SPEECH INTELLIGIBILITY IN HIGHER EDUCATION INSTITUTE LECTURE THEATRES

P Rutherford University of Nottingham, School of the Built Environment, Nottingham, UK.
R Wilson University of Nottingham, School of the Built Environment, Nottingham, UK.
V Hickman University of Nottingham, School of the Built Environment, Nottingham, UK.

1 INTRODUCTION

Recent years have borne witness to a significant growth in student numbers within UK universities. Although the bulk of this expansion has occurred through the recruitment of home students, significant emphasis has been placed on expanding the numbers of those from overseas. As a result, larger class sizes with increasing numbers of non-native English speakers are becoming more familiar to teaching academics. When combined with the substantial financial cost of education, this demographic change has resulted in increasing student expectation with respect to the quality of teaching and learning environment provided. Given that the lecture is often the forum for information delivery, the venue itself must cater for its target population. It is imperative therefore that room acoustics are seen as playing no part in reduced student academic achievement.

This study therefore investigates the influence that both excellent and degraded speech intelligibility characteristics have on word recognition by native and non-native student populations. Based on a simple investigation of speech transmission within lecture theatres through correlating its effect on listeners, it asks the following questions: (1) what quality of room facilities does a nationally leading university provide for its students? (2) what influence do degraded speech conditions have on intelligibility for native and non-native listeners? (3) which population of listeners is most severely affected by such degraded conditions? (4) what factors may be influencing these results? (5) what can teaching academics and institutions learn from such a study?

Several studies have investigated the performance of native and non-native speech perception in the presence of noise and reverberation[1,2,3,4,5,6]. Although the scope, aims and methodologies of these studies differ, three basic findings have emerged. Firstly, whilst native listener performance in degraded speech conditions decreases (through decreasing the signal to noise ratio or increasing the reverberant conditions), non-native listeners demonstrate considerably lower degrees of accuracy in speech recognition tasks than their native counterparts[1]. Secondly, context and word difficulty has been demonstrated to favour native as opposed to non-native listeners[2,4], suggesting that native listeners may be able to recover more easily than non-native listeners from degraded speech conditions. Finally, and in a similar vein to the second point, it has been mooted that degree of familiarity with the native language may play a significant part in speech recognition processes.

In order to investigate these criteria and questions, a general sweep of 15 lecture theatres was performed in early 2003 to paint a broad picture of the quality of facilities offered by the University of Nottingham. Two lecture facilities were identified as offering extremes of intelligibility and were investigated in more detail. Impulse responses representing the best and worst speech conditions from these rooms were acquired and convolved with word lists. These were subsequently presented to a sample group of native and non-native English listeners, with the findings forming the basis of this paper.

It must be noted however that this paper does not seek to directly correlate speech transmission (or any other articulation-based index) to the results gained from this study as this has been done extensively elsewhere (for example[3,4]). Instead, this paper seeks to paint a realistic picture of the effects that degraded speech conditions have on room occupants, therefore promoting the requirement for excellent acoustics within further education facilities.

2 ROOM MEASUREMENT AND IDENTIFICATION

An initial measurement sweep to rate 15 unoccupied university lecture theatres for their unaided speech acoustical qualities was performed, representing the general occupancy capacities and conditions of teaching facilities as offered by the University of Nottingham. Neither classroom equipment nor speech reinforcement systems were in operation at the time of measurement[7].

Room impulse responses were obtained using Acoustic Engineering's DIRAC software, a Brüel & Kjær 4292 Omnisource and 2260 Sound Analyser. Whilst the source was constrained to a position close to the lectern, receiver positions were distributed uniformly throughout each room, the number of these receiver positions proportional to the occupancy of the room and positioned at typical seated head height. A-weighted background noise levels varied between 32 and 39dB(A) across all rooms. Figure 1 illustrates the physical correlates obtained from these measurements.

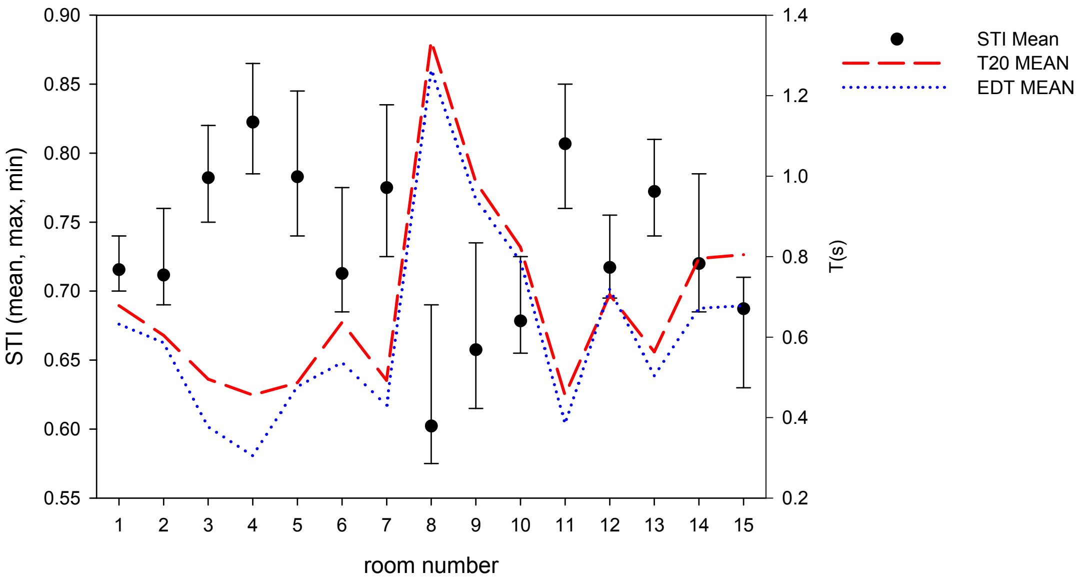

Figure 1. Room metrics obtained from measurement

This sweep revealed that facilities as offered by the university were, on the whole, satisfactorily fulfilling their teaching function with 41% of all seat positions having an STI score in excess of 0.75 (therefore classed as excellent), 52% in the 'good' range, and 7% in the 'fair' range. Those which performed worst generally had the highest occupancy (in excess of 250 seats), and were close to their original design condition, this reflected in their T20 (1KHz) being well in excess of that considered to be ideal for lecturing (0.7s). Many rooms exhibited large variations in their speech transmission properties, with all rooms, except one, being classed either as 'good' or 'excellent' on the STI quality class scale. This room, number 8, contained solely the 7% of total transmission scores in the 'fair' range and was investigated in more detail with the view to auralizing its impulse responses for listener testing. This and room 11, which was representative of the other extreme was short listed for further investigation.

The two short listed rooms were examined in great detail, with room acoustic parameters such as EDT and T20 extracted using the method described previously, with speech intelligibility measurements done in accordance with IEC 60268-16 using an artificial mouth, once again situated at the lectern position projecting perpendicular to the audience. Table 1 illustrates the key acoustic characteristics of both rooms.

	Lecture Room 11 (l) 15m, (w) 8m, (h) 4.85m, Tiered			Lecture Room 8 (l) 16m, (w) 17m, (h) 9.4m, Tiered		
	EDT (1KHz)	T20 (1kHz)	STI	EDT (1KHz)	T20 (1kHz)	STI
mean	0.31	0.40	0.84	1.28	1.29	0.62
stdev	0.054	0.028	0.012	0.083	0.042	0.036
range	0.23-0.45	0.34-0.46	0.8-**0.86**	1.11-1.45	1.21-1.39	**0.55**-0.71

Table 1. Results of in-depth analysis of selected rooms

This investigation yielded two contrasting listener positions for use in the intelligibility study. Lecture room 11 comprising of a deep, narrow plan (capacity 144) exhibited generally excellent acoustics across all listener positions with little variation in intelligibility, probably as a result of a recent fit-out. Lecture room 8 on the other hand, with its almost square plan, exhibited a wide variation in intelligibility across listener positions, with the worst positions corresponding to some front row seats. This was particularly alarming as seating position within educational environments has been demonstrated to affect grades and attitudes of students[8], and the positions identified were known by the authors' experience to be occupied primarily by overseas students. The positions relating to the extremes of these speech transmission scores were carried forward in the study for auralization.

3 SPEECH INTELLIGIBILITY TESTING IN SIMULATED SOUNDFIELDS

Having gained two monaural room impulse responses (MRIRs) relating to the best and worst-case scenarios, the purpose of this part of the study was to investigate the effect that the listening position, with its corresponding STI score, had on the perception of word units by both native [L1] and non-native [L2] listeners. It is widely accepted that whilst the speech transmission index is a reasonably robust indicator of the degree to which the transmission channel affects speech recognition potential for talkers and listeners in a native language, it by no means takes into consideration the language ability of the listener. It must be restated, however, that this study does not aim to provide correction factors that correlate speech transmission scores to non-native listening potential, but seeks to explore the effect that degraded speech conditions have on both [L1] and [L2] listeners through investigating the dominant characteristics that are affecting such listening perception, such as base language proficiency or room conditions. By exploring these effects, an understanding of appropriate teaching methods can be gained that will help enhance the teaching process under various listening conditions.

3.1 Experiment Design

Word Lists and Auralization Two lists of fifty unrelated monosyllabic, phonetically balanced words representative of the vocabulary of a built environment professional were recorded using native female clear speech in a hemi-anechoic vocal booth using a DPA 4060 omnidirectional microphone and DigiDesign Protools hardware. Each word was recorded with an inter-word interval of 8 seconds to allow sufficient time for later orthographic transcription by the listeners. Although debate exists as to the appropriateness of predictable and semantically anomalous sentences and word lists in speech recognition studies, this simple study deemed that with simple word lists, neither [L1] nor [L2] listeners could use contextual information to resolve ambiguities.

The two word lists were convolved with the MRIRs representing the two listener positions selected on the basis of their speech transmission index scores (0.86 in room 11 and 0.55 in room 8), therefore yielding 4 lists of convolved words. Although MRIRs are known to marginally under-predict speech intelligibility[9] due to the possible lack of binaural effects such as head shadowing, spatial separation etc., the applicability of such a technique was appropriate for this study as it was a constant to all listeners throughout the exercise.

Subjects Twenty native [L1] listeners (13 male, 7 female, average age 26 years, range 20-40 years,) and twenty non-native [L2] listeners (8 male, 12 female, average age 26 years, range 19 to

33 years) were selected from the staff and postgraduate population of the school and characterised into two groups:

- Five L2[E] from Europe representing German, French, Spanish and Bulgarian languages.
- Fifteen L2[A] from Asia representing Thai, Mandarin, Cantonese, Japanese and Korean languages.

All [L2] listeners had spent at least 6 months in the UK and were deemed proficient at English language listening comprehension, with L2[A] listeners having a total IELTS score of at least 6. No listener reported any prior history of hearing impairment.

Procedure Listeners were placed in a sound attenuating booth and asked to wear headphones (Sennheiser HD600). The convolved word lists were presented diotically at 70dB (SPL). Each subject was presented with one of the convolved word lists representative of a particular room condition, taking approximately 10 minutes to complete the study therefore minimising experimental fatigue. Having completed this task, they returned several days later to complete the list of words that was representative of the other room condition; therefore never being exposed to the same word list, negating any learning effects.

Prior to the test, it was explained to each listener that must undertake two simultaneous tasks. With task 1, the listener was expected to transcribe exactly what they heard, even if they only had partial perception of the spoken word; that is they were required to write phonetically their response. Having written down the word, listeners embarked on task 2, which rated the degree of difficulty they had in perceiving the word, whereby a score of 1 meant the word was easily recognised and a score of 5 meant it was exceptionally difficult to recognise. In total, each subject transcribed 100 words during the two experiment sessions.

Data Analysis To minimise any bias through lack of understanding of the method employed, the first 10 words from each word list were discounted in the analysis, giving a total of 800 words in each room condition. All errors were counted for further linguistic breakdown, but for the purpose of this paper, only total word errors will be described. Data analysis (both word error count and difficulty rating) sought to identify differences within listener groups and between varying speech transmission conditions in the two room conditions and is therefore presented in both relative and absolute terms.

3.2 Results of Word Error Analysis - Overall Errors and Percentage of Correct Words, [L1] and Combined [L2] Listeners

Figure 2 reveals clearly that the combined [L2] listening population made far more word errors overall in response to both excellent and degraded room acoustic conditions than the [L1] listeners.

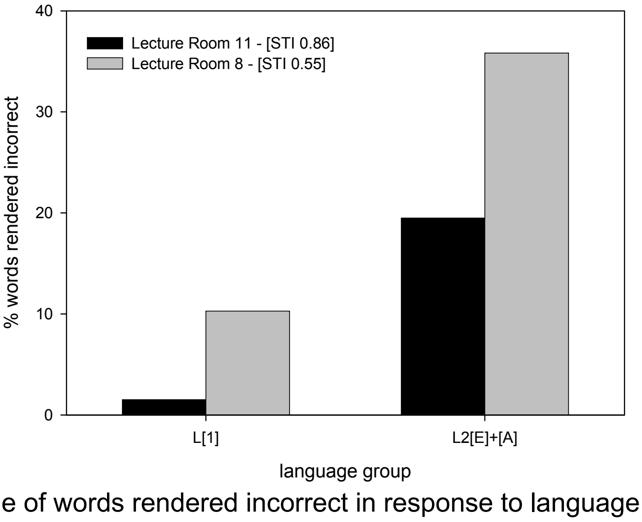

Figure 2. Percentage of words rendered incorrect in response to language group and room type

The results, summarised in table 2, show that under excellent speech conditions, [L2] listeners made nearly 13 times as many errors as [L1] listeners. Whilst [L1] listeners found the exercise under excellent speech conditions relatively easy (difficulty rating of 0.51 out of 5), [L2] listeners found it only marginally more difficult with a mean difficulty score 0.56 points higher. Under degraded speech conditions, [L2] listeners made nearly 3.5 times as many errors as [L1] listeners, but interestingly found the exercise no more difficult. This suggests that there is a significant difference between languages for the exercise.

Investigating the interaction between single languages and speech conditions (e.g. the difference in performance between [L1] listeners in excellent and degraded conditions) reveals the obvious fact that both groups made considerably more errors under degraded listening conditions as compared with excellent conditions. Interestingly, [L2] listeners made (relatively) fewer errors (1.84 times) in the excellent vs. degraded listening conditions than [L1] listeners (6.8 times). Additionally, when looking at the difficulty ratings, both groups reveal the obvious fact that they found the degraded listening conditions more difficult than the excellent listening conditions. [L1] listeners found the degraded conditions to be considerably more difficult (almost 1 difficulty point harder) than the [L2] listeners (less than ½ difficulty point harder).

When looking at the results with respect to absolute percentages of incorrectly rendered word units, [L1] listeners under excellent conditions rendered 1.5% words incorrectly, and under degraded conditions, 10.3%. [L2] listeners on the other had showed a substantial performance decrement under both listening conditions, rendering 19.5% words incorrect under excellent conditions and 35.8% under degraded conditions (as shown in figure 2).

	LR1 [STI 0.86]		LR2 [STI 0.55]		ratios			
	L1$^{(a)}$	L2$^{(b)}$	L1$^{(c)}$	L2$^{(d)}$	(a):(b)	(c):(d)	(a):(c)	(b):(d)
mean errors per person	0.6	7.7	4.05	14.15	12.8	3.49	6.8	1.84
stdev	1	7	1.7	6.84				
range	0 – 4	1 - 24	0 - 9	7 - 27				
mean difficulty score	1.51	2.07	2.5	2.52				
stdev	0.54	0.69	0.6	0.66				
range	1.1-2.2	1.3-3.3	1.7-3.7	1.5-3.8				

Table 2. Overall errors and difficulty score [L1] vs combined [L2] listeners

3.2.1 Brief discussion

The results indicate that [L2] listeners perform significantly worse than [L1] listeners in both excellent and degraded room conditions. Indeed, under excellent conditions, [L2] listeners had almost 13 times as many errors as [L1] listeners. When put into context, [L2] listeners made on average only 1.84 times as many errors between the excellent and degraded conditions, whereas [L1] listeners made almost 7 times as many errors. This suggests that for this combined [L2] listening population, responding to the task of word recognition and transcription was itself very difficult under excellent, nevermind degraded speech conditions, suggesting that the this ratio was disproportionately affected by these base errors under excellent conditions. This mirrors the work of Mack[10] who found similar error statistics for native and non-native populations when presented with natural and synthesized (vocoded) speech. This discussion is reinforced by the small variation in difficulty rating and large variation in range and standard deviation of combined [L2] results, suggesting that there was indeed a considerable range in ability to perform the task. To investigate this range of ability (or suitability of task), data were subsequently broken down representing the two [L2] listening groups described in section 3.1.

3.3 Results of Word Error Analysis - Overall Errors and Percentage of Correct Words, [L1] and Separated [L2] Listeners

Figure 3 reveals clearly that whilst combined [L2] listeners made considerably more word errors in response to both excellent and degraded room acoustic conditions than [L1] listeners, those most affected were L2[A] listeners.

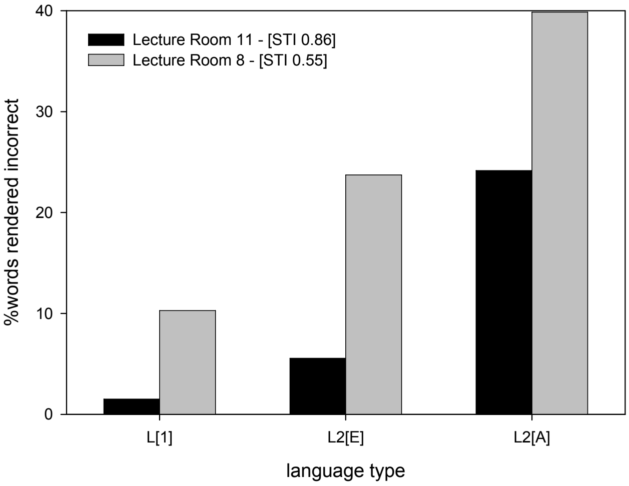

Figure 3. Percentage of words rendered incorrect in response to separated language group and room type

Under excellent listening conditions, L2[E] listeners made nearly 4 times as many errors as [L1] listeners, whilst L2[A] listeners made nearly 16 times as many errors (see Table 3). In a similar pattern to that described in section 3.2, under degraded listening conditions, L2[E] listeners made 2.3 times more errors than L1 listeners, with L2[A] listeners making nearly 4 times the number of errors. Once again, this suggests that there is a significant difference between language ability for the exercise.

When investigating the interaction between single languages and room acoustic conditions, it is obvious that all groups made considerably more errors under degraded than excellent listening conditions. Investigating this phenomenon in more depth by looking at the percentage disadvantage that degraded speech conditions have on listening within a particular language group:

- [L1] listeners make on average 8.8% more errors (equating to an additional 3.45 more words wrong),
- L2[E] listeners make on average 18.14% more errors (equating to an additional 7.2 more words wrong),
- L2[A] listeners make on average 15.7% more errors (equating to an additional 6.2 more words wrong).

Therefore, in this instance with a speech transmission difference of 0.31 between two rooms, L2 listeners make almost double the number of word errors as L1 listeners, however, the ratio between these word errors decreases with language group (6.8 to 4.3 to 1.7).

	LR1 [STI 0.86]			LR2 [STI 0.55]			ratios						
	L1[(a)]	L2[E][(b)]	L2[A][(c)]	L1[(d)]	L2[E][(e)]	L2[A][(f)]	(a):(b)	(a):(c)	(d):(e)	(d):(f)	(a):(d)	(b):(e)	(c):(f)
mean errors	0.6	2.2	9.53	4.05	9.4	15.73	3.7	15.9	2.3	3.9	6.8	4.3	1.7
stdev	1	1	7	1.7	2.1	7							
range	0 - 4	1 - 3	1 - 24	0 - 9	7 - 11	7 - 27							

Table 3. Overall errors and difficulty score [L1] vs [L2] listeners: breakdown by language type

4 DISCUSSION

Reinforcing the discussion from sections 3.2.1 and results from section 3.3, there is little doubt that all [L2] listeners, irrespective of their language origin, performed significantly worse than L1 listeners under both excellent and degraded listening conditions. To understand this, it is helpful to break these results down into the two room conditions experienced.

Under excellent speech transmission conditions, [L1] listeners recorded very few errors (1.5%), with L2[E] listeners, although performing considerably worse (5.6%), still performing adequately. L2[A] listeners on the other hand showed great difficulty with this exercise, getting over 24% of words wrong in what could be classed as an 'easy' exercise. Mack[10] reported a similar trend (although not to the extreme of the L2[A] group), with work by Bradlow and Pisoni[2] suggesting that [L2] listeners have problems discriminating fine phonetic detail, something that was expected of them in this speech intelligibility task. However, the score differences between the two [L2] groups suggest that the listener's linguistic background may have fundamentally affected the results of the 'easy' listening test. van Wijngaarden et al.[3], as mooted in the introduction, suggest several variables that may play a role in this difference. Firstly, languages that bear some similarity may affect word comprehension differently from languages that have very little in common. Secondly, the [L2] average experience with the second language may have a major bearing. In this case, L2[E] listeners are exposed continually in their home countries to spoken English, such as through visual and aural media. Age of acquisition of the target language is another important variable, suggesting that those who learned English at an earlier stage in their education proved more proficient at the task (i.e. L2[E] participants generally begin English language learning in primary school, whereas L2[A] begin in secondary school). The final potential factor in this debate centres around continued native language usage, therefore suggesting (and confirmed through interview) that many L2[A] listeners had spent only a brief period of study in the UK, this being their first real exploration of English language. As the base line for this experiment was an IELTS score of 6 (representing 57.5% correct responses in listening tests), it suggests that whilst all L2[A] listeners in this experiment exceeded their IELTS listening scores, many did not poses the base ability to undertake this experiment, a point that is evident in the range and standard deviation of the word scores.

If it is assumed that the ratios described throughout this paper are fundamentally influenced by either language or task competency and can therefore be accounted for, it is evident that both groups of [L2] listeners made approximately (and absolutely) twice as many word errors as [L1] listeners did under degraded listening conditions (as represented by the difference in STI of 0.31 between the two rooms) as reported in section 3.3. This suggests that, irrespective of English proficiency level, [L2] listeners perform significantly worse under degraded speech intelligibility conditions than [L1] listeners[1,5,11]. Indeed, Nábělek and Donahue[12], in their study of the influence of reverberation on the perception of consonants by non-native listeners found that even short reverberation times had a more profound impact in consonant recognition than on native listeners (whereby consonant recognition is affected more by decreased intelligibility than is vowel recognition), with degraded conditions affecting listening more in the non-native than native language. Mack[10] and Florentine[13] attribute this performance decrement to dissimilar word processing strategies at a base language level, whereby native listeners were able to make more effective use of their language knowledge. Indeed, Florentine[13] suggests that non-native listeners may need to recognise nearly every word clearly in order to understand English speech. Upon closer linguistic examination of word errors in the experiment, it was evident that whilst there were very few inexplicable linguistic mistakes with words, the majority of mistakes were simple phonemic substitutions, these substitutions more prevalent under degraded conditions (such as 'quit' becoming 'quick') for all listeners, but increased in proportion for [L2] listeners, suggesting a major interaction with the temporal and intensity smearing associated with lower STI scores. Additionally, several word errors could be attributed to inter-language interference (for example 'did' was often perceived as 'deed' by native French and Spanish listeners). If such word-by-word processing by [L2] listeners is an essential component of their speech recognition ability (and hence their learning ability), it is essential that lecture theatres do not degrade word recognition accuracy.

5 CONCLUSIONS

Although this paper presents a modest investigation into the effects of lecture theatre acoustics on both native and non-native listeners, it has highlighted several points that should be borne in mind when teaching students. Firstly, and obviously, not all students have the same base language ability, therefore lecturers must seek to explore techniques that ensure the taught message is being delivered in an effective manner that caters for all listeners. Secondly, it is evident from this (and other) studies that degraded speech conditions have a fundamental influence on perceived intelligibility for all listeners. Whilst native listeners may be able to recover from such degradation through access to redundant linguistic, semantic and social information[1], those most severely disadvantaged are the non-native population who may not have access to this information. It is essential therefore that university lecture theatres are designed with great care, with a full understanding of the potential effects that degraded listening conditions have on all those who occupy its spaces. Additionally, lecturers must be made aware of issues that they themselves can control, such as facilitating visual cueing and being aware of speaking directionality by avoiding the age old practice of 'chalk and talk', the use of speech reinforcement systems where available, the provision of handouts, and a general level of control over the class where self-noise is often problematic. Finally, having the same lecturer throughout a lecture series may offer some advantage as studies have shown that familiarity with the talker's voice enhances word recognition accuracy under difficult listening conditions (for example[14]).

Fortunately, work by van Wijngaarden et al.[4] seeks to quantify non-native speech intelligibility with the aim of providing correction factors that may be applied to the Speech Transmission Index, and it is hoped that at some stage, such data will be an integral component of room acoustics prediction and measurement software.

6 REFERENCES

1 Takata, Y., and Nábělek, A.K., "English consonant recognition in noise and in reverberation by Japanese and American listeners," Journal of the Acoustical Society of America. 88, 663-666. (1990).

2 Bradlow, A.R., and Pisoni, D.B., "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors," Journal of the Acoustical Society of America. 106, 2074-2085. (1999).

3 van Wijngaarden, S.J., Steenken, H.J.M., and Houtgast,T., "Quantifying the intelligibility of speech in noise for non-native listeners," Journal of the Acoustical Society of America. 111, 1906-1916. (2002).

4 van Wijngaarden, S.J., Bronkhorst, A.W., and Steenken, H.J.M., "Using the Speech Transmission Index for predicting non-native speech intelligibility," Journal of the Acoustical Society of America. 115, 1281-1291. (2004).

5 Cutler, A., Weber, A., Smits, R., and Cooper, N., "Patterns of English phoneme confusions by native and non-native listeners," Journal of the Acoustical Society of America. 116, 3668-3678. (2004).

6 Garcia Lecumberri, M.L., and Cooke, M., "Effect of masker type on native and non-native consonant perception in noise," Journal of the Acoustical Society of America. 119, 2445-2454. (2006).

7 Hodgson, M., "Rating, ranking, and understanding acoustical quality in university classrooms," Journal of the Acoustical Society of America. 112, 568-575. (2002).

8 Stires, L., "Classroom seating location, student grades and attitudes: Environment or selection?," Environment and Behaviour. 12, 241-254. (1980).

9 Peng, J., "Feasibility of subjective speech intelligibility assessment based on auralization," Applied Acoustics. 66, 591-601. (2005).

10 Mack, M., "Sentence processing by non-native speakers of English: Evidence from the perception of natural and computer-generated anomalous L2 sentences," Journal of Neurolinguistics. 3, 293-316. (1988).

11 Bergman, M., Aging and the perception of speech, University Park Press, Baltimore. (1980).

12 Nábělek, A.K., and Donahue A.M, "Perception of consonants in reverberation by native and non-native listeners," Journal of the Acoustical Society of America. 75, 632-634. (1984).

13 Florentine, M., "Non-native listeners' perception of American English in noise," in Proceedings of Inter-Noise '85. 1021-1024. (1985).

14 Nygaard, L.C., and Pisoni, D.B., "Talker-specific learning in speech perception," Percept. Psychophys. 60, 335-376. (1998).