his forefinger. Owing to probable rotational movements of the coil in this process, Fig. 1 very likely gives a crude approximation of the location and shape of the palate in the midsagittal plane.

[3]Such a coordinate system can be considered as being more articulatorily relevant than the original coil-independent one, since it is possible to assume that one of the two axes of the ellipse will be parallel to the vocal-tract midline and the other one perpendicular to that midline, in the region where the coil is located (see Perkell and Cohen, 1989, for a similar approach).

[4]Note that a spectral mean is defined as a linear combination of the various components of a spectrum, just as an acoustic factor in this work.

Berry, W. D., and Feldman, S. (1985). *Multiple Regression in Practice* (Sage, Beverly Hills).

Cazes, P. (1976). "Régression par boule et par l'analyse des correspondances," Rev. Statist. Appliq. **XXIV**, 5–22.

Faber, A. (1989). "Lip protrusion in sibilant production," J. Acoust. Soc. Am. Suppl. 1 **86**, S113.

Fant, G. (1960). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Harshman, R., Ladefoged, P., and Goldstein, L. (1977). "Factor analysis of tongue shapes," J. Acoust. Soc. Am. **62**, 693–707.

Heinz, J. M., and Stevens, K. N. (1961). "On the properties of voiceless fricative consonants," J. Acoust. Soc. Am. **33**, 589–596.

Honda, M., and Kaburagi, T. (1993). "Comparison of electromagnetic and ultrasonic techniques for monitoring tongue motion," Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München 31, 121–136.

Hoole, P. (1993). "Methodological considerations in the use of electromagnetic articulography in phonetic research," Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München 31, 43–64.

Hoole, P., Ziegler, W., Hartmann, E., and Hardcastle, W. (1989). "Parallel electropalatographic and acoustic measures of fricatives," Clin. Linguist. Phon. **3**, 59–69.

Hughes, G., and Halle, M. (1956). "Spectral properties of fricative consonants," J. Acoust. Soc. Am. **28**, 303–310.

Jackson, M. T. T. (1988). "Analysis of tongue positions: Language-specific and cross-linguistic models," J. Acoust. Soc. Am. **84**, 124–143.

Lindau-Webb, M., and Ladefoged, P. (1989). "Methodological studies using an x-ray microbeam system," UCLA Working Papers Phon. **72**, 82–90.

Liljencrants, J. (1971). "A Fourier series description of the tongue profile," Q. Prog. Status Rep., Speech Transmission Lab., Royal Inst. of Technology, Stockholm, Sweden, STL-QPSR **4**, 9–18.

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 Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 131–149.

Nguyen, N., and Hoole, P. (1993). "Frequency variations of the lowest main spectral peak in sibilant clusters," in *Proceedings of Eurospeech '93*, 21–23 Sept. 1993, Berlin, Germany, pp. 81–84.

Nguyen, N., Hoole, P., and Marchal, A. (1991). "Articulatory-acoustic correlations in the production of fricatives," in *Proceedings of the XIIth Int. Cong. Phon. Sci.*, Aix-en-Provence, France, 19–24 Aug. 1991, Vol. 3, pp. 18–21.

Perkell, J. S., and Cohen, M. H. (1989). "An indirect test of the quantal nature of speech in the production of the vowels /i/, /a/ and /u/," J. Phon. **17**, 123–133.

Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., and Jackson, M. T. T. (1992). "Electro-magnetic midsagittal articulometer

(EMMA) systems for transducing speech articulatory movements," J. Acoust. Soc. Am. **92**, 3087–3096.

Perkell, J. S., Boyce, S. E., and Stevens, K. N. (**1979**). "Articulatory and acoustic correlates of the [s-š] distinction," J. Acoust. Soc. Am. Suppl. 1 **65**, S24.

Soli, S. (**1981**). "Second formants in fricatives: acoustic consequences of fricative-vowel coarticulation," J. Acoust. Soc. Am. **70**, 976–984.

Stevens, K. N. (**1989**). "On the quantal nature of speech," J. Phon. **17**, 3–45.

Stone, M. (**1991**). "Toward a model of three-dimensional tongue movement," J. Phon. **19**, 309–320.

Tuller, B., Shao, S., and Kelso, J. A. S. (**1990**). "An evaluation of an alternating magnetic field device for monitoring tongue movements," J. Acoust. Soc. Am. **88**, 674–679.

Zahorian, S. A., and Rothenberg, M. (**1981**). "Principal-components analysis for low-redundancy encoding of speech spectra," J. Acoust. Soc. Am. **69**, 832–845.