EVALUATION OF AN OBJECTIVE METHOD FOR MEASURING SPEECH INTELLIGIBILITY

L.J. Webb

Department of Applied Acoustics, University of Salford, SALFORD M5 4WT

INTRODUCTION

Traditional subjective methods of measuring speech intelligibility are time consuming, expensive and the results are open to wide variation due to human error. In comparison, an objective measurement system should be more consistent, quicker and (after the initial purchase price), cheaper.

Early studies in this field carried out by French and Steinberg [1] yielded the Articulation Index (AI), which was further developed by Kryter [2]. However, the complexity of the AI calculation scheme makes it cumbersome. Probably the most important recent development is the application of the Modulation Transfer Function (MTF) to speech intelligibility measurement. The MTF of a transmission channel is the reduction in modulation index of an initially 100% modulated signal, as a function of frequency. This work was pioneered by Houtgast and Steeneken [3], whose efforts resulted in the Speech Transmission Index (STI). Their concept has now been embodied in a measuring system, RASTI (RApid Speech Transmission Index), made by Bruel and Kjaer. The STI is calculated from a large number of MTF's in different octave bands in a similar way to the AI scheme, and RASTI uses a simplified version of this.

Owing to the complexity and cost of the RASTI method there appears to be scope for the development of a simpler and cheaper system. This paper describes the evaluation and use of such a system, which was first developed by Kihlman and Nordlund [4], and recently used by Barron [5].

THEORY

To understand how the MTF of a room relates to speech intelligibility it is necessary to look at why intelligibility is not always perfect. In a theatre, the sound arriving at a listener's ears will have been affected in some way by background noise and reverberation. Speech may be thought of as a sound with a specific distribution pattern of sound intensity over time and frequency [6]. When this sound reaches the listener its distribution pattern is much less clearly defined, and the degree of this "smearing" is an indication of the reduction in intelligibility. The MTF measures the degree of loss of clarity of a signal, and so must relate in some way to intelligibility.

The MTF is defined by Schroeder [7] as:

> "the expected value of the complex amplitude of the squared output of a system at radian frequency ω, divided by the average value of the squared output at ω".

where the system is fed with a test signal of cosine modulated (frequency $\omega/2$) white noise.

It is possible to measure the MTF of a speech signal, but this is not necessary provided that the modulation rate of the replacement test signal is similar to that of the intensity envelope of continuous discourse. Other features of the normal transmission channel should be reproduced as closely as possible. In the case of a theatre, the loudspeaker radiating the test signal should have the same directional characteristics as a person, and the signal should be received by a dummy head containing two microphones.

The modulation frequencies most prevalent in speech have been shown to lie between 0.25Hz and 25Hz [3]. The method used by Houtgast and Steeneken for measuring the MTF (to obtain the STI) was to use cosine wave modulation at the relevant frequencies and analyse the response of the room in $^2/_3$ rd octave bands, giving a table of 98 values.

A simpler method used by Kleiner [8] was to modulate white noise with a rectangular wave. The fourier analysis of this signal reveals its usefulness; for example with a period of 140ms it contains components at frequencies of 7, 14 and 21Hz, which are similar to those required. The MTF was then measured according to Schroeder's definition: the complex amplitude divided by the average value. As the system used was intended to give a quick measurement of speech intelligibility analysis was carried out for a single frequency band between 500Hz and 2kHz.

Kleiner measured the modulation index reduction (or MTF) m as:

$$m = (L_{m1} - L_{Dm1}) - (L_{m2} - L_{Dm2}) \quad dB.$$

where L_{m1} and L_{m2} are the average values of the test signal before and after transmission respectively.
L_{Dm1} and L_{Dm2} are the amplitudes of the signal before and after transmission.

L_{Dm1} and L_{Dm2} were measured by a simple envelope detector. This may be simplified by designing the envelope detector so that it may be switched to give either the signal or the amplitude at its output, and calibrated to make $L_{m1} = L_{Dm1}$ before transmission. This leaves m as a simple ratio: $-(L_{m2}-L_{Dm2})$ in db (or, V_{Dm2}/V_{m2} with values in volts).

EXPERIMENTAL METHOD

The equipment used may be split into two parts (fig 1): the transmitter and the receiver. The transmitter was required to generate a test signal of noise, modulated with a rectangular wave, and radiate it into a room to replace a person speaking. The receiver picks up the altered signal and provides a measure of its amplitude and average value from which the MTF can be calculated.

A Bruel and Kjaer white noise generator was used to provide a wideband noise signal, which was then chopped at the correct frequency and filtered by the 'filter and gating' unit. The signal was then amplified by a Quad II power amplifier, which drove the loudspeaker. In an attempt to achieve the required directional characteristic, the speaker drive unit was chosen for its size - approximately that of a human mouth. Considerable difficulty was encountered in finding a speaker of this size whose frequency response covered the 500Hz to 2kHz range of the signal. The unit used was a Peerless 2" midrange dome, which very nearly met the frequency requirements. This was mounted in the lower third of a small wooden enclosure, designed to have dimensions similar to those of a human head. The speaker cabinet was supported by a tripod at the approximate height of a person's head for the measurements.

The receiving section should ideally consist of a dummy head, but this was not possible within the scope of this project. Instead an omni-directional capacitor microphone attached to a CEL sound level meter was used. This was connected directly to the envelope detector (fig 1). The resulting levels were read from a measuring amplifier. The envelope detector was calibrated before testing by connecting it after the 'filter and gating' unit, and adjusting the calibration control until the amplitude and average levels were equal.

Measurements were carried out in two lecture theatres in the Chapman Building at the University of Salford. These particular theatres were chosen because they are known to have intelligibility problems. In the smaller of the two, Chapman 3, which seats 150 people, intelligibility is normally good. However, there is a ventilation system that is switched on and off periodically which increases the degree of concentration required from the listener. Chapman 1 (seating 500 people) usually requires the use of a public address (PA) system. Tests were carried out to evaluate the sound reinforcement system by comparing intelligibility with and without the PA. The results are shown in the next section. This theatre was also used to evaluate the objective measuring system against articulation tests (results will be presented at the conference).

RESULTS

Articulation tests were performed in Chapman 3. These indicated good intelligibility throughout the room with the ventilation system switched on. The front row yielded the highest scores, which would be expected due to the strong direct sound. STI values were measured with and without the interfering noise. The results are shown in table 1. Fig 2 shows the measuring positions.

Seat No Row	No Ventilation Noise			With Ventilation Noise		
	2	8	13	2	8	13
A	0.49	0.62	0.46	0.46	0.58	0.42
C	0.45	0.48	0.45	0.45	0.47	0.42
F	0.40	0.41	0.38	0.43	0.41	0.40
H	0.45	0.42	0.44	0.43	0.38	0.36
K	0.35	0.35	0.36	0.33	0.34	0.34

Table 1: STI Values, Chapman 3

Most seats show a decrease in intelligibility with the ventilation system on, which is as expected. From table 1 it can be assumed that an STI of 0.62 represents good intelligibility, and that values down to 0.32 are fair.

STI measurements have been made in Chapman 1 (fig 3); the results are displayed in table 2. Seats near the front show reasonable intelligibility even without the PA. All readings taken near the back of the hall are very low, which corresponds to the comments of students using the room regularly. The measurements taken in the back rows using sound reinforcement are particularly interesting as the STI values are lower than those without. The average signal level was higher with the PA in all cases, but the amplitude readings were smaller, resulting in low STI's. This suggests that intelligibility will always be low in these seats. Large background noise levels were often observed at the rear of the hall, due to the fact that it backs onto the gallery, which is in constant use. The largest improvement in intelligibility with the PA was observed in seats at the sides of the theatre. This is what would be expected of an efficient system as more sound is required in areas receiving least direct sound.

Row+ Seat No	STI	STI with PA	Row+ Seat No	STI	STI with PA
A14	0.47	0.44	I18	0.22	0.23
A25	0.32	0.33	I25	0.23	0.28
E1	0.28	0.33	N1	0.26	0.25
E10	0.33	0.35	N10	0.29	0.26
E25	0.29	0.35	N25	0.28	0.25
I6	0.26	0.28	R1	0.28	0.26

Table 2: STI Values, with and without PA, Chapman 1.

CONCLUSIONS

Word tests carried out in a small lecture theatre (Chapman 3) correspond well with the high STI values measured, particularly at the front of the room. Lower STI values were obtained in the presence of high background noise as would be expected. Measurements carried out in Chapman 1 to evaluate the sound reinforcement system showed improved intelligibility when the PA was used, but only in certain areas. The biggest improvement was observed at the sides of the theatre, indicating that the PA performed as required. Results obtained so far with this objective measuring system indicate that it correlates well with expected speech intelligibility in the rooms tested.

REFERENCES

[1] French, N R and Steinberg, J C. "Factors Governing the Intelligibility of Speech Sounds". J.Acoust.Soc.Am. 19, 90-119 (1949).

[2] Kryter, K D. "Methods for Calculation and Use of the Articulation Index". J.Acoust.Soc.Am. 34, (11), 1689-1697 (1962).

[3] Houtgast, T and Steeneken, H J M. "The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Intelligibility". Acustica, Vol 28, 66-73 (1973).

[4] Kihlman, T and Nordlund, L. "Speech Intelligibility in Auditoria", BFR Report R61, (1973).

[5] Barron, M. "Speech in Theatres: What are the Important Considerations?" Proc. IOA, Vol 8: Part 3, 371-378 (1986).

[6] Houtgast, T and Steeneken, H J M. "A Review of the MTF Concept in Room Acoustics and its use for Estimating Speech Intelligibility in Auditoria". J.Acoust.Soc.Am. 77 (3), 1069-1077, (1985).

[7] Schroeder, M R. "Modulation Transfer Functions: Definition and Measurement". Acustica, Vol 49, 179-182, (1981).

[8] Kleiner, M. "Speech Intelligibility in Real and Simulated Sound Fields". Acustica, Vol 47, 55-71, (1981).

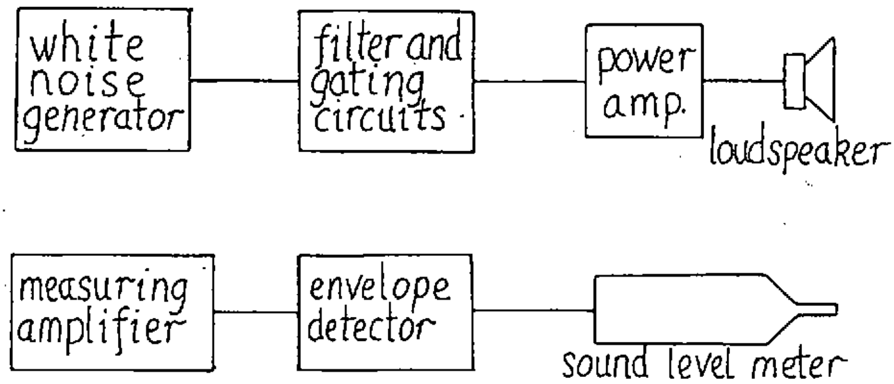

Figure 1: Block Diagram of Measuring System.

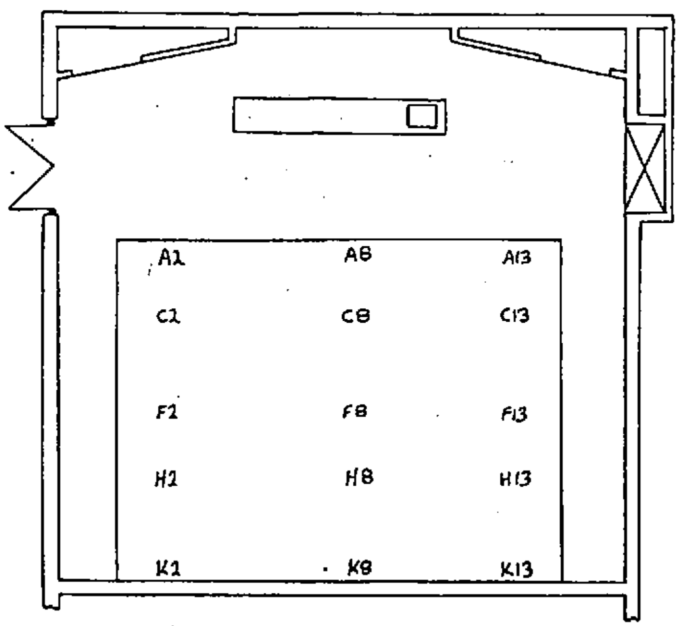

Figure 2: Plan of Chapman 3 Showing Measuring Positions.

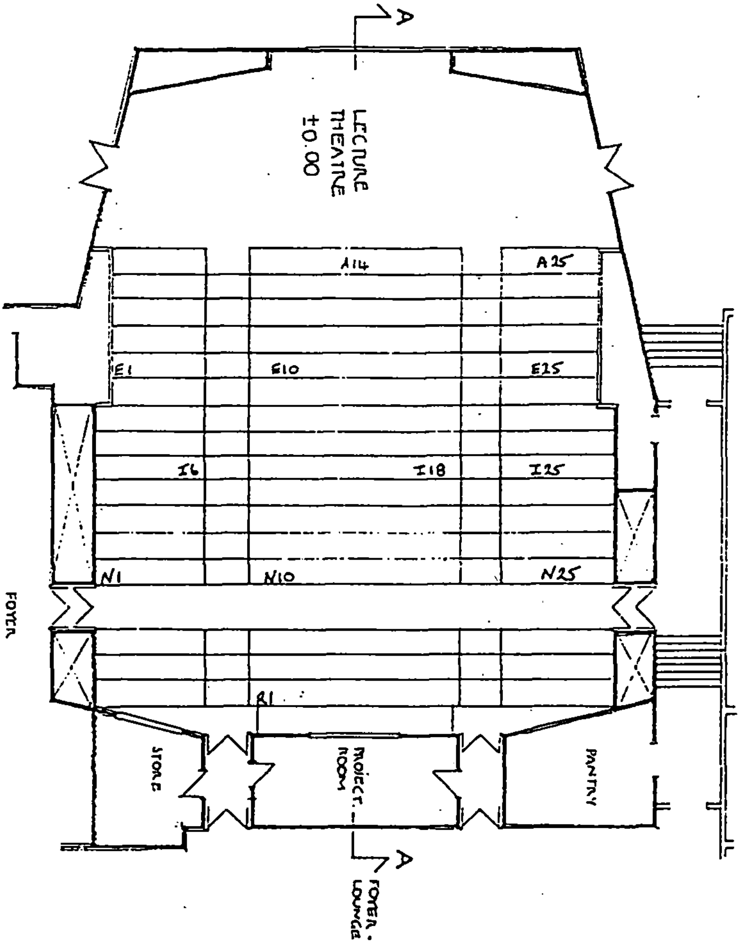

Figure 3: Plan of Chapman 1 Showing Measuring Positions.