Journal of Phonetics (1991) **19**, 387–395

Measurement of the acoustic transfer function of the vocal tract: a fast and accurate method

A. Djeradi,* B. Guérin, P. Badin and P. Perrier

Institut de la Communication Parlée, INPG/ENSERG and Université Stendhal, 46 Avenue Félix Viallet, F-38031 Grenoble Cedex, France

Received 15th November 1990

Knowledge of vocal tract acoustic transfer functions is important for the understanding of phenomena occurring during speech production. We present here a new measurement method based on the external excitation of the vocal tract with a known pseudo-random sequence, where the transfer function is obtained as the FFT of the cross-correlation between the sequence and the signal measured at the lips. The advantages of this method over methods based on sweep-tones or white noise excitation are (1) a much shorter measurement time (about 100 ms), (2) the possibility of speech sound production during the measurement, and (3) a more accurate definition of bandwidths. This method has been checked against classical methods through systematic comparisons on a small corpus of vowels and fricative consonants. Moreover, it has been verified that simultaneous speech sound production does not perturb significantly the measurements. This method should thus be a very helpful tool for the investigation of the acoustic properties of the vocal tract in various cases such as vowels, fricatives and nasals.

1. Introduction

Measurements of the vocal tract transfer function have induced a better knowledge of acoustic phenomena occurring during speech production (Van den Berg, 1955; Fujimura & Lindqvist, 1971; Castelli & Badin, 1988). Classical techniques for direct investigation of the acoustic characteristics of the vocal tract are based on the transcutaneous excitation of the tract near to the glottis. The most common method, employed by Van den Berg (1955) and later by Fujimura & Lindqvist (1971), uses a pure tone signal swept in frequency. Castelli & Badin (1988) used instead a white noise excitation (WNE) in order to improve the experimental process. Whatever the technique, the measurement duration is about 10 s. This duration is due either to the time required to sweep from 100 Hz to 5000 Hz (considering the sharpness of vocal tract resonances) or to the need to average a large number of FFT spectra of the lip signal for the WNE method. The tract configuration has to be kept as rigid as possible during this long measurement time. Furthermore, these techniques do not

* Guest researcher from the Institut d'Electronique, USTHB, El Allia – Bab Ezzouar – BP 32, Algiers, Algeria.

0095-4470/91/030387 + 09 $03.00/0

© 1991 Academic Press Limited

allow any kind of speech sound production during the measurement process. This last point could be neglected in the case of vocalic configurations, but it is very important for the acoustic characteristics of fricatives. In the case of fricative consonants, the area and place of constriction need to be accurately controlled. This control is mainly obtained through the kinaesthesic sensations of pressure and flow perceived by the articulators, and especially the tip of the tongue.

This paper describes a new method based on the excitation of the vocal tract by a pseudo-random sequence. This method (denoted PRE) has two advantages: it allows the accurate measurement of acoustic transfer functions within a few hundreds of milliseconds, and it permits speech sound production during the measurement process. This technique was derived from a technique developed earlier for acoustic room characterization (Jullien, Gilloire & Saliou, 1984).

2. Theoretical basis

The vocal tract can be considered as a linear acoustic filter of which the impulse response is denoted $h(t)$. The sampled counterpart of this response will be denoted $h(n)$ and the filter transfer function will be H. We then express the output signal $y(n)$ as:

$$y(n) = [b(n) + x(n)] * h(n) \tag{1}$$

where $b(n)$ denotes any undesirable noise superimposed on the excitation $x(n)$.

The correlation, R_{xy}, of $x(n)$ with the output signal $y(n)$, can be expressed as:

$$R_{xy}(n) = R_1(n) + R_2(n) \tag{2}$$

where

$$R_1(n) = \sum_{k=-\infty}^{k=\infty} h(k) \left[\sum_{m=-\infty}^{m=\infty} x(m) \cdot x(m+n-k) \right] \tag{3}$$

and

$$R_2(n) = \sum_{k=-\infty}^{k=\infty} h(k) \left[\sum_{m=-\infty}^{m=\infty} x(m) \cdot b(m+n-k) \right] \tag{4}$$

Moreover, let $\varphi_{xx}(k)$ be the autocorrelation of $x(n)$ and $\varphi_{xb}(k)$ the correlation between $x(n)$ and $b(n)$. We can rewrite Eqn (2) as:

$$R_{xy}(n) = [h * \varphi_{xx}](n) + [h * \varphi_{xb}](n) \tag{5}$$

Assuming that $x(n)$ and $b(n)$ are uncorrelated,

$$\varphi_{xb}(n) = 0$$

and then:

$$R_{xy}(n) = [h * \varphi_{xx}](n) \tag{6}$$

It is clear that, the closer $\varphi_{xx}(n)$ is a Dirac impulse, the better will be the approximation of $h(n)$ by $R_{xy}(n)$. A good approximation of the statistical properties of white noise is the so-called pseudo-random sequence (Jullien *et al.*, 1984; Schroeder, 1983), of which the autocorrelation is an impulse train $e_N(n)$ of period N. Then

$$R_{xy}(n) = [h * e_N](n) \tag{7}$$

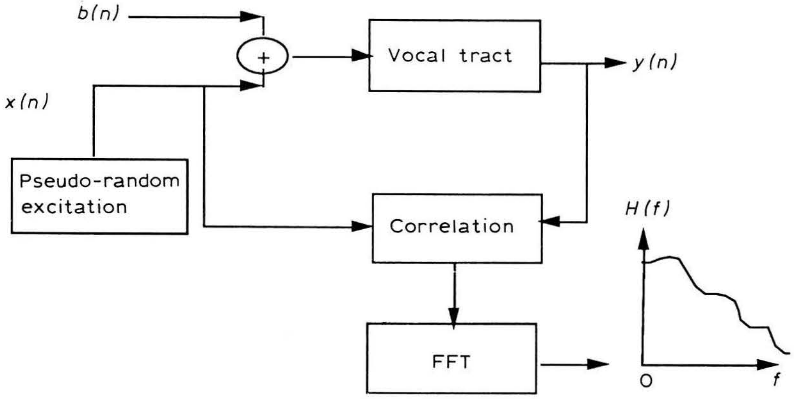

Figure 1. Block diagram of the principle of the transfer function measurement of the vocal tract.

If the length of the impulse response $h(n)$ is smaller than N, the sequence $R_{xy}(n)$ corresponds exactly to $h(n)$ for n varying from 0 to $N-1$. In this case, if $\mathrm{FT}[R_{xy}](k)$ is the Fourier transform of length N of $R_{xy}(n)$, one would get

$$\mathrm{FT}[R_{xy}](k) = H(k)$$

Let $\Phi_{xx}(k)$ be the discrete Fourier transform of $\varphi_{xx}(n)$:

$$\mathrm{FT}[R_{xy}](k) = H(k) \cdot \Phi_{xx}(k) \tag{8}$$

Since $\Phi_{xx}(k)$ is equal to a constant Φ_0:

$$\mathrm{FT}[R_{xy}](k) = H(k) \cdot \Phi_0 \tag{9}$$

The complete block diagram describing the principle of measurement of the transfer function is represented in Fig. 1.

The sampling frequency was 20 kHz. In order to get a frequency resolution of 10 Hz, the length of the pseudo-random sequence was set to 100 ms. Furthermore, to obtain a useful range for the transfer function, between 0 and 5 kHz, a clock frequency of 10 kHz was chosen for the sequence. Finally the length of the sequence was set to 1023 ($N = 2^n - 1$) clock periods (about 100 ms).

3. Simulations

In order to test this method, simulations were performed on a microcomputer. The vocal tract was simulated by an all-pole filter corresponding to the first three formants, and the signal $b(n)$ (see Fig. 1) consisted of a voiced pulse source. The correlation $R_{xy}(n)$ was computed between the pseudo-random excitation $x(n)$ and the output signal $y(n)$, and the transfer function of the vocal tract was computed as the discrete Fourier transform of R_{xy}.

Two cases had to be considered: (1) excitation by the pseudo-random sequence only; (2) a speech source signal added to the pseudo-random sequence (which is the situation corresponding to the speech sound production conditions).

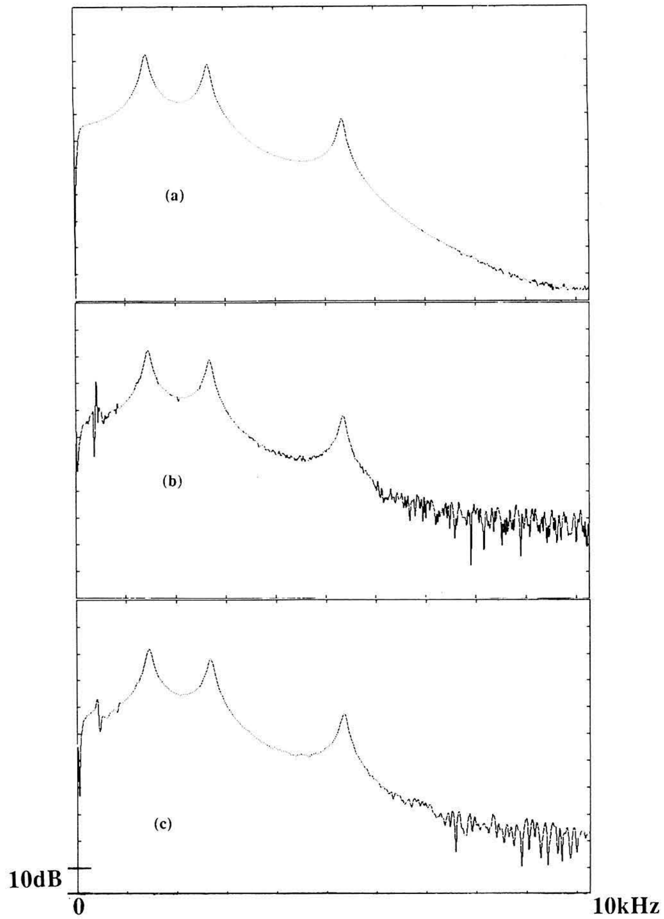

Figure 2. Transfer functions obtained by simulation (a) for a pseudo-random excitation alone; (b) a perturbation added to the excitation, without window; (c) a perturbation added to the excitation, with window.

Figure 2(a) shows the transfer function computed in the case where the pseudo-random sequence alone is used. As expected, this transfer function is identical to that of the simulated filter. Figure 2(b) shows the transfer function obtained when a voiced excitation of fundamental frequency F_0 is added to the pseudo-random excitation. A perturbation appears on the transfer function in the vicinity of F_0. This is due to the fact that $\varphi_{xb}(n)$ is not exactly equal to zero. That induces unimportant errors when $h(n)$ is large (at the beginning of the impulse response), but can lead to non negligible errors when $h(n)$ is small. These errors result in an extra peak in the transfer function at F_0 frequency. In order to reduce this unwanted peak, the correlation is limited to the first part of the signal corresponding to large values of $h(n)$ by the use of a Hamming window (with the effect seen in Fig. 2(c)). It is well known that the effect of windowing is to widen the bandwidth of the resonances. However, our simulations showed that for a 50 ms window this widening was small, of the same order of magnitude as the resolution.

4. Experimental setup and procedure

The experimental setup is described in Fig. 3. The acoustic part is very similar to that developed by Castelli & Badin (1988). The subject presses the loudspeaker membrane externally at the level of the thyroid cartilage, checking that there is no noticeable sound leakage at the junction between the loudspeaker and the skin. A condenser microphone picks up the sound pressure signal at a distance of about 1 cm from the mouth opening. It was, however, necessary to make certain modifications to the original setup because of the lower output sound pressure level due to the constriction which is typical of fricatives. In order to reduce the acoustic leakage due to possible direct radiation from the loudspeaker towards the microphone, a fibre-glass board was constructed to fit around the subject's chin between the loudspeaker and the microphone.

The simultaneous digital-to-analog and analog-to-digital conversions are performed by a standard signal processing board plugged into the microcomputer.

The procedure for measuring one transfer function consists of four steps:

(1) the loudspeaker is fed with white noise in order to allow the subject to position the loudspeaker correctly by listening through earphones to the signal picked up by the microphone;
(2) the loudspeaker is made silent, so the subject can initiate steady state speech sound production (if needed) in normal conditions;
(3) the pseudo-random excitation is switched on and simultaneously the response signal from the mouth is recorded;
(4) the cross-correlation and the FFT are computed.

It was checked that the effect of the fiber-glass board was to reduce the noise leaking from the loudspeaker and measured by the microphone, down to a level almost identical to that of the background noise in the room. The recordings may thus be considered as uncontaminated by noise leaking directly from the loudspeaker.

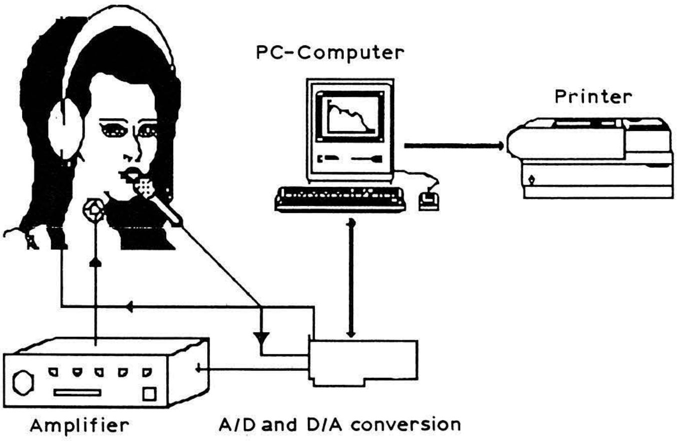

Figure 3. Experimental setup.

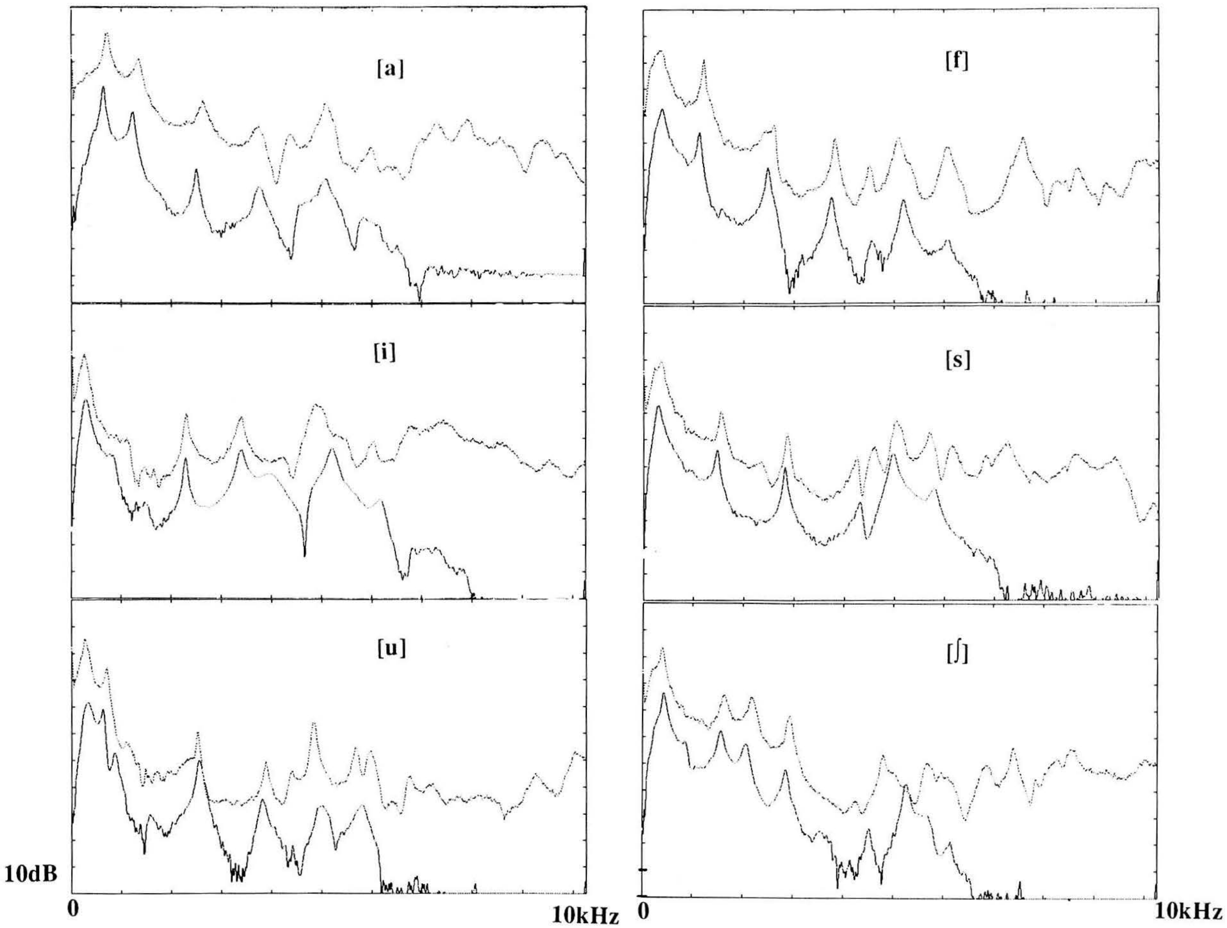

Figure 4. Transfer functions in closed glottis conditions for the vowels [a, i, u] (left) and the fricatives [f, s, ʃ] (right), using WNE method (upper line) and PRE method (lower line).

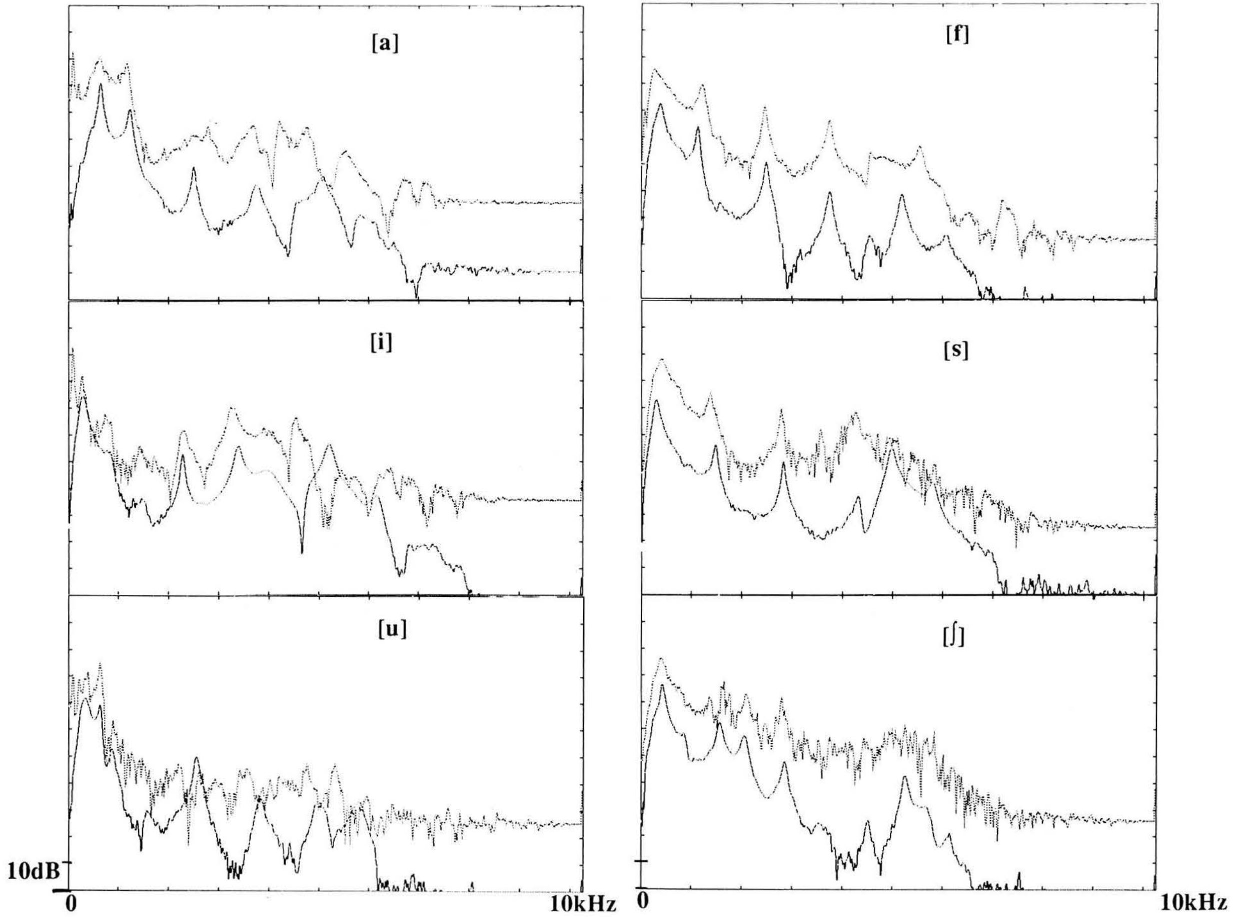

Figure 5. Transfer functions for the vowels [a, i, u] (left) and the fricatives [f, s, ʃ] (right) under speech sound production conditions (upper line) and closed glottis conditions (lower line).

5. Results

5.1. *Transfer functions measured in closed glottis conditions*

Typical examples of vowel transfer function have been processed for three extreme configurations of the vocal tract, [a, i, u], in closed glottis conditions. Figure 4 (left) shows that the first four formants are well defined. In order to compare the PRE method and the WNE method, we have superposed in Fig. 4 the transfer functions obtained for the same subject with both methods. The results are quite similar. There are however differences: (1) the valleys between formants are deeper with the new method, and at the same time more noisy; (2) the bandwidths are narrower with the new method. An explanation can be proposed for these differences. The principle of the WNE method is to average a number of FFT spectra over a long period of time (6.6 s); during measurement, the vocal tract of the subject cannot be expected to be perfectly rigid; slight physiological movements will lead to small variations of the poles and zeros of the transfer function, and this will widen the formants and smear out zeros and valleys.

Note that there are no significant spectral components above 6 kHz for the PRE method: this is inherently due to the fact that the pseudo-random excitation signal had a clock frequency of only 10 kHz.

Similar measurements were performed with fricative configurations. In spite of the lower output level due to the oral constriction in the vocal tract typical for fricatives, in this case also we have obtained good results with clear resonances (see Fig. 4 right). We should mention that the transfer function measured contains only poles, since the vocal tract is artificially excited at the end.

5.2. *Transfer functions measured in speech sound production conditions*

Transfer functions were measured for the vowels [a, i, u] with simultaneous speech sound production. These transfer functions (see Fig. 5 left) are clearly more noisy than those obtained under the closed glottis conditions. A first explanation for this is that the pseudo-random excitation and the glottal signal can be correlated in the short term, so the hypothesis that their cross-correlation is zero does not strictly apply. But a better explanation is certainly that the level of the excitation by the loudspeaker was too low in comparison with that of the glottal source.

The bandwidth of the first formant is wider for [a], the other higher resonances being also damped (Fig. 5 top panel, left). The first sharp peak that appears in the transfer function of vowel [a] in speech sound production conditions is due to the contribution of the periodic vocal source signal. For vowels [i] or [u], the second harmonic of the glottal source is very probably in the region of the first formant. It is thus difficult to determine the origin of the widening of the first formant bandwidth: it arises partly from damping by the periodic glottis opening, and partly as an artéfact due to the traces of glottal excitation in the measurement.

Measurements were also made for fricative consonants under speech sound production conditions (see examples in Fig. 5 right). A significant widening of the bandwidths should be noticed: this is most likely due to losses induced by the wide glottis opening (coupling with subglottal cavities) and by the turbulent air flow (constriction resistance related to turbulences).

In the case of fricatives, the possibility of simultaneous speech sound production during the transfer function measurement is an important feature (Badin, 1991). In this case, the area and place of constriction need to be accurately controlled. This control is mainly achieved through the kinaesthesic sensations perceived from pressure and flow: these aerodynamic conditions are incompatible with silent articulation.

6. Conclusion

We have described a new method for vocal tract acoustic transfer function measurements, based on a pseudo-random excitation (PRE). The first results obtained with the PRE method have been shown to be at least as accurate, in terms of the definition of peaks, as those obtained by the WNE method. Furthermore, the PRE method exhibits two attractive features, compared with the WNE method: first, the measurement duration is reduced from about 6.6 s to 100 ms; second, the measurements are possible even in speech sound production conditions.

We have shown by simulation that the excitation signal must be three or four times more powerful than the vocal signal itself in order to obtain correct results in speech sound production conditions. The excitator that is used currently should therefore be replaced by another more powerful excitator.

Measurements of acoustic transfer functions have already been very useful in the case of nasals (Castelli, Perrier & Badin, 1989) and for fricatives (Badin, 1991). The new method should enable the processing of more extensive corpora, in better experimental conditions, and thus help to increase our understanding of the acoustics of speech production, especially for fricative consonants.

This work has partly been funded by the *French Centre National de la Recherche Scientifique* and by an academic contract of cooperation between Algiers and Grenoble (grant 89 MES 108). We are very grateful to Gang Feng for his suggestions on signal processing and to Philippe Verdier for his help in implementing the input/output and signal processing software, and to Andrew Morris for revising the English text.

References

Badin, P. (1991) Fricative consonants: acoustic and X-ray measurements, *Journal of Phonetics*, **19**, 397–408.

Castelli, E. & Badin, P. (1988) Vocal tract transfer function measurements with white noise excitation. Application to the naso-pharyngeal tract. In *Proceedings of the Seventh FASE Symposium, SPEECH'88* (W. A. Ainsworth & J. N. Holmes, editors), pp. 415–422.

Castelli, E., Perrier, P. & Badin, P. (1989) Acoustic considerations upon the low nasal formant based on nasopharyngeal tract transfer function measurements. In *Proceedings of the European Conference on Speech Communication and Technology* (J. P. Tubach & J. J. Mariani, editors), Vol. 2, pp. 412–415.

Fujimura, O. & Lindqvist, J. (1971) Sweep-tone measurements of vocal-tract characteristics, *Journal of the Acoustical Society of America*, **49**, 541–558.

Jullien, J. P., Gilloire, A. & Saliou, A. (1984) Mesure de réponses impulsionnelles en acoustique, Note technique NT/LAA/TSS/181, Centre National des Télécommunications, Lannion, France.

Schroeder, M. R. (1983) *Number Theory in Science and Communication*, pp. 248–258. Berlin: Springer-Verlag.

Van den Berg, J. (1955) Transmission of the vocal cavities, *Journal of the Acoustical Society of America*, **27**, 161–168.