STIPA: a comparison between direct and indirect measuring methods

Constant Hak[1], Thomas King[2], Han Vertegaal[3]
[1] *Eindhoven University of Technology, E-Mail: c.c.j.m.hak@tue.nl*
[2] *Hottinger Brüel & Kjær UK Ltd, E-Mail: thomas.king@hbkworld.com*
[3] *Acoustics Engineering, Groenling 43-45, Boxmeer 5831 MZ, E-Mail: h.vertegaal@acoustics-engineering.com*

Introduction

The IEC 60268-16 [1] standard describes how to measure the STI (Speech Transmission Index) [2] and its simplified derivative STIPA (STI for Public Address systems) [3], using either of two measurement techniques. The first technique is based on modulated noise (MN) signals and often referred to as the **direct method**. The second or **indirect method** uses impulse responses (IR), obtained from MLS (Maximum Length Sequence) or swept sine stimuli, and applying Schroeder's modulation transfer function (MTF) [4][5].

The difference between the direct and indirect method with regard to the accuracy of the resulting STIPA values was investigated by Zhu et al. [6], using different measurement chains for each method. Although on average they did not find substantial differences, a more detailed study could provide more insight into the conditions in relation to the accuracy.

For this study the used acoustic measurement program DIRAC 7 [6] features both techniques, thereby enabling a nearly simultaneous comparison and evaluation of both methods, using exactly the same equipment and room, and under exactly the same environmental conditions. This paper describes a comparative study, consisting of a mathematical part using various well-defined synthetic room impulse responses, and an experimental part that has been carried out using an acoustic measuring room. The repeatability of the real STIPA measurements and the minimum practically achievable STIPA values for various types of background noises has been examined.

Background and theory

The STI(PA) uses low frequency modulations ranging from 0.6 to 12.5 Hz in noise bands from 125 to 8 kHz as a model for human speech. Where the STI uses 14 modulation frequencies in each noise band, the STIPA only uses 2. The preservation of these modulations in the transmission path from speaker to listener is a measure of speech intelligibility. Modulation (or speech intelligibility) reduction is caused by background noise and by the acoustic properties of the room, as modelled in equation 1, where $m(f_m)$ is the modulation reduction, *fm* the modulation frequency in Hz, *T* the reverberation time in s and *SNR* the signal to noise ratio in dB.

$$m(f_m) = \frac{1}{\sqrt{1 + \left(\frac{2\pi f_m T}{13.8}\right)^2}} \cdot \frac{1}{1 + 10^{-SNR/10}} \qquad (1)$$

Other factors that can influence speech intelligibility such as compression and distortion can sometimes be modelled as additional noise sources, but they are not part of this study. Previous research [7] has compared both direct and indirect measurement methods and found no significant differences between them in 'ordinary' rooms. This study aims to explore the limits of both methods and show where either method outperforms the other, with particular focus on the measurement time and repeatability.
The IEC 60268-16 standard suggests a minimum measurement time of 15 s for the direct STIPA method. This should result in a maximum deviation of around 0.03 STI (or 1 JND) for repeated measurements. The indirect method requires a minimum measurement length of 1.6 seconds and no less than half the reverberation time.

In this study, the maximum levels used in the experimental parts are far below the distortion level of the measuring equipment. The auditory threshold and masking, mentioned in the standard, are not relevant here and therefore not included in the measurement results.

Research methods

Mathematical method using synthetic impulse responses

For the mathematical research, synthetic room impulse responses (RIRs) have been convolved with the STIPA stimulus to get STIPA responses. A synthetic background random noise signal (BGN), having the same frequency spectrum as the response, has been added to each response. Four measurement lengths 2^n/48kHz (n = 16, 17, 18, 19), or 1.37, 5.46 10.9 and 21.8 s, have been combined with 11 different SNRs (Signal to Noise Ratios) from -25 through 25 dB in 5 dB steps. This has been done with each of three reverberation times (RTs), being 0.01, 1 and 5 s.

Experimental method using real measurements

For the experimental research, an empty acoustic measuring room with a volume of 90 m^3 and stable temperature and humidity conditions has been used. The RT ranges from 6 s in the 125 Hz octave frequency band to 1.5 s in the 8 kHz band. The background noise level exceeds 60 dB at 125 Hz down to 20 dB at 8 kHz. The measurements aim at comparing the (single number) STIPA value and its frequency dependent MTI (Modulation Transfer Index) components, using the direct and indirect methods, with the same measuring equipment and under the same environmental conditions. For this exploratory experiment, three stimulus types have been used: Modulated Noise, MLS and the exponentially swept sine (ESS). Each of the measurements is repeated 10 times at the same source-receiver distance and measuring time.

Repeatability of STIPA(MN)

For this experimental setup, part of the experiments is carried out in a highly reverberant empty room, and another part in the same room, but with absorbent elements on the floor. The average RT is 4 s in the empty room and 1 s in the absorbent room. In both situations an omnidirectional sound source has been used. Figure 1 shows the interior of the room in each situation. For both the direct and indirect method, the STIPA is measured at a distance of 4 m. For the indirect method, an ESS stimulus is used; for the direct method, a standard STIPA stimulus. The IR length is 10.9 s, in accordance with ISO-3382; the STIPA stimulus length is 21.8 s, in accordance with IEC 60268-16. Each measurement is performed 10 times per situation.

Figure 1: Measuring room.
(Left: $RT_{average}$ = 4 s; Right: $RT_{average}$ = 1 s)

STIPA measurements without stimulus

From the IEC 60268-16 standard it is known that even in the absence of stimuli, measured STIPA values can be higher than 0. In this part of the study, experimental measurements have been carried out without stimuli. In addition to random white and pink noise, solo and multi voice recordings have been used as a BGN signal, to find out the resulting measured STIPA values, using both the direct and indirect method. During 5.46 and 21.8 s respectively, the STIPA(MN) and STIPA(IR) have been measured 100 times in the absence of stimuli.

Used software and hardware

DIRAC 7 has been used for both the mathematical and the experimental parts of this study. The list of equipment used is presented in table 1.

Table 1: Used measuring equipment.

Description	name	Type	manufacturer
Software	DIRAC 7.0	7841	Hottinger Brüel & Kjær
Omnidirectional Sound Source	Omnipower	4292-L	Hottinger Brüel & Kjær
Measuring Power Amplifier	Smart Power Amplifier	HBK 2755	Hottinger Brüel & Kjær
Omnidirectional Microphone	½ inch	4189	Hottinger Brüel & Kjær

Results and Discussion

Synthetic file results

Table 2 shows the theoretical STIPA values and the measured values from synthetic room impulse responses (RIRs). These RIRs were generated, convolved with the STIPA stimulus and mixed with various BGN signal levels. The applied BGN files have the same spectral distribution, thus obtaining the same SNRs for all octave frequency bands. All room acoustic processing actions and calculations have been carried out using Dirac 7.

Table 2a to 2c: STIPA values calculated from synthetic Room Impulse Responses with various reverberation times. Values in red: deviating from theory by more than the JND (Just Noticeable Difference) of 0.03. Th = Theoretical, MN = Modulated Noise technique, IR = Impulse Response technique.

Table 2a: STIPA values for RT = 0.01 s

Measuring Time [s]	SNR [dB] -25	0	25
	Th MN IR	Th MN IR	Th MN IR
1.37	0.00 0.21 0.14	0.50 0.47 0.51	1.00 0.83 1.00
2.73	0.00 0.15 0.09	0.50 0.50 0.51	1.00 0.94 1.00
5.46	0.00 0.11 0.00	0.50 0.50 0.50	1.00 0.98 1.00
10.9	0.00 0.07 0.00	0.50 0.50 0.50	1.00 0.98 1.00
21.8	0.00 0.03 0.00	0.50 0.50 0.50	1.00 0.99 1.00
43.7	0.00 0.01 0.00	0.50 0.50 0.50	1.00 1.00 1.00

Table 2b: STIPA values for RT = 1 s

Measuring Time [s]	SNR [dB] -25	0	25
	Th MN IR	Th MN IR	Th MN IR
1.37	0.00 0.17 0.12	0.36 0.37 0.36	0.59 0.58 0.59
2.73	0.00 0.21 0.08	0.36 0.38 0.37	0.59 0.62 0.59
5.46	0.00 0.14 0.00	0.36 0.33 0.36	0.59 0.60 0.60
10.9	0.00 0.10 0.01	0.36 0.35 0.35	0.59 0.58 0.58
21.8	0.00 0.05 0.00	0.36 0.35 0.35	0.59 0.60 0.58
43.7	0.00 0.04 0.00	0.36 0.36 0.36	0.59 0.59 0.60

Table 2c: STIPA values for RT = 5 s

Measuring Time [s]	SNR [dB] -25	0	25
	Th MN IR	Th MN IR	Th MN IR
1.37	0.00 0.02 0.12	0.16 0.18 0.19	0.27 0.31 0.30
2.73	0.00 0.13 0.09	0.16 0.26 0.18	0.27 0.33 0.26
5.46	0.00 0.08 0.00	0.16 0.21 0.17	0.27 0.30 0.29
10.9	0.00 0.07 0.00	0.16 0.19 0.16	0.27 0.29 0.27
21.8	0.00 0.03 0.00	0.16 0.18 0.17	0.27 0.29 0.27
43.7	0.00 0.03 0.00	0.16 0.18 0.16	0.27 0.29 0.26

The direct method (MN) shows the most errors exceeding the JND, occurring at high BGN levels (SNR = -25 dB) and short measurements. IEC 60268 rightly prescribes little background noise and a minimum measuring time of 15 s when applying the direct method.

The indirect method (IR) shows less errors exceeding the JND, occurring only at high BGN levels combined with low measuring times and where the measurements are too short. ISO 18233 [8] rightly prescribes the measuring time to exceed RT.

Of all results in Table 2, those with the minimum measuring times meeting the respective standards (21.8 and 5.46 s) have been depicted in more detail in Figure 2. This figure shows STIPA versus SNR obtained in 3 ways: by calculation (green dotted), the direct method (red) and the indirect method (blue). The blue and green curves nearly coincide, which means that the indirect method is accurate for all conditions in the figure. This is unlike the required conditions for accurate direct method results, represented only by the rightmost quarter of the figure, and being a high measuring time (21.8 s) and a low BGN level (SNR > 0 dB).

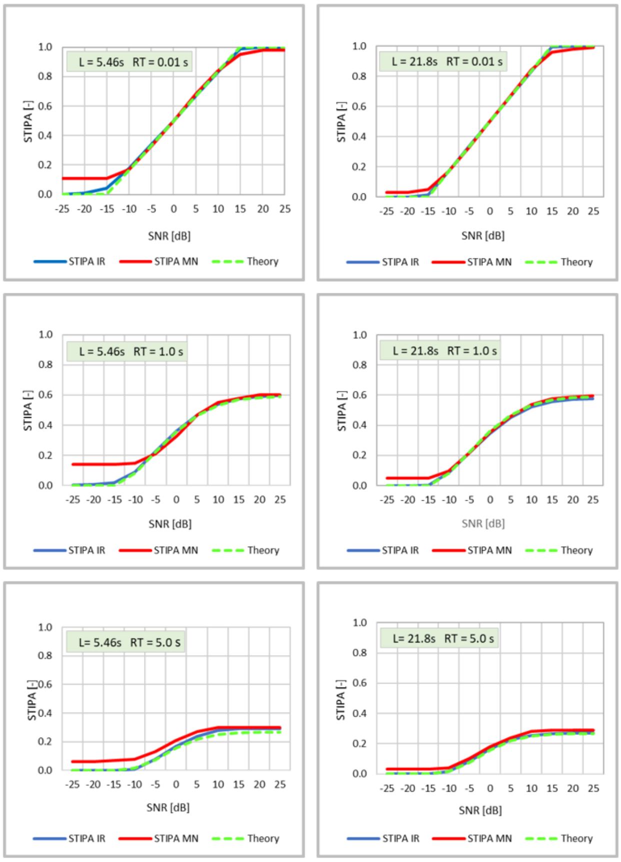

Figure 2: STIPA as a function of the measurement length L, Reverberation time RT and signal to noise ratio SNR.

Experimental measurement results

Table 3 and Figure 3 show the average MTIs from the direct and indirect methods, over 10 measurements at the same sound and source positions in the empty acoustic measuring room. The MTI(IR)-curves for MLS and e-Sweep coincide, while the MTI(MN)-curve fluctuates around them. The average STIPA values from the direct and indirect methods however show a mutual difference less than the JND [9].

Table 3: MTIs and STIPAs, measured in a reverberation room.

n =10	Modulated Transfer Index MTI (n = 10)			
	MN		IR	
Frequency	Avg	StDev	Avg	StDev
125	0.550	0.0021	0.383	0.0001
250	0.310	0.0025	0.295	0.0000
500	0.463	0.0007	0.307	0.0001
1000	0.221	0.0026	0.301	0.0000
2000	0.284	0.0031	0.331	0.0001
4000	0.488	0.0087	0.404	0.0001
8000	0.413	0.1734	0.512	0.0001
STIPA	0.353	0.0215	0.365	0.0001

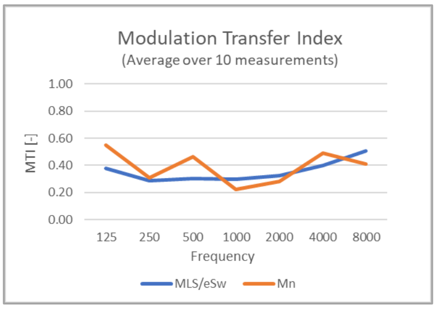

Figure 3: Average over 10 MTI-curves for IR and MN.

Lower limit in STIPA measurement values

Table 4 shows the averages, standard deviations and the min/max values of 100 STIPA measurements with various BGN signals, but without stimuli. Two types of BGN signals were used: stationary generated random noise (white and pink noise) and recorded speech sound (chatter and single voice).

In all cases, pink and white noise as BGN signal keep the STIPA within an error of the JND. Chatter and single voice as BGN signals are clearly more problematic, in particular with the direct method, as these lead to STIPA errors exceeding a multiple of the JND.

Table 4: STIPA caused by background noise only.

4a) Direct method, L = 21.8 s

STIPA(MN), Background Noise Only L= 21.8 s				
n = 100	Pink N	White N	Chatter	Voice
Avg	0.018	0.020	0.395	0.516
StDev	0.007	0.007	0.043	0.035
Min	0.047	0.008	0.293	0.442
Max	0.037	0.036	0.478	0.742

4b) Indirect method, L = 21.8 s

STIPA(IR), Background Noise Only L= 21.8 s				
n = 100	Pink N	White N	Chatter	Voice
Avg	0.000	0.000	0.116	0.079
StDev	0.001	0.001	0.051	0.039
Min	0.000	0.000	0.000	0.000
Max	0.004	0.003	0.206	0.173

4c) Indirect method, L = 5.46 s

STIPA(IR), Background Noise Only L= 5.46 s				
n = 100	Pink N	White N	Chatter	Voice
Avg	0.003	0.003	0.196	0.239
StDev	0.003	0.003	0.059	0.133
Min	0.000	0.003	0.011	0.000
Max	0.014	0.012	0.311	0.404

Conclusions

1. If applied under conditions prescribed by the standards, both the direct and indirect method results in STIPA values within the JND, as compared to the theoretically expected values.
2. For a given accuracy of the resulting STIPA, the direct (MN) method requires significantly longer measuring times than the indirect (IR) method.

References

[1] IEC 60268-16: 2020: International Standard, Sound System Equipment – *Part 16: Objective rating of speech intelligibility by speech transmission index*. International Organization for Standardization, (2020).

[2] T. Houtgast, H.J.M. Steeneken, H.J.M. plomp: Predicting the speech intelligibility in rooms from the modulation transfer function. I. General room acoustics. Acustica **46** (1980) 60-46.

[3] P. Mapp: Is STIPA a robust measure of speech intelligbility performance? AES 118th Convention, Barcelona, Spain, 2005.

[4] D.D. Rife: Modulation transfer function measurement with maximum-length Sequences. J. Audio Eng. Soc. **40** (1992) 779 – 790.

[5] M.R. Schroeder: Integrated impulse method measuring sound decay without using Impulses. J. Acoust. Soc. Am. 66 (1979) 497-500.Modulation transfer function measurement with maximum-length Sequences. J. Audio Eng. Soc. **40** (1992) 779 – 790.

[6] Dirac 7.1, URL: https://www.acoustics-engineering.com/html/blog-012.html

[7] P. Zhu, F. Mo, J. Kang: Experimental comparison between direct and indirect measurement methods for the objective rating of speech intelligibility. 21th International Congress on Sound And Vibration, Beijing, China, 2014.

[8] ISO 18233: International Standard ISO 18233: Acoustics Application of new measurement methods in building and room acoustics. International Organization for Standardization, 2006.

[9] J.S. Bradley, R. Reich, S.G. Norcross: A just noticeable difference in C_{50} for speech. Applied Acoustics **58** (1999) 90-108.