Proceedings of The Institute of Acoustics

THE TECHNICAL ASPECTS OF HUMAN SPEECH

N. T. BOWMAN

SCHOOL OF ARCHITECTURE, LEICESTER POLYTECHNIC.

1. SPEECH MATERIAL AND INTELLIGIBILITY TESTS

In most studies connected with speech reception, the experimenter is interested in obtaining an estimate of the intelligibility of speech under particular listening conditions. Furthermore "speech" is normally assumed to be conversational speech. In any listening situation there are many factors which can, and often do, affect the resulting intelligibility. There are effects due to vocal force (1), visual factors (2), talker-listener angle (3), binaural time delay (4), speech context (5) and training (6), to mention but a few. Many such factors are utilised quite naturally by the listener to improve intelligibility, being part of the speech recognition learning process. In order to study the true information transfer due to a speech link it is necessary to try to exclude or control many such variables.

Speech tests, or rather intelligibility and articulation tests, are therefore normally studied under carefully controlled conditions with trained talkers and listeners, and specially prepared test material. The test material is usually presented in lists which are arranged to be representative of conversational speech and equi-difficult. Among the more significant materials which have been developed, and proven, are those of Egan (7), Beranek (8), Hirsh (9) and Richards (10). A more detailed account of test arrangements and procedure has been given elsewhere (11).

2. PREDICTIVE MODELS FOR SPEECH INTELLIGIBILITY

The intelligibility test, although not too difficult to arrange, is very time consuming and as such is not a very flexible design tool. A need exists therefore for a model which can express the performance of a communication link in terms of its physical parameters which can be either measured or calculated.

Various workers (12, 13) have made measurements of speech levels within the useful speech frequency range (approx. 200Hz - 7000Hz) and it is possible to represent speech as a map on a pressure-frequency graph: such a map is shown in FIG. 1. The spectrum reaches its maximum at about 500Hz and above 1000Hz the spectrum reduces at approximately 8dB/octave. Speech peaks are about 12dB greater than the average level, and speech minimums about 18dB below the average level. The position of the map relative to the sound pressure level ordinate is somewhat arbitrary and depends on

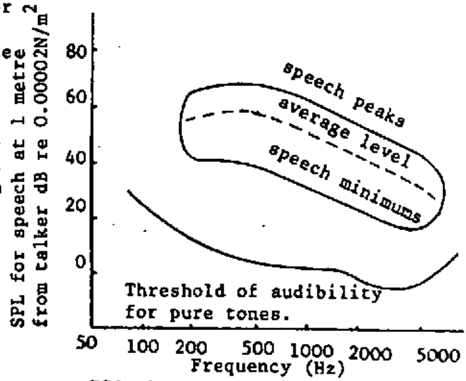

FIG. 1. The Speech Map.

vocal effort. In theory, provided all of the speech map is above the threshold of hearing, then conditions will be optimum for listening. With this model it is also relatively easy to represent the phenomenon of masking.

Within the theoretical framework developed by several of the pioneer workers (14, 15, 16, 17) the speech frequency scale was transformed to a new scale where frequency bands of equal width contributed equally to a new quantity called the articulation index (A). These predictive models are discussed in detail elsewhere (11), but the underlying principle may be understood through reference to the French and Steinberg model. They established 20 frequency bands in the range 250Hz - 7000Hz, which each contribute a factor of 0.05 to A ($0 \leq A \leq 1.0$) when all bands are at their optimum levels. A value of A = 1.0 represents the condition for optimum intelligibility.

The most modern version of the articulation index is that proposed by Kryter (18, 19) which has also been published as a standard (20). This model is similar to the previous models but offers either a one-third octave band method involving fifteen bands, or an octave band method involving five bands, each covering the range 179Hz - 5600Hz. A useful technique for estimating the upward and downward spread of masking of a noise upon speech is also incorporated. For good intelligibility, the articulation index ≥ 0.7, although 0.5 is sometimes considered satisfactory. Typical relationships between articulation index and speech material are illustrated in FIG. 2.

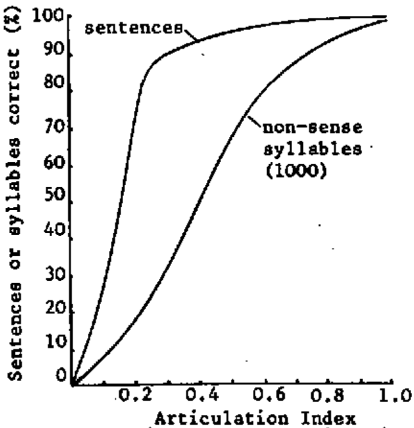

FIG.2. Approximate A.I. relationships

3. WHAT ABOUT REVERBERATION?

Much work has been carried out to study the effects of reverberation time and reverberation decay characteristics on the intelligibility and clarity of speech. Early studies by Beranek (21) suggest that the useful speech is the speech power contained in the first 50m secs after arrival of the first speech pulse at the ears of the listener. Later work by Lochner and Burger (22, 23, 24) suggests that only those reflections arriving within 95m secs after the direct sound have a positive contribution to speech intelligibility. Mankovsky (25) suggests that 63m secs is the period when useful energy may be extracted from the reflections. All of these studies demonstrate the importance of the steepness of the decay characteristic during the first stage. This conclusion contrasts with the previously held view of requiring a short overall reverberation time for good speech intelligibility. Mankovsky (25) concludes that it is sufficient to increase the steepness of the decay curve by 5-6 dB in order to obtain the same clarity for a total reverberation of 3 secs., as for a total reverberation of 1 sec.

4. PRACTICAL COMPUTATION OF ARTICULATION INDEX

The articulation index for a large enclosure eg. theatre, may be computed using the one-third octave band method as follows:

(a) Derive the reverberant speech spectrum at the receiving position by applying the usual formula, at each one-third octave band centre frequency, viz.

$$SPL_f = PWL_f + 10\log_{10}\left[\frac{Q_f}{4\pi r^2} + \frac{4}{R_f}\right] \text{ dB,}$$

where all symbols have their usual meaning. Values for Q_f and PWL_f can be obtained from most standard acoustics texts eg. Beranek (21).

(b) Determine the background (masking) noise for each one-third octave band, and correct for any spread of masking effects.

(c) Determine the speech peak-to-noise ratio for each one-third octave band, and multiply each by the appropriate frequency band weighting factor (20). This weighting factor allows for the maximum contribution from each band and also corrects for the frequency transformation from the original 20 equally contributing frequency band model to the 15 one-third octave band model.

(d) Add together the numbers obtained from (c). The resultant is the articulation index.

5. CONCLUSION

The articulation index is not going to replace the intelligibility test which has a unique role to play in speech work. It is a supplementary process, which can be used to rank-order the performance of listening situations and give a good indication of the resulting intelligibility, and as such it is an important tool for use in room acoustics design.

References

(1) J. M. PICKETT 1956 J. Acoust. Soc. Amer., 28, 902-905. Effects of vocal force on the intelligibility of speech sounds.

(2) W. H. SUMBY and I. POLLACK 1954 J. Acoust. Soc. Amer., 26, 212-215. Visual contribution to speech intelligibility in noise.

(3) P. O. THOMPSON and J. C. WEBSTER 1963 Acustica, 13, 319-323. The effect of talker-listener angle on word intelligibility.

(4) E. D. SCHUBERT 1956 J. Acoust. Soc. Amer., 28, 895-901. Some preliminary experiments on binaural time delay and intelligibility.

(5) G. A. MILLER, G. A. HEISE and W. LICHTEN 1951 J. Exp. Psychol., 41, 329-335. The intelligibility of speech as a function of the context of the test materials.

(6) C. W. STUCKEY 1960 Rome Air Development Center, Report TN-60-198. Development of a method to evaluate learning in articulation testing.

(7) J. P. EGAN 1948 The Laryngoscope 58, 955-991. Articulation testing methods.

(8) L. L. BERANEK 1949 Acoustic Measurements, Ch. 17, John Wiley, New York.

(9) I. J. HIRSH et al 1952 J. Speech Hear. Disorders, 17, 321-337. Development of materials for speech audiometry.

(10) D. L. RICHARDS and J. BERRY 1952. Post Office Research Station, Report No. 13596. Tables for the generation of logatoms based on conversational speech.

(11) N. T. BOWMAN 1974 J. Sound Vib., 32, 109-129. The articulation index and its application to room acoustics design.
(12) H. K. DUNN and S. D. WHITE 1940 J. Acoust. Soc. Amer., 11, 278-288. Statistical measurements on conversational speech.
(13) D. L. RICHARDS and J. N. SHEARME 1954 P. O. Elect. Engrs. J., 47, 159-161. The measurement of speech level.
(14) D. L. RICHARDS and R. B. ARCHBOLD 1956 Proc. IEE, 103B, 679-691. A development of the Collard principle of articulation calculation.
(15) N. R. FRENCH and J. C. STEINBERG 1947 J. Acoust. Soc. Amer., 19, 90-119. Factors governing the intelligibility of speech sounds.
(16) L. L. BERANEK 1947 Proc. IREE, 35, 880-890. The design of speech communication systems.
(17) H. FLETCHER and R. H. GALT 1950 J. Acoust. Soc. Amer., 22, 89-151. The perception of speech and its relation to telephony.
(18) K. D. KRYTER 1962 J. Acoust. Soc. Amer., 34, 1689-1697. Methods for the calculation and use of the articulation index.
(19) K. D. KRYTER 1962 J. Acoust. Soc. Amer., 34, 1698-1702. Validation of the articulation index.
(20) AMERICAN NATIONAL STANDARD ANSI S3.5 1969. Methods for the calculation of the articulation index.
(21) L. L. BERANEK 1954 Acoustics. McGraw-Hill.
(22) J. P. A. LOCHNER and J. F. BURGER 1958 Acustica, 8, 1-10. The subjective masking of short time delayed echoes by their primary sounds and their contribution to the intelligibility of speech.
(23) J. P. A. LOCHNER and J. F. BURGER 1960 Acustica, 10, 394-399. Optimum reverberation time for speech rooms based on hearing characteristics.
(24) J. P. A. LOCHNER and J. F. BURGER 1961 Acustica, 11, 195-200. The intelligibility of speech under reverberant conditions.
(25) V. S. MANKOVSKY 1971 Acoustics of Studios and Auditoria. Focal Press.