Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 287–291
(S. Karger, Basel/New York 1965).

From the Department of Phonetics, University College London

A Note on the Spectral Analysis of Unvoiced Sounds

By A. J. Fourcin

The aim here is to describe a particular frequency spectrum analysis technique which can gather information for a closer approximation to a representative sample of an unvoiced sound than is normally available with a sound spectrograph (the Kay Sonagraph) and to illustrate some of the initial results obtained from the application of this analysis technique to the assessment of the sounds produced from a self excited acoustic model of the vocal tract.

Possibly a primary difficulty in the analysis of speech arises from our relative ignorance of the characteristics of the source of excitation applied to the vocal tract. For voiced sounds this difficulty is somewhat tempered by the quasi-periodic nature of the excitation, which leads to a useful degree of spectral coherence, and the well defined physical position of the source. For unvoiced sounds however the excitation is temporally random, which hampers spectral analysis, and often at least partially distributed in position along the vocal tract, which hinders structural interpretation.

The analysis of random signals has received a great deal of attention in other fields and the theoretical basis for their adequate spectral analysis is well established. The result to be expected in the simple case, for example, of a noise of gaussian amplitude distribution applied to a rectangular filter driving a linear rectifier followed by a finite integration has been given succinct expression by *Bennett* (1960, p. 149) who shows how the analysis tends to become more consistent as the filter bandwidth and the integration time are increased. For speech analysis *Fant, Fintoft, Liljencrants, Lindblom* and *Mártony* (1963) have considered the properties of a practical bandpass filter the effective output of which is averaged by a low pass filter. In the Kay Sonagraph however there is essentially no post filtering averaging either by integrator or low pass filter. In consequence the spectrum, or section, obtained from the frequency analysis of any unvoiced sound whose steady state duration is greater than the response time of the analyzing filter is less representative than need be the case and the spectral envelope of the analysis made by the machine is more disturbed than is theoretically necessary. An improvement can be obtained simply by averaging the filter output, before employing it in the sectioning circuit, for the maximum time permitted by the stability of the sound to be analyzed.

A simple arrangement for making this averaging possible is shown in figure 1. Once in every drum revolution an integrator is reset to zero by a short circuiting contact which can be placed at any point around the periphery of the drum base to discharge the capacity, C. After this resetting, integration of the analyzing filter output commences and the instant at which the normal sectioning microswitch is actuated determines the end of the integration interval. In this way an integration time between less than 0.1 sec and up to about 2.3 sec can readily be obtained. The only additions to the basic Sonagraph are the resistance R and the capacity short circuiting arrangement. It is an advantage, however, to use a more sophisticated integrator having a gain which may be adjusted to correspond to the integrating period and this has been employed to derive the lower spectra of figure 1, the integration intervals for which are shown below the spectrogram of the parent sound sequence. A more extreme indication of the improvement in spectral delineation which this post filtering integration affords is given in figure 2 where the ordinary narrow band sections for two fricatives are shown above and the corresponding long term spectra (2 sec integration intervals) are shown below. Narrow band analysis has been used for all the examples in the present note and this sets a limit to the degree of spectral smoothing which may be obtained, unless the speech sound analyzed contains spectral features of comparable or smaller bandwidth.

It might appear from the clarity of these spectra that this technique could profitably be employed for the determination of vowel formant frequencies; since the smooth spectral form which may be associated with a breathed excitation could overcome the disadvantage accompanying the discrete frequency spacing of voiced harmonics. This is correct for the case in which the breathed excita-

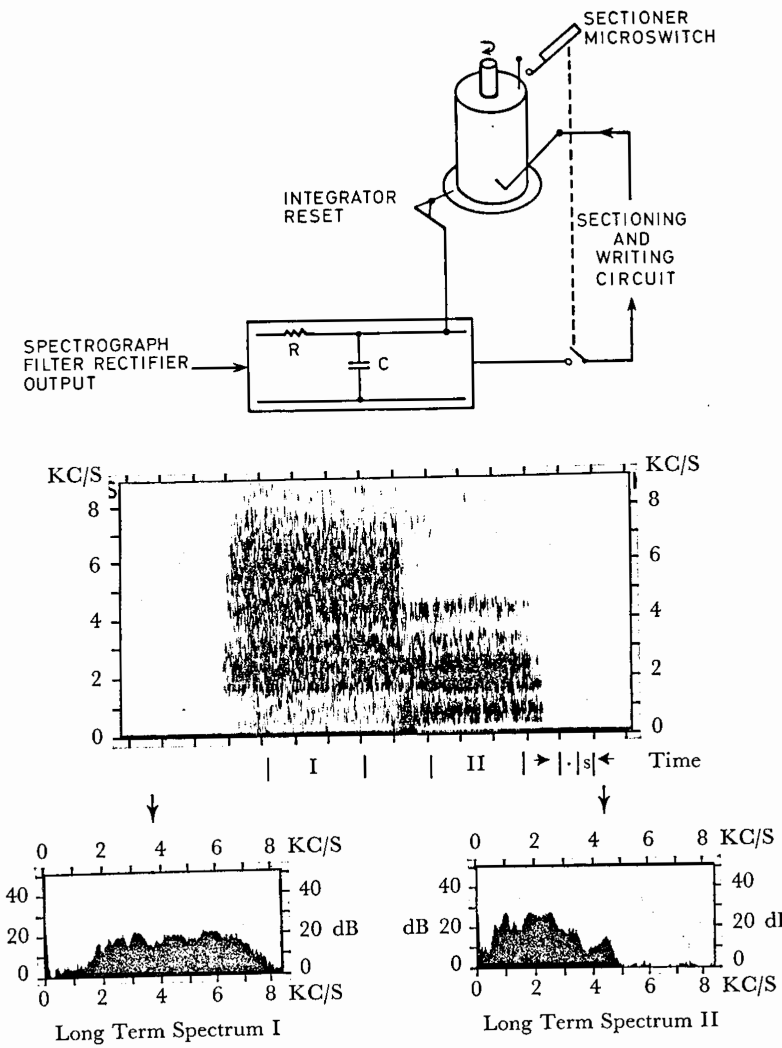

Fig. 1

tion is located at the glottis but when, as is often the case, friction is present along an appreciable length of the vocal tract the vowel spectrum will be seriously deformed. Results which are of more immediate interest can be derived from the analysis of unvoiced sounds which are associated with a single well defined source of turbulence in the vocal tract, such as those whose analyses are shown in figure 2. These two sounds were each produced with the same degree of lip rounding and an attempt was made to keep the tongue-alveolar

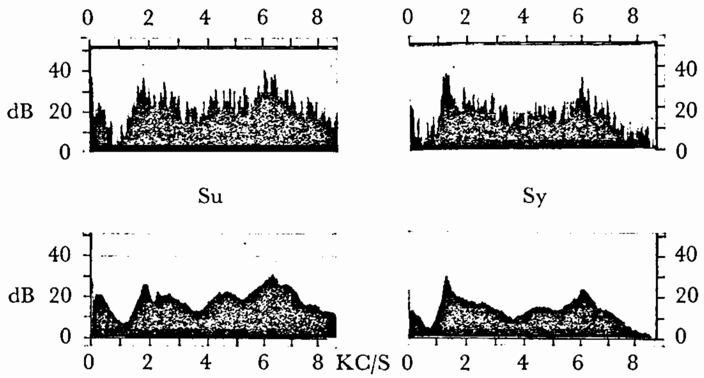

Fig. 2. Normal and Long Term Spectra.

constriction the same although the position of the back of the tongue was altered. It is evident from the long term spectra that the alteration of the pharyngeal cavity has an appreciable influence on the sound emerging from the lips. If the excitation of the vocal tract is entirely due to the sound produced by turbulence at the alveolar constriction it follows that the pharyngeal cavity is acting to absorb acoustic energy which would otherwise appear at the mouth.

The situation may be explored in a more controlled fashion by the use of acoustic models of the articulatory configuration involved in which the passage of an air stream through an appropriate internal constriction produces the acoustic excitation. In figure 3 (a) is shown a lateral medial view of an actual vocal tract in the configuration corresponding to a palatalized palato-alveolar fricative (taken from *Fant*, 1960, p. 170) and in figure 3 (b) is the outline of a corresponding planar acoustic model. This model has parallel transparent rigid cheeks spaced by 2 cm, so that the separation between the walls – which are curved in only one plane and, here, made of plasticene – is directly proportional to the cross sectional area of the corresponding section in the human original. The constriction was 1 cm long and of 0.1 cm² cross sectional area. The long term spectral form of the output from the lips when an air stream is applied from a simulated glottis is shown in figure 3 (c); an integration time of 2 sec was employed. In figure 3 (e) the same model has been modified to diminish the volume of the pharyngeal cavity whilst keeping the remainder of the artificial vocal tract unchanged. Once again the long term spectrum of the output, figure 3 (f), integration time of 2 sec, has been altered and a distinct spectral trough has been

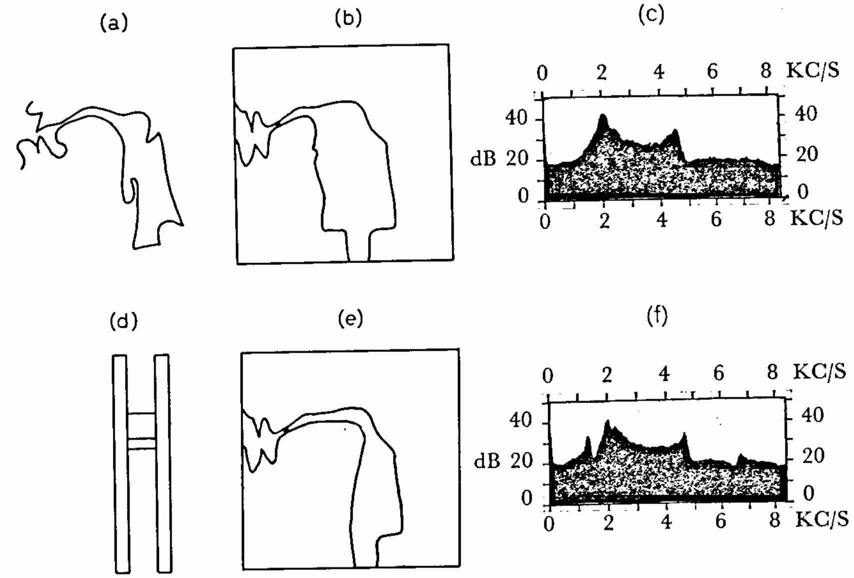

Fig. 3

introduced. Since one can now be sure that the constriction and the cavity arrangement which follows it are unchanged, this change in spectral form must be related to pharyngeal modification.

These spectral results could have been obtained by other means but it is of some practical convenience even with models to employ the same means of analysis that is used for speech, and with speech it is essential to be able to choose the integration time interval.

Acknowledgement. I have derived benefit in the preparation of this summary from discussion with Mr. *D. C. Bennett* and other colleagues in the Department of Phonetics and have been very much helped in the experiments by Mr. *J. E. West.*

References

Bennett, W. R.: Electrical noise (McGraw-Hill, New York 1960).

Fant, G.; Fintoft, K.; Liljencrants, J.; Lindblom, B. and *Mártony, J.:* Formant amplitude measurements. J. acoust. Soc. Amer. *11:* 1753–1761 (1963).

Fant, G.: Acoustic theory of speech production (Mouton, 's-Gravenhague 1960).

Author's address: A. J. Fourcin, University College London, Department of Phonetics, Gower Street, *London W.C. 1* (England).

Discussion

O'Connor (London): Would the material of which the teeth of the model are made have any effect upon the resulting friction for, say, an [s]?

(Mr. *Fourcin* thought not; and that the constriction itself would be responsible.)