ARTICULATORY-ACOUSTIC CORRELATIONS IN THE PRODUCTION OF FRICATIVES

N. Nguyen-Trong[1], P. Hoole[2], & A. Marchal[1]

1. CNRS, URA 261, Univ. Provence, Aix-en-Provence, France
2. Institut für Phonetik, München Univ., FRG

ABSTRACT
This work is aimed at exploring the relationships between a set of articulatory parameters, and the acoustic output, in the production of French fricatives /s, S/. More precisely, we attempt to find out whether the dimensions of maximal contrast among the fricative spectra, are correlated with movements of lingual transducers monitored by means of an electromagnetic (EMA) system. Our results show that the EMA measurements can be considered to be very reliable. It appears that the spectra can be regenerated with a good accuracy from these measurements, with the help of a statistical method the advantages of which are pointed out. In conclusion, implications of this work in the domain of articulatory modelling are discussed.

1. INTRODUCTION

In a recent work [3], Hoole et al. have shown that the EPG tongue-contact patterns in the production of the fricatives /s/ and /S/ in English, were strongly correlated with a set of acoustic parameters extracted from the corresponding spectra by means of a factorial analysis. It has appeared that this relationship was close enough to allow a prediction of the acoustic data from the EPG data, with the help of a multiple linear regression. The results supported the conclusion that an empirical investigation of this kind, was suitable for providing information on the articulatory-acoustic correlations, which could be fruitfully incorporated into a model of fricative production [2,5]. The present experiment was based on the same methodological principles, and was aimed at investigating in a more extensive way two specific points. First, the question could be raised to know whether it is possible to relate the articulatory parameters with the spectra themselves by regenerating these spectra from the acoustic factors. Second, it seemed important to compare the results of the multiple linear regression, with those of a method giving the possibility to detect non linear articulatory-acoustic relationships similar to the ones which are described in Stevens' quantal theory [6].

2. MATERIAL

The experiment has been carried out at the Institute of Phonetics of Munich University. It consisted of an audio recording synchronized with a parallel EMA tracking. The electromagnetic system used is a commercially available device (Articulograph AG 100, Carstens, Göttingen, FRG) which has been recently assessed in [4,7], and which allows monitoring of articulatory movements with the help of five electromagnetic transducers (coils). In the present experiment, three coils were attached to the mid-line of the tongue, one was attached to the lower incisors, and one reference coil to the upper incisors. The tongue rearmost coil was placed as far back as possible, the frontmost coil about 1 cm back from the tip, and the third coil in between these two. The output of the EMA system was digitized (sampling rate 250 Hz) and transferred online to a PC AT-386 computer where software compensation for the effects of possible tilt of the receiver coils was applied. The digital signal, which represented the displacements of each of the five coils in the x-y plane, was finally stored on a hard disk. The audio signal was recorded by means of a B&K microphone on a DAT recorder, digitized on a LSI 11/73 computer (sampling rate 16000 Hz, LP filtered at 7500 Hz), and aligned with the EMA signal thanks to a set of synchronization pulses recorded on the second track of the DAT tape. The estimated accuracy of the alignment was +/-3 ms. The speech material consisted of the following combinations: /aS#/, /aSa/, /iSa/, /#sa/, /#si/, /as#/, /asa/, /asi/, /isa/, /isi/ embedded in 9 sentences which have been pronounced from 8 to 9 times by two male native speakers of French (AM, NN). In this paper, results will be presented for speaker AM.

3. ANALYTICAL PROCEDURES

3.1. EMA measurements

To minimize the variations in the articulatory signal which could have been generated by any head movements, the coordinates of the coil affixed to the upper incisors has been subtracted from those of the other coils. Moreover, for each repetition, the whole cloud of data has been rotated around the origin in the x-y plane, so as to achieve a vertical orientation of the first principal axis of the jaw movement. Figure 1 displays the positions, averaged over all the repetitions, of the tongue-back, tongue-mid, tongue-tip and jaw coils, at the mid-point of the fricative (which has a coordinate on the time axis determined with the help of the acoustic signal) for each item.

3.1. Acoustic analysis

The acoustic analysis consisted in calculating an FFT spectrum within a 32 ms Hamming window centered at the fricative mid-point. This spectrum was next reduced to 21 components by averaging the spectral energy over 1 Bark intervals from 0 to 8 kHz. Moreover, the information below 8 Barks has been ignored, in such a way that the acoustic data were finally made up in the present experiment of a set of 13-dimensional vectors. For reasons that are given below, we have chosen to proceed to a new data reduction by means of a principal-components analysis, which proved that 4 linear combinations of the 13 original parameters could account for more than 90% of the variance among the spectra. It has appeared that the 2 first factors were sufficient to differentiate the fricatives /s/ and /S/ from each other. Factor 1 is interpretable as a dimension of average energy; factor 2 can be considered as underlying an opposition between the spectra which show a local maximum within the 12-15 Barks range, and those in which the energy is relatively higher above 15 Barks.

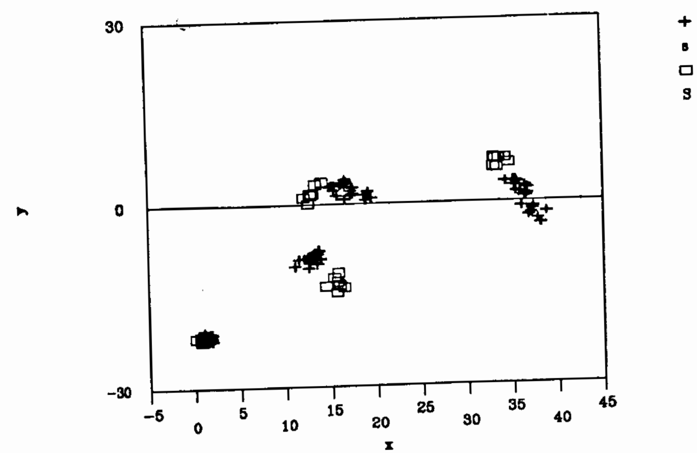

FIGURE 1

4. RESULTS

4.1. empirical regression of the acoustic factors from the EMA data

One approach to exploring EMA-acoustic relationships consists of analyzing separately the way in which the data are distributed in the EMA space, and in the acoustic space, to examine whether the sources of variation can be considered as being the same in the two cases (pronounced consonant, immediate context, carrier sentence, etc.). But it is also possible to check the existence of such relationships, by attempting to predict the acoustic parameters from the articulatory ones. At the present time, the information obtained with EMA on the midsagittal section of the vocal tract, doesn't allow calculation of the area function required by a standard acoustic model to resynthesize the output signal. In this experiment, the predictions have been based on the so-called statistical regression, which has been performed in two different ways, since we have compared the results of a classical, multiple linear regression, with those of an empirical, non linear variant **[1]**. In the second case, the predicted value of a given acoustic parameter **y** for a given articulatory input (a tongue «profile» composed of 3 points) was simply defined as the **y** mean value

TABLE 1: correlation coefficients between measured and predicted values for the first 4 acoustic factors (results given for two different regression techniques).

	REG.1	REG.2
fac.1	0.765	0.882
fac.2	0.851	0.925
fac.3	0.722	0.877
fac.4	0.384	0.536

associated with the input k-nearest neighbours in the articulatory space (k being determined by the user). It can be easily shown that the regression achieved by means of such a local approximation, is suitable for modelling non linear relationships, between any number of independent variables and the to-be-predicted one.

The calculations have involved the articulatory and acoustic data relative to all the (V)C(V) sequences recorded by speaker AM. The prediction quality has then been assessed on a test set which was composed of the tongue «profiles» averaged over all the repetitions for each of the two consonants, in each possible context. In table 1 are given the *r* 's corresponding to the correlation between the measured and the predicted values for factors 1, 2, 3 and 4. It appears clearly that the *r*'s are high whatever regression technique is used, and reflect a close relationship between the EMA parameters and the dimensions of maximal variance among the spectra, while the empirical regression (referred to as REG.2 in table 1) produces results which are systematically superior to those of the multiple linear regression (REG.1). The number of neighbours, on the basis of which a value has been given to each tongue profile for each of the 4 factors in REG.2, was fixed to 5.

4.2. Regeneration of the original spectra from the acoustic factors

The fricative spectra have been finally regenerated from the output parameters of the empirical regression, through the usual operation based in the present case on the eigenvectors of the covariance matrix relative to the 14 original acoustic variables **[8]**. Figure 2 proves that the accuracy of this regeneration is quite good (average *r* between the measured and predicted components for each of the spectra in the test set >0.9). It is especially encouraging to note that the shape shown by a spectrum is restored in a very satisfying way.

6. CONCLUSION

Our results could be explained in part by the fact that speaker AM has pronounced /s/ and /S/ in a rather stable way across repetitions and across contexts. Consequently, the variability among the spectra was likely to be accounted for by a relatively small

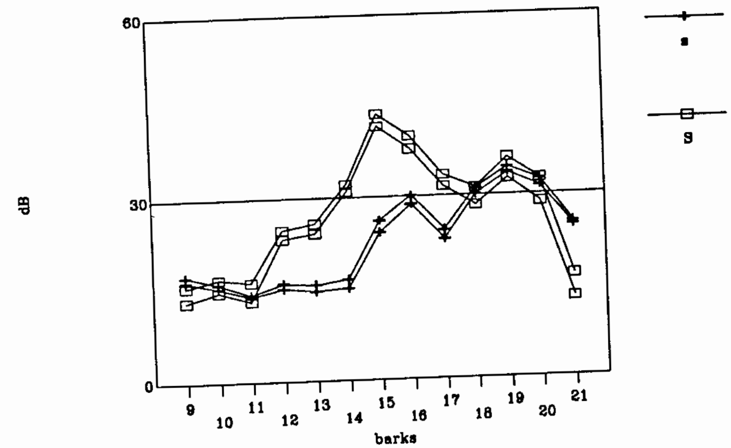

FIGURE 2: average spectra for /s/ and /S/ displayed together with the output of the empirical regression from the corresponding EMA profiles.

number of factors which in return allowed to regenerate these spectra without any major distortion. It remains that the factors themselves (which can be considered as dimensions of maximal contrast among the acoustic data) have proven to be accurately predictable from the EMA measurements. Therefore, it can be said that under the conditions adopted in this experiment, the articulatory parameters extracted by EMA are closely correlated with the acoustic output.

The empirical regression is very interesting in that it is gives the possibility to make predictions which have an accuracy calculated for each point in the articulatory space (while this calculation wouldn't have much sense in the case of the linear regression, in which the criterion to be optimized concerns an average accuracy). From our point of vue, this issue should give rise to a more systematic investigation. An experiment carried out according to the same methodological principles as those of the present work, on a more extensive material, would probably allow to determine whether a given articulatory neighborhood is stable (with respect to the acoustic output), or unstable. It seems to us that this kind of empirical exploration of the articulatory space, could constitute an quite interesting way to verify hypotheses on the mechanisms of speech production, such as the ones which are supported in the quantal theory **[6]**.

7. ACKNOWLEDGMENTS

This work was supported by ESPRIT II/ BRA n°3279 ACCOR.

9. REFERENCES

[1] CAZES, P. (1976). Rev. Stat. Appl., 24, 5-57.
[2] FANT, G. (1960). Acoustic Theory of Speech Production (Mouton, The Hague).
[3] HOOLE, P., ZIEGLER, W., HARTMANN, E., and HARDCASTLE, W.J. (1989). Clin. Ling. & Phon., 3, 59-69.
[4] SCHÖNLE, P. MÜLLER, C., WENIG, P., HÖHNE, J., SCHRADER, J., and CONRAD, B. (1987). Brain & Lang., 31, 26-35.
[5] SHADLE, C.H. (1985). «The acoustics of fricative consonants», Ph. D. Diss. (MIT).
[6] STEVENS, K.N. (1989). J. Phon., 17, 3-45.
[7] TULLER, B., SHAO, S., and KELSO, J.A.S., (1990). J. Acous. Soc. Am., 88(2), 674-679.
[8] ZAHORIAN, S.A., and ROTHENBERG, M. (1981). J. Acous. Soc. Am., 69, 832-845.