MEASURING SPEECH PRIVACY – AN INVERSE INTELLIGIBILITY PROBLEM ?

Peter Mapp

Peter Mapp Associates, Colchester, Essex C06 1LG

ABSTRACT

The results of a number of experiments using STI and Speech Privacy Index to measure the degree of speech privacy / confidentially between rooms are reported. Both STI and STIPa measurements are compared. Large discrepancies were found to occur not only between the metrics but also between different measurement equipment. The paper discusses these errors, their causes and the potential for using STI to measure speech privacy.

INTRODUCTION

The need to be able to guarantee that confidential speech occurring in adjacent rooms could not be overheard or understood caused the author to review and experiment with various methods of speech privacy / lack of intelligibility measurements. In particular, being somewhat familiar with STI and its derivative STIPa (Ref 6,8,9), these two metrics were further investigated in an attempt to see if an 'on the spot' reading of privacy / confidentiality could be attained. Comparisons with traditional measures such as the Articulation Index (Ansi S3.5 1969) and a newer but still indirect Speech Privacy index were also made. It was also hoped that the STI / STIPa approach would enable the additional degradation of intelligibility due to room reverberation components to be accounted for – a factor not incorporated into either the traditional AI or Speech Privacy Index techniques. (Though the very much more sophisticated complex Speech intelligibility Index (SII) method would allow this). Furthermore, it was hoped that by using an STI based technique that a direct, on the spot reading of intelligibility / confidentiality could be obtained rather than a later computation as is the case with AI for example. The author also wanted to ensure that the absolute speech Sound Pressure Levels were also taken into account, as this factor is also of considerable significance.

Of particular concern to the author was the ability of the instrumentation to not only to be able to pick out the STIPa test signal modulation from the background room noise, but also to be able to do this at relatively low sound levels. Assuming for example that speech is typically produced at around 60 – 65 dBA in an office or similar environment, and that a screen or partition is likely to achieve at least 25 - 30 dB attenuation, then the received signal levels are likely to be in the order of only 30 – 40 dBA and may be buried in up to 10 – 15 dB of noise.

In order to rate the level of Speech Privacy / confidentiality and potential distraction, the author devised the following 5 point scale or categories : (1) Not audible – total confidentiality (2) Speech sounds audible (eg cadence) but no words discernable – i.e. confidential (3) Occasional words discernable – reasonable privacy & little distraction (4) words fully understandable - no confidentiality / privacy and strong possibility of distraction. (5) Words completely intelligible & content potentially distracting.

Scale Rating	1	2	3	4	5
Subjective speech	Not audible	Audible / unintelligible	Occasional words intell	Most or all words intell	Words intell & distracting

From a review of the STI scale, it would seem likely that in order to attain reduced distraction and confidentiality, that STI values of below at least 0.3 and 0.2 respectfully would be required, indicating an equivalent SNR of at around at least -7 dB.
In order to be able to accurately track the results and potential problems, experimentation began using two different sets of rooms with known poor speech privacy. The results of the initial experiments using a fairly simple set up and test procedure are the ones reported here.

The speech source employed was a fully characterised 4inch cone based loudspeaker that was equalised to exhibit a very flat axial frequency response. Although in previous papers, (Ref 2 & 3) the author has shown the directivity of the test loudspeaker is a significant factor to accurately determining the STI of a space or system, for the purposes of the initial experiments the non directionally ideal source was favoured due to its impeccable frequency response characteristics. (Fig 1). (It is likely that in fact that for room-room tests an omnidirectional source should be used)

Both test room sites exhibited low background noise levels (20 & 36 dBA respectively), which meant that little natural noise masking was present. In site 1 therefore, the background noise level was also artificially increased in order to reduce the potential intelligibility. In order to obtain a subjective feel for the situation, speech was auditioned at different a series of different levels ranging from 50 – 70 dBA (at 1 metre from the source).

STI / STIPa test signal levels of 60 & 70 dBA (1m) were employed. Six different measurement platforms were tested. Two of the platforms were software based and four were instrument based.

INITIAL TESTS

The first test room had a very low (but by no means unusual) background level of 20 dBA. Partition loss was approximately 26 dBA. Subjectively, speech at 70 & 60 dBA in the transmission room was not only audible but completely intelligible in the adjacent receiving room. Even with speech at 55 dBA, most words and certainly the overall meaning of a sentence were discernable, so subjectively confidentiality had not been achieved.

The initial series of tests gave rise to an unexpectedly wide range of results varying from 0.82 down to 0.27 STI - for a completely identical condition ! This clearly required further investigation before the rest of the measurement programme could continue. Table 1 below summarises the initial test results.

Table 1 STI / STIPa Values for Room 1 with low background noise

Meter	P1	P2	A	B	C	D	D*
STI 70 dBA	0.82	0.82/0.77	0.58	0.62	0.62	0.39	0.65
STI 60 dBA	0.82	0.82/0.77	0.40	0.41	0.57	0.27	0.50
BGN 20 dBA							

Two results are shown for Platform P2. The higher value of 0.82 is in agreement with P1 but this did not incorporate a speech-weighting filter. The lower value of 0.77 is the result obtained when using an appropriate filter. Meter D* which is in general agreement with B & C at 70 dBA speech level, employed a different microphone to meter D, which exhibited the poorest agreement and was completely at odds with subjective impression. Figure 2 below shows the above results graphically

Figure 2 STIPa results for 4 meters – at 60 & 70 dBA transmission SPL

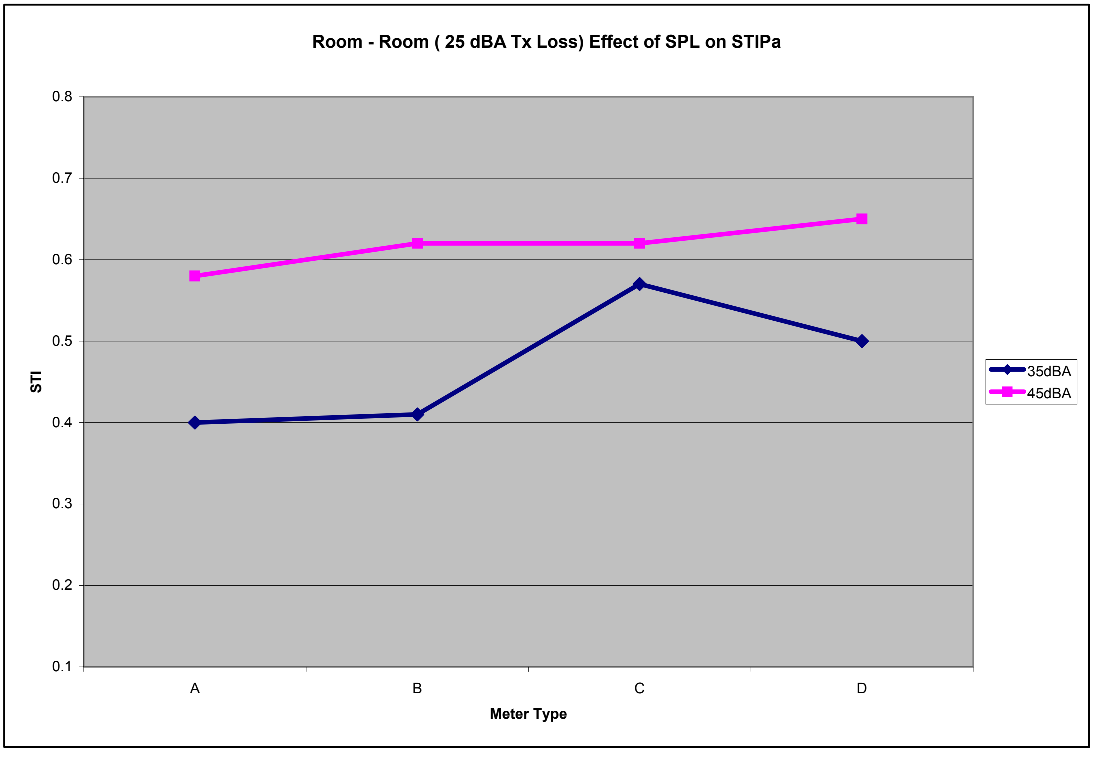
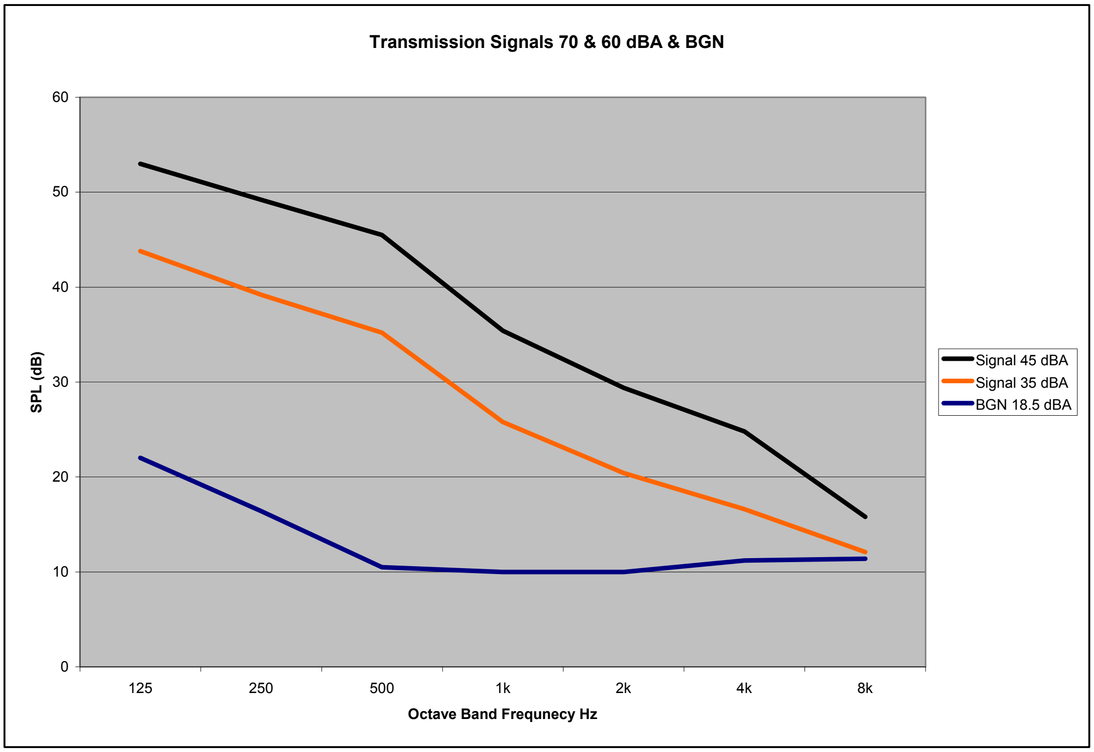


Figure 3 Received speech spectra & background noise

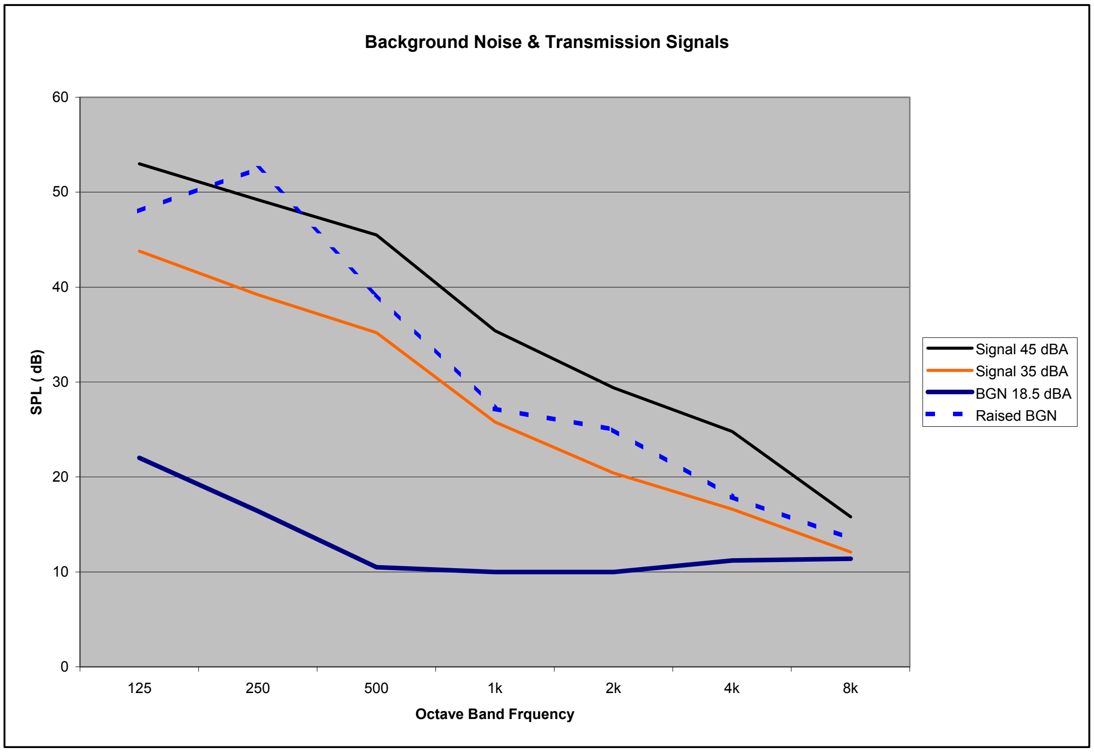


The background noise level in the receiving room was then increased from its normal value of around 20 dBA to 44 dBA using a speech like noise masking spectrum (see figure 4). The Results obtained with the raised background noise level condition are presented in table 2 and figures 4 & 5 below.

Table 2 STI / STIPa Values for Room 1 with raised background noise

Meter	P1	P2	A	B	C	D	D*
STI 70 dBA	0.83 (R)		0.51	0.53	0.55	0.34	0.55
STI 60 dBA			0.29	0.26	0.34	0.23	0.31
BGN 44 dBA							

Figure 4 Speech Transmission and background noise levels



As can be seen from the above table, with a speech “talker” level of 70 dBA, meters A, B, C & D’ are in good agreement, but meter D is clearly at odds with the others - as is the result obtained from the software based measurement. Subjectively, with speech production levels of 60 & 70 dBA, the received speech was still audible and intelligible. The Meter STI readings at 70 dBA therefore may be about right. However, at 60 dBA, the speech heard in the receiving room was still intelligible, with most words being discernable as indeed was the understanding of test sentences. This suggests that all the meters were underestimating the intelligibility. However, without, full blown word score testing, it s not possible to verify this. Meter C gave the highest value (at 0.34 STI). Interestingly, from a technical point of view, the author would have expected meter C to probably be the most accurate under these conditions.

Figure 5 STIPa results with a raised background noise level of 44 dBA

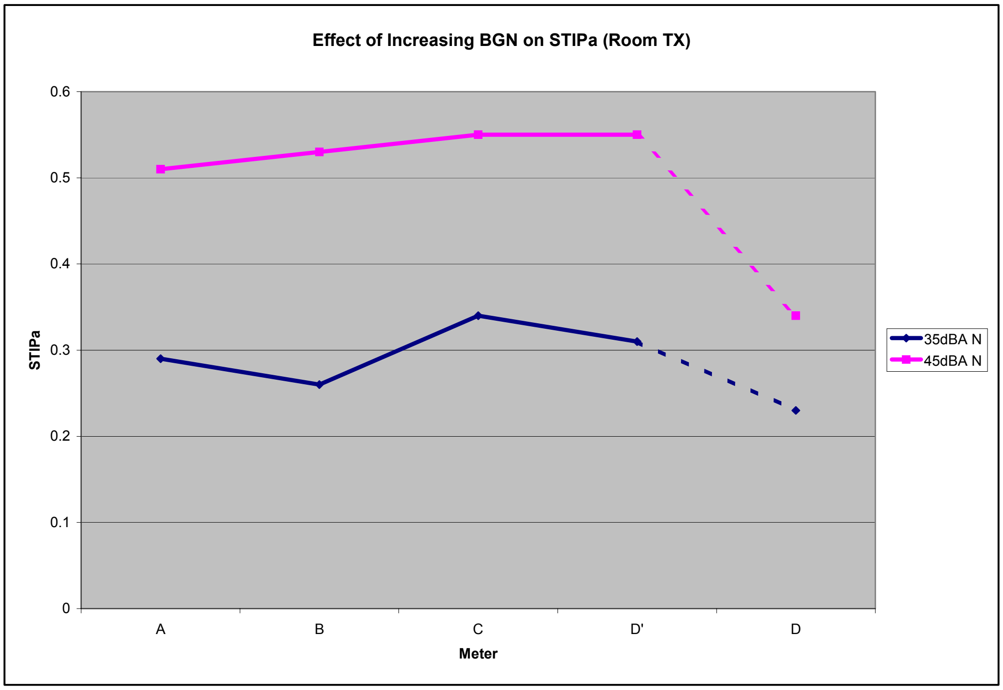
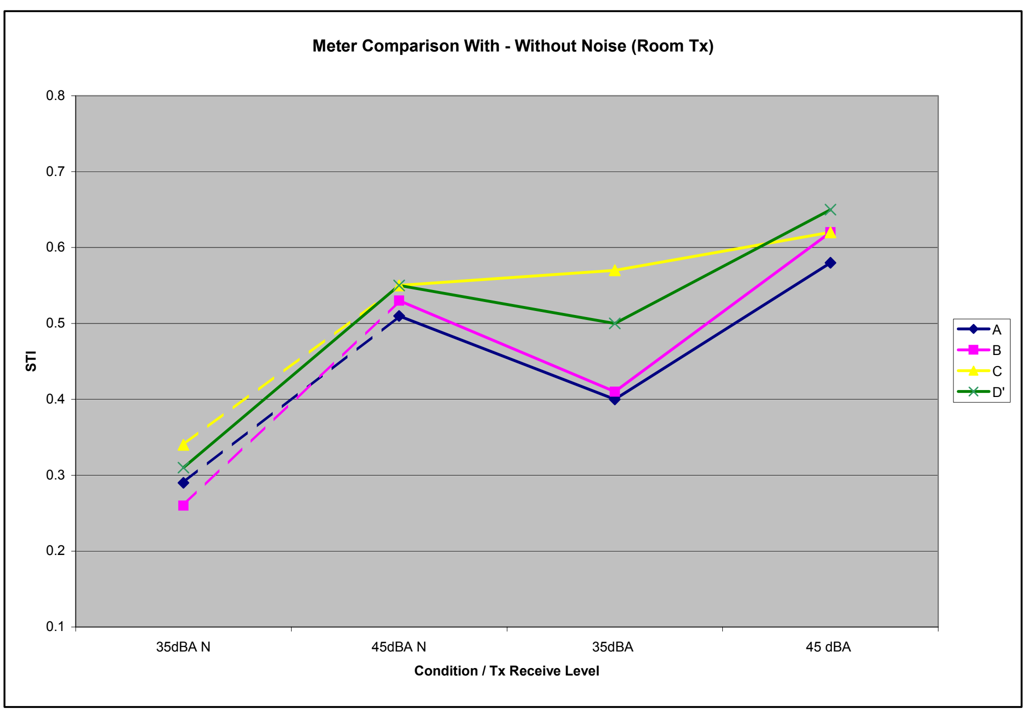


Figure 6 Summary - comparing meter results for 4 transmission conditions



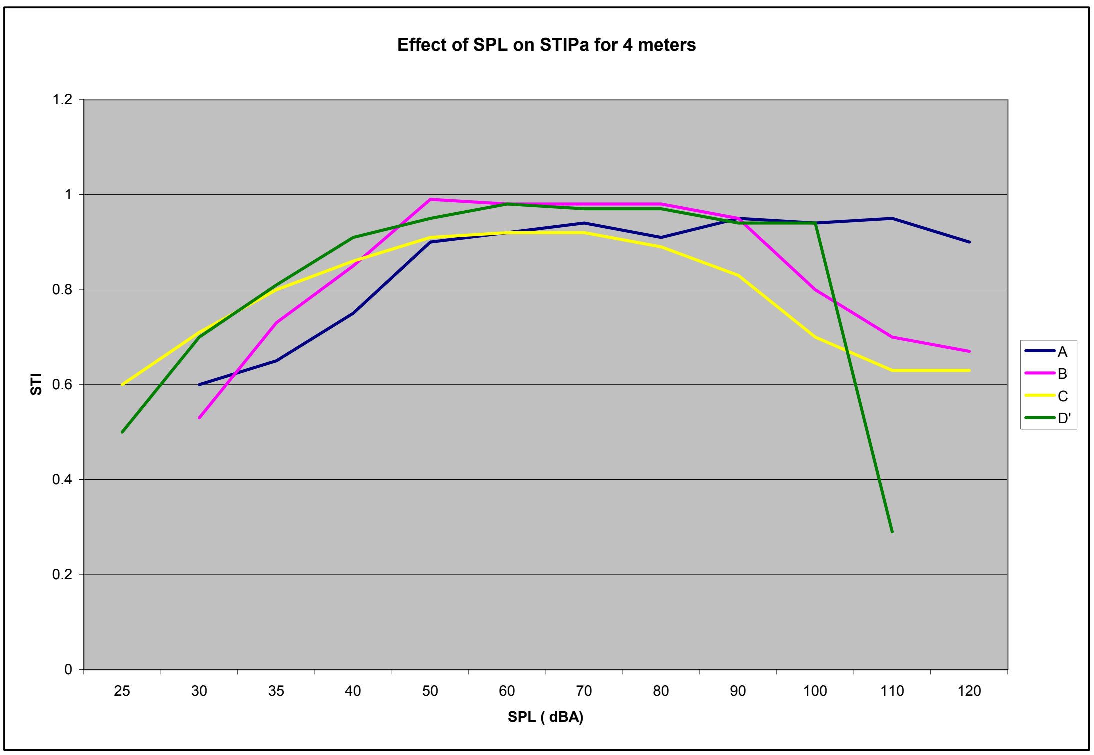
From figure 6 above, the largest discrepancy between the meters (ignoring meter D which was clearly incorrect) occurred with the lower voice level of 60 dBA, where a variation of 0.17 STI occurred. With the 70 dBA STIPa signal the variance was 0.07 STI. Interestingly the lowest level of variation occurred at the 70dBA STIPa level and the raised background noise condition. (variance was 0.04 STI).

STIPa METER TESTS & CHARACTERISTICS

The large discrepancy noted above, between the various STIPa meters, was investigated to see if the reasons for this could be identified. The meters were therefore lab tested under a range of identical SPL and signal to noise ratio conditions.

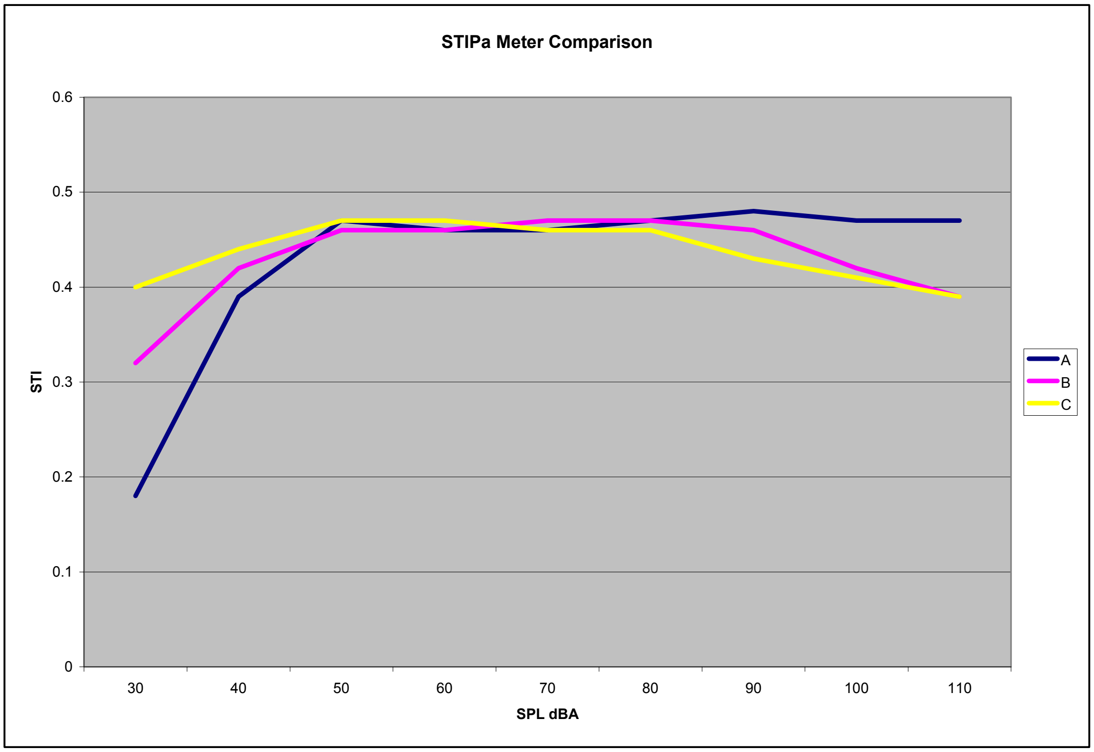
Figure 7 below, shows how the meters varied over an Sound pressure Level range of 25 to 120 dBA. As the figure shows, a number of significant differences and discrepancies were found.

Figure 7 Comparison of STIPa meter characteristics over range of 25 – 120 dBA



Over the range 50 – 80 dBA there is reasonable good agreement between the meters, with two distinct trends occurring. Meters B & D’ tracked each other well as did meters A & C though there was an apparent offset of around 0.06 STI between the two sets of readings. However, the variations noted between the 2 groups of meters shown in figure 7 above, are not true offsets, as shown in Figure 8 below. Here, the nominal STI has been reduced to 0.46 and now the meters agree very much more closely.

Figure 8 – Comparison of STIPa meters at 0.46 STI



The discrepancies between the meters can be explained by a number of factors, but primarily relate to the dynamic range capabilities of the meters, their self noise characteristics and the masking algorithms employed.

The differences between the meters at low signal levels is of critical importance to the accurate measurement of speech privacy /confidentiality. From figure 8 above it can bee seen that at 40 dBA there is a variation of around 0.05 STI, which increases to 0.22 STI at 30 dBA. A series of further tests were therefore carried out that varied the signal to noise ratio of the modulated STIPa signal which was held constant at a reference level of 40 dBA.

The results of this experiment are shown in Figure 9. All 3 meters can be seen to work satisfactorily over the test range of +/- 15 dB SNR, though whereas meters B & D track each other well, meter C consistently produces higher readings. (eg at 0 dB SNR, meters B & D computed an STI of 0.48 whereas meter C computed 0.55.

Whereas, theoretically one might expect the STI to be 0.5 at 0 dB SNR, this assumes that the noise and STIPa spectra match exactly. Although the two spectra were close (see figure 14), the match was not exact and so some apparent discrepancy would be expected. Meter C, is in fact in fact providing a more accurate answer. This is shown by the curves plotted in figure 10, which shows what happened when the test was repeated at 60 dBA. (There is also an indication of the potential error by the plots shown in figures 7 & 8).

Figure 9 Effect of different Signal to Noise ratios at a fixed SPL of 40 dBA.

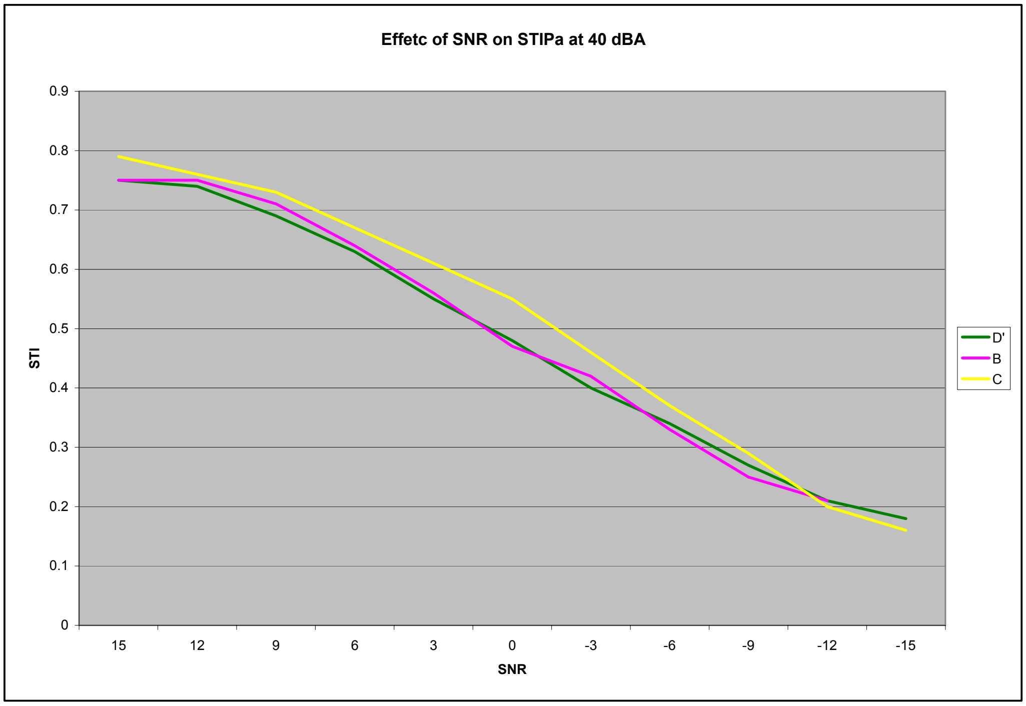
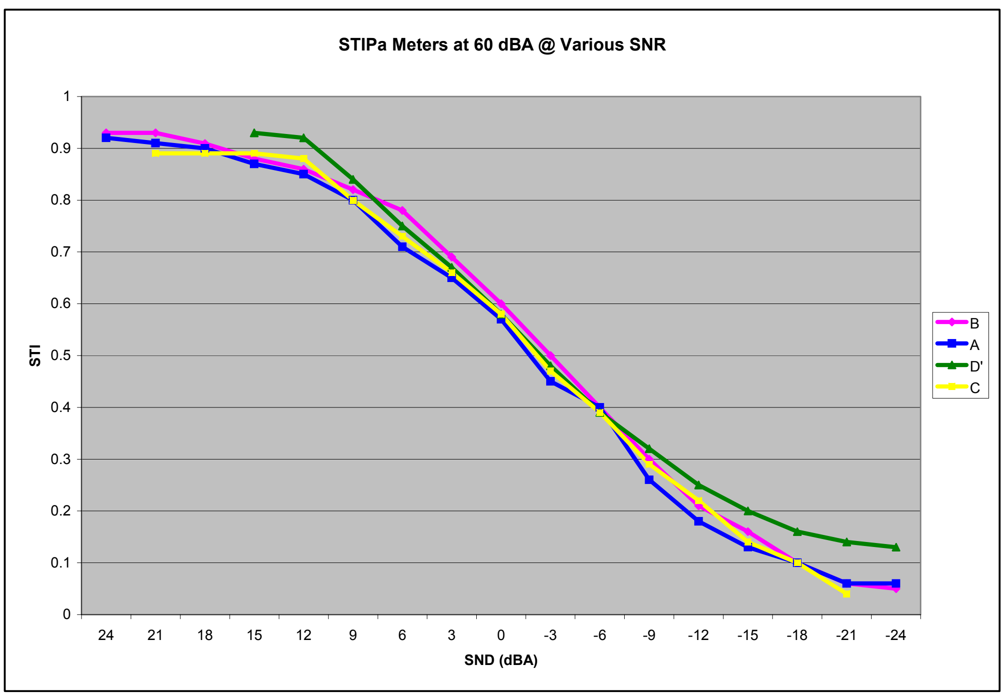
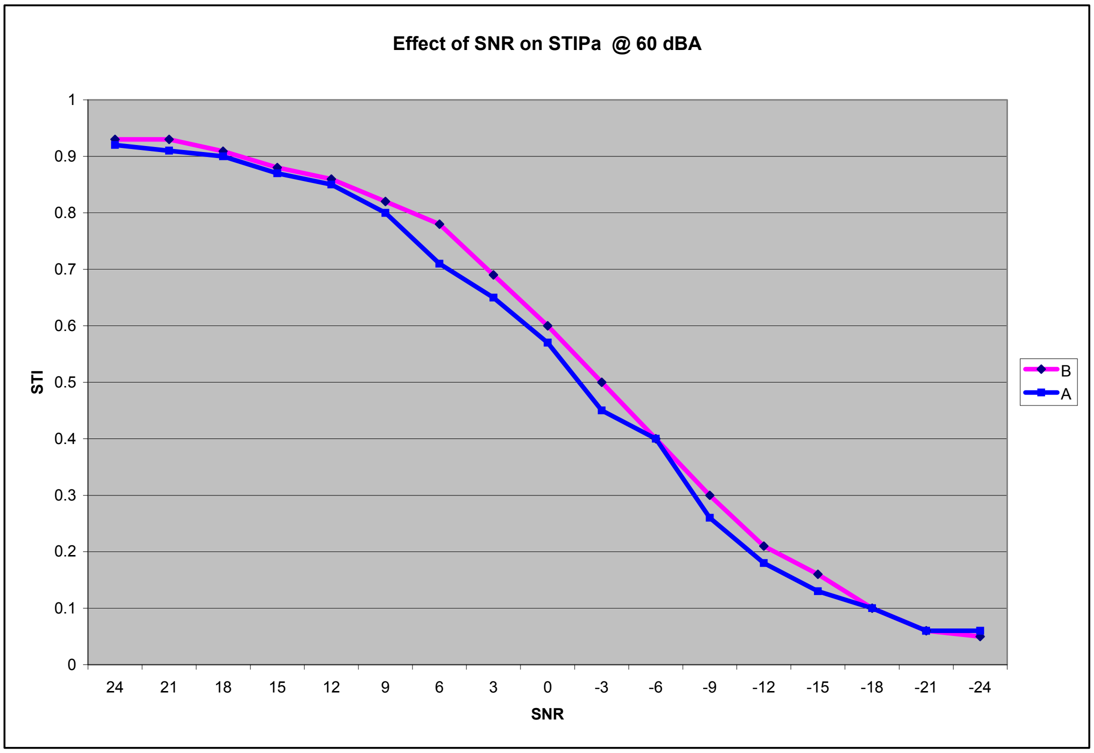


Figure 10 Effect of different Signal to Noise ratios at a fixed SPL of 60 dBA



At 60 dBA signal level (figure 10), the meters agree quite well, particularly over the range encompassing 0.8 to 0.3 STI. At the extremes of the scale, the variance increases. As can bee seen from the figure, to be below 0.2 STI, which may be taken as the threshold for privacy, one needs a signal to noise ratio of at least –15 dB. In order to ensure that the STIPa meters could cope with a non-constant background noise level, the simulated background noise used a random noise signal. Most of the meters coped well this, but meter A did exhibit a greater fluctuation in readings than most. This can be seen for example by comparing it to meter B over the same range and conditions as shown in figure 11 where the multiple curves tend to hide this.

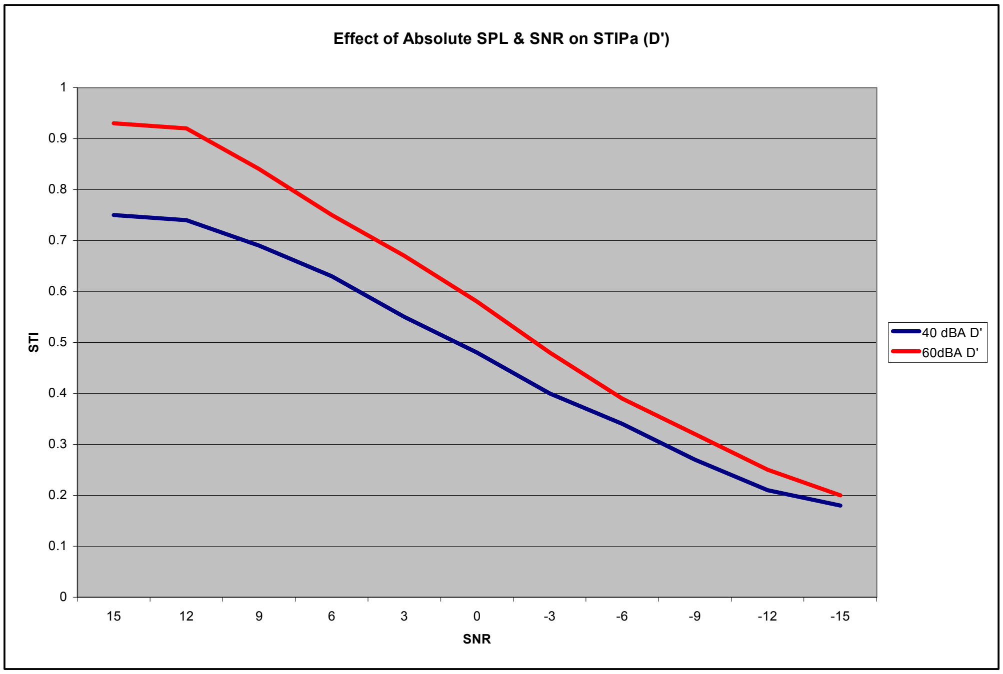
Figure 11 – Comparison of meters A & B at 60 dBA for range of +/- 24 dBA SNR



As already noted, figure 9 shows the STI / SNR plots for meters B, C & D' at 40 dBA. It can immediately be seen from the figure, that meters B & D' agree very closely with each other whilst meter C produces significantly higher readings, except at the extremes of the scale, where all the meters agree.

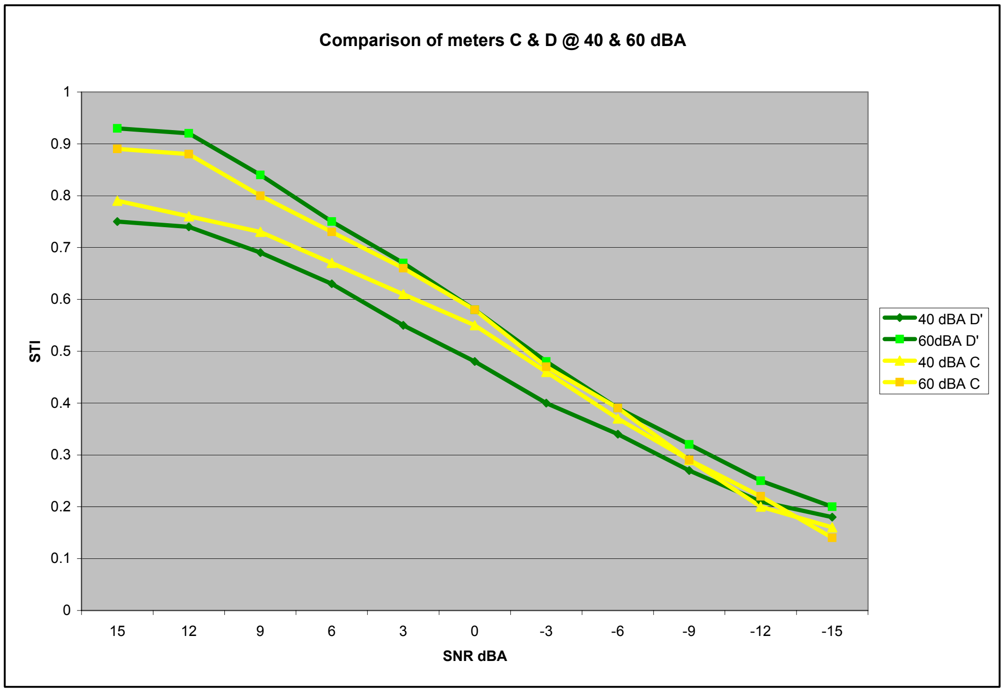
The difference between the SNR curves obtained at 40 & 60 dBA is shown in figure 12 – for meter D'.

Figure 12 Difference between the SNR curves obtained at 40 & 60 dBA for meter D'



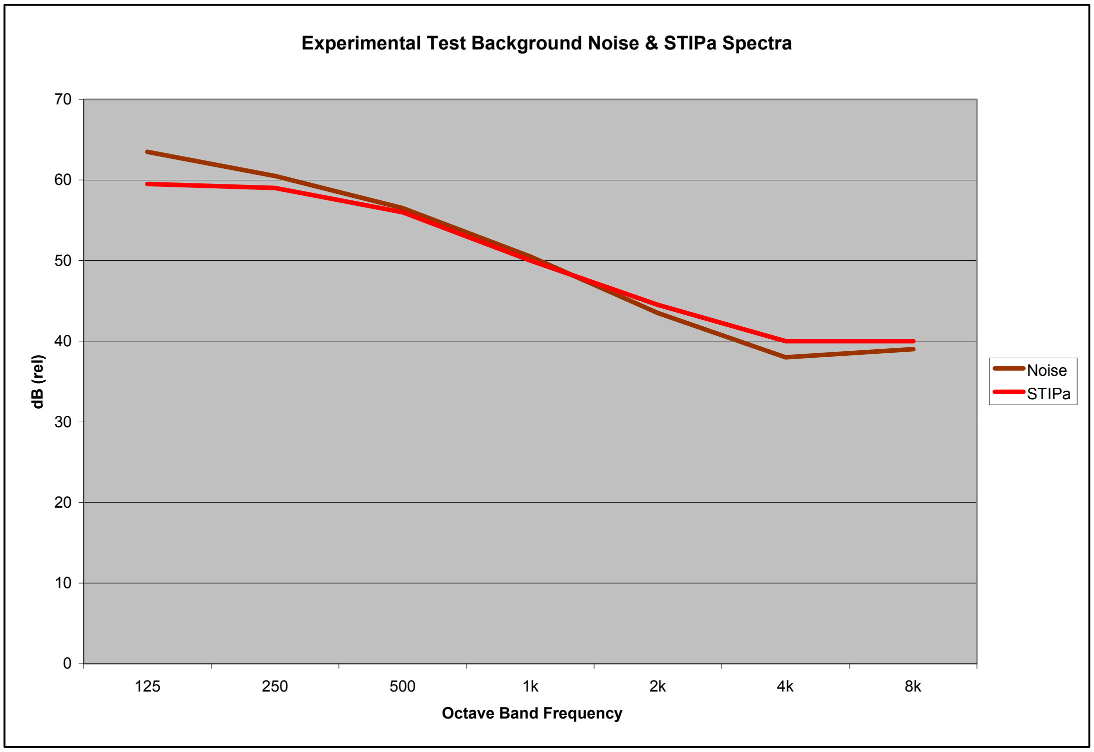
Perhaps surprisingly, the difference between the curves is not constant but decreases with decreasing SNR. To see if this was an oddity of meter D', the experiment was repeated with meter C. The results are shown in Figure 13 below.

Figure 13 Comparison of meters C & D' at 40 & 60 dBA



Although the basic trend is the same for both meters, meter C shows slightly less difference, particularly negative signal to noise ratios.

Figure 14 Test Background Noise & STIPa Spectra



CONCLUSION & COMMENTS

The foregoing analysis shows that in principle, STIPa should be capable of measuring relatively low speech signal levels over a wide range of signal to noise ratios. However, the self noise of the instrumentation (the detailed analysis is not shown here) as well differences between the meter algorithms did give rise to fairly significant variations in the reported STI values computed by the various meters. The way in which the meters dealt with a random background noise signal was instructive and some further work in this area in respect to some of the meters would be beneficial. Clearly, considerable further research is required before this technique can be more widely adopted, but it would appear that the method should be viable. Again the need for software programs to implement correct speech spectrum filters and absolute sound level calibration is highlighted.
How low STI scores relate with the subjective impression of privacy and confidentially also requires considerable further research. A comparison study between AI, PI and STIPa is currently underway by the author, and it is hoped to report more fully on this at a future conference.

REFERENCES & BIBLIOGRAPHY

1 ANSI standard S3.5 1969 Methods for calculation of the Articulation Index

2 Mapp P – Measuring Speech Intelligibility in classrooms. Proc IOA Vol 25 Pt 7 Sound-Bite conference November 2003

3 Mapp P – The acoustic and Intelligibility Performance of Assistive Listening and Deaf Aid Loop (Afils) Systems. AES 114th Convention Amsterdam, March 2003.

4 Mapp P, The measure of Intelligibility, S&VC Vol 20 No 4

5 Houtgast, T, Steeneken, H & Plomp, R. Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. Acustica vol. 46, 1980.

6 Mapp P Limitations of current sound system intelligibility verification techniques. AES 113th Convention Los Angeles 2002

7 Jacob, K, Steeneken, H, Verhave, J, McManus, S, (2001) ' Development of an Accurate, Handheld Simple-to-Use, Meter for the Prediction of Speech Intelligibility. Proc IOA Vol 23 Pt 8

8 Mapp P Some further thoughts on STI – How accurate are the measurements in practice ? IOA Vol 24 Pt 8, RS 18 Stratford on Avon 2002

9 Mapp P Is STIPa a robust measure of speech intelligibility performance? AES 118 Convention, Barcelona, May 2005