BRITISH ACOUSTICAL SOCIETY

"SPRING MEETING" at Chelsea College, London, S.W.3 on
Wednesday 25th April / Friday 27th April, 1973

SPEECH AND HEARING: Session 'B': Speech Analysis and Transmission

Paper No:

73SHB1

Noise Resistant Larynx Synchronous Speech Processing

A.J. Fourcin
Department of Phonetics and Linguistics
University College London

(Invited paper)

SUMMARY

For ordinary speech communication, the most important class of sounds is the one for which the speaker's vocal tract is excited by the interruption to the lung air stream produced by the nearly regular closing of his vocal folds. Work at University College on the development of a laryngograph, an instrument giving an electrical output corresponding to the closing of the vocal folds, has resulted in a practical device capable of providing a precise indication of the beginning of vocal fold contact.

This information about vocal fold activity can be used as the basis of a voiced-voiceless detector which, since the laryngograph operates from an electrical input, can be relied on even in the presence of high ambient noise. In consequence the externally generated acoustic noise which may accompany a spoken message can easily be eliminated during voiceless intervals and during voicing periods selective temporal gating and, or, selective comb filtering controlled by the accurate noise-free larynx closure information is possible. A digital computer has been programmed to provide this processing and both its outputs and those of similar analogue circuits operating in a practical situation will be discussed in the oral presentation. In what follows, the physical features in the laryngograph waveform which correspond to different speech activities are described and examples of selective temporal gating given both with and without acoustic interference.

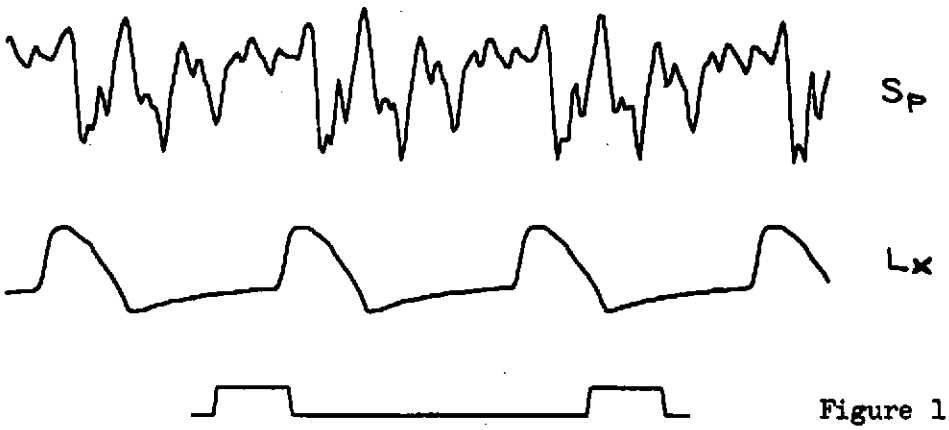
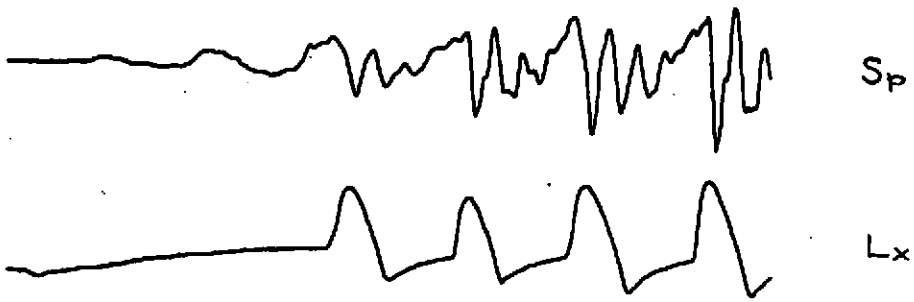


Figure 1

Although, in general, only a version of the acoustic pressure waveform produced by a speaker is employed in speech communication systems, there may be an advantage in special situations in supplementing, or even replacing, Bell's original invention with information derived directly from the speaker's larynx and vocal tract. The particular example of this approach described here utilises the output of a non-acoustic detection of vocal fold contact during normal phonation. Since this information is impervious to acoustic interference it can be employed to reduce the importance of ambient acoustic noise in the speaker's normal acoustic output. The top waveform, Sp, in Figure 1 corresponds to a speech pressure waveform - for the vowel ə: ; male adult, 10 msec time marker - and it is representative of the most important type of speech signal normally transmitted. This is both because of its high energy content and because of the essential transition information which it carries. The speaker's vocal tract has been excited by the quasi regular interruption of the tracheal air stream resulting from the vibration of his vocal folds and each complete pressure cycle is marked by the change from small to large acoustic activity which occurs just after the leading edge of the Lx waveform. This middle waveform has been derived from a qualitative electrical measure of the variation in vocal fold contact area during the corresponding laryngeal activity (Fourcin & Abberton, 1971) and uses superficial electrodes placed over the wings of the speaker's thyroid cartilage. The positive rise corresponds to the closing of the vocal folds. Closure occurs more rapidly than opening because of the positive feedback action associated with the Bernoulli effect in the glottis during the closing phase, and consequently produces the greatest acoustic effect in the vocal tract. The rise in Lx precedes the change in Sp as a result of the delay from glottis to microphone.

Figure 2



In Figure 2, the Sp and corresponding Lx waveforms are shown for the voice onset of ə2ɔ (conditions of Figure 1) and it is important to notice that the small regular vocal fold vibration which is responsible for the precursive part of the Sp waveform, and which does not involve fold contact, is not associated with a corresponding Lx activity. The Lx waveform can, in consequence, be used as the basis for a conservative estimate of the voiced parts of a speech signal. If the first substantial vocal fold closure is detected to be above an assigned threshold then the speech can be gated so that those portions which are voiced are transmitted, with all successive Lx peaks being employed to maintain the decision. The decision procedure can be fixed to correspond to the largest period which it is required to conserve or it can easily be made selfadjusting so that hang-on at the end of a voiced sequence is minimal. Speech processed in this way is entirely intelligible in context but when isolated words depend for their recognition on the identification of voiceless elements then they cannot be correctly perceived. In restricted communication situations this obstacle can be overcome in some measure by the substitution of voiced for all voiceless fricatives. The situation for ordinary speech is illustrated in Figure 3(b). The fricative portion in the centre of Sp is in the sequence "to shine" and the breaks in Lx correspond to it and the [t] closure.

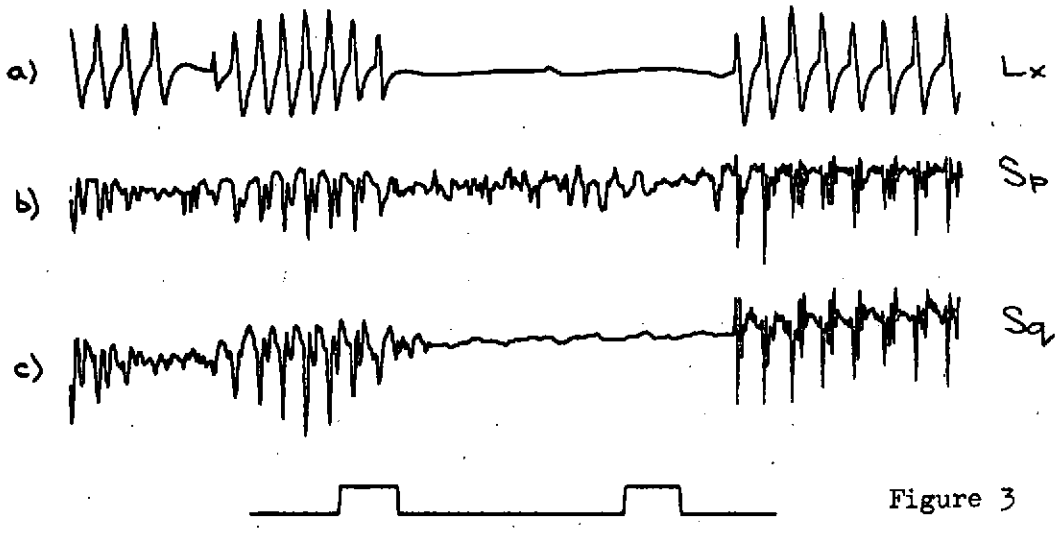


Figure 3

A fixed hang-on of 10 mS has been employed in the detection of the voiced sequence, Sq, and this results in the elimination of [S] but in the retention of the [t] sequence in Figure 3.

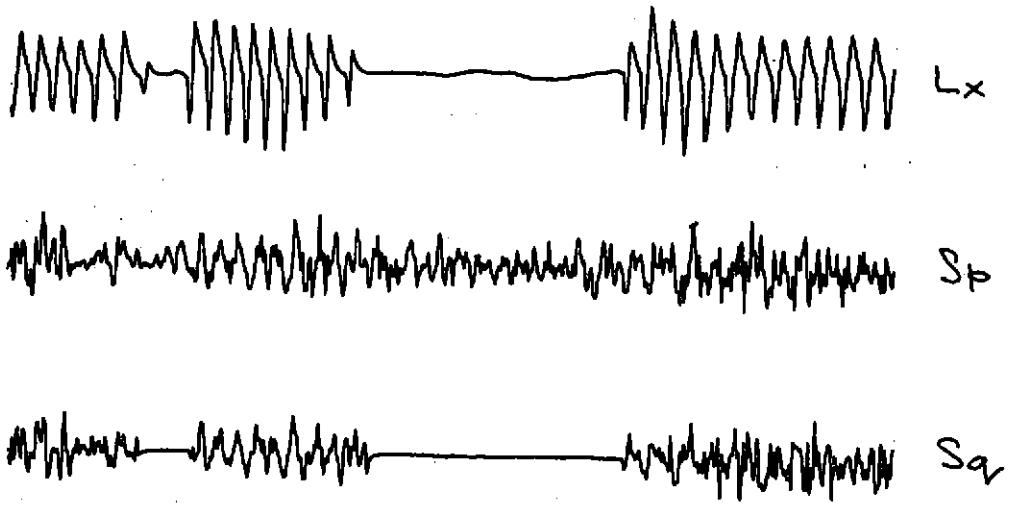


Figure 4

In Figure 4, the hang-on time has been reduced and the speech sequence for Figure 3 repeated with greater ambient noise (and plotted with some time scale). When speech is produced in conditions of high ambient noise the distinctions carried by its low intensity voiceless components are lost and the perception of the voiced elements made all the more difficult by the masking of the temporal structure of the sequence. The use of Sq signals of the type shown in Figure 4 can reduce these difficulties and the Lx information can be employed fairly easily in the control of gating and filtering to provide the basis for further signal enhancement.*

REFERENCES

Fourcin, A.J., Abberton, E., First applications of a new laryngograph. Med. biol. Illust., 21, 3, 172-182, 1971.

*British Provisional Patent