SPEECH INTELLIGIBILITY IN HIGHER EDUCATION TEACHING FACILITIES

Iain Paterson-Stephens The University of Nottingham, Nottingham, UK
Peter Rutherford The University of Nottingham, Nottingham, UK
Robin Wilson The University of Nottingham, Nottingham, UK

ABSTRACT

This paper expands upon the initial work conducted by Rutherford, Wilson and Hickman[4] and explores the relationship between the Speech Transmission Index (STI) and its application within the context of higher education teaching and learning facilities. As is well known, the modern learning environment comprises a diverse student population of both native [L1] and non-native [L2] listeners & speakers and, as has been evidenced in research and recognized within BS EN 60268-16:2011[16], such [L2] listeners provide a significant challenge when predicting STI performance in any given space.

The purpose of the research presented here is to delve deeper into the relationship between STI and both native and non-native listening groups. Data is presented that extends the findings from the original study, particularly with respect to the relationship between the STI value and [L2] listener performance. The paper concludes that whilst STI over predicts [L2] listener performance, the level of over-prediction itself is fundamentally dependent upon the STI value. For high STI values (>0.8), a relatively small over prediction was observed during intelligibility experiments (approx. 6%) however at low STI values (<0.5), a much higher over prediction was observed (approx. 40%).

Such findings clearly point to the need to look more critically at Speech Transmission and Speech Intelligibility as metrics for evaluating room acoustic performance for diverse, international populations.

INTRODUCTION

It is fair to say that UK higher education institutions have experienced significant changes to their student population over the past 15 years or so. The widespread introduction of tuition fees in the mid to late 1990s combined with significant emphasis placed on overseas and national recruitment has resulted in not only greater student numbers, but increasingly fierce competition between higher education (HE) institutions. As a result, many departments and courses have seen both growth in class sizes and a much higher proportion of non-native English speakers than in previous years.

When applying for courses, many applicants rank institutions based on nationally published performance metrics such as the National Student Survey (NSS), which is based on many parameters including teacher performance. Consequently, teaching staff are under increasing pressure to improve delivery of teaching and demonstrate this by providing evidence, such as student ratings, these often serving as key performance indicators used as part of annual staff appraisal processes. This throws into focus the relationship between teacher, learning environment and the influence these have on a student's learning experience. For example, if a lecture theatre used for teaching does not promote good quality transfer of speech, then students are less likely to follow and engage with the material taught in-class. Their performance overall may suffer and their perception (and rating) of the teacher may be reduced. The teaching environment therefore has the potential to impact negatively both on lecturer and student and by implication the institution offering taught programmes.

Good acoustic design is an essential aspect of any indoor space used for teaching and learning. The ability of the space to support effective transfer speech is of prime consideration and the key factors that determine this are the level of background noise present within the space and the reverberant qualities of the space itself. This study investigates a number of typical teaching spaces used within an HE establishment and attempts to assess their suitability for supporting speech transfer in the context of a typical cohort of mixed native and non-native English students.

Guidelines given in Building Bulletin 93, BB93[1], make clear recommendations about the appropriate level of background noise and reverberation times for different types of activity. The scope of BB93 is however limited to LEA funded Nurseries, Schools and FE Colleges. No agreed standard or set of guidelines exist that relate specifically to the design of HE teaching spaces. Despite the limited regulatory scope of BB93, the nature of the spaces it addresses means many of its recommendations relating to speech transfer characteristics may be reasonably applied to the HE context[2]. Some of the recommendations given in BB93 are based on standard objective measures of speech transmission (e.g. STI), which are based on communication between native speakers and listeners. However, given that in a typical UK University the population will consist of a wide mix of native [L1] and non-native [L2] English speaking students and staff, it is unlikely that native-to-native communication will be the norm. This being the case, it is likely that teaching spaces will need to be designed to meet more demanding STI criteria than the base level recommendations of BB93 would suggest. Indeed, a number of studies[3,4,5,6,7] have shown that STI may incorrectly predict the performance of an acoustic space by as much as 20-30% for non-native listeners.

In response, this study sought to address two fundamental questions; (a) '*how well does a modern HE teaching environment perform in the context of different listener groups?*' and (b) '*how well do the accepted Speech Transmission metrics predict this performance?*'. In attempting to answer these questions, two specific aims were set as follows; 1) to evaluate a cross section of teaching spaces within a UK University with regard to Speech Transmission capabilities and acoustic characteristics in order to determine the range that exists and its quality relative to published guidelines. 2) to assess both the accuracy and limitations of the conventional Speech Transmission Index, STI, as a metric when used to predict the speech intelligibility characteristics of teaching spaces used by a typical cohort of university students i.e. a cohort containing a mix of native [L1] and non-native [L2] English speaking participants.

EXPERIMENTAL PROCEDURE

Procedure Overview

For this study, two main experiments were developed; Objective Measurements and Subjective Listening Tests (Figure 1). These were then used to explore five different HE teaching spaces.

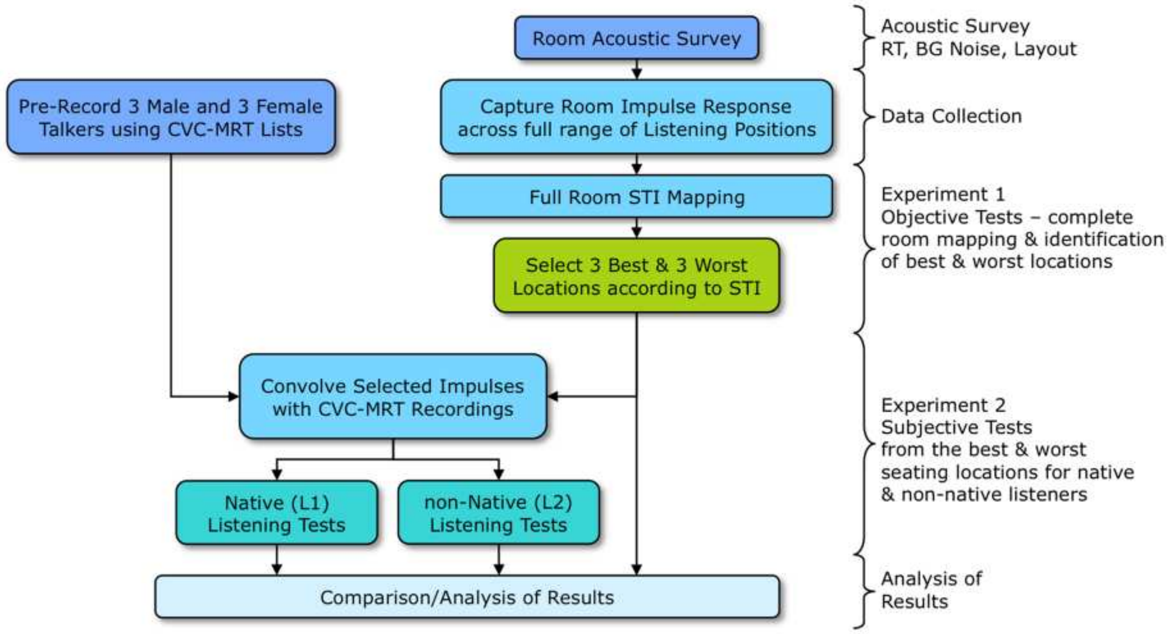


Figure 1   Experimental Procedure – *repeated for each teaching space*

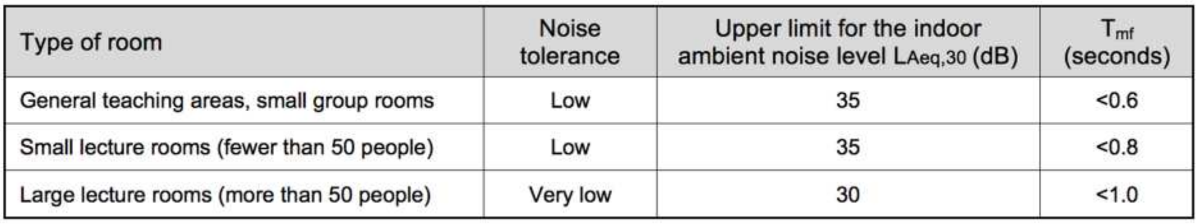
After an initial room acoustic survey of each teaching space to determine reverberation time (RT) and A-weighted background noise levels, the following investigations were carried out:

1) Experiment 1 - A comprehensive, objective investigation of each of the five teaching spaces that sought to determine the spread of speech transmission capability across the full range of listening positions within each room. This comprised a suite of STI measurements at multiple locations that sought to identify the three ‘best’ and three ‘worst’ locations, for each room.
2) Experiment 2 - For the six selected locations in each room, subjective assessments were conducted using two groups of listeners i.e. a native English speaking group [L1] and a non-native English group [L2].
3) From the two experiments, the ability of STI to predict speech transmission capability was compared with the observed performance of the [L1] and [L2] groups.
4) The relative performance of the [L1] compared with the [L2] group was also assessed and the extent to which STI under or over predicts the performance of the [L2] group was determined.

Method in Detail & Relevant Standards

Room Acoustic Survey

The relevant recommendations, to this investigation, from BB93 are shown in Table 1.



Type of room	Noise tolerance	Upper limit for the indoor ambient noise level $L_{Aeq,30}$ (dB)	T_{mf} (seconds)
General teaching areas, small group rooms	Low	35	<0.6
Small lecture rooms (fewer than 50 people)	Low	35	<0.8
Large lecture rooms (more than 50 people)	Very low	30	<1.0

Table 1 Recommended performance criteria for teaching spaces as Defined in BB93

For the Acoustic Survey, three assessments were carried out; visual inspection, background noise assessment and reverberation time assessment. Noise level assessments were carried out in terms of the 30 minute A-weighted average i.e. $L_{Aeq,30}$, and reverberation times in terms of the mid-frequency average, i.e. T_{mf} in seconds. Both of these measures were chosen so as to be comparable to the values specified in BB93 and shown in Table 1.

Reverberation time measurements were carried out as described in BS-EN-ISO:354(2003)[8] and the method referred to as the ‘Indirect version of the Integrated Impulse Response Method’ was chosen. All RT measurements were carried out using B&K DIRAC software with the test source signal set to an exponential sweep sequence, which was presented using a high power omnidirectional loudspeaker.

Objective Assessments (STI)

For the STI measurements, Impulse responses were captured between a source 'teaching position' and a number of receiver ‘listener’ positions selected from the full range of available positions within the room as defined in BB93. Typically, impulse responses were captured for around 50% of the available seating positions. This gave a good cross section of the positions available and generated sufficient data to judge the variation in performance for each teaching space. Impulse response measurements were made using a dual source high quality (non-ported) loudspeaker, approximating[2,9] an artificial mouth[10], which should lead to an STI error of no more than the typical standard deviation of STI i.e. <0.02[17,18]. For each of the measurements, the relative height and placement of the source (loudspeaker) and receiver (microphone), along with calibration protocols all followed the guidelines as set out in the relevant standards[1,16] and published research[3,4,17,18].

All impulse responses were captured using B&K DIRAC software via an Earthworks M30 omnidirectional measurement microphone moved to each 'listening' position in turn. The source loudspeaker was located at 1.65m above floor level and the receiving microphone was located at head height for a typical adult in a sitting position, i.e. 1.2m. STI calculations were performed using B&K DIRAC software and various STI functions written for MATLAB used to produce room 'maps' of STI capability, an example of which is shown in figure 2. The room STI maps were useful in terms of helping to identify areas within each teaching space where STI performance was particularly good or poor – an interesting exercise in itself in relation to the architectural or acoustic features present.

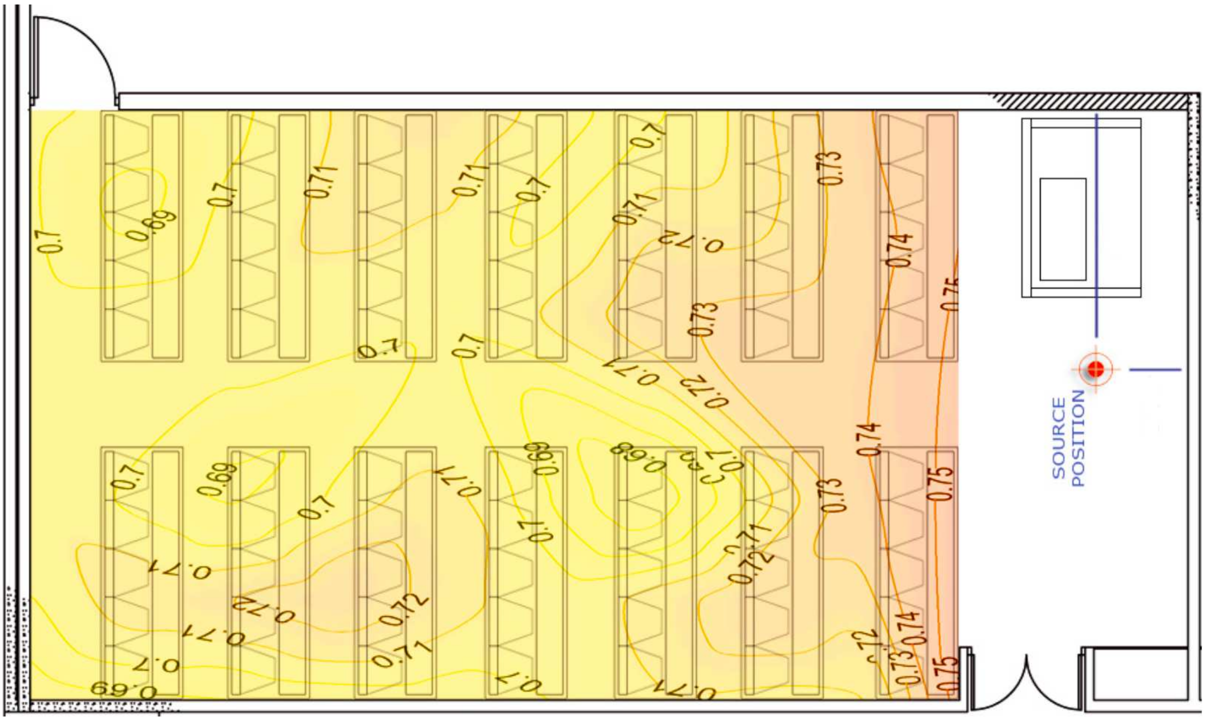


Figure 2 – Room STI Map produced using MATLAB STI functions

Subjective Assessments (Word Scoring)

For each of the rooms tested, the three 'best' and three 'worst' seating positions, according to STI were selected for further investigation using groups of [L1] and [L2] listeners. Listening tests were conducted in a listening booth and in-line with the approach already established by other researchers[4,11]. Also, as far as possible, all subjective listening tests followed the guidelines as given in ISO:9921(2003)[12] and were designed to be in accordance with ISO:TR-4870(1991)[13]. This approach offers great flexibility for the listening group in terms of when and for how long listening tests are conducted. Also, the access time required in the acoustic space is kept to a minimum.

A total list of 300 words, organised as six sub-sets of 50, were used in this study. The list was designed as a phonetically balanced set of CVC rhyming words, which were presented in a closed-set form. This type of listening test lends itself well to automated data collection which was a necessary consideration due to the number of tests to be conducted. The complete 300 word list was vocalised by three male and three female speakers. Prior to vocalising the list, each speaker was given some training and an opportunity to practice. Speakers were asked to practice vocalising the sounds at a consistent rate and level of presentation and to avoid putting any intonation or emphasis on words as they are spoken. To help with this a sound level meter and a visual metronome (flashing led) were placed in front of the speakers. This helped them to monitor their levels as they spoke and to fall into a rhythmic pattern of vocalising the words at a set rate. Speakers were asked to vocalise each word within an agreed carrier phrase, i.e.: "You will mark *'test word'* now". All vocalised words were recorded in a quiet semi-anechoic environment using a measurement microphone connected to a computer with recording software.

After being convolved with the impulse responses of the selected listening positions, the vocalised lists were presented to the listeners in random order so as to reduce the likelihood of listeners learning a set pattern of words over a series of tests. All vocalisations were presented to the listener in a quiet listening booth over closed-back headphones, which were set up for a listening level of 70 dB (SPL). Printed machine readable *Speedwell* response sheets were prepared and shown to the listeners prior to each test. Listeners were asked to respond to each vocalisation by identifying the word spoken from the listed alternatives. Listeners were informed that they could stop the experiment at any point, take a break and/or come back on different days in order to complete the tests. Two groups [L1] and [L2] of eight subjects each were assessed, details of which are:

Group 1 – Native Speaking English Students

[*L1] Listeners: L1-1 to L1-8.* Age range: 18-25. Average age: 19 years. Sex: 4 male and 4 female equally distributed between the two groups. No reported Hearing Impairment and all students were subsequently assessed as having 'normal' hearing. All native students had <8 months at the University.

Group 2 – Non-Native English Speaking Students (4 Saudi, 4 Chinese)

[*L2] Listeners: L2-1 to L2-4 (Saudi Arabia) and L2-5 to L2-8 (Chinese).* Age range: 18-22. Average age: 19 years. Sex: 4 male and 4 female distributed evenly between the two groups. No reported hearing impairment and all students were subsequently assessed as having 'normal' hearing. All students had <8 months at the University and had been resident in the UK for <9 months. All non-native listeners held an English Language qualification of IELTS 6.0 with all assessed elements (Reading/Writing/Listening and Speaking) at IELTS 5.5 or above.

RESULTS & DISCUSSION

Room Acoustic Measurements

For this study, a cross section of teaching spaces was selected for assessment ranging from small flat lecture rooms to large tiered lecture theatres. In total five spaces were chosen for detailed analysis, as outlined in Table 2. For each teaching space, detailed room acoustic parameters and multi-position STI, measurements were taken, also presented in Table 2. At the time of taking the measurements, the room was unoccupied and speech reinforcement systems and other classroom equipment were not in operation.

Room	Type	Seats	Features	-	T30,mf	STI	LAeq,30
1	Large Tiered Lecture Theatre	250	Old build. Diffusion panels & reflectors distributed throughout the room. Carpet and soft furnishings throughout.	***mean***	1.0s	0.64	36dB(A)
				stdev	0.063	0.046	
				range	0.81-1.08	0.51-0.78	
2	Large Flat Lecture Room	160	Refurbished in the late 90's. Acoustic absorbers above head height on walls. Carpet, curtains and freestanding tables & soft covered chairs.	***mean***	1.2s	0.64	44dB(A)
				stdev	0.054	0.044	
				range	0.93-1.26	0.51-0.75	
3	Medium Tiered Lecture Theatre	95	Modern design, diffusion panels on back wall, carpet and curtains. Soft furnishings throughout.	***mean***	0.8s	0.72	42dB(A)
				stdev	0.076	0.043	
				range	0.64 - 0.92	0.65-0.82	
4	Medium Tiered Lecture Theatre	80	Carpet. Reflecting panel above 'stage' area. No other treatment.	***mean***	1.0s	0.71	38dB(A)
				stdev	0.071	0.019	
				range	0.69-1.10	0.67-0.75	
5	Small Flat Lecture Room	45	Carpet and curtains. No acoustic treatment. 'Hard' painted concrete walls and ceiling. Narrow 'corridor' design to the room with a high ceiling.	***mean***	1.7s	0.54	42dB(A)
				stdev	0.134	0.039	
				range	1.08-1.95	0.47-0.65	

Table 2 Results of analysis of selected teaching rooms

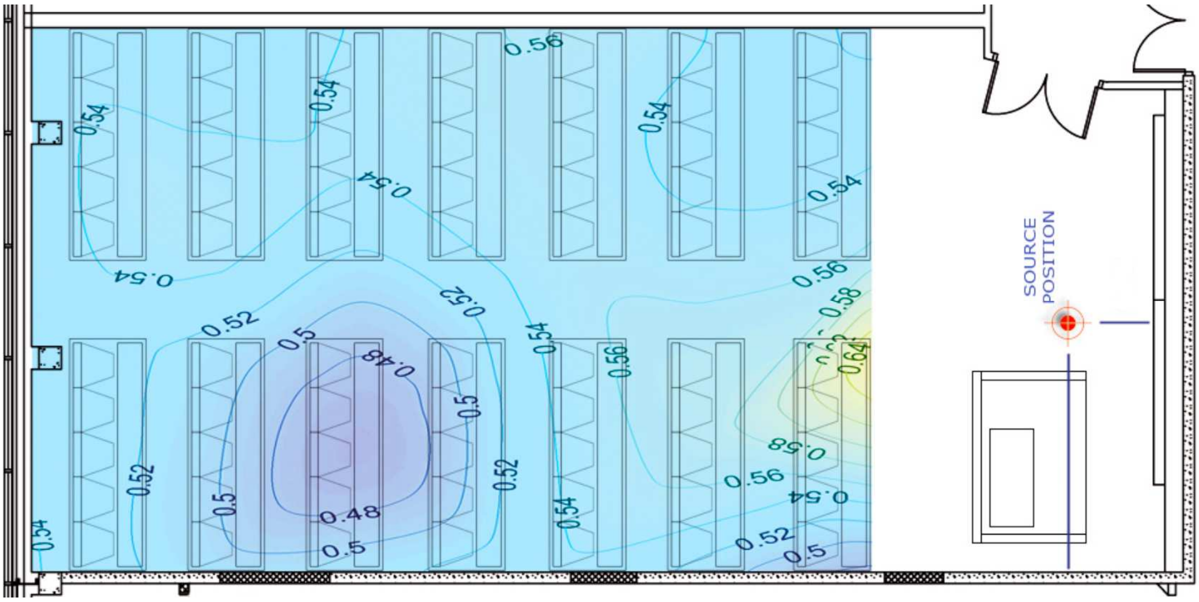


Figure 3 Room STI map for Room 5, the ‘worst’ performing room according to STI

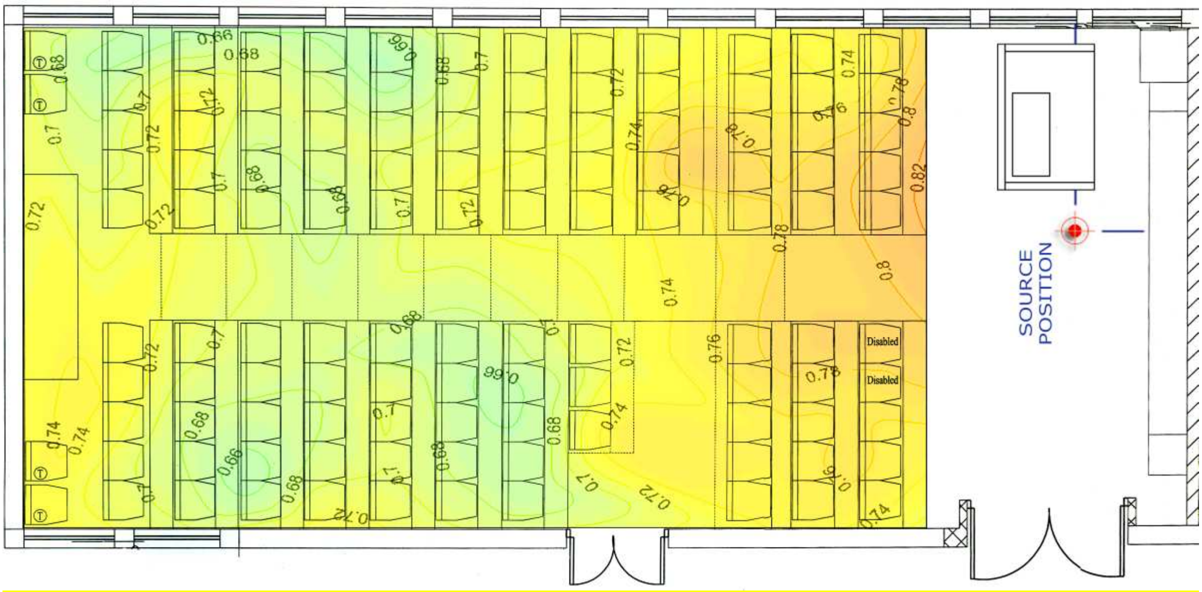


Figure 4 Room STI map for Room 3, the ‘best’ performing room according to STI

Comparing the results shown in Table 2 with the expected BB93 criteria shown in Table 1, it is clear that not all of the teaching spaces assessed fully meet the criteria. Background noise level measurements for all rooms were higher than the BB93 expectation. A predominant issue appeared to be extraneous noise infiltration via either/or both of the following key mechanisms: 1) poor isolation from neighboring corridors and/or poorly isolating windows, and 2) poorly designed forced air ventilation systems which appeared to be significant generators of background noise within each space. In terms of the room acoustic measurements, the worst performing room was room 5, which had an average STI figure of 0.54 and a relatively long reverberation time of 1.7 seconds, this room is illustrated in the room STI map of Figure 3. The room with the best overall performance, taking into account Noise level, Reverberation Time and STI was room 3 and this supports an anecdotal view held amongst a number of teaching staff using the different spaces on a regular basis. The room 3 STI map is shown in Figure 4.

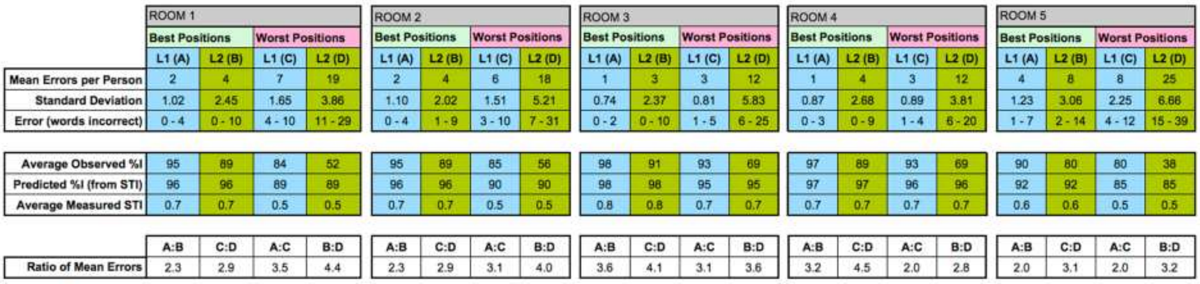
Room	Bad & Poor (0 to <0.3 to <0.45 STI)	Fair (0.45 to <0.6 STI)	Good (0.6 to <0.75 STI)	Excellent (0.75 to 1.0 STI)
1	0 %	11 %	87 %	2 %
2	0 %	11 %	89 %	0 %
3	0 %	0 %	78 %	22 %
4	0 %	0 %	97 %	3 %
5	0 %	95 %	5%	0 %

Table 3 Percentage of measurement points within each STI Band

Despite the acoustic performance issues, all of the rooms apart from room 5, contained at least some listening positions, that met the speech intelligibility qualification of ‘GOOD’ according to the ISO:9921(2003)[12] scale. However the spread of STIs in any room was observed to be quite large and dependent upon position within the room. Table 3 shows the percentage of measured STI points that fell within each qualification interval and this leads us to the notion of the 'best' and 'worst' listening position within each room.

[L1] and [L2] Listener Performance – Results and Discussion

Putting the measured acoustic performance data from Tables 2 & 3 into context of real listeners, Table 4 compares the average listener performance made by the [L1] and [L2] groups in each of the different rooms.



	ROOM 1				ROOM 2				ROOM 3				ROOM 4				ROOM 5			
	Best Positions		Worst Positions		Best Positions		Worst Positions		Best Positions		Worst Positions		Best Positions		Worst Positions		Best Positions		Worst Positions	
	L1 (A)	L2 (B)	L1 (C)	L2 (D)	L1 (A)	L2 (B)	L1 (C)	L2 (D)	L1 (A)	L2 (B)	L1 (C)	L2 (D)	L1 (A)	L2 (B)	L1 (C)	L2 (D)	L1 (A)	L2 (B)	L1 (C)	L2 (D)
Mean Errors per Person	2	4	7	19	2	4	6	18	1	3	3	12	1	4	3	12	4	8	8	25
Standard Deviation	1.02	2.45	1.65	3.86	1.10	2.02	1.51	5.21	0.74	2.37	0.81	5.83	0.87	2.68	0.89	3.81	1.23	3.06	2.25	6.66
Error (words incorrect)	0 - 4	0 - 10	4 - 10	11 - 29	0 - 4	1 - 9	3 - 10	7 - 31	0 - 2	0 - 10	1 - 5	6 - 25	0 - 3	0 - 9	1 - 4	6 - 20	1 - 7	2 - 14	4 - 12	15 - 39
Average Observed %I	95	89	84	52	95	89	85	56	98	91	93	69	97	89	93	69	90	80	80	38
Predicted %I (from STI)	96	96	89	89	96	96	90	90	98	98	95	95	97	97	96	96	92	92	85	85
Average Measured STI	0.7	0.7	0.5	0.5	0.7	0.7	0.5	0.5	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.6	0.6	0.5	0.5
	A:B	C:D	A:C	B:D	A:B	C:D	A:C	B:D	A:B	C:D	A:C	B:D	A:B	C:D	A:C	B:D	A:B	C:D	A:C	B:D
Ratio of Mean Errors	2.3	2.9	3.5	4.4	2.3	2.9	3.1	4.0	3.6	4.1	3.1	3.6	3.2	4.5	2.0	2.8	2.0	3.1	2.0	3.2

Table 4 Comparative results for the L1 & L2 listener performance for each room.

Figure 5 depicts actual room performance in terms of the average number of words interpreted incorrectly as a percentage of the total words observed. This gives a tangible appreciation of what the differences in STI mean for real listeners in each situation.

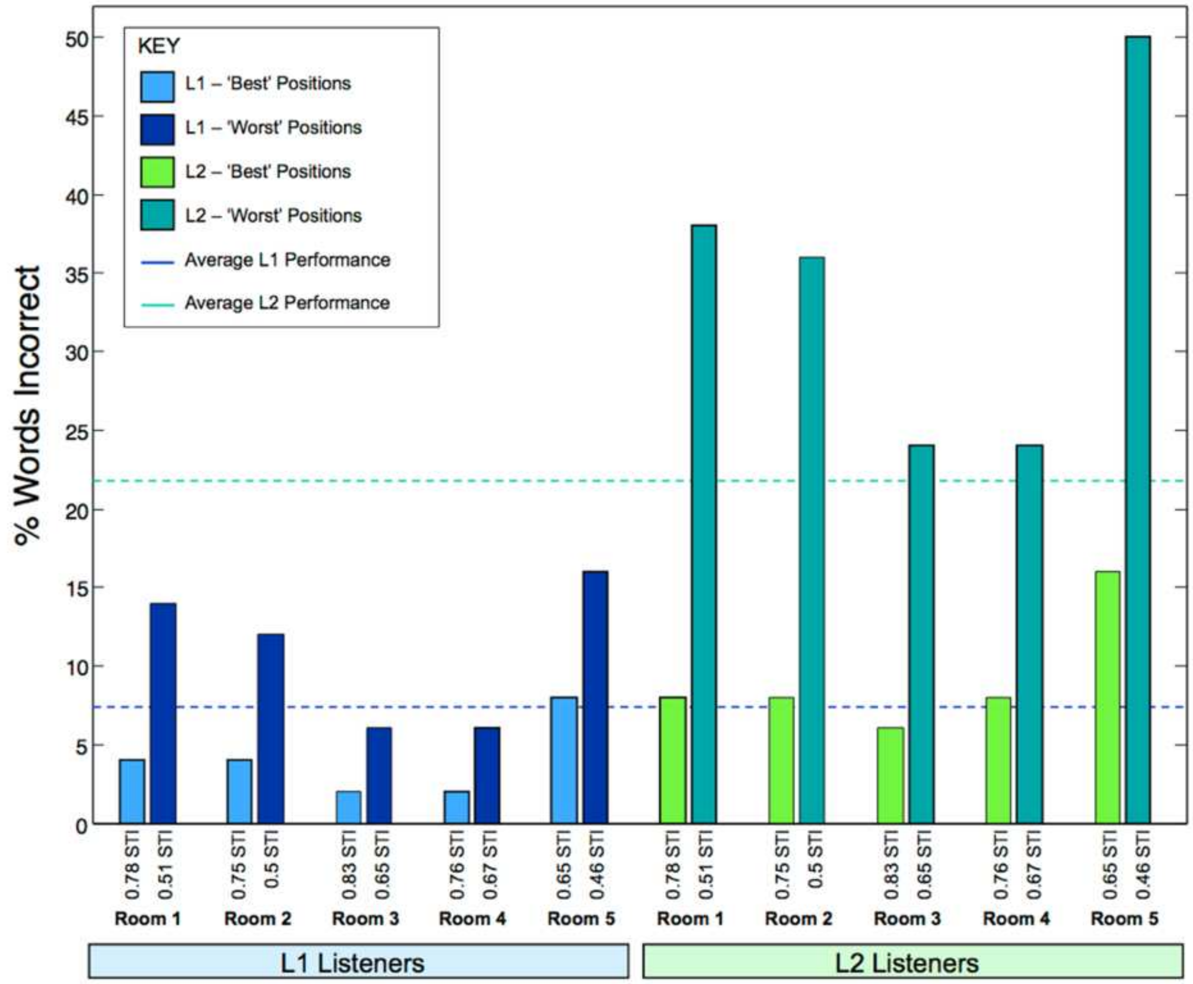


Figure 5 Comparison of [L1] to [L2] listener performance in terms of the average number of words incorrect (as a percentage) for the 'best' and 'worst' positions in each room

From Figure 5, in the apparently best room according to STI, room 3, and the best seats, [L1] listeners made on average 2% errors while the [L2] listeners averaged at around 6% errors. Going to the apparently worst performing room, room 5, [L1] listener performance was 7% errors at best and 15% at worst while the [L2] listeners achieved 16% errors at best and 50% errors at worst. The results indicate that the [L2] listeners tend to cope well in the best performing rooms (and best positions) and even perform almost on a par with the [L1] listeners. However, as the quality of the acoustic environment is reduced, the [L2] listeners are progressively less able to cope and make a significantly higher proportion of errors compared with the [L1] group.

The data presented here suggest that on average [L2] listeners are at a disadvantage to some extent in all rooms tested, in terms of their experienced speech intelligibility. However, listener performance also varies within each room and more often than not, the best performance can be achieved close to the front and center and/or as near as possible to the lectern. In a sense this is not surprising that listeners can 'hear better at the front' but what the results also suggest is that the [L2] listeners are much more affected by sitting away from the front than the [L1] group. While seating position may not necessarily affect student performance in terms of grades achieved at the end of a course[19,20], numerous studies[20,21] have shown that student engagement, enjoyment and evaluation of the learning experience (and the teacher) is very much affected by seating position. A major factor in this is the extent to which the student is able to attend to and engage in dialogue with the teacher - factors that are influenced significantly by the quality of the communication channel.

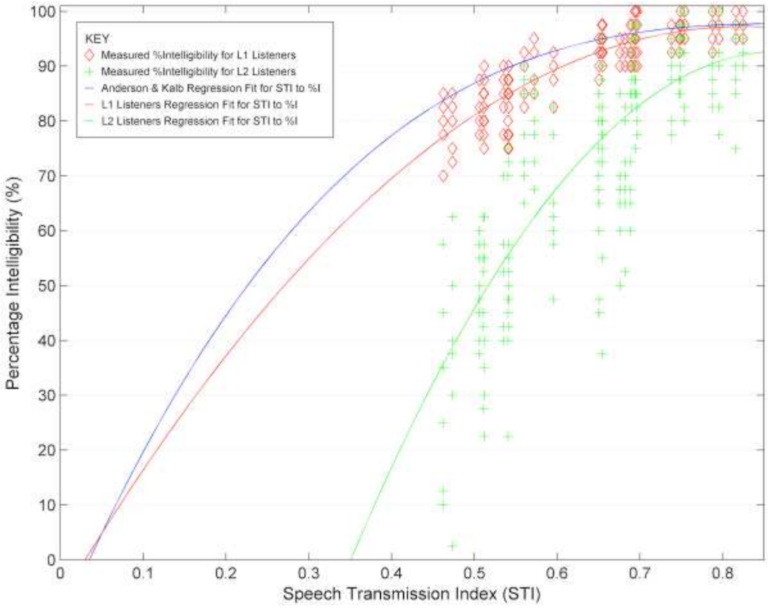


Figure 6 %Intelligibility for [L1] & [L2] Listeners -vs- Measured STI for all Rooms and Locations

Figure 6 depicts the observed speech intelligibility for the [L1] and [L2] groups across all of the rooms and listening positions plotted against the measured STI in each location. Of considerable importance in this study is the discovery that the degradation in performance for [L2] listeners is much more rapid than for [L1] listeners, this is clearly illustrated by the best fit curves for the [L1] and [L2] data sets shown. For the rooms evaluated the intelligibility tests show that STI is a reasonably reliable predictor of intelligibility for [L1] subjects since the intelligibility scores and STI match quite closely. This point is further illustrated by the best fit observed by Anderson & Kalb[15] for their experiment involving [L1] listeners only - shown in Figure 6 as the upper (blue) curve.

Referring to the [L2] listener data shown in Figure 6, it is apparent that STI is not a reliable predictor of perceived intelligibility for the [L2] group. For this group, Figure 6 shows that there is a very clear and marked difference in performance particularly at the lower end of the STI range. Furthermore, the [L2] best fit curve has a considerably steeper decay rate than that observed for the [L1] group and/or by Anderson & Kalb[15] in their [L1] only experiment. Taking this further and splitting the data into the 'best' and 'worst' listeners within the [L1] category, as shown in Figure 7, it is apparent that there are variations, particularly towards lower STIs, between the 'best' and 'worst' listeners but in general their performance is quite similar – this suggests some inter-subject variation as would be expected within a normal listening population. However, also shown in Figure 7, are the 'best' and 'worst' [L2] listeners for whom there is a much greater inter-subject variation within this group.

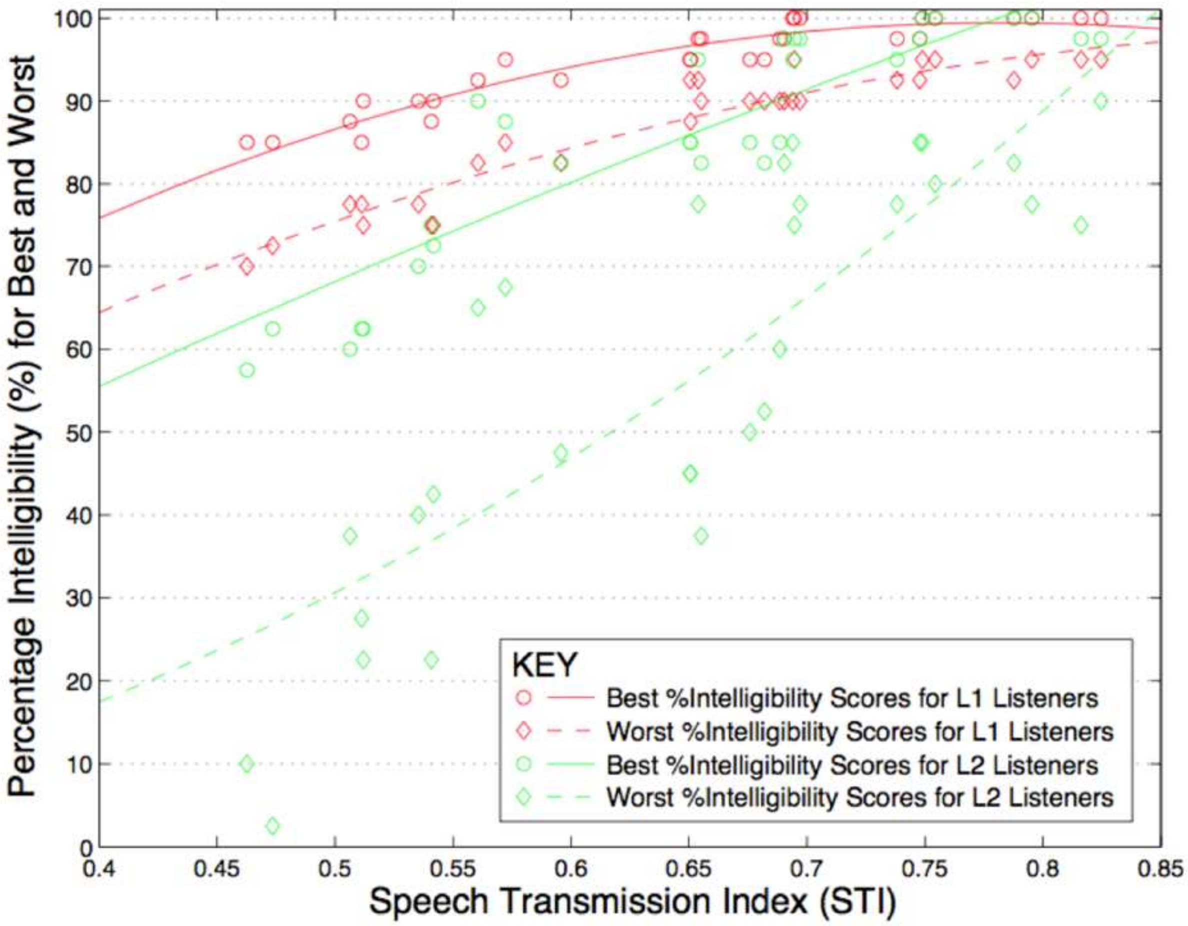


Figure 7 %Intelligibility for 'best' & 'worst' performing [L1] & [L2] Listeners, all Rooms & Locations

Looking at the 'best' [L2] listeners shown in Figure 7, it can be seen that at STIs above about 0.7, their performance is comparable to the general [L1] population, this evident where the best fit curves cross between the best L2 listeners and worst L1 listeners. However, for the 'worst' [L2] listeners in that group, they require the STIs to be in excess of about 0.8 for their performance to be comparable. What is also very apparent is the degree of separation between the best fit curves for the [L2] listeners. For the 'best' [L2] listeners, below an STI of 0.7 the trend clearly shows that the room is definitely affecting their performance to a greater extent than for the 'worst' [L1] listeners. But the problem really shows for the 'worst' [L2] listeners for whom their performance drops off very rapidly. Clearly, STI is not a reliable predictor of performance for the 'worst' [L2] listeners.

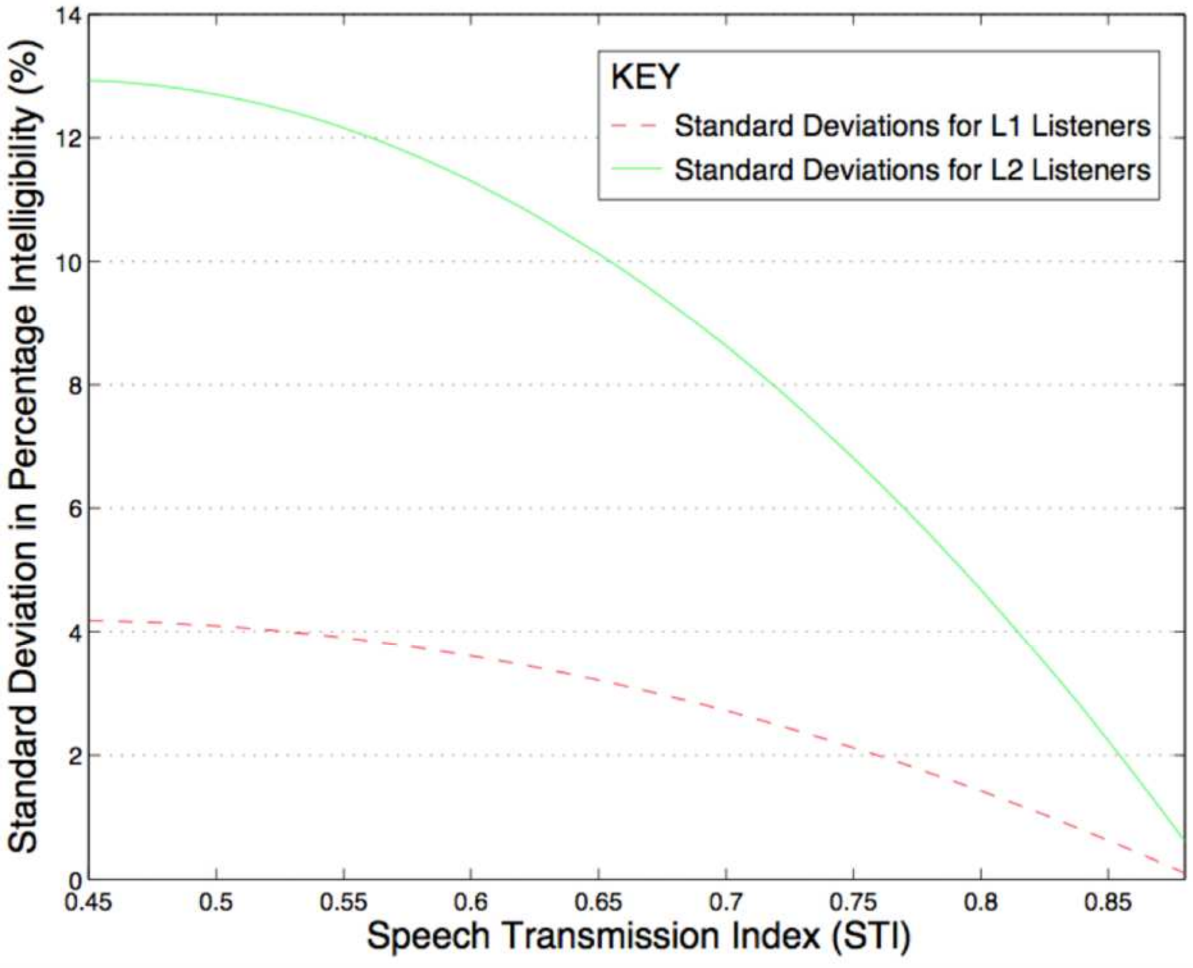


Figure 8 Standard Deviations in %Intelligibility for [L1] & [L2] Listeners

Looking at the standard deviations in [L1] and [L2] listener performance, as shown in Figure 8, the differences in inter-subject variation are apparent. For [L1] listeners, there is relatively little inter-subject variation overall, especially towards higher STIs although this does increase towards lower STIs as would be expected. For [L2] listeners however, the inter-subject variation is much larger. If we take the 4 standard deviations in percentage intelligibility as a benchmark, this is reached at an STI of 0.45 for [L1] listeners but, for the [L2] listeners the same standard deviation is reached at an STI of about 0.82. At lower STIs the standard deviation for [L2] listeners increases rapidly - yet these students all have similar/equivalent comprehension/listening scores based on an internationally accepted metric for language comprehension (IELTS).

CONCLUSIONS

One might expect that [L2] listeners with similar IELTS scores would have decay curves not too dissimilar to that of the [L1] population, i.e. you would expect the standard deviations to be roughly the same for both groups as this is an indicator of inter-subject variability. However, this was not the case, for the [L2] students there was a massive discrepancy in intelligibility. There was also some variation in the [L1] student population but this variation was considerably less at high STIs. One might conclude from this that the STI rating of 'Good' is just not acceptable for the [L2] population and even for the [L1] population there is some impact. Indeed anything lower than an STI of 0.75 or so is just not good enough. Considering that only 10% of all measured positions in this study exceeded this criteria, then the current aspirations in the design of teaching spaces appear not to be high enough.

In answering the question; '*How well does the accepted Speech Transmission metric predict [L1] and [L2] performance?*', the study suggests that for some (including the [L2] population at certain STIs) the metric is a good predictor of performance. However at low STIs there are some significant problems for [L2] listeners in particular. The BS-EN-60268-16 standard does discuss correction factors for [L2] groups which are broadly graded into various listening abilities. However this study tested for one of these specific groups and found that even within one group of listeners with apparently similar ability (according to IELTS) there was wide variability. This does raise some questions about either the recommendations within the standard concerning those specific listener groups and / or the robustness of currently accepted English Language proficiency tests such as IELTS. In both cases, whilst they are obviously useful they do need further clarification.

For the [L1] and [L2] population, it was shown that above about 0.75 STI, the 'worst' of the [L2] population lies within the STI of about 0.6 for the [L1] population. Below this point, [L1] performance decreases but in comparison the [L2] population is highly disadvantaged. However, this is not consistent across the whole [L2] population as the 'best' [L2] listeners (whilst performing worse than the 'worst' [L1] listeners) still find things to be reasonably intelligible at lower STIs. The 'worst' [L2] listeners on the other hand really struggle - and given our duty of care to be inclusive and provide learning environments that are suitable for all, we are obviously failing on those duties. Universities therefore need to increase their aspirations and design environments with very high STIs thus not disadvantaging our general [L2] population.

REFERENCES

1. BB93, Building Bulleting 93, DfES – UK, ISBN:0-11-271105-7. (2003).
2. Shield, B.M. and Dockerell, J.E., External and internal noise surveys of London primary schools. Journal of the Acoustical Society of America, 115, 730-738. (2001).
3. Wijngaarden, van S.J. and Verhave, J.A., Advances in STI Measuring Techniques, Proceedings of the Institute of Acoustics. (2005).
4. Rutherford, P., Wilson, R. and Hickman, V., (2006), IOA – Intelligible Measurements Conference Paper, London. (2006)
5. Steeneken, H.J.M. and Houtgast, T., A Physical Method for Measuring Speech Transmission Quality, Jnl of the Acoustical Society of America, Vol:67, pp:318-326. (1980)
6. Winjgarden, van S.J., Steeneken, H.J.M. and Houtgast, T., Quantifying the Intelligibility of Speech in Noise for Non-Native Listeners, Jnl of the AS of America, Vol:111(4). (2002)
7. Winjgarden, van S.J. et al., Using the Speech Transmission Index for Predicting Non-Native Speech Intelligibility, Jnl of the Ac' Soc' of America, Vol:115(3), pp:1281-1291. (2004).
8. BS-EN-ISO:354:2003, British Standard, Acoustics – Measurement of Sound Absorption in a Reverberation Room. (2003).
9. Mapp, P., Measuring Speech Intelligibility in Classrooms, with without Hearing assistance, IOA Proceedings, Vol:25(7) pp:156-160. (2003).
10. ITU-T:P.51-1993, Telephone Transmission Quality Objective Measuring Apparatus: Artificial Mouth, International Telecom's Union, ITU, Recommendation P.51. (1993).
11. Houtgast, T. and Steeneken H.J.M., Past, present and future of the Speech Transmission Index, TNO Human Factors, Soesterberg, The Netherlands, ISBN 90-76702-02-0. (2002).
12. ISO:9921-2003, Int' Standard, Ergonomics – Assessment of Speech Comm's. (2003).
13. ISO:TR-4870-1991, International Standard, Acoustics – The Construction and Calibration of Speech Intelligibility Tests. (1991).
14. Jacob, K.D., Correlation of Speech Intelligibility Tests in Reverberant Rooms with Three Predictive Algorithms, Jnl. of the Audio Engineering Society, Vol:37, pp:1020-1030. (1989).
15. Anderson, B.W. and Kalb, J.T., English Verification of the STI Method for Estimating speech Intelligibility of a Comm's Channel, Jnl of the Ac' Soc' of America, vol:81, (1987).
16. ISO:EN-60268-16-2011, International Standard, Sound System Equipment – part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index. (2011).
17. Mapp, P., Introduction to Intelligibility Measurements and Common Error Mechanisms, IOA Conference Paper, London. (2006).

18. Mapp, P., Error Mechanisms in Speech Intelligibility Measurements, IOA – Intelligible Measurements Conference Paper, London. (2006).
19. Williams, R.L., Relationship of Class Participation to Personality, Ability and Achievement, Jnl of Social Psychology, 83, 193-198. (1971).
20. Stires, L., Classroom Seating Location, Student Grades and Attitudes: Environment or Selection? Jnl of Environment & Behavior, 12, 241-254. (1980).
21. Montello, D. R., Classroom Seating Location and its Effect on Course Achievement, Participation and Attitudes, Jnl of Environmental Psychology, 8, 149-157. (1988).