SOME FURTHER THOUGHTS ON STI – HOW ACCURATE ARE THE MEASUREMENTS IN PRACTICE ?

Peter Mapp
Peter Mapp Associates, Colchester, UK, CO3 4JZ. Petermapp@btinternet.com

Abstract

Over the past ten years, there has been an ever increasing awareness of the need for Public Address and Voice Alarm systems to provide emergency as well as general announcements with a high degree of intelligibility. RaSTI (and STI) have been at the forefront of this revolution as the need to measure and verify system performance has gathered importance and momentum. However, measuring PA system performance is quite different to assessing the natural intelligibility of an auditorium or classroom for which the measures were originally intended. The paper discusses the relationship between STI and Rasti, based on examination of the data taken from over 80 sound systems. It is shown that RaSTI is generally an inaccurate predictor of STI for a wide range of conditions. Some of the practical complications and limitations when testing sound systems are also discussed. In particular, it is shown that some forms of signal processing and irregular sound system frequency responses can give rise to STI results with reduced accuracy. Equally, it is shown that different measurement platforms can give significantly differing results. These aspects therefore raise the question ‘ How accurate is an STI measurement ? ’

1. Introduction and background

The past few years has seen an increasing demand for the installation of Voice Alarm systems rather than traditional fire alarms using bells or sirens. The ability of a Voice Alarm system to provide specific information concerning safety related incidents, potentially improves the evacuation response and successful egress from buildings and public spaces. However, to be effective, voice alarm systems need to be able to provide intelligible announcements and messages. The introduction in 1989 of a standard for emergency sound systems (IEC 8049 / BS 7743 - 1991) has had a significant positive impact – particularly in the UK. For the first time, the required intelligibility of the sound system was specified (0,5 STI) with Rasti being the designated measurement / verification method. Other standards and codes of practice codes soon followed including CAA specification No 15 1989 and RTCA/DP-214, 1993, which for example specify given RaSTI performance values for the PA systems of all commercial aircraft. Many military aircraft, also adopt these specifications and RaSTI requirements. Contractually, many systems became liable to meeting a given RaSTI performance and the implications for both sound system design and the acoustic environment in which a given system has to operate, slowly emerged. The result in the UK has been a significant improvement of the quality and intelligibility of PA and VA systems, though at times this has been a painful process, as initially the acoustic implications were not fully understood. Indeed, it still comes as a surprise to many, that the acoustic conditions of a building or space, are the limiting and often the determining factor of the resultant intelligibility of an installed sound system.

Whereas, subject based word score tests are the most accurate method of ensuring that a given sound system is intelligible, this is an extremely expensive and cumbersome technique to conduct and is not a practical assessment option for most sound system installations. The requirement for a simple to use, electroacoustic method, whereby a specific test signal is either electronically or acoustically injected into the sound system and the resultant acoustic response measured was manifestly needed. Whereas the most common technique in Europe is RaSTI, in the USA, a Direct to Reverberant ratio measure relating to percent Alcons is still currently the most frequently employed. [Though the recent introduction in the USA of NFPA 72 and the adoption of the requirements of EN60849 is likely to change the USA’s reliance on % Alcons].

Concern has been expressed over the accuracy of both the above techniques when applied to sound systems. (Mapp, 1997, 2001,2002) As early as 1991 Mapp showed that it was possible to radically alter the frequency response of a sound system and for RaSTI not to recognise this. Figure 1 shows an example taken from this early investigation.

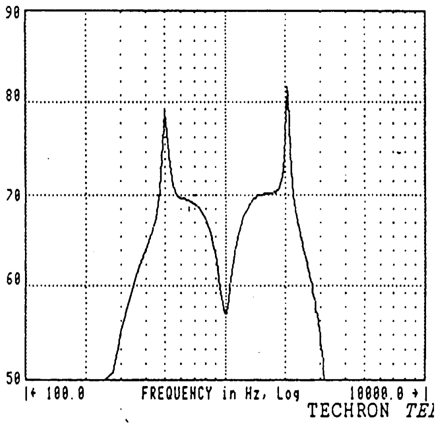

Fig 1. Effect of Limited frequency information on Rasti

Mapp also pointed out at this time that system non-linearity such as deliberately introduced signal compression could give rise to misleading results. It is interesting to note that both RaSTI and % Alcons were originally developed to assist with the assessment of natural speech transmission in auditoria or other acoustic spaces and not sound systems (eg IEC 268 pt 16, ‘The objective rating of speech intelligibility in auditoria by the Rasti method’ - 1988). The application to the measurement of PA and sound systems followed later - without the realisation of the unique environment and conditions under which such systems operate and the potential for error that this might introduce.

The revised version of EN 60849 (1998) – Sound systems for Emergency Purposes, recognised some of the potential problems with RaSTI and introduced the concept of the ‘Common Intelligibility Scale’ or CIS. The concurrent revision of IEC 268-16 also recognised some of the problems and added some additional guidance with regard to measurement validity. (EN 60286-16 : 1998).

Although there is currently considerable opinion that RaSTI should be abandoned with respect to sound system verification and STI or STIPa employed instead, (eg the current draft revisions of EN 60849 and EN60268-16), there are many standards and codes that still directly cite RaSTI and it will take several years to revise these. It therefore seemed timely (and this symposium offers an ideal opportunity) to present some hitherto unpublished research carried out by the author which illustrates the typical errors associated with RaSTI and it accuracy in predicting STI.

2 RaSTI – STI Error Analysis

As it is the intelligibility performance of sound systems as opposed to rooms or auditoria, that is the current interest to the author, an analysis of a wide range of ‘real world’ sound systems was conducted. In total, the measurement data for eighty-one sound systems was examined and the RaSTI, STI and % Alcons values have been compared, together with a number of other parameters. Subsets of the data (25 systems) have also been subject to further experimentation and analysis. One of the first investigations was to look at the typical variation in RaSTI results one might expect to see when using different measurement systems / platforms. Ideally, of course, one should expect to get identical results between techniques / equipment. The variations encountered were larger than expected. An accuracy of 0.02 STI would seem both reasonable and achievable. It was found that controlling the measurement variables was extremely critical and difficult to achieve under normal field conditions. A set of experiments was therefore conducted under well controlled but non laboratory conditions. Figure 2 shows a typical plot of the variations obtained using three different measurement platforms.

The difference between the measurement systems is not only significant but legally

Fig 3 Measured Rasti Values for 25 reverberant spaces - Using 3 Different Platforms

and contractually worrying ! As has been said many times by the author before, “what was the answer you wanted” ?

In order to reduce the above error, the subsequent tests were essentially restricted to two measurement platforms. The STI – RaSTI comparisons reported, relate to the same MLS based platform. However, other verification techniques have also been carried out in order to ensure that the results relate to RaSTI – STI differences rather than programme software errors.

Review of 81 Sound Systems

The sound systems involved covered a very wide range of types and applications, ranging from rail and transportation terminus systems to churches, theatres, concert halls, museums and stadiums. The range of associated STI values is also very wide but offers a skewed distribution as fewer systems exhibited very low performance results. Figure 3 shows a regression plot of the STI & RaSTI data taken for the situation where the signal to noise ratios were sufficiently high such as to ensure that reverberation is the dominant distortion or speech distractor.

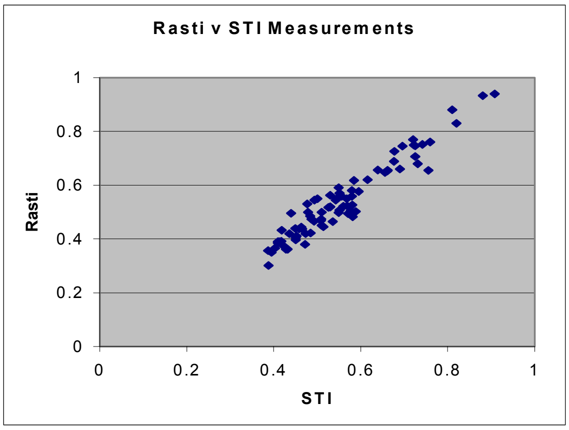

Fig 3. Correlation between RaSTI & STI

Although there is a strong relationship between the data, as would be expected, the correlation is only 0.9097, a significant discrepancy clearly being present.

Figure 4 plots the data in terms of the error between the two measures. The mean error for the 81 systems is 0.08, which might at first glance seem fairly small, though both contractually and subjectively, this is certainly not the case. Examination of the curve shows that at around 0.5 STI (the target criteria for many VA systems) the error is as much as +/- 0.1.

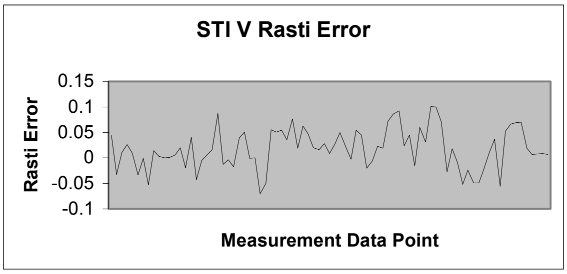

Fig 4. STI Vs RaSTI error curve

Inspection of the error curve (figure 4) shows that the error distribution is not linear but in approximately 2/3 of the cases, the Rasti values underestimate the STI but over estimate it in the other 1/3. An initial review of the data shows no immediate factor that causes either the under or over estimation of the STI by RaSTI. Figure 5 shows the measured STI values versus the range of the Reverberation Times of the spaces employed (approximately 0.2 to 3.5 seconds). It is interesting to note from the figure that no relationship exists between the mean reverberation time (over the range of 500 Hz to 4 kHz) and the STI, although there is clearly a trend showing higher STI values with lower reverberation times – a not unsurprising finding (correlation Coeff = 0.4983).

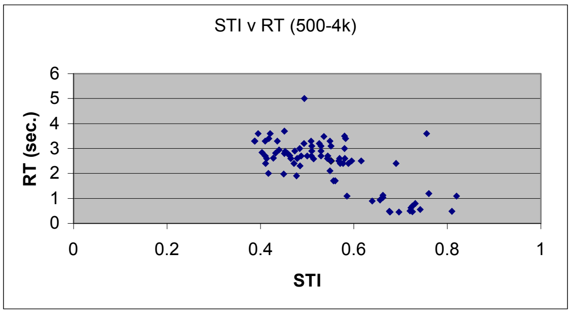

Fig 5 Range of RT values and STI

The effect of noise and reverberation as distractors was also examined at two different signal to noise ratios (+6 & +10 dBA) – but using a smaller sample of 25 systems. These again encompassed a wide range of initial STI values and reverberation times. Figure 6 shows the error graph for the reverberation only case. Again 2/3 of the errors are such as to be underestimating the actual STI, as shown by positive error values.

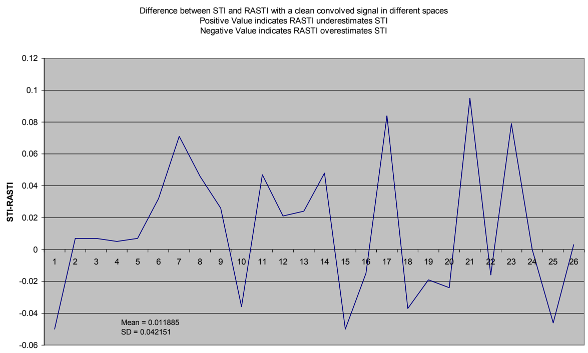

Fig 6 STI – Rasti error (25 systems) Reverberation only

However, in the presence of noise and reverberation, the trend completely changes as shown in figure 7 - the error results for the +10 dBA S/N case. Now the error analysis shows that Rasti is over estimating the STI in 90 % if the cases. The mean error is –0.05. To put this into context, if we assume that Rasti was predicting a value of 0.50 STI ie a 'pass' as far as most emergency sound systems are on concerned, the actual value would in fact be just 0.45 – not only on the limit of practical intelligibility but also a clear 'fail' with respect to many current codes and system requirements ! (The legal and contractual ramifications of this are of significant interest). At +6 dBA S/N, a similar trend was found to occur with the mean error increasing to – 0.065. (Further details can be found in ref. Mapp, AES 112 convention 2002).

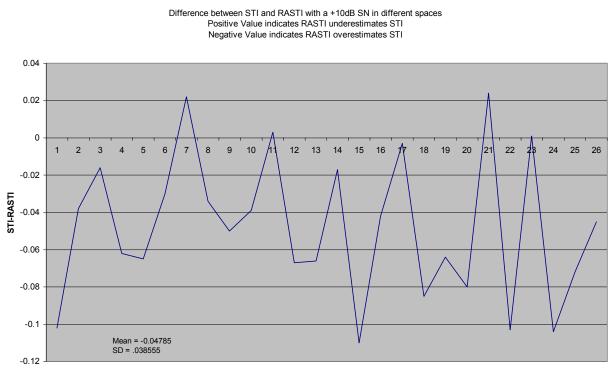

Fig 7 STI – RaSTI error for 10 dBA S/N ratio & reverberation

The effect of noise only as the intelligibility reducing parameter was also investigated. This particular study related to the performance of aircraft PA systems. The reverberation time of the cabin was very short (< 0.2 seconds) this coupled with the short distance between the loudspeaker and listener resulted in very high initial STI and RaSTI scores – as shown in table 1 – column 1.

Table 1 : Aircraft PA System RaSTI & STI Measurements

LS Type	**RaSTI** (no noise)	**RaSTI** (with noise)
A	0.93	0.76
B	0.91	0.65
C	0.95	0.71
	STI (no noise)	**STI** (with noise)
A	0.88	0.60
B	0.92	0.64
C	0.93	0.57

The agreement for the 'no noise' case between Rasti & STI is good, the variations being within 0.02 to 0.05 (mean = 0.03). However under typical aircraft operational background noise conditions, the discrepancy between RaSTI and STI increased markedly. In this case the error range was between 0.1 and 0.16 – again very significant errors. This is particularly so when it is considered that in the case of loudspeaker type 'c' the situation changes from a good pass of 0.71 Rasti against the CAA requirement of 0.60 and the real STI value of 0.57 ! Again technically a failure, but as the CAA certification is in terms of RaSTI and not STI, this becomes a contentious point. Particularly when it is realised that the better performing loudspeaker (B) both subjectively and in terms of STI was the worst performer in terms of RaSTI ! Word score testing carried out in conjunction with the measurements showed good agreement with the STI values. (Further details can be found in Mapp, AES 111 Convention 2001)**.**

The effect of non linear behaviour in terms of signal clipping was also investigated for RaSTI & STI. An interesting discrepancy was found to occur between laboratory tests and real world sound systems, where a 'straight wire' set up underestimated the effect found under field conditions. In each case the signal level was increased to clip the same component in the signal chain. Noticeable differences were found to occur between the reductions in STI & RaSTI. Figure 8 shows the effect of clipping on STI for five different sound systems. The behaviour is highly non linear, with the same degree of clipping having a different effect on the different systems. Comparing the results with figure 9**,** which presents the data plot for the effect on RaSTI, immediately shows the reductions for Rasti to be far smaller. The associated errors range from around 0.04 to 0.14 STI.

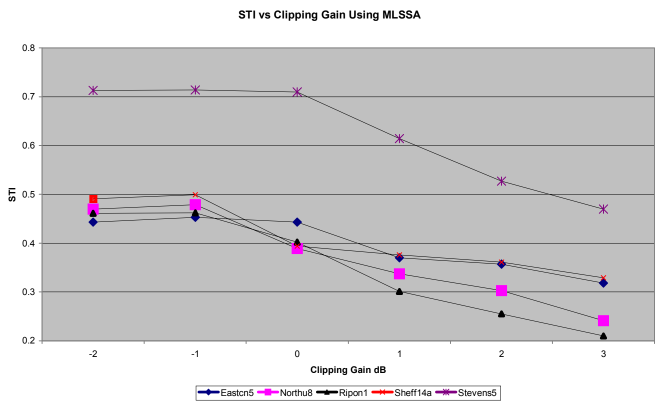

Fig 8 Effect of clipping on STI (MLS measurement)

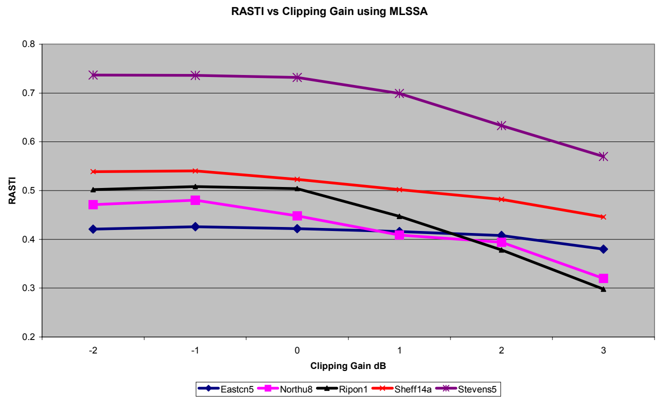

Fig 9 Effect of clipping on RaSTI (MLS measurement)

The above test was conducted using an MLS test signal. An experiment was also run to compare the effects of signal stimulus.

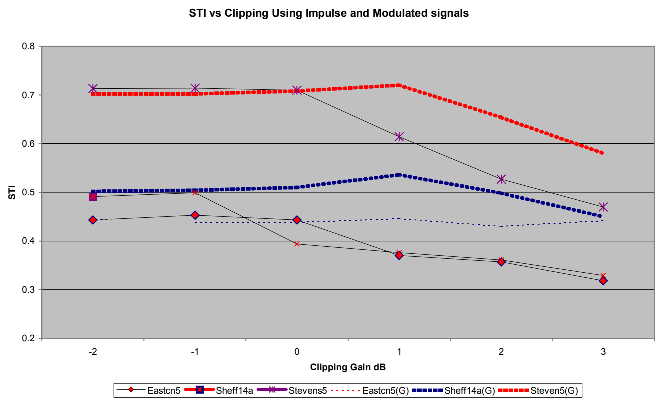

Figure 10
Effect of clipping on RaSTI
(Modulated signal)

Figure 10 shows the comparison between the results obtained using MLS STI and a modulated test signal (STIPa). As can be seen, the STIPa signal is far less sensitive to clipping.

The above examples show not only the potential magnitudes of the errors between RaSTI & STI under typical working conditions but clearly also show the need to retire RaSTI for all forms of PA system assessment. However, the above measurements and results assume that STI is correct and infallible – but how robust is STI ? can it be fooled ? Some preliminary investigations have therefore been carried out and are reported below.

According to EN 6840 –16, STI should be able to cope with a wide range of sound system parameters, including distortion and frequency response anomalies. However, it is the author’s experience that on occasion, the measured system STI does not agree with a subjective evaluation. In particular this seemed to relate to issues regarding the frequency response of a system. This is particularly noticeable when equalising a system, as almost instantaneous comparisons of different frequency response balances and speech clarity / intelligibility can be appraised. Whereas the improvement in the clarity of speech broadcast over the system after appropriate equalisation could generally be readily heard, the measured STI did not appear to reflect this, often not showing any measurable difference at all !

3.1 Residual Measurement Errors in STI

It has previously been noted by the author that electronic measurements of a signal chain, ie before an acoustic output occurs, can result in residual errors of typically up to 0.05 STI. Although this is a small residual error, a loss of 0.05 STI say, when measuring a sound system with a nominal performance of for example 0.50 STI, could become crucial. This is particularly relevant where modulated signals are used (on which STI is fundamentally based). For example, STIPa employs modulated signals recorded on CD. The CD is then played into the system under test. Previous audio measurements made by the author on a range of CD players have shown that there can be significant performance differences. It is beyond the scope of this particular paper to discuss such effects, but the following table summaries the effect on STI (STIPa) using a direct loop back (straight wire) measurement. The table also shows the typical degree of variation between readings obtained when using a pseudo-random based test signal. Though as can be seen, this may also be CD player dependent.

Table 1

CD Player	STIPa Run1	STIPa Run2	STIPa Run3	STIPa Run4	STIPa Run5	STIPa Run6	STIPa Run7	Mean
Marantz 1	0.97	0.97	0.96	0.94	0.96	0.93	0.95	0.95
Marantz 2	0.97	0.94	0.94	0.96	0.98			0.96
Sony Discman 1	0.95	0.99	0.94	1.00	0.95	0.99	- -	0.97
Sony Discman 2	0.93	0.96	0.92	0.96	0.94	0.94	0.96	0.94
Philips 1	0.98	0.99	0.93	0.99	0.99	- -	- -	0.98
Philips 2	0.81	0.81	0.83	0.82	0.81	0.81	0.83	**0.82**
Technics	0.97	0.99	0.98	0.99	0.98	- -	- -	0.98
Goodmans	0.93	0.95	0.92	0.96	0.94	- -	- -	0.94
ALBA	0.97	0.96	0.97	0.97	0.99	0.98	- -	0.97
Citizen	0.92	0.94	0.94	0.94	0.95	0.94	0.94	0.94
DELL	0.99	0.95	0.97	0.99	0.98	0.98	- -	0.98

With the exception of the 'Philips2' machine, where the residual error was 0.82, the other machines in practice gave the correct STIPa, when operating under normal conditions below 0.8 STI. The tests showed that the measured results need not be corrected for a residual error when producing a loop-back result of 0.95 or above.

3.2 Effects of Signal Processing

Most modern sound systems employ some form of signal processing, which can be either digital or analogue or a combination of the two. Typical processing includes, equalisation, compression, filtering, and limiting. Such processing may not necessarily be obvious, as it may be hidden within either a general DSP unit also employed for routing and signal distribution or within devices such as dedicated loudspeaker controllers. AGC (automatic gain control) and automatic noise sensing and gain adjusting circuits may also be incorporated into paging and voice alarm systems. The microphone stations on such systems may also incorporate compression, limiting or expansion circuitry. Some sound systems may also incorporate feedback suppression circuitry or processors. Level dependent frequency response shaping may also occur. The result is that many systems are highly non linear. Whereas it may be possible to bypass or switch out some processing, this is often not possible in practice when auditing a completed system. (Equally, is it reasonable to be expected to have to do this ? An accurate intelligibility measuring system should automatically take such processing into account). Some brief experiments were therefore carried out in order to obtain an idea of the potential problem. Different measurement techniques were also tested so that it would be possible to differentiate between measurement errors or artefacts as opposed to underlying discrepancies with STI itself.

The effects of equalisation and frequency response limiting have been reported elsewhere by the author and have been found to have a very a significant effect.

Compression is another very common form of processing. In that it deliberately changes the dynamic range and crest factor of the signal, it would seem likely that such processing could affect the STI. Compression was applied to a variety of STI test stimuli including MLS, TDS sine sweep and the modulated STIPa signal. A maximum of 6 dB was initially applied as this represents a typical value used in practice. No effect was found to occur with either the MLS or TDS signals but a noticeable reduction in STI did occur with the STIPa signal. Typically, on average, the STIPa of a 'straight wire system', reduced to around 0.74 ie a reduction of 26 %. This is a very significant reduction, particularly as it occurs before the transmission of the test signal to the loudspeaker and out into the space. It means for example that an intelligible system, with a nominal STI of say 0.55 – well inside the usual safety system criteria of 0.5 STI (0.7 CIS) would reduce to 0.29 STI or 0.46 CIS – a dismal failure with indications of being almost totally unintelligible ! The need to check that the system is behaving totally linearly is therefore essential when using such a signal. (A similar effect was also found to occur with an amplitude modulated RaSTI signal). Some further tests are currently being conducted to quantify the problem for a wider range of situations.

Feedback is a limitation of almost any sound reinforcement system and various techniques have been developed to improve the gain before feedback margin. Currently, broadband equalisation used in conjunction with narrowband notch filtering is probably the most common technique and generally employs digital filters. Other techniques include phase and frequency shifting and signal decorrelation. Tests on typical digital filter set ups resulted in STI reductions of around 0.05 STI when using MLS signals and so again care clearly needs to be taken when evaluating systems with nominal values in the region of 0.5 STI. However, little or no difference was found when using the STIPa signal. A number of tests were also conducted on phase / frequency shifters. Using the impulse based techniques eg MLS / TDS led to completely anomalous results being created with the received signal either being regarded purely as noise or as a highly distorted artefact. MLS based signals for example computed STIs of around 0.03 to 0.04 ! Clearly, there is a need to either bypass or deactivate such devices when testing. However, at least in this case the result is obviously wrong, whereas with other forms of processing, this may well not be the case. Interestingly, no problem was encountered when the tests were repeated with the STIPa signal and analyser, which completely ignored the frequency / phase shifting processing.

3.4 Data Compression and digital signal stores

Most emergency sound systems and many PA systems incorporate pre-recorded messages. These days they almost exclusively rely on some form of digital recording and storage. It would seem both logical and desirable to be able to test not only the effectiveness of such stores but to also incorporate an 'intelligibility test signal' within the system enabling the system effectiveness to be checked or verified. A number of formats were checked with both MLS and STIPa amplitude modulated signals. Table 2 summaries the results of the tests.

Table 2

Format	MLS direct	STIPa direct	STIPa reverb	STIPa noise
Direct	0.99	0.98 - 0.99	0.46	0.56
DAT	0.98 – 0.03 *	0.98	0.46	0.56
CD	0.74 to 0.99	0.98	0.46	- - -
minidisc	0.03 to 0.99 *	0.98 – 0.96	0.46 – 0.47	0.57
Wav pcm 44kHz	0.99 *	0.98	0.46	- -
Mpeg 3 128 KBps	- -	0.98	0.46	- -
Mpeg 3 64 KBps	- -	0.98	0.46	- -

* Spectrum distorted

From the above table it can be clearly seen that the STIPa signal is very much more robust than its MLS counterpart. The results suggest that the STIPa signal should be able to be stored or incorporated directly into emergency and PA systems as a test signal.

Further information on potential system and measurement errors can be found two other papers by the author (Mapp, AES 100 Convention 1996 & Mapp, AES 113 Convention 2002).

3.5 Differences between Measurement Platforms & Equipment Validation

Considering the potential legal and contractual implications that underlie many STI measurements, it is surprising and potentially scandalous that the majority of the measurement platforms and programmes on the market have never been properly verified and validated ! Indeed, the only systems currently available that have been validated by TNO and verified against their reference system are the B&K RaSTI and Goldline STIPa instruments. Given this lack of validation, it is surprising, how many people unquestionably use a particular platform and can not believe it may be giving them an inaccurate answer. When environmental noise measurements are made, it is accepted without question that the meter being used must be calibrated and proof of this calibration should be available. Yet there is no such requirement when making Intelligibility performance measurements ! It therefore seemed appropriate to compare a number of the current platforms under 'controlled conditions'. In order to protect the guilty, the various systems are identified purely by a designatory letter. Two experiments were carried out. The first employed six different measurement systems, set up in a, reasonably good, acoustic environment to measure the performance of a sound system at exactly the same position. The second experiment, took four of the platforms and carried out a far more rigorously controlled investigation. This was conducted using three different positions in a highly reverberant church. Tables 3 & 4 set out the results.

Table 3

Platform	A	B	C	D	E	F	Variation
STI	0.76	0.70	0.74	0.64	0.68	0.61	0.15 STI

As can be seen from the above table, the measured results have a range of 0.15 STI. This is an appalling variation, but shows exactly what is happening in practice.

Table 4

Platform	A	B	C	D	Variation
STI Pos1	0.76	0.67	0.77	0.80	0.13
STI Pos 2	0.61	0.61	0.62	0.65	0.04
STI Pos 3	0.47	0.42	0.48	0.49	0.07

Table four shows similar variations and puts the accuracy and reliability of current measurement systems clearly into perspective.

Since the introduction of STIPa last year at RS 17, there has been considerable interest in not only the method but its accuracy – particularly as compared to Mlssa. It therefore seemed timely to carry out some comparison methods between the two techniques. (NB It is the accuracy of Mlssa that is being tested, as STIPa is the validated system). Under laboratory conditions, using reverberation as the destructive element to intelligibility, the two systems agreed extremely well, being within 0.02 STI for a wide range of cases tested. Under carefully controlled, steady state noise conditions the measures also tracked well. Under normal field operating conditions some slightly greater variations were observed. The variations were generally found to stem either from the differences between the speech weighting filters employed by the systems or due to the differing measurement sampling & integration times, a problem that has also been identified by other researchers (Steeneken 2002, Hodgson 2002). System non-linearities will of course also affect the Mlssa stimulus to a significantly greater extent than the modulated STIPa signal. Table 5 sets out some typical comparison measurements made in a reverberant space, using a high density, distributed sound system.

Table 5

Position	1	2	3	4	5	6	7	8	9	10
STIPa	0.37	0.39	0.47	0.45	0.44	0.52	0.48	0.33	0.40	0.48
Mlssa	0.40	0.40	0.46	0.42	0.44	0.51	0.44	0.31	0.42	0.48

It is interesting to note that the mean values for the space, as measured by the two different methods, are identical at 0.43 STI. The standard deviations are also comparable.

4 Conclusions

1. It has been shown that considerable discrepancies occur between the STI and RaSTI values when measuring or verifying the performance of typical sound systems. Of the 81 systems studied, under reverberant and high S/N ratio conditions, Rasti underestimated the STI in 66% of the cases. However a brief analysis of the data showed there to be no immediately apparent consistent mechanism responsible for this. It would therefore appear that it is not possible to predict the conditions whereby RaSTI will either under or overestimate the STI for the case of reverberant reduction of intelligibility.

2. In the presence of background noise at +10 and +6 dBA and reverberation, the above trend was completely reversed and in the vast majority of cases Rasti overestimated the STI.

3. In the case of low reverberation and background noise, Rasti was found to significantly overestimate the STI.

4. When high levels of harmonic distortion were present (ie signal clipping) RaSTI was found to underestimate STI.

5 The findings clearly show that RaSTI is an unreliable predictor of STI when measuring the performance of sound systems. This has been shown to be the case for a wide range of typical conditions and system topologies.

6 Both anecdotal experience and limited subjective testing using 3 different sets of conditions and test subjects, indicate that STI does not appear to be able to accurately predict intelligibility for low noise, reverberant conditions with irregular or falling high frequency responses.

7 Good agreement was found to occur between impulse based STI measurements (as measured using Mlssa and TEF TDS techniques) and the recently introduced STIPa modulated signal with a sparse mtf analysis matrix. This indicates that the discrepancies noted above are inherent within STI as opposed to being a measurement artefact or systematic measurement error. (However, under field measurement conditions slightly greater variations were noted).

8 Good agreement was found to occur between STI and word scores for systems operating under low reverberation conditions but in the presence of background noise.

9 Little difference was found to occur between conventional STI and the male weighted STI under reverberant and reverberation with noise conditions. However the female weighting did produce slightly improved values, with a mean improvement of 0.02 for the low noise reverberant condition. In the presence of background noise and reverberation, the scatter increased significantly. Increases of 0.1 STI were noted in twenty percent of the cases when background noise was present and overall there was an increase in the STI for sixty percent of the samples when the female weighting was applied.

10 Typical sound system signal processing was found to have stimulus dependent effects and anomalies. Whereas for example, simple amplitude compression affected a modulated stimulus (as would be expected) up to 6 dB of such compression had little or no effect on MLS or TDS based signals and measurement systems. Equally however, time variant processing had no discernable effect on STI Pa but very badly affected MLS and TDS FFT based analysis to the extent that no useful signal could be recovered.

11 The research shows that although STI is a good predictor of sound system intelligibility under many conditions, measurements made under low noise, reverberant conditions, on systems exhibiting irregular frequency responses may give rise to erroneous results with the potential intelligibility being over estimated.

12 It was found that when tested under controlled conditions, significant differences in the results could occur when using different measurement platforms. Some platforms were clearly less prone to error than others. The accuracy of several systems currently on the market would appear to be highly suspect.

13 The research findings published in this paper raise a number of interesting contractual and legal implications relating to system verification and performance measurement.

14 Systematic calibration and measurement procedures are required for carrying out measurements for verification or contractual purposes. Current standards need to be updated to include further guidance on such measurement methods and calibration procedures.

15 Further research is required to increase the accuracy of STI for a wider range of typical sound system responses and operating conditions.

5 References

CAA, (**1989**) 'Public Address Systems' Specification no. 15.

IEC 268 pt 16 (**1988**) The objective rating of speech intelligibility in auditoria by the Rasti method. Later to become EN 60268-16, (**1998**). Objective Rating of Speech Intelligibility by Speech Transmission Index.

IEC 849 (**1989**) (later EN60849) and BS 7443 1990 (later BS EN6084 : 1998) 'Sound systems for emergency purposes'.

Barnett, P, (**1997**) Implications of Amplitude Compression on RaSTI Performance. Proc IOA Vol 19, Pt 6. p. 213

Hodgson M (**2002**) – Private communication October 2002.

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.

Mapp, P (**1991**), reprinted in Handbook for sound engineers 2nd edition, Ed Ballou, Pub. Focal Press, Chapter 32, Speech Intelligibility by Davis & Davis.

Mapp, P (**1996**), 'A Comparison between STI and RaSTI Speech Intelligibility Measurement Systems'. 100th Convention AES, Copenhagen.

Mapp, P, (**1997**), 'Limitations of Speech Intelligibility Methods.' 133rd Meeting ASA, Pennsylvania.

Mapp, P, (**1997**), Some Effects of Equalisation and Spectral Distortion on Sound System Intelligibility. Proc IOA Vol 19, Pt 6. p. 245

Mapp, P, (**2001**), 'Limitations of current sound system intelligibility performance measurement techniques and metrics'. 142nd Meeting ASA Fort Lauderdale, Florida.

Mapp, P, (**2001**), 'Improving the Intelligibility of Aircraft PA Systems'. 111th AES Convention New York.

Mapp, P, (**2002**) 'The measure of Intelligibility', Sound &Video Contractor, Vol 20 No 4,

Mapp, P, (**2002**) 'Relationships between Speech Intelligibility Measures for Sound Systems', 112th AES Convention Munich.

Mapp, P, (**2002**) 'Limitations of current sound system intelligibility verification techniques' 113th AES Convention Los Angeles.

Mapp, P, (**2002**) Practical Application of STI to Assessing Public Address and emergency Sound Systems. International Symposium on the Past, Present & Future of STI, Soesterberg, Netherlands, October 2002.

Steeneken, H, (**2002**) International Symposium on the Past, Present & Future of STI, Soesterberg, Netherlands, October 2002.