ICSV14
Cairns • Australia
9-12 July, 2007

14th International Congress on Sound & Vibration

COMPARISON OF SPEECH INTELLIGIBILITY MEASUREMENTS IN A DIFFUSE SPACE

L. GOMEZ[1], C. NESTORAS[2], S. DANCE[2], S. MURANO[1]

1. Telent plc, Harbour Exchange Square, Docklands, London E14 9GE, UK
2. Acoustics Group, School of Engineering, London South Bank University, London SE1 0AA, UK

louis.gomez@telent.com

Abstract

An on-site experiment was undertaken in the ticket hall of London's Heathrow Terminal 4 Underground station. The purpose of the investigation was to evaluate the speech intelligibility obtained from two distinct measuring systems characterized by their open-loop and closed-loop configurations. Previous acoustic measurements in the space of interest showed significant diffuse field characteristics. Two additional comparative analyses were also undertaken, contrasting intelligibility results between multi and single source as well as derived STI values against STI-PA meter readings.

The station's public address system (multi-source) and a single omni-directional source were utilized in turn for the tests. Both sound sources were driven by an amplified e-swept sine signal. Impulse response derived RT and STI parameters were obtained by using WinMLS and Dirac measuring and post-processing systems. WinMLS in closed loop mode requires long source and receiver cables, while Dirac in open loop mode reduces the requirement for hazardous cabling, particularly when the existing station's PA system is used as a source, by employing a hand-held sound recording SLM as a receiver.

Similar STI values were obtained for both open and closed loop systems. The close agreement in the STI values obtained from multi-source and single-source configurations suggested that both methods may be valid to excite an acoustic diffuse field space to measure speech intelligibility. STIPA readings were in agreement with derived STI values.

1. INTRODUCTION

Speech intelligibility issues have been increasingly gaining importance over recent years. Applications of the topic can be nowadays highlighted in a multitude of working environments, ranging from classroom conditions to voice alarm systems for emergency announcements. Concentrating on the fact that measurement of intelligibility can be finally acknowledged as a key process, this paper considers the acoustical conditions of underground train stations in particular, for which speech intelligibility would unavoidably comprise a critical characteristic of the quality of service provided to the public (i.e. commuters). Focus is on the measurement implementation.

Currently, several measurement systems and procedures exist and can be used for this purpose [1]. In the current investigation, two measurement systems were employed and results compared for the ticket hall of Heathrow Terminal 4 Underground station; the systems characterized by their closed and open loop configurations. For the latter case, measurements could be performed without the use of long cables, requiring however post processing of the data for any results to be obtained. For the closed loop configuration, the measurement system could provide with a basic output instantly, having however the need for cables to close the loop. The choice of one of the two configurations could be determined by the environment to be measured since both systems had good and bad points. Nonetheless, a comparable performance needs to be assumed in this case. This paper is based on a comparison of results between the two systems (for the current conditions) in the attempt to validate the measurement procedure.

Furthermore, two different source configurations were investigated for both systems; using a single omni-directional loudspeaker (single source) and the station's P. A. system (multi-source). In this case, it was implied that measurements using the existing multi-source configuration of a space (i.e. no additional equipment needed by the acoustic consultant) would comprise a more efficient procedure in terms of time and resources. In addition, no long cables are required for the sound source, for either measurement system. Results obtained in this sense, are presented and discussed here as a comparative performance evaluation of the P.A. vs. the omni directional source.

2. THEORETICAL BACKGROUND

Speech Intelligibility is defined as a measure of the proportion of the content of a speech message that can be correctly understood [2]. This metric is one of the most important attributes of a PA system. It can be measured by direct methods and indirect methods.

Indirect methods (also called subject based) involve a panel of listeners who are to identify spoken material broadcasted through the PA system or communication channel system under test. Indirect methods are based on objective acoustical metrics which combined or alone have a reasonable degree of correlation with speech intelligibility.

The main factors having a significant effect on a PA system's speech intelligibility include reverberation time (RT), speech level to noise ratio, listener's distance from loudspeakers, system's bandwidth and frequency response, geometry and volume of the space, directivity and aim of loudspeakers.

An *Acoustic diffuse field* is mainly characterized by a uniform sound pressure level (SPL) and a consistent RT throughout the space beyond the source's direct field. At any point in the diffuse field sound propagates and is received from all directions. The RT decay curves are of perfect exponential decay and of very similar character for all frequency bands.

The acoustical *impulse response* of a system to an impulse input signal provides at the output all the information on how the system would behave to any other excitation signal. Hence from acoustical impulse responses many acoustical parameters can be derived, such as RT and speech transmission index (STI). However, other input signals than a pure impulse can also be used to extract the impulse response of a system, for instance the e-sweep.

An *e-sweep* input signal is an exponential sweep of a sine wave through the desired frequency range (audio spectrum in this study).

Reverberation Time (RT) is defined as the time taken for the sound level to fall by 60 dB, after the sound source is turned off. The RT is directly proportional to the volume of the room and inversely proportional to the sound absorption present in that room. *RT* is also affected by the room shape and contents (furniture or fittings).
A versatile and convenient method to calculate the RT was developed by M. Schroeder [3] in 1965. This method uses the impulse response of the enclosure at the investigation point to derive the RT by means of a reverse integration technique. This method allows RT to be computed swiftly on a computer by means of dedicated acoustic measuring software like the ones compared in the present experiment (WinMLS and Dirac).
The *Speech Transmission Index (STI)* has shown since its conception in 1971 by Houtgast and Steeneken [4] to be possibly the most valuable and accurate metric for objective rating speech intelligibility through a communication channel.
The STI is based on the concept of modulation transfer function *m(F)*. When the speech signal is transmitted through an enclosure, its amplitude modulation, rather than the carrier, contains the important information. The room reverberation and noise cause a decrease in the amplitude modulation. Here *m(F)* quantifies the degree of preservation of the original speech amplitude modulations (from 0.63 to 12.5Hz) as a function of modulation frequency, as it is transmitted throughout the room. As originally proposed, the *m(F)* only reflects the amplitude characteristics of modulation transfer and the phase characteristics are disregarded.
M. Schroeder showed [5] how the *m(F)* may be calculated from the octave-band filtered impulse response of the space.

$$m(F) = \frac{\left| \int_0^\infty p(t) e^{-j2\pi Ft} dt \right|}{\int_0^\infty p(t) dt} \tag{3}$$

where p(t) is the room sound power density impulse response. Here the p(t) is nothing more than the time interval derivate of the sound decay curve. This *m(F)* function is the complex Fourier Transform of the squared impulse response divided by its total energy.
Each m(F) value is converted into an apparent signal to noise ratio, averaged, and normalized, resulting in a single figure of merit called the speech transmission index STI, ranging from zero to one, for each seven octave frequency band relevant of the speech envelope (from 125Hz - 8KHz).
The apparent signal to noise ratio, at frequency f is as follows,

$$(S/N)_{app,f} = 10\log\left(\frac{m(F)}{1-m(F)}\right) \text{ (dB)} \tag{4}$$

Modern measuring software can quickly and conveniently calculate STI from the impulse response. Different STI score ranges were agreed to correlate to subjective descriptors of perceived intelligibility. See table 1.

STI-PA is a simplified version of the full STI which allows easy and fast measurement of speech intelligibility of PA systems. It can be conveniently measured by a hand held meter provided the appropriate excitation signal is reproduced through the PA system. Due to its high degree of correlation with the full STI [6] and convenience, STI-PA was incorporated in international standards (BS 60268-16:2003) and has experienced wide acceptance by the industry concerned.

Table 1. Correspondence between STI or STI-PA scores and perceived speech intelligibility

Subjective descriptor	Unintelligible	Poor	Fair	Good	Excellent
STI score	0 - 0.3	0.3 - 0.45	0.45 - 0.6	0.60 - 0.75	0.75 - 1.0

3. TEST METHODOLOGY

3.1 Test room

The ticket hall of Heathrow Terminal 4 Underground station (Figures 1-2) closely approximates a rectangular shape of 4000m^3 of volume, having an open end (half-way to the ceiling) at the platform end. Surfaces could be described as hard and reflective, since the majority of the latter consisted of marble, glass and plastered wall. The P.A. loudspeaker arrangement comprised of fourteen column loudspeakers around the perimeter of the room, located at a height of 3m (angled 45˚ downwards) and spaced every 4m. Seven and twelve ceiling loudspeakers were also in use in the entrance area and access to platform area respectively. Background noise level (BGNL) at the time of measurements was $L_{eq,\ 5min}45dB_A$.

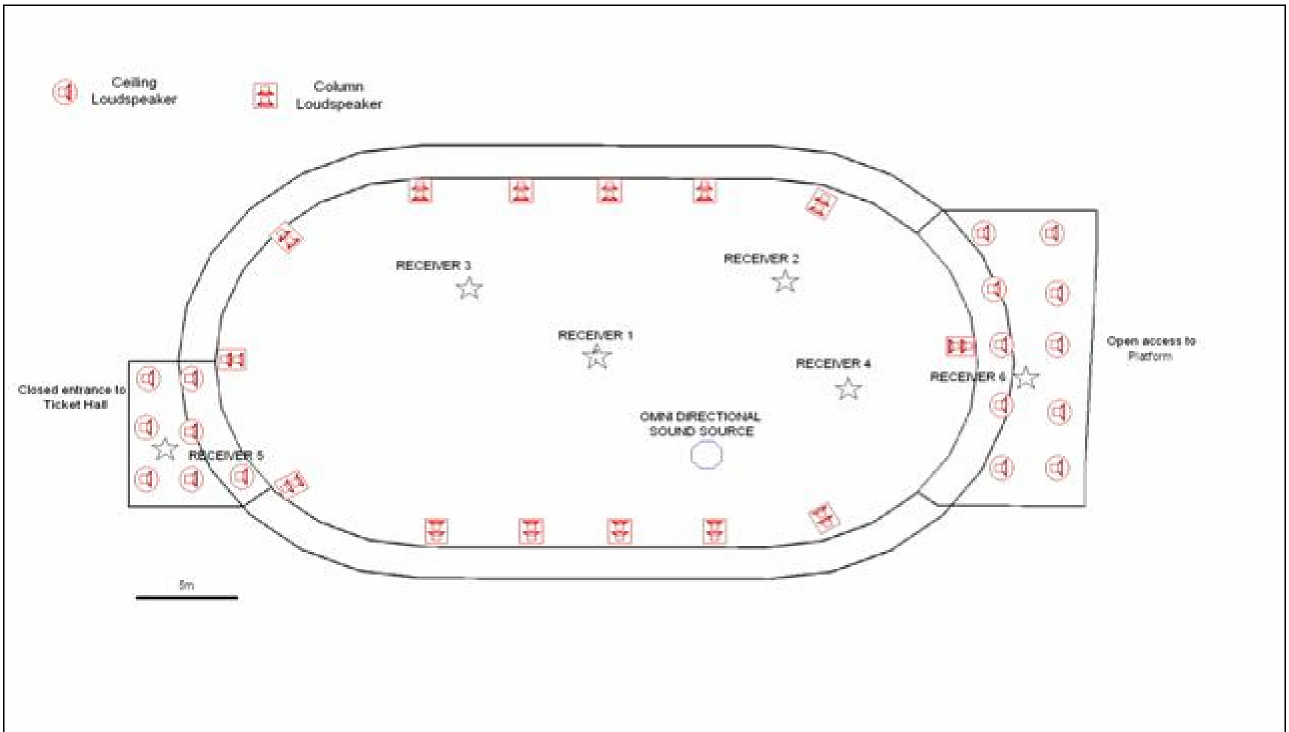

Figure 1. Heathrow Terminal 4 Underground station, Ticket hall plan

Four measurement sessions took place (during engineering hours 1am: to 4am); the overall procedure divided in two main parts for the two source conditions (single, multi). Measurements for the closed and open loop configurations were performed simultaneously.

Figure 2. Heathrow Terminal 4 Underground station, Ticket hall

3.2 Closed-open loop measurement systems

For the closed loop configuration (Figure 3) a portable computer was used to run the measurement software (WinMLS). A pair of omni directional source and receiver was employed at the sound card's output and input, respectively. Using a 10 second swept sine, impulse response measurements of the room were taken for a representative number of positions, as seen in figure 1.

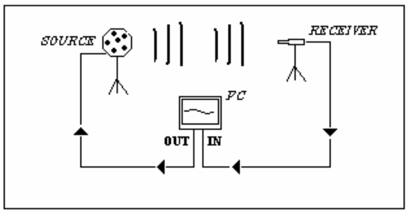

Figure 3. Closed loop configuration

Typically, an open loop system (Figure 4) consists of a sound source and a portable receiver with recording capabilities. For the purposes of this investigation, the receiving end of the closed loop system was complemented by a B & K type 2250 sound level meter (SLM) to separately represent an open loop measurement setup. Measurements were performed simultaneously for the two systems (using the same source), with the SLM positioned next to the receiver of the original setup. Post processing of the data in this case was performed at a later stage in B & K's Dirac software.

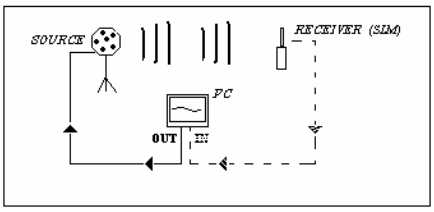

Figure 4. Open loop configuration

3.3 Use of single omni directional source and station P.A system

For the two main parts of the measurement session the source configuration was changed, in turn, for the omni directional source and the station's P.A. system. Using the two setups, as previously described, the source type was changed for the second (multi source) part of the session to the station's P.A. system. An NTI TalkBox, featuring human speech directivity, was positioned 10 cm from the station supervisor's office microphone and was used to reproduce the swept sine generated from the portable computer (approximately at a 'normal talker' level). The level of the signal transmitted to the ticket hall through the P.A. system was then adjusted to a realistic level for normal station conditions, achieving a favorable S/N ratio ($>15dB_A$). Measurements were performed for the same receiver positions as in part one, simultaneously for the closed and open loop configurations.

3.4 STIPA Measurements

An additional point of reference adding confidence to the main procedure was created by a complementing session of STIPA measurements. Using the NTI TalkBox as a source (as previously positioned), a STIPA test signal [7] was reproduced and transmitted through the station's P.A. system. Readings were taken with an NTI Acoustilyzer hand held meter for the same receiver positions. Three measurements were performed per position and the average stated in the results.

4. RESULTS

RT values shown below are the average of RT measured at 500Hz and 2KHz bands.

Table 2. Open Loop vs Closed Loop RT comparison, at the six measuring points.

Single source						
RT (sec)	position 1	position 2	position 3	position 4	position 5	position 6
Open Loop	2.9	2.9	2.8	2.8	3.1	2.8
Closed Loop	2.7	2.7	2.8	2.7	2.7	2.8

Table 3. Open Loop vs Closed Loop STI comparison, at the six measuring points.

Single source						
STI	position 1	position 2	position 3	position 4	position 5	position 6
Open Loop	0.33	0.31	0.32	0.31	0.32	0.33
Closed Loop	0.44	0.44	0.43	0.43	0.39	0.43
STI-PA	0.40	n/a	0.41	0.42	0.45	0.46

Table 4. Open Loop vs Closed Loop RT comparison, at the six measuring points.

Multi source						
RT (sec)	position 1	position 2	position 3	Position 4	position 5	position 6
Open Loop	2.4	2.4	2.5	2.4	2.3	2.3
Closed Loop	2.7	2.6	2.7	2.7	2.7	2.7

Table 5. Open Loop vs Closed Loop STI comparison, at the six measuring points.

Multi source						
STI	position 1	position 2	position 3	position 4	position 5	position 6
Open Loop	0.44	0.40	0.42	0.38	0.40	0.47
Closed Loop	0.45	0.42	0.44	0.43	0.40	0.47
STI-PA	0.40	n/a	0.41	0.42	0.45	0.46

5. RESULTS' DISCUSSION

From the apparent physical characteristics of the Ticket Hall: non-extreme dimensions, regular geometry and type of surface materials; it would be reasonable to expect acoustic diffuse field characteristics. The highly consistent RT results obtained throughout the space corroborated the diffuse field initial supposition.

Close loop measuring system gave highly consistent RT values irrespective of the receiver position (single source Standard Deviation STD=0.05, multi source STD=0.04) or sound source configuration, see tables 2 and 4.

Open loop system measured consistently lower RT values using the multi-source configuration than when single source was utilised. However RT values from Open loop system using the single source, closely agreed with Close loop RT values of either sound source configuration (STD = 0.05- 0.2 at six receiver positions).

Looking at the Open loop-multi source RT data in more detail, it was found from all the e-sweep samples post processed that the RT values at 500Hz were less consistent than values at 2000Hz. This detail leads to believe that Open loop measuring system is possibly more susceptible to poor lower frequencies reproduction from the multi source system.

Speech intelligibility measured by the Close loop system gave very consistent STI values irrespective of the receiver position (single source STD=0.01, multi source STD=0.02). Typical STI standard deviation is 0.02 [7]. Although, Open loop system using multi source gave slightly dispersed STI values (STD =0.03), individual position values did not differ significantly from their Close loop counterparts in both source configurations.

Open loop-single source configuration yielded STI values at the six positions consistently 0.1 lower than values in any other 3 configuration combinations (see tables 3 and 5).This consistent discrepancy may suggest that the Open loop was more sensitive to lower signal to noise ratio (S/N) than Close loop system, due to the longer source-receiver distances than in the multi-source situation.

Although receiver position 2 was placed on-axis with the nearest PA loudspeaker, its STI score was not significantly higher than other off-axis receiver positions. This suggests that the receiver was positioned outside the loudspeaker's direct field and therefore its on-axis position did not benefit its STI score, see table 5. Receiver position 6, although positioned off-axis between two low ceiling loudspeakers, benefited from its close distance to the loudspeakers which affected favourably the S/N ratio in the STI calculation. This is suggested by its higher STI score when compared to the other 5 receiver positions, see table 5.

Most STI-PA scores correlated well with their corresponding positions' STI scores in all measurement configurations except in the single source & open loop. STI-PA standard deviation was 0.02, which is the typical standard deviation of STI-PA [7].

6. CONCLUSIONS

The London underground Heathrow terminal 4 ticket hall space showed clear quasi diffuse acoustic field characteristics. Most of the RT values measured in any of the two sources configurations (multi source or single source) or by any two measuring systems (Dirac, WinMLS) were in good agreement across the six measuring positions.

Measurements of STI using Close loop or Open loop measuring systems gave similar marks for either single or multi source configuration. Results from this investigation showed that in a space of similar characteristics as the ticket hall under investigation, objective speech intelligibility (STI) may be measured utilising the installed multi source PA system by either Open loop or Close loop measuring systems.

7. FURTHER WORK

It is intended that similar investigation to the work here presented, will follow in other spaces in order to confirm that the findings summarised above also hold with various degrees of field diffusion, different multi source types and loudspeaker characteristics. Similarly, the same comparison between the two measuring systems presented in this work should be carried out under different degrees of acoustic diffusion and sound sources configurations to ascertain the conclusions above extracted. It is also proposed to investigate more in detail STI-PA repeatability, accuracy and correlation with full STI in the different conditions mentioned above.

REFERENCES

[1] "Ergonomics - Assessment of speech communication", *British Standards,* BS EN ISO 9921: 2003

[2] "Code of practice for the design, planning, installation, testing and maintenance of sound systems", *British Standards,* BS 6259:1997

[3] M.R. Schroeder, "New method of measuring reverberation time". *Journal of the Acoustical Society of America* vol **37**, 409-412 (1965).

[4] T.Houtgast and H.J.M. Steeneken, "Evaluation of speech transmission channels by using artificial signals" *Acustica* vol **25**, 355-367 (1971).

[5] M.R. Schroeder, "Modulation transfer functions: Definitions and measurements", *Acustica*, vol **49**,179-182 (1981).

[6] Ole-Herman Bjor, "STIPA, The Golden mean between full STI and RASTI", *Proceedings of the Institute of Acoustics* Vol.**25**, Pt 7. (2003)

[7] "Sound System Equipment - Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index", *British Standards,* BS ISO 60268-16 (2003)