SPEECH INTELLIGIBILITY RATING USING A HEAD AND TORSO SIMULATOR

J Battaner-Moro Solent Acoustics, Southampton Solent University, UK
M Hickling Solent Acoustics, Southampton Solent University, UK

1 INTRODUCTION

Methods for objective rating of speech intelligibility are described in the standard IEC 60268-16:2011[1]. In this standard, the Speech Transmission Index (STI) and STIPA measurement methodologies predict intelligibility by analysing a modulation transfer function at a number of frequency bands, determined from the comparison of a synthesised reference signal (instead of a human's speaker voice) with a received signal recorded at a listener's position using a measurement microphone.

The reference signal's modulation is based on typical fluctuations in speech, comprising of octave band filtered noise, modulated at certain frequencies ranging between 0.5 Hz and 16 Hz which are designed to represent the spectrotemporal variations in a voice signal. The received signal is then analysed taking into account the modulation depth deviation, threshold of hearing, and sound pressure level in each band. The weighted sum of each band gives the overall STI value[2].

This methodology is based on monoaural listening and studies suggest that this may cause an underestimation of intelligibility, particularly when speech and noise arrive from different directions[3]. IEC 60268-16:2011 goes on to recommend using STI results from the best ear when using an artificial head instead of a single measurement microphone.

It is known that amplification from the pinna, subtle spectral distortions and binaural unmasking, can lead to improved signal-to-noise ratios[4]. Acoustic shading provided by the mass of the head, and the physical distance between ears provides interaural time and amplitude differences that can potentially increase the signal-to-noise ratio[5].

Good intelligibility is critical in situations in which emergency messages such as scape directions have to be broadcasted over a public address system. For example, in the UK the *Code of Practice for the Design, Installation, Commissioning and Maintenance of Voice Alarm Systems*[6] has a minimum STI requirement for fire voice alarms systems which has to be objectively measured as per IEC 60268-16:2011.

In ensuring intelligibility in emergency situations it is common to consider a "worst-case scenario". Thus if single microphone "monaural" STI measurements were to underestimate intelligibility, as they disregard the binaural advantage, then is fair to assume that monoaural measurements are adequate. In principle a person with hemi lateral hearing loss, such as one of the authors of this paper, should at least achieve intelligibility on par with that the measured monoaural STI.

The findings discussed here arise from a preliminary study comparing azimuthal binaural STIPA measurements against standardised monaural measurements taken using a single source masked with at various diffuse noise levels.

The results show a significant measured difference between "best" and "worst" ear STIPA scores as a function of head orientation relative to the source. This binaural STIPA scores are also quite different from monoaural STI scores determined under the same conditions.

2 MEASUREMENT PROCEDURE

For the purposes of this experiment, the STIPA methodology was used due its relatively short measurement duration of 15 seconds and the availability of a commercial STIPA measurement system (NTi XL2). The source signal was reproduced using an artificial mouth (NTi TaklBox), at a calibrated level of 60 dBA at 1 m.

2.1 Instrumentation

The NTi XL2 Sound Level Meter (SLM) with the appropriate software package was used to analyse the signal at the receiver's position and to provide a reading for STIPA and $L_{A,eq}$. Initially, these measurements were taken using an omnidirectional Class 1 measurement microphone. The process was then repeated but this time replacing the measurement microphone with a Brüel & Kjær type 4100 Head and Torso Simulator (HATS).

Data was collected for both the left and right channels of the HATS separately for each measurement situation given by orientation angle to the source and masking noise level. The two STI results scales mentioned in this paper are shown in Table 1, which is current at the time of writing, and Figure 1 which shows the deprecated previous scale but which is still in use in some measurement systems such as the NTi XL2.

Band	STI Range	Examples of typical uses
A+	> 0.76	recording studios
A	0.72 - 0.76	theatres, speech auditoria, parliaments, courts
B	0.68 - 0.72	theatres, speech auditoria, parliaments, courts
C	0.64 - 0.68	teleconference, theatres
D	0.60 - 0.64	class rooms, concert halls
E	0.56 - 0.60	concert halls, modern churches
F	0.52 - 0.56	PA in shopping malls, public offices, cathedrals
G	0.48 - 0.52	PA in shopping malls, public offices
H	0.44 - 0.48	PA in difficult acoustic environments
I	0.40 - 0.44	PA in very difficult spaces
J	0.36 - 0.40	not suitable for PA systems
U	< 0.36	not suitable for PA systems

Table 1 – STI qualification scale, after IEC60268-16: 2011

STI	0 – 0.3	0.3 – 0.45	0.45 – 0.6	0.6 – 0.75	0.75 – 1.0
	BAD	**POOR**	**FAIR**	**GOOD**	**EXCELLENT**
CIS	0 – 0.48	0.48 – 0.65	0.65 – 0.78	0.78 – 0.88	0.88 – 1.0

Figure 1 – STI qualification scale, after IEC60268-16: 2003

2.2 Measurement Environment

The HATS was placed in the centre of a reverberant, approximately cuboid room, with the source directly on axis, 1 m from the front of the dummy head (fig. 2). The noise masker was presented using four noise generators and speakers, projecting into the corners of the room with the intention to provide a diffuse and uncorrelated broadband masking field.

The noise field's random incidence was checked at the head's position by performing level measurements at a number of adjacent positions to the dummy's reference point (centre point between the artificial ears) as described in ISO 4869-1.

This is achieved by taking a reference sound pressure level measurement in place of the test fixture, and comparing this value with six measurements taken 150 mm from the reference point on the front-back, right-left and up-down axis. The diffuse-field check requires a difference of no more that ±2.5 dB relative to the centre point, and no more than 3 dB between left-right positions. This check was performed for two different noise rating levels, corresponding with the "Good" and "Bad" categories (fig. 1), and the maximum difference between opposing measurements never exceeded 2 dB.

The broadband masker signal is introduced as an interferer relative to the STIPA signal. Five different masking levels were chosen, correlating with the five different categories for STI: Excellent, Good, Fair, Poor, Bad (as per the 2003 STI categories in IEC 60268-16).

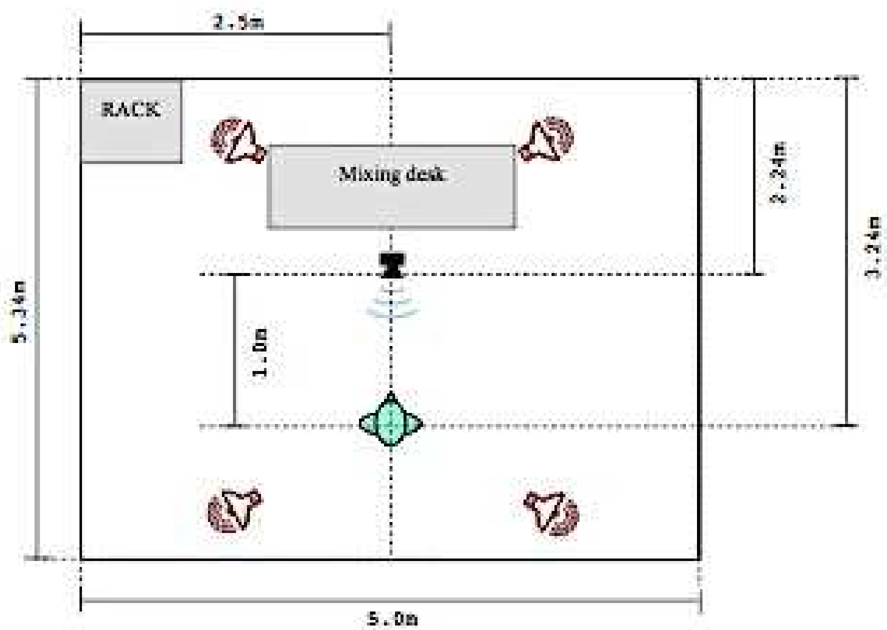


Figure 2 – Equipment setup in reverberant room

2.3 Methodology

The source's level was set to a 60 dB(A) at 1 m. The diffuse noise levels were set so that the single microphone STIPA result averaged in the middle of each of the 2003 STI categories i.e. "Good", "Fair", "Poor" and "Bad". The noise masking was muted to satisfy the "Excellent" criteria.

A Noise Rating measurement was taken for each masking level using a Brüel & Kjær Type 2270 sound level meter. A spatial average of six positions in the room was chosen to be representative.

A Class 1 omnidirectional microphone was used to measure a reference STIPA value for each of the different diffuse noise levels. This part of the measurement is designed to represent the current methods of rating speech intelligibility using a single microphone method. This was compared with binaural STIPA measurements using the Head and Torso Simulator at different angles with respect to the source's position. Individual left and right channel STIPA measurements were taken for each of the five noise rating levels, at rotational increments of 45°. This resulted in a dataset of 40 STIPA measurements per ear channel, totalling 80 binaural intelligibility measurements.

As the data from the single channel omnidirectional microphone is to be compared to rotational HATS data, four initial STIPA measurements were taken with the omnidirectional microphone to ensure there wasn't any rotational variation, verifying that a single traditional STIPA measurement is acceptable as a comparison. The orientations are outlined in the table of results below:

The monoaural STIPA measurements were taken with the microphone's diaphragm parallel to the floor. A check was performed to verify that the monoaural STIPA measurements were orientation-independent (Table 2).

Diaphragm Orientation	Rotational Angle	STIPA Result
Parallel with ground plane	N/A	0.99
Perpendicular to ground plane	0°	0.98
Perpendicular to ground plane	90°	0.98
Perpendicular to ground plane	180°	0.98

Table 2 – Omnidirectional microphone measurement variations

With the diaphragm on-axis with the single driver source, there was no noticeable deviation in STIPA evaluation with the microphone at 0° on-axis, 90°, or 180°, and therefore for the purposes of this project it was decided that an on-axis measurement in a single orientation with the reference microphone (0°) will suffice as a comparison with the binaural data.

3 RESULTS

Figures 3 to 5 show measurement results at the five noise conditions, ranging from a noise rating NR27 (no masking noise) to NR60. STIPA scores are plotted against the artificial head's orientation relative to the source. Also plotted are the monoaural (single microphone) STIPA score and the left-right ears average score.

For the four noise ratings where background noise was introduced (NR48, NR52, NR58, NR60), each ear shows a favourable increase in STIPA values during the section of rotation where that particular ear is angled toward the STIPA source signal, compared with the omnidirectional reference methodology. Similarly, as that same ear rotates away from the source, the STIPA signal drops below that of the reference omnidirectional microphone. The binaural scores coincide with the monoaural score at 180° azimuth angle.

For the test run without noise introduced into the room (noise rating NR27) at the majority of angles the dummy head slightly underestimated the intelligibility value compared with the omnidirectional reference, but remained within the "Excellent" category.

When an ear was rotated 45-90° toward the source signal, a STIPA score gain in the region of 12-55% with respect to the single microphone score was measured across for all masking levels.

Similar percentage gains of 13-50% were achieved in the right ear for the complementary angles of 270-315°. This corresponds with a positive increase in STI category in each case i.e. "Poor" to "Fair", or "Fair to Good" using the 2003 metrics, and an increase between two-four categories using the 2011 metrics. The inverse is also true of the opposite ear at the angles above, the STI category was decreased in each case.

The arithmetic mean of the left and right ear data for each angle and masking noise level, averages in all cases to the same 2003 STI metric, and within ±1 category using the 2011 STI metric, with respect to the omnidirectional reference measurement.

Using the “better ear” data in all cases yielded equal or higher STIPA values at all the measurement azimuth angles. This indicates that relying on “better-ear” scores may overestimate the STI value as this assumes the listener has equal sensitivity in both ears.

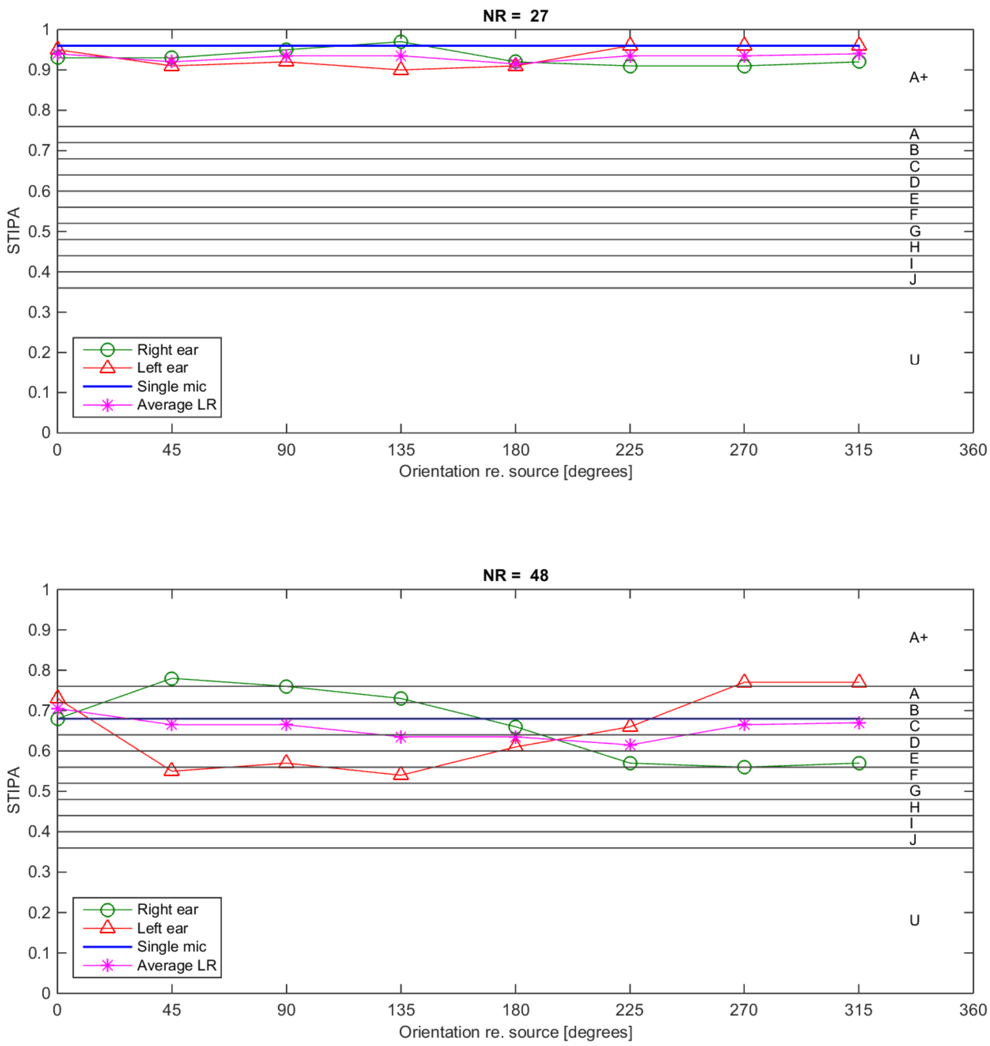
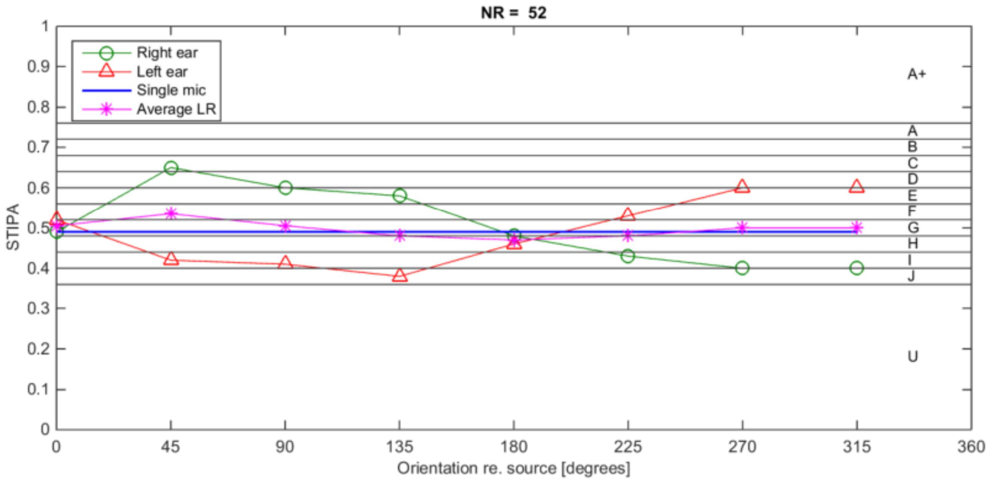


Figure 3 – Binaural STIPA scores for conditions NR27 and NR48.



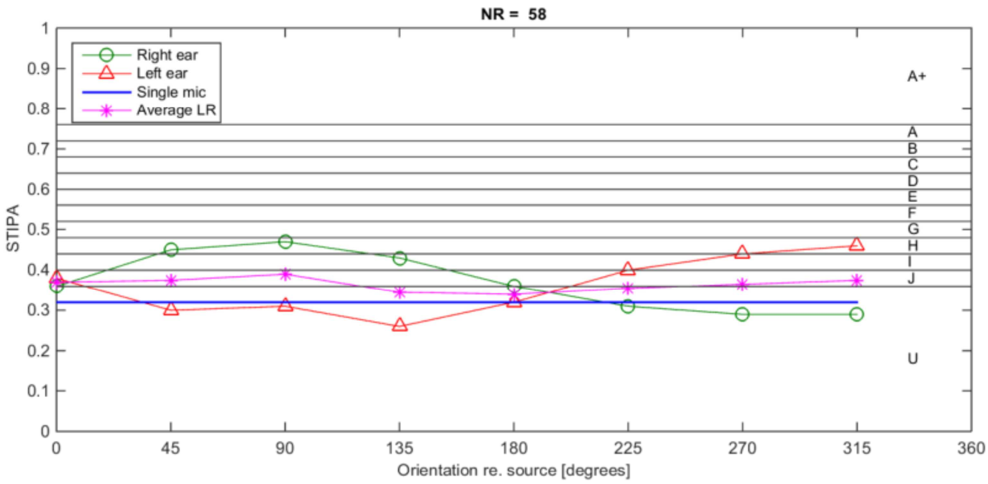


Figure 4 – Binaural STIPA scores for conditions NR52 and NR58.

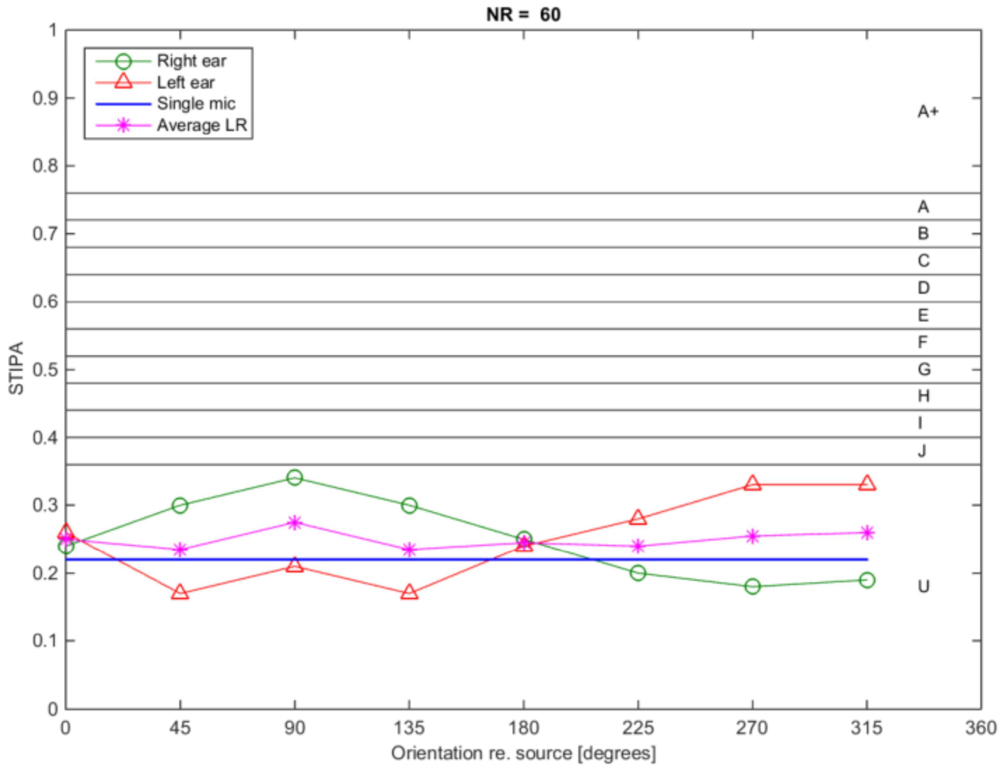


Figure 5 – Binaural STIPA scores for conditions NR60.

4 CONCLUSIONS

The motivation for this short study arises from the interest of the authors to explore the validity of objective intelligibility measurements using single microphone measurements, in particular in cases where the effectiveness of voice alarm systems is concerned.

The study was limited to diffuse field noise masking and a single source with the receiver located in its direct field. No reverberation or distortion conditions were tested, although these will be incorporated in future work.

It would not be adequate to extract many conclusions from this limited case-study. However, it suggests that more research is required regarding the assertion that better-ear scores should be used when performing binaural STI in lieu of the significant azimuth-dependent differences in STI scores observed.

The measurements were performed manually and this made the whole process somewhat laborious. The authors are currently working in a fully automated measurement rig in the university's new hemi-anechoic facility, mounting the artificial head on a controllable turntable and using a multi-channel data acquisition system.

5 REFERENCES

1. IEC 60268-16:2011. *Sound System Equipment. Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index*

2. Steeneken, H. J. M., and Houtgast, T., 1980. *A physical method for measuring speech transmission quality*, J. Acoust. Soc. Am. 67, 318–326

3. Wijngaarden, S., & Drullman, R., 2008. *Binaural Intelligibility prediction based on the speech transmission index. J. Acoust. Soc. Amer.* 123, 2008, p.4514-4523

4. Dillon, H. (2012). *Hearing Aids,* 2nd ed. Boomerang Press, Sydney, Australia, pp. 430–467.

5. Moore, D. (1991). *Anatomy and physiology of binaural hearing*, Int. J. Audiol. 30, 125–134.

6. BS 5839-8:2013 *Fire detection and fire alarm systems for buildings. Code of practice for the design, installation, commissioning and maintenance of voice alarm systems*.