

# A COMPARISON OF STI-PA AND SWEPT SINE WAVE STI MEASUREMENT METHODS AND RESULTS

Tony Stacey    AMS Acoustics Ltd

## 1. INTRODUCTION

The ability to predict the speech transmission index (STI) of a voice alarm system is an essential part in the design of that system. In fact, the primary aim of the acoustic consultant or engineer involved is to ensure that it will be compliant with a standard or specification, or to alert the client that this is not possible due to some constraints (usually architectural). There seems little point in having a method for measuring the STI if the results of the measurements cannot be predicted in advance. To this end, the test signals, test methodologies, and the implications these have on the measured results must be fully understood and appreciated.

This paper considers two methods for measuring the STI of a voice alarm system as follows:

1. Measurement using the NTI STI-PA meter and associated modulated test signal,
2. Measurement using WinMLS 2004 and a swept sine wave.

The above methods are then compared with mathematical predictions using other measured or predicted acoustic parameters. Note that for the purposes of comparison with STI-PA, only the revised STI<sub>r,male</sub> will be considered.

The paper considers only the direct injection of the test signals into a PAVA system via CD. Acoustic injection of the signals via a system microphone requires tight control of additional parameters such as the frequency response of the mouth simulator (or reference loudspeaker), the distance to the system microphone, and the angle of the loudspeaker with respect to the microphone. Note that the CD player showed no measurable time shift measured in accordance with the NTI STI-PA manual.

## 2. METHODS

### 2.1. Mathematical Prediction

The revised STI (STI<sub>r,male</sub> or female) is a function of the following parameters:

- (i.) Reverberation time (T),
- (ii.) Signal-to-noise ratio (SNR),
- (iii.) Direct-to-reverberant ratio (DRR),
- (iv.) Receiver Q (Q<sub>r</sub>),
- (v.) Auditory masking,
- (vi.) Absolute hearing reception threshold,
- (vii.) Male or female weighting factors as appropriate.

Equation 1 shows the formula used for predicting the 98 indices ( $m(F)$ 's) as a function of the above parameters. The first two factors in this equation have previously been introduced<sup>1</sup>.

$$m(F) = \frac{\sqrt{\left(Q_r 10^{DRR/10} + \frac{1}{1+(\omega\tau)^2}\right)^2 + \left(\frac{\omega\tau}{1+(\omega\tau)^2}\right)^2}}{Q_r 10^{DRR/10} + 1} \cdot \frac{1}{1+10^{-SNR/10}} \cdot \frac{I_k}{I_k + I_{am,k} + I_{rs,k}} \quad (1)$$

where:

$DRR$  is the direct-to-reverberant ratio at the receive position in dB

$Q_r$  is the receiver  $Q$  and acts as a  $DRR$  modifier

$\omega$  is the radial frequency  $2\pi F$  in Hz

$F$  is the modulation frequency in Hz

$\tau$  is the exponential time constant  $T/6\ln(10)$  in seconds

$T$  is the reverberation time at the receive position in seconds

$SNR$  is the signal-to-noise ratio at the receive position in dB

$I_k$  is the mean intensity in octave band  $k$

$I_{k-1}$  is the mean intensity in octave band  $k-1$

$I_{am,k} = I_{k-1} amf$  is the mean intensity of the audio masking in octave band  $k$  where  $amf$  is the auditory masking factor (shown in table 1) for octave band  $k$  derived from the intensity of octave band  $k-1$ .

$I_{rs,k}$  is the mean intensity of the absolute hearing reception threshold in octave band  $k$  (see table 2)

Auditory masking factor as a function of level

Octave band level [dB]	46-55	56-65	66-75	76-85	86-95	> 95
Slope of masking	-40	-35	-25	-20	-15	-10
Auditory masking factor ( $amf$ )	0.0001	0.000316	0.003162	0.01	0.031622	0.1

Table 1

Absolute hearing reception threshold

Octave band [Hz]	125	250	500	1k	2k	4k	8k
Absolute reception threshold ( $L_{rs,k}$ ) [dB]	46	27	12	6.5	7.5	8	12

Table 2

In order to find the  $DRR$ , the octave band levels of the direct and total fields were measured in AMS Acoustics' anechoic and reverberant chambers respectively at a known and constant distance from the loudspeaker used in the measurements.

It is important to remember that all sound pressure levels (and there associated intensities) are the total levels in the octave band and so can include direct, reverberant and noise components not necessarily associated with the test signal itself.

When all the  $m(F)$ 's have been determined, the  $STIr_{male}$  is found using the formulae and weighting factors given in BS EN 60268 Pt16: 2003.

## 2.2. NTI STI-PA Measurements

The STI-PA signal is a modulated noise signal with a bandwidth covering the octaves 125Hz to 8kHz. The crest factor is approximately 14dB as shown in figure 1. The frequency response is that of the male spectra defined in BSEN60268 Pt16: 2003. For comparison, figure 2 shows the time line of a short passage of speech where the crest factor is in excess of 20dB. Each of the seven octave bands is modulated with two modulation frequencies making 14 modulation indices in total. Note that this is not in agreement with BSEN60268 Pt16: 2003 where the 125Hz and 250Hz octave bands are said to be combined in some way so making only 12 modulation indices. Figure 3 gives an example of the STI-PA text file output available from the NTI meter.

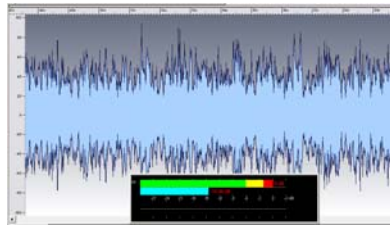


Figure 1 – STI-PA Signal (time line)

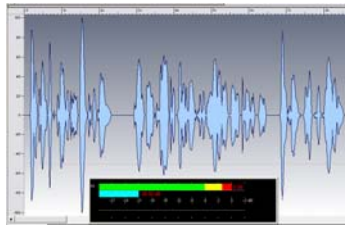


Figure 2 – Short passage of speech (time line)

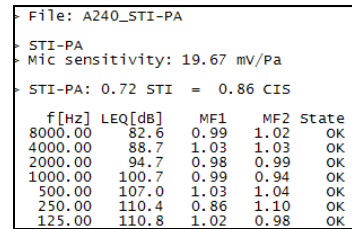


Figure 3 – NTI STI-PA Text file output

It has been the understanding of the author that the advantage of the modulated STI-PA signal is its ability to be used under the 'normal' operating conditions of a PAVA system. Further, that the NTI STI-PA meter is capable of being used by an acoustic layperson<sup>2</sup>. However, both BS EN 60268 Pt16: 2003 and the NTI STI-PA meter manual state that the method must not be used if:

1. the background noise is impulsive,
2. strong non-linear distortion components are introduced into the signal.

The NTI manual states that speech itself should be considered impulsive background noise and since any PAVA system is (under normal circumstances) designed to address the public, it wouldn't be unreasonable to expect some of the surrounding people to be talking during a measurement, potentially causing that measurement to be in error.

The first exclusion would undoubtedly make measurements in train stations during normal operational times very difficult or even impossible. In particular, London underground ticket halls and platforms generally exhibit severe impulsive background noise caused by the continuous opening and shutting of the automatic ticket gates and the thumping of the train wheels as it moves over uneven tracks. Table 3 below shows measurements made during a typical train cycle in a London underground station (i.e. the noise profile between two trains and including the arrival and departure of one train). Note that the cycle was played back into AMS Acoustics' reverberant chamber and so the actual STI-PA values are not realistic, only the variation in STI-PA and the errors (FAULTS) encountered during the train cycle are of interest here. The level of the STI-PA signal was set 10dBA above the  $L_{A10}$  of the 200 second cycle.

NTI STI-PA Measurements During a Train Cycle

Time [secs]	STI-PA	Status
0 - 20	0.60	OK
20 - 40	0.59	OK
40 - 60	0.60	OK
60 - 80	0.47	OK
80 - 100	<b>0.56</b>	<b>FAULT</b>
100 - 120	0.55	OK
120 - 140	<b>0.51</b>	<b>FAULT</b>
140 - 160	0.61	OK
160 - 180	0.60	OK
180 - 200	0.59	OK

Table 3

Table 3 shows that the meter has reported two measurement errors during the cycle and neither is associated with the lowest STI-PA value. The NTI manual states STI-PA measurements are usually overestimated when in the presence of impulsive noise. If a design solution is to be found, or a risk assessment carried out, a single number value is required but what value would an acoustic layperson assign to this platform given the results in table 3? The variation in STI over time is a fact for any train station platform but the time between trains is rarely constant. As an example, to

determine a single noise level and hence single STI value for a typical (or more precisely, published) train cycle, a recording of an actual cycle may need to be truncated or extended in a sound editing package before a single noise level (usually the  $L_{A10}$ ) is found. The octave band noise levels are then used to mathematically contaminate the noiseless STI matrix. This is not something an acoustic layperson should be expected to do. A more worrying prospect is that a single STI-PA measurement taken in space and/or time is used to represent the STI value of an area by a layperson who has been informed that all is needed is to press the button on the meter. Of course, training can always be given but the potential for misuse is very high when compared to more convoluted methods of obtaining the STI.

It should be noted that NTI appreciate the difficulty in measuring STI-PA during normal operational times in public places and so have provided a very useful spreadsheet to allow out of hours STI-PA measurements to be mathematically corrupted with noise taken during normal operational times.

The second exclusion refers to 'strong non-linear distortion components' but neither BSEN60268 Pt16: 2003 nor the NTI STI-PA meter manual offers guidance as to what is meant by 'strong'. Table 4 along with figure 4 show measured NTI STI-PA values (measured directly via a high quality sound card after copying the STI-PA CD to a wave editing package) for a number of compression settings applied to the STI-PA signal and to the real speech passage shown in figure 2 above.

NTI STI-PA Measurements vs Compression (electronic)

Compression scenario	NTI STI-PA (live)	SPL dBAS	Wavelab Compressor settings				Crest factor [dB]	
			Threshold	Ratio	Attack	Release	Speech	STI-PA
C8	0.98	69	-	1:1	-	-	20.9	14.5
C9	0.90	69	-20	2:1	1ms	10ms	17.7	11.6
C10	0.86	69	-20	3:1	1ms	10ms	17.3	10.6
C11	0.84	69	-20	4:1	1ms	10ms	17.0	10.2
C12	0.94	69	-14	2:1	1ms	10ms	18.2	11.8
C13	0.92	69	-14	3:1	1ms	10ms	17.9	11.1
C14	0.90	69	-14	4:1	1ms	10ms	17.8	10.7
C15	<b>0.83 ???</b>	69	-20	8:1	1ms	10ms	16.6	9.7
C16	0.89	69	-14	8:1	1ms	10ms	17.6	10.1

Note that scenario C15 returned an error

Table 4

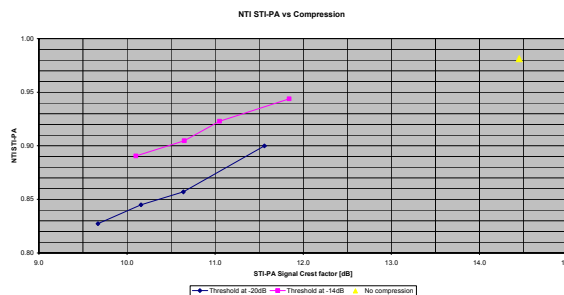


Figure 4 – NTI STI-PA vs. Compression

As can be seen from the plot, the tendency is for the STI-PA to decrease as the crest factor of the signal decreases. Not an unsurprising result given the modulation depth is likely to be reduced by the compression.

Although compression wouldn't be out of place in a PAVA system, it is more common to find limiters, usually on the microphone inputs. Tables 5 and 6 along with figure 5 show measured NTI STI-PA values as a function of limiting and, for completeness, digital clipping.

NTI STI-PA vs Hard Knee Limiting (electronic)

Limit Threshold	Release setting	NTI STI-PA (live)	SPL dBAS	Crest factor	
				Speech	STI-PA
No limiting	-	0.98	69	20.9	14.5
-3dB limiting	10ms	0.98	69	18.1	11.5
-6dB limiting	10ms	0.97	69	15.5	9.0
-9dB limiting	10ms	0.89	69	13.7	7.5
-12dB limiting	10ms	0.84	69	12.3	6.8

Table 5

NTI STI-PA vs Digital Clipping (electronic)

Clipping scenario	NTI STI-PA (live)	SPL dBAS	Crest factor	
			Speech	STI-PA
No clipping	0.98	69	20.9	14.5
3dB clipping	0.98	69	18.0	11.5
6dB clipping	0.97	69	15.1	8.6
8.5dB clipping	0.90	69	12.5	6.0
10dB clipping	0.78	69	10.1	4.1

Table 6

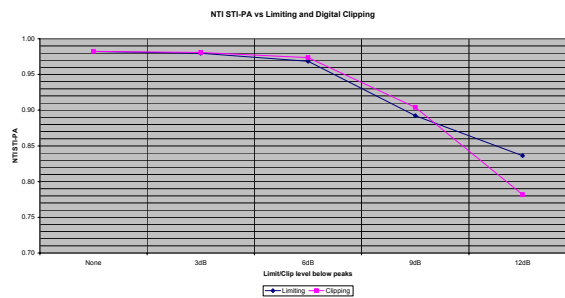


Figure 5 – NTI STI-PA vs Limiting and Digital Clipping

Again, we can see from figure 5 that reducing the limit or clip threshold (so reducing the crest factor) reduces the STI-PA value. Figures 6 and 7 show the resulting waveforms of the STI-PA signal and speech passage following 10dB of digital clipping. Compare these to figures 1 and 2. Table 6 shows that for this scenario the STI-PA value has reduced by about 20%. The question must be whether this corresponds to a reduction of 20% in actual intelligibility. Whilst there is no doubt that the speech passage now has audible distortion, the author believes it is a question of debate as to what degree the intelligibility has reduced, if at all.



Figure 6 – STI-PA Signal at 10dB Clipping

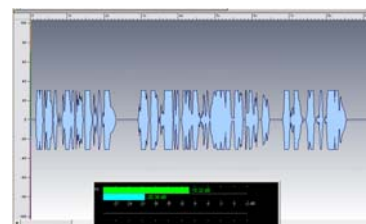


Figure 7 – Short passage of speech at 10dB Clipping

It should be noted that the authors' subjective assessment is based on the clipping applied to the particular passage of speech shown in figure 2 and that the assessment may well be different if the speech passage had a lower starting crest factor. It is the opinion of the author that any non-linearity that reduces the STI-PA value may not reduce the intelligibility of real speech to the same degree.

Based solely on the results given in tables 4, 5 and 6, the author now assumes 'strong non-linear distortion' to mean any compression whatsoever or more than 3dB of limiting or clipping applied to the STI-PA signal.

### 2.3. WinMLS Swept Sine Wave Measurements

The WinMLS signal used in the investigation was a 12 second logarithmically swept sine wave with flat frequency response ranging from 40Hz to 22kHz. From this signal, WinMLS derives the impulse

response of the system from which the modulation transfer functions (among other parameters) are determined.

A swept sine wave with a flat frequency response has little or no relationship to real speech, in particular with regard to its crest factor of only 3dB. Therefore, we cannot use the signal to determine speech levels or to show the effects of non-linearity distortion on the STI. The speech levels along with the ambient noise levels must be found using other methods and the results entered into the WinMLS speech settings dialog box as a post processing exercise (figure 8 shows the dialog box). However, the low crest factor is in fact one of the strengths of the swept sine wave method since it allows greatly increased signal-to-noise ratios (without overdriving the PA system) during the measurements. A second strength is that any generated distortion components are confined to the end of the impulse response where they can be truncated and discarded if not required<sup>3</sup>. These two factors are conducive to a clean, stable and hence repeatable impulse response in environments where a maximum length sequence derived impulse response, for example, would be badly corrupted and unusable. It must be understood however, that obtaining the STI using a swept sine will always require additional post processing if noise is to be considered.

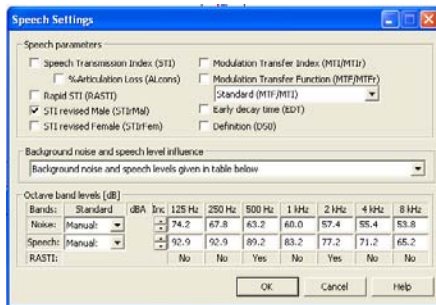


Figure 8 – WinMLS Speech Settings dialog box

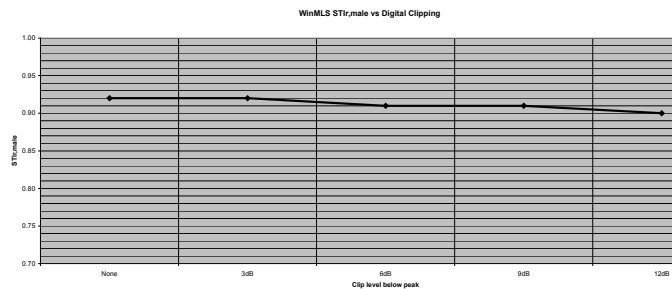


Figure 9 – WinMLS STI<sub>r,male</sub> vs Digital clipping

Figure 9 shows the results of digital clipping on the swept sine wave STIr,male. As can be seen there is only 0.02 reduction in STIr,male (in this case 2%) with 12dB of digital clipping. Moreover, it appears to be very easy to tell when the signal is being clipped due to the characteristic whistle caused by aliasing at the end of the sweep. Even 0.5dB of clipping is obvious.

### 3. RESULTS OF COMPARISON MEASUREMENTS

Given the issues regarding non-linearity outlined above, all measurements from here on have been carried out with great care to ensure the transmission paths used were operating linearly for both the swept sine wave and the STI-PA signal.

In this investigation the following scenarios are compared:

- 1) STIr,male/STI-PA vs. Reverberation time (electro-acoustic domain),
- 2) STIr,male/STI-PA vs. Steady pink noise (electro-acoustic domain),
- 3) STIr,male/STI-PA vs. SPL (electronic domain).

In addition, the STIr,male vs. SPL was predicted for the case of no reverberation or noise components, i.e. when the first two factors in equation (1) are always 1.000.

For the purpose of the investigation and a fair comparison of the measurements, all speech levels have been taken from the STI-PA signal and the Leq's displayed in the STI-PA output text files (figure 3). In so doing, any frequency response anomalies in the class 2 NTI MiniSPL microphone can be ignored. Figures 10 to 13 show the results for each scenario.

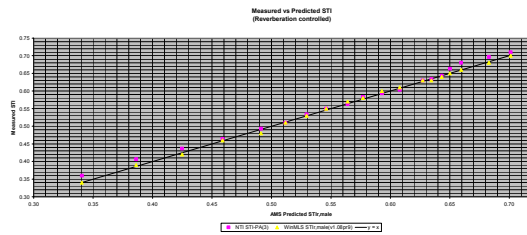


Figure 10 – STI vs. RT

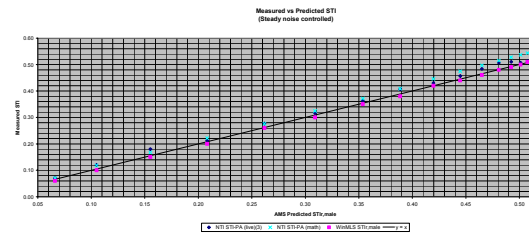


Figure 11 – STI vs Steady pink noise

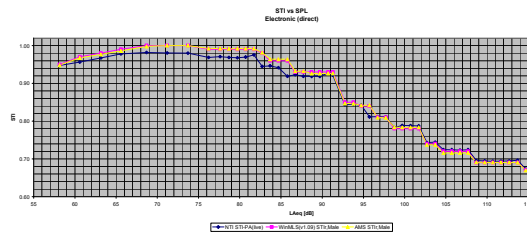


Figure 12 – STI vs. SPL

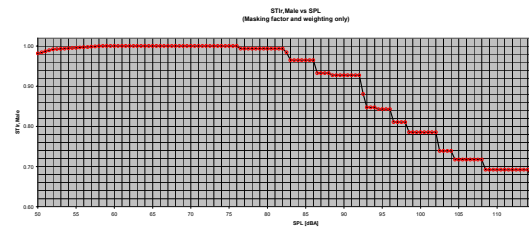


Figure 13 – STI vs. SPL  
(theoretical with masking and weighting factors only)

Note that for STI vs. Reverberation time and STI vs. Steady pink noise each point on the graphs represents the average of three STI-PA measurements. For the STI vs. SPL scenario, only a single measurement was taken. This was done to allow a better comparison of the effect of the weighting and masking factors using the octave band Leq's of the STI-PA output text files.

## 4. CONCLUSIONS

Figure 10 shows there is good agreement between the measured STI and the AMS predicted STI for both the NTI STI-PA meter and the WinMLS swept sine wave when only reverberation time is the controlling factor. Note that each STI-PA value on the graph is an average of 3 measurements and that the maximum STI-PA range found for any set of 3 in this investigation was 0.03 with a mean range of 0.02

For the scenario STI vs. Steady pink noise (figure 11) there is excellent agreement between the WinMLS swept sine wave and the AMS STI predictions. The agreement with the predictions is also good using the live (average of 3) STI-PA measurements. However there appears to be a small (0.01) but consistent overestimation of the STI-PA values. The reason for this discrepancy is not yet understood by the author but a potential avenue worth investigating may be found by inspection of equation (1). In equation (1) both the AMS predictions and WinMLS apply the second and third factors of the equation in a purely mathematical way to the matrix obtained via the first factor. On the other hand, STI-PA contaminated with pink noise and measured in real time encompasses both the first and second factors within the matrix and so only the last factor is mathematically applied. Note that the overestimation does not occur until the A-weighted signal-to-noise ratio is 15dB or less. The graph of figure 11 also shows the STI-PA results when the initial STI-PA text file (at a high signal-to-noise ratio) is mathematically contaminated with the measured noise (from a type 1 SLM) using the supplied NTI STI-PA excel worksheet. As can be seen, the overestimation has now increased to an average of 0.03 above the AMS predicted values. Again, the reason is not yet understood by the author and requires further investigation.

Figure 12 shows the results of the comparison exercise when the only varying factor is the SPL. This was carried out in the electronic domain with no noise added, however a small residual noise did exist and was considered in the calculations. The graph shows the WinMLS swept sine wave results are in very agreement with the AMS predictions. This is not surprising since the same formulae are applied to a matrix with all indices at 1.000 or nearly so in both cases. The reason for

the differences shown in figure 12 between the AMS predictions and the NTI STI-PA values may be twofold. Firstly, there appears to be an apparent upper limit of 0.98 to the achievable STI-PA value that can be measured. The author was unable to achieve higher values with the exception of a single 0.99 no matter what the conditions. The reason for this is not clear. Secondly, the STIr can be very sensitive to changes in octave band SPL's of only 0.5dB. The decision by a manufacturer to either truncate or round a SPL can cause some of the differences shown in figure 12. This is due to the coarseness of the SPL vs Slope of masking curve 'picking' procedure (see table 1) where the steps in SPL is 10dB and the steps in consecutive masking curves is 5dB. This causes sudden and significant changes in the STIr and, using a male spectrum, is particularly prevalent in the area around 90 to 95dBA, a common maximum SPL target for voice alarm systems. Figure 13 shows the same information when factors one and two in equation (1) are set to 1.000. i.e. only the masking and weighting are controlling the STIr,male values. This shows that the graph of figure 12 is very definitely a consequence of the formula and will be true for any measurement platform. Significant changes in subjective intelligibility with small changes in frequency response have been investigated and compared to the original STI<sup>4</sup>. A further study using the STIr,female (a female voice was used in the investigation<sup>4</sup>) should be undertaken.

To put the graph in perspective let us assume we wish to design a voice alarm system to achieve a STIr,male value of 0.50. This means that the system should be designed to achieve 0.54 STIr,male at levels between 60 and 75dBA. Then, at 90dBA the system should achieve 0.50 (i.e.  $0.54 \times 0.93$ ). However, let us suppose we predict that the system will only achieve an average of 89dBA with the 100V line loudspeakers tapped at 3W. Being prudent, we ask all loudspeakers are tapped at 6W knowing that a 2dB increase is below the just noticeable difference. On commissioning we find that the system is actually working at 93dBA (only 1dBA out here!) and the STIr,male average has now dropped to 0.46 (i.e.  $0.54 \times 0.85$ ) and the system has failed. Notwithstanding the above, it is not unreasonable to expect to have a range of SPL's across an area of voice alarm coverage of  $\pm 2$ dBA. Therefore, we must all be sure to choose our average SPL target very wisely.

Given that STI-PA cannot be used for non-linear systems (as defined above), there seems little advantage over impulse response methods of obtaining the STIr,male. It appears to be no more or less accurate. Although impulse response methods cannot be considered perfect, the swept sine wave signal has the not to be underestimated advantage of being able to achieve significantly higher signal-to-noise ratios during the measurements. This is particularly useful when measuring an existing underpowered system. A further advantage of the impulse response is the ability to determine other parameters allowing us to analyse and possibly solve a problem.

## 5. REFERENCES

1. A. Stacey, Checking the accuracy of MLSSA STI measurements. Proceedings of the IOA, Reproduced Sound 18, 2002.
2. K Jacob et. al, Development of an accurate, handheld, simple to use meter for the prediction of speech intelligibility. A symposium for the "Past, present and future of the speech transmission index", 2002.
3. A. Farina, Simultaneous measurement of impulse response and distortion with a swept sine technique. 108<sup>th</sup> AES Convention, 2000.
4. G. Leembruggen and A. Stacey, Should the matrix be reloaded? Proceedings of the IOA, Vol. 25, Part 8, 2003.
5. P. Mapp, Systematic & common errors in sound system STI and intelligibility measurements. 117<sup>th</sup> AES Convention, October 2004.
6. P. Mapp, Is STIPa a robust measure of speech intelligibility performance? 118<sup>th</sup> AES Convention, May 2005.
7. NTI Acoustilyzer with STI-PA option manual.
8. WinMLS electronic user manual.
9. BE EN 60268 Pt16 : 2003. Objective rating of speech intelligibility by speech transmission index