

# Intelligibility of Reverberant Speech with Amplification: Limitation of Speech Intelligibility Metrics, and a Preliminary Examination of an Alternative Approach

Doheon Lee\*, Eunju Gong, Densil Cabrera, Manuj Yadav, William L. Martens

The Faculty of Architecture, Design and Planning, The University of Sydney, Sydney, Australia  
Email: [\\*dlee7117@uni.sydney.edu.au](mailto:dlee7117@uni.sydney.edu.au)

Received 15 October 2014

---

## Abstract

This study examines the effect of speech level on intelligibility in different reverberation conditions, and explores the potential of loudness-based reverberation parameters proposed by Lee *et al.* [J. Acoust. Soc. Am., 131(2), 1194-1205 (2012)] to explain the effect of speech level on intelligibility in various reverberation conditions. Listening experiments were performed with three speech levels ( $L_{Aeq}$  of 55 dB, 65 dB and 75 dB) and three reverberation conditions ( $T_{20}$  of 1.0 s, 1.9 s and 4.0 s), and subjects listened to speech stimuli through headphones. Collected subjective data were compared with two conventional speech intelligibility parameters (Speech Intelligibility Index and Speech Transmission Index) and two loudness-based reverberation parameters ( $EDT_N$  and  $T_N$ ). Results reveal that the effect of speech level on intelligibility changes with a room's reverberation conditions, and that increased level results in reduced intelligibility in highly reverberant conditions.  $EDT_N$  and  $T_N$  explain this finding better than do STI and SII, because they consider many psychoacoustic phenomena important for the modeling of the effect of speech level varying with reverberation.

## Keywords

Speech Intelligibility, Speech Level, Room Reverberation, Loudness

---

## 1. Introduction

Speech intelligibility measurements are considered to be important in establishing acoustical performance for public buildings, as most of them are intended for speech communication rather than music performance. Buildings and rooms may require a different degree of speech intelligibility depending on their intended use. For example, high speech intelligibility is desirable for classrooms and conference venues, whereas speech privacy and confidentiality (*i.e.*, low speech intelligibility) are often required in open offices, hospitals and lawyer's of-

---

\*Corresponding author.

fices. Given that buildings that do not function properly are often demolished or renovated before reaching their planned life spans, speech intelligibility also influences the longevity of buildings. For speech intelligibility to match the intended use of a room, it is important to accurately estimate or predict intelligibility at the design stage. While a room's reverberation condition and background noise are known to be the two most important factors governing speech intelligibility (at a given speaker-to-listener distance), the present study investigates the effect of speech level on intelligibility in various reverberation conditions—which is relevant to practical situations in which amplification is used. In an effort to better explain the results, this study explores the potential of psychoacoustic parameters to explain how speech level in conjunction with a room's reverberation condition affects intelligibility.

To estimate speech intelligibility, ISO 3382-1 [1] recommends clarity index ( $C_{50}$ ) and definition ( $D_{50}$ ). These parameters are computed based on energy ratios of time periods in the squared sound pressure decay envelope of a room impulse response, which represent a room's reverberation condition. Because the effect of background noise is not incorporated, these parameters are not suitable for situations where background noise is the dominant interference. The speech transmission index (STI) specified in IEC 60268-16 [2] is derived from modulation transfer function (MTF) of the transmission channel (e.g. from talker to listener position in a room), which quantifies the degree of reduction in the envelope modulation depth, resulting from sound travelling from one position to another. Because both background noise and reverberation affect the modulation depth, this approach often yields a good representation of speech intelligibility in noisy and reverberant conditions. The speech intelligibility index (SII) [3] is a complex form of a weighted speech-to-noise ratio, but a room's reverberation condition is taken into account when SII is computed from MTFs. While SII is less commonly used in room acoustics contexts, it does have the advantage of a more detailed approach to auditory modeling (including the use of critical-band filters). The STI and SII have functions for auditory spectral masking and hearing threshold, but do not consider auditory temporal masking in their calculation.

In addition to background noise and reverberation, it is obvious that speech level also affects speech intelligibility, but this relationship is subtler than might be initially assumed. Clearly, very quiet speech (near the hearing threshold) will have reduced intelligibility relative to mid-level speech (e.g. at 60 dBA), and this is modeled by STI and SII. However, in some circumstances, increasing the speech level results in the opposite effect—a reduction in intelligibility, an effect shown in studies by Fletcher [4], Kryter [5], Pollack and Pickett [6], and Hagerman [7]. Kryter found that intelligibility starts to fall when the speech level exceeds 75 dB in a reverberant room, but the reduction of intelligibility was not observed in the same experiment performed in an anechoic condition. Pollack and Pickett conducted listening experiments with a range of speech-to-noise ratios, and reported that the change in intelligibility associated with speech level is more evident for lower speech-to-noise ratios. Hagerman also concluded that intelligibility decreases when speech level is greater than 55 dB when a speech-to-noise ratio remains constant.

The negative relationship between high speech level and intelligibility is often explained by auditory spectral masking—because it becomes stronger with louder sound. In STI, the masking intensity at an octave band,  $k$ , is modeled by considering speech and noise level observed in the adjacent previous octave band,  $k-1$  (*i.e.* the octave just below). SII employs a similar method, but functions in SII also consider bandwidth (or center frequency depending on band-pass filters), so that the equivalent masking spectrum level increases for higher frequency bands. In addition to spectral masking, auditory temporal integration (and masking) is also likely to affect intelligibility at high speech level, but (as mentioned previously) this is not included in STI or SII.

The effect of auditory temporal masking may be somewhat analogous to the effect of reverberation in reducing intelligibility (both could be considered to reduce amplitude modulation depth of the transmitted signal due to temporal smearing), and this could be a basis for an interaction effect. Therefore, in this study we consider a model that includes temporal masking (at least, in a simple way): the loudness decay analysis approach proposed by Lee *et al.* [8] [9] for the prediction of the subjective extent of reverberation (hereafter, reverberance). In this approach, computerized loudness models (the Time-varying Loudness Model by Glasberg and Moore [10] or the Dynamic Loudness Model by Chalupper and Fastl [11]) are used to analyze room impulse responses, and parameters are derived from the resulting loudness decay functions. The loudness-based parameters, named  $T_N$  and  $EDT_N$ , are found to outperform the conventional reverberation time (RT) and early decay time (EDT) specified in ISO 3382-1 [1] in predicting reverberance for various types of stimuli. Importantly, the study of Lee and Cabrera [8] found that functions for temporal integration within the loudness models substantially contribute to the performance of  $T_N$  and  $EDT_N$ -based reverberance predictions. Furthermore, the amount of temporal integration

also varies with the playback level of stimuli. According to Poulsen [12], Sone *et al.* [13] and Florentine *et al.* [14], the maximum temporal integration was observed at moderate levels (e.g. 40 dB to 60 dB for a 1 kHz tone and 60 dB to 80 dB for broadband noises). Lee *et al.* (2010, 2012) also reported that subsequent sound is more masked when the same sound is listened at higher level in a reverberant condition. Given that reverberance has a strong effect on intelligibility, the inclusion of temporal integration has the potential to be beneficial for predicting the effect of speech level on speech intelligibility in a reverberant condition.

To the authors' knowledge, there have been only a few studies systematically investigating the effect of speech level varying with reverberation conditions on intelligibility. Therefore, in the present study, listening experiments are performed with a range of speech levels ( $L_{Aeq}$  of 55 dB, 65 dB and 75 dB) and reverberation conditions ( $T_{20}$  of 1.0 s, 1.9 s and 4.0 s). Collected subjective data are analyzed with two conventional intelligibility parameters (STI and SII) and two psychoacoustic-based reverberation parameters ( $EDT_N$  and  $T_N$ ), to explore whether incorporating psychoacoustic phenomena beyond spectral masking is beneficial for modeling the subjective speech intelligibility varying with speech level in reverberation conditions.

## 2. Computation of $EDT_N$ and $T_N$

$EDT_N$  and  $T_N$  are computed with the following procedures. First,  $L_{AFmax}$  of a room impulse response (RIR) or a binaural room impulse response (BRIR) is adjusted to match with  $L_{Aeq}$  of music or speech.  $L_{AFmax}$  is the A-weighted sound pressure level (SPL) with a temporal integration of 125 ms, and  $L_{Aeq}$  is the power averaged SPL over a given time period. Second, the level-adjusted RIR is input to a computerized loudness model for the calculation of loudness decay envelop of the RIR. Third, a linear regression line is drawn on the loudness decay envelope of the RIR over 0.707 to 0.178 of loudness of a direct sound for  $T_N$ ; and over loudness of a direct sound to a half loudness of a direct sound for  $EDT_N$ . Based on the Stevens' power law [15], these evaluation ranges correspond to the  $-20$  dB and  $-10$  dB evaluation ranges of the conventional  $T_{20}$  and EDT, respectively. Note that the Schroeder's reverse integration method [16] is not applied to the loudness decay of the RIR because the loudness summation is different to the sound pressure summation. For this reason, the direct sound is not necessarily the greatest value of the loudness decay envelope. Last, similarly to the conventional reverberation parameters,  $EDT_N$  and  $T_N$  are calculated by multiplying the time taken over the linear regression line by 6 and 3, respectively.

Lee and Cabrera [8] reported that the Time-Varying Loudness Model [10] and the Dynamic Loudness Model [11] perform equally well for  $T_N$  and  $EDT_N$ . However, the latter model extended with functions for binaural loudness summation (as per Moore and Glasberg [17]) is chosen for the present study because BRIRs are convolved with anechoic speech for the listening experiment.

## 3. Subjective Experiments

### 3.1. Subjects

Twenty subjects were recruited for the listening experiment (fourteen male and six female) on a volunteer basis, and thirteen of the subjects participated in the experiments twice. The subjects were aged from 22 to 46 years old with a median age of 28 years old. Nine subjects had professional or educational background in acoustics and five subjects had previously participated in similar listening experiments.

### 3.2. Stimuli

A synthesized voice in MAC OS X software, Text-to-Speech, spoke 55 sets of words listed in Annex A of AS 2282 [18]. Each set consists of 6 words having similar pronunciations (e.g. mop, hop, cop, top, shop and pop) for speech intelligibility experiments. The synthesized voice spoke the words with a carrier sentence, "Can you choose the word \_\_\_\_?" The voice named "Lee" was chosen because "Lee" speaks with an Australian accent, which was familiar to the most of the subjects.

A BRIR was measured in a reverberant high-ceiling auditorium (the Great Hall of the University of Sydney) and was convolved with the dry speech spoken by "Lee". In that measurement, a logarithmic swept sinusoid (with a sampling of 48 kHz, a duration of 60 s, and a frequency range of 50 Hz to 20 kHz) was played through a loudspeaker, Meyer sound UPA-1P, placed on a stand on stage. A Brüel & Kjær head and torso simulator (type 4128C) was set up in the audience area—at 20 m away from the source on stage—to record the sine sweep. The

resulting BRIR has  $T_{20}$  of 1.9 s, averaged left and right ear values over the 250 Hz to 4 kHz in octave band range. To simulate more reverberation conditions, the sound pressure decay of the BRIR was modified in a way suggested by Cabrera *et al.* [19]. First, the BRIR was octave-band filtered from 125 Hz to 8 kHz, and the noise floor of the band-pass filtered RIRs was decayed out at the same rate of the dominant decay envelope (see **Figure 1** for an example). Second, the noise-free BRIR was multiplied by exponential functions so that its  $T_{20}$  value changed to 1.0 s, and to 4.0 s. The sound pressure level of the speech stimuli convolved with the original and modified RIRs was adjusted for the headphone presentation of the stimuli to be 55 dB, 65 dB and 75 dB in  $L_{Aeq}$ . These presentation levels were based on  $L_{Aeq}$  averaged over 990 stimuli (55 sets of 6 words multiplied by 3 reverberation conditions), so differences in  $L_{Aeq}$  between the stimuli (due to the particular speech content) are preserved in the listening experiment.

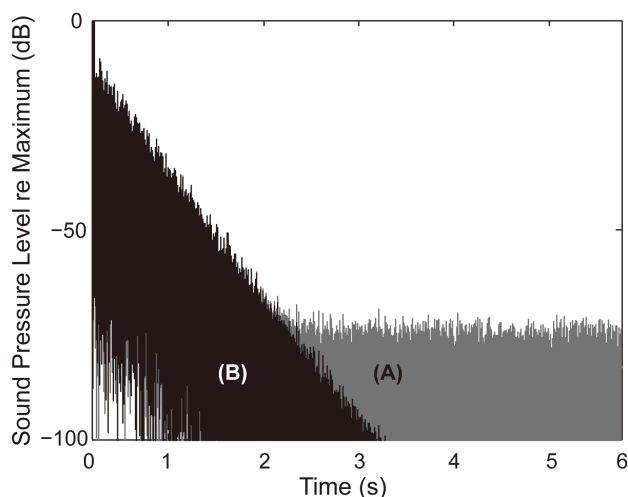
### 3.3. Procedure

One word was randomly chosen from each set of 6 words, and the subjects listened the chosen word with the carrier sentence through headphones (Sennheiser HD600) in a quiet listening environment. The task was to identify the target word in the stimulus sentence from 6 words displayed on the graphical user interface (GUI) (which was implemented with MATLAB). The synthetic voice “Lee” did not correctly pronounce 12 of 330 words listed in Annex A of ANSI 2282 [18], so these were excluded from the random word selections. A training experiment was performed prior to the actual experiment for the subjects to be familiarized with the GUI and fully understand the task. The experiment took about 45 minutes, and the subjects were free to take a short-break at anytime during the experiment.

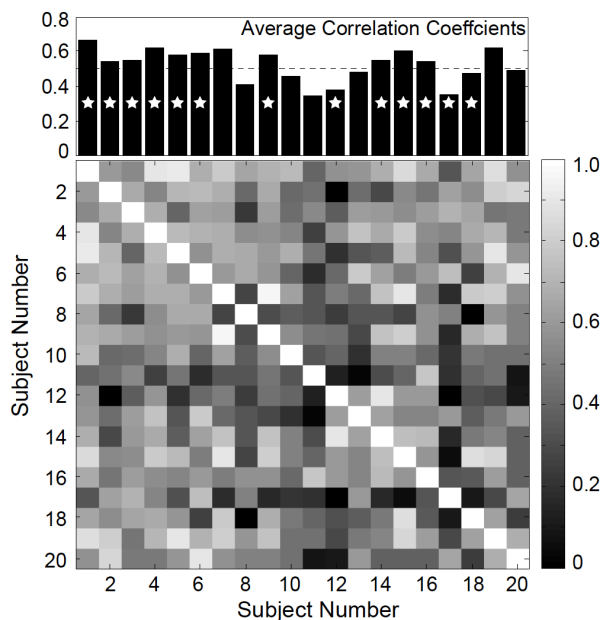
## 4. Results

### 4.1. Reliability of Subjective Responses

The reliability of the collected subjective data was examined to exclude responses of atypical subjects from further analyses. For the present study, atypical subjects were determined based on a degree to which each subject’s responses correlate with the responses of other subjects. First, subjective responses were z-scored so that the responses of every subject have a mean value of 0 and a standard deviation of 1. Second, correlation coefficients between the z-scored subjective responses (hereafter, simply “subjective responses”) of every subject were calculated. As seen in the upper chart of **Figure 2**, eight subjects (Subjects 8, 10, 11, 12, 13, 17, 18 and 20) have an averaged correlation coefficient of less than 0.5, so they were excluded from analyses. Stars in the same figure indicate the subjects that participated in the experiment twice. Given that the averaged difference in the response between the two attempts is only 1.3%, subjects 12, 17 and 18 appear to consistently respond in a different way to the other subjects.



**Figure 1.** The normalized sound pressure level of a room impulse response (A) and the same room impulse response with its noise floor decaying out at the same rate of the dominant decay (B).



**Figure 2.** Correlation coefficients of z-scored subjective responses. The bar chart illustrates the correlation coefficients averaged over the subjects. The subjects participated in the experiment twice are marked with stars.

#### 4.2. Effect of Room Reverberation

The subjective responses are plotted as a function of reverberation condition (**Figure 3(A)**) and as a function of reverberation condition for each speech level (**Figure 3(B)**). Visual inspection of **Figure 3(A)** reveals that speech intelligibility has an apparent negative relationship with a room’s reverberation condition (as expected). According to a one-way ANOVA, the effect of reverberation on intelligibility is significant at a confidence level of 99% ( $F = 98.01$ ,  $p < 0.001$ ). The same analyses performed for each speech level also lead to the same conclusion (see the values tabulated in **Figure 3(B)**). A Tukey/Kremer *post hoc* test was executed to investigate whether the subjective responses are significantly different between the three reverberation conditions. The results show that the significant difference is found for all possible pairs of the three reverberation conditions at each speech level.

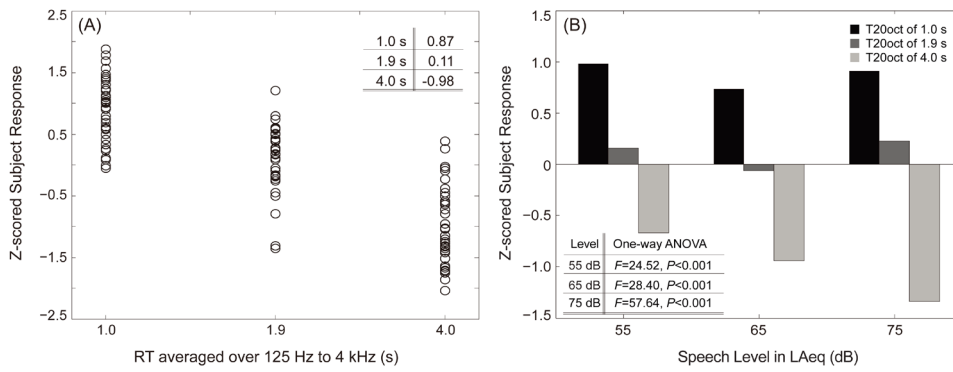
#### 4.3. Effect of Speech Level

The subjective responses are plotted as a function of speech level (**Figure 4(A)**) and as a function of speech level for each reverberation condition (**Figure 4(B)**). Results of a one-way ANOVA indicate that the effect of speech level is only significant when  $T_{20}$  is 4.0 s ( $F = 4.07$ ,  $p < 0.05$ ) at a confidence level of 95%. A Tukey/Kremer *post hoc* test shows that the significant difference in the subjective responses is only observed between speech level of 55 dB and 75 dB when  $T_{20}$  is 4.0 s, as indicated with a star in **Figure 4(B)**. Therefore, the effect of speech level appears to be contingent on a room’s reverberation condition.

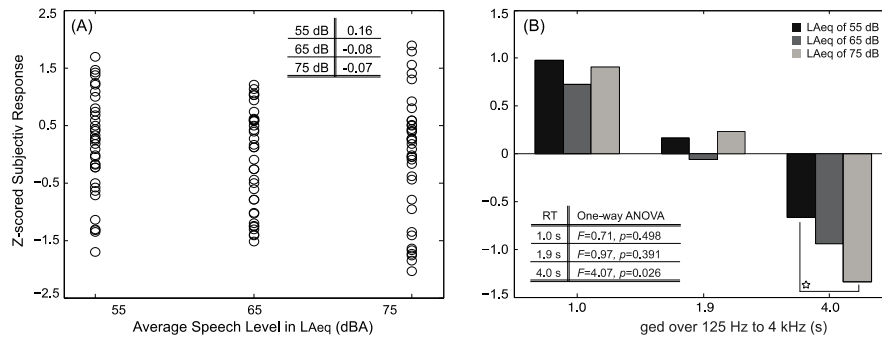
#### 4.4. Parametric Analyses

Two conventional speech intelligibility parameters (STI and SII) and two loudness-based reverberation parameters ( $T_N$  and  $EDT_N$ ) were computed to see if the parameters effectively model the subjective results as observed in the previous sections. For STI computations, octave-band values of speech level were derived as per Annex J.2 in IEC 60286-16 [2] using the convolved speech stimuli. For both STI and SII computations, background noise level was set to zero as the listening experiment was performed in quiet conditions. These and other acoustic parameters were calculated using a suite of functions in MATLAB, known as AARAE [20] [21].

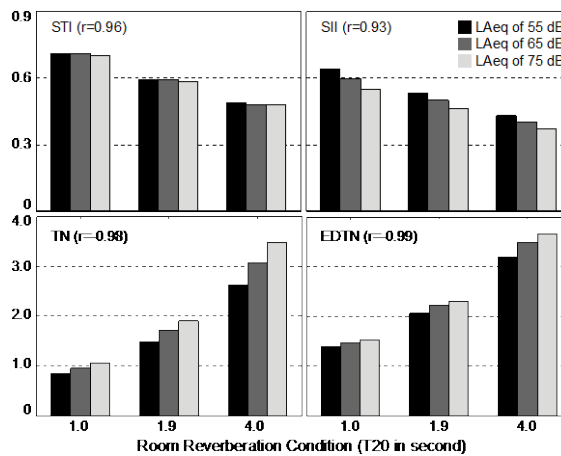
As seen in **Figure 5**, STI and SII decrease as  $T_{20}$  increases, but  $T_N$  and  $EDT_N$  have a positive relationship with  $T_{20}$ —because they are reverberation parameters. Apart from STI, the parameters are sensitive to the variation of



**Figure 3.** (A) Z-scored subjective responses as a function of reverberation condition. Averaged values of the responses for each reverberation condition are tabulated inside the figure; (B) Z-scored subjective responses as a function of reverberation condition for three speech levels. Black color is the responses collected for  $T_{20}$  of 1.0 s. Dark gray color is the responses collected for  $T_{20}$  of 1.9 s, and light gray color is the responses collected for  $T_{20}$  of 4.0 s. A one-way analysis of variance (ANOVA) is executed on the z-scored subjective data and reverberation conditions. Results of a one-way ANOVA are tabulated inside the figure.



**Figure 4.** (A) Z-scored subjective responses as a function of speech level. Averaged values of the responses for each speech level are tabulated inside the figure; (B) Z-scored subjective responses as a function of speech level for three reverberation conditions. Black color is the responses collected for speech level of 55 dB. Dark gray color is the responses collected for speech level of 65 dB, and light gray color is the responses collected for speech level of 75 dB. A one-way analysis of variance (ANOVA) is executed on the z-scored subjective responses and listening level. Results of a one-way ANOVA are tabulated inside the figure.



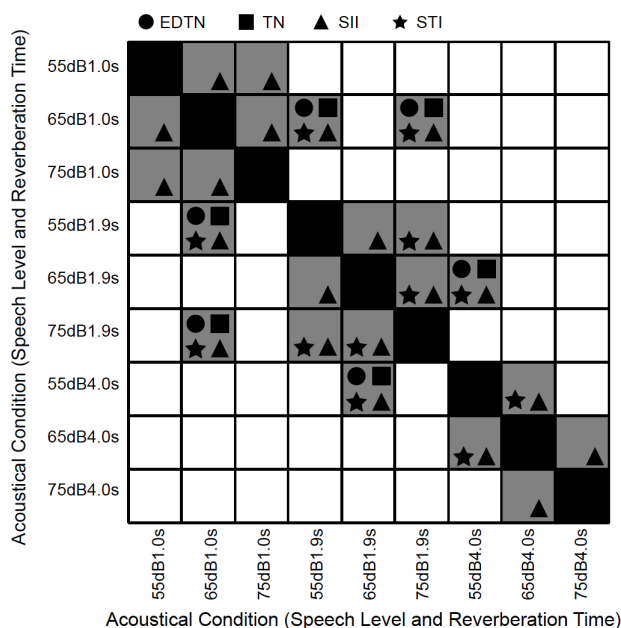
**Figure 5.** Speech Transmission Index (STI), Speech Intelligibility Index (SII) and two loudness-based reverberation parameters ( $T_N$  and  $EDT_N$ ) are computed from three BRIRs having  $T_{20}$  of 1.0 s, 1.9 s and 4.0 s. Three speech levels used for the calculations are indicated with colors. Correlation coefficients between corresponding parameters and z-scored subject responses are tabulated inside the figure.



speech level. SII varies almost equally for the change of speech level in three reverberation conditions, but the loudness-based parameters change more in higher reverberation conditions. In other words, STI and SII do not explain the significant difference in the subjective response that is only observed between speech level of 55 dB and 75 dB at  $T_{20}$  of 4.0 s (see **Figure 4(B)**). Because  $T_N$  and  $EDT_N$  more accurately model the masking effect of speech level varying with reverberation conditions, values of these parameters better correlate with subjective intelligibility than do STI and SII (see the averaged correlation coefficient values given in **Figure 5**). Some of parameters specified in ISO 3382-1 [1]— $T_{20}$ , EDT,  $C_{50}$  and  $D_{50}$ —were also computed and compared with the subjective data.  $T_{20}$  and EDT yields a correlation coefficient of  $r = 0.97$  with the subjective data, which is slightly higher than STI.  $C_{50}$  and  $D_{50}$  also highly correlate with intelligibility ( $r = 0.96$ ).

More analyses were performed to find acoustical conditions where subjective speech intelligibility is significantly different, and parameters that accurately predict such conditions. For this, the subjective responses were grouped into nine acoustic scenarios (3 reverberation conditions and 3 speech levels), and a Tukey/Kremars *post hoc* test was executed. Results are shown in **Figure 6**, of which *x*- and *y*-labels represent speech level and  $T_{20}$ . White color indicates pairs of conditions where intelligibility is significantly different at a confidence level of 90% and gray color indicates pairs of conditions where speech intelligibility is subjectively indistinguishable at the same confidence level. As seen in **Figure 6**, “55 dB 1.9 s” and “55 dB 4.0 s” have subjectively same intelligibility as “65 dB 1.0 s” and “65 dB 1.9 s”, respectively. This indicates that increasing speech level by 10 dB and shortening a room’s reverberation by  $T_{20}$  of 0.9 s or 2.0 s seem equally beneficial to intelligibility (albeit the former is considerably more labor and cost effective than the latter in most situations). However, this trend is not observed when speech level is 75 dB. For example, the subjective responses at “75 dB 1.0 s” and “75 dB 1.9 s” have significantly different mean values to those at “65 dB 1.9 s” and “65 dB 4.0 s”. Furthermore, it is hard to explain why “75 dB 1.9 s” has intelligibility equal to “65 dB 1.0 s”.

The symbols in **Figure 6** indicate values of corresponding parameters do not correctly predict the similarity of speech intelligibility. The smallest change of the parameters resulting in significantly different intelligibility is set as limens, *i.e.*, 2% for STI, 4% for SII, 15% for  $EDT_N$ , and 33% for  $T_N$ —so the all the significantly different conditions are assumed to be correctly predicted. While Bradley *et al.* [22] reported that 0.03 is the JND of STI, a 2% change of STI in the present study corresponds to 0.01. Please note that the JND of  $EDT_N$  and  $T_N$  is unknown due to their recent development. With these limens, it is found that  $EDT_N$  and  $T_N$  are 98% consistent with the results of Tukey/Kremars *post hoc* tests, while STI and SII are 92% and 69%, respectively.



**Figure 6.** Results of a Tukey/Kremars *post hoc* test performed on the *z*-scored subjective responses. White color indicates pairs of conditions where the subjective data has a significant mean difference at a confidence level of 90%. Gray color indicates two conditions where the *z*-scored subjective data is not significantly different at the same confidence level.

## 5. Discussion

The results of the listening experiments show that a room's reverberation condition has a significant effect on intelligibility. However, the negative effect of speech level on intelligibility is contingent on a room's reverberation condition—only significant at high speech levels in a high reverberation condition, *i.e.*, between 55 dB and 75 dB at  $T_{20}$  of 4.0 s. Since Kryter [5] reported that speech intelligibility starts to decrease at a speech level of 75 dB in a moderate reverberant condition (RT of 1.6 s for a 500 Hz tone), if speech levels higher than 75 dB had been tested in the present study, a significant negative level effect might have been expected at  $T_{20}$  of 1.0 s and 1.9 s. The negative effect of speech level varying with reverberation is also supported by the studies of Studebaker *et al.* [23] and Pollack and Pickett [6]. These studies concluded that speech intelligibility in a quiet anechoic condition does not change for speech levels higher than 70 dB.

The subjects participated in a short interview after the listening experiments. They reported greater difficulty in listening at higher speech level, although they could still correctly recognize the words in speech. Given that the speech stimuli tested in the present study are in a form of simple sentences, testing more complex sentences with the Listening Difficulty Rating Method suggested by Morimoto *et al.* [24] has the potential to be more sensitive to the effect of speech level in the moderate reverberation conditions. In the Listening Difficulty Rating Method, subjects rate the listening difficulty into one of four categories: 1) not difficult, 2) a little difficult, 3) fairly difficult, and 4) extremely difficult, rather than finding correct words. Morimoto *et al.* [24] posited that correctly repeating and finding words in speech does not necessarily mean there is no difficulty in listening, which is consistent with what was conveyed by the subjects in the short interviews.

The present study uses  $EDT_N$  and  $T_N$  to investigate if incorporating more psychoacoustic phenomena is beneficial for modeling the effect of speech level contingent on reverberation. As  $EDT_N$  and  $T_N$  are derived using a computerized loudness model, the parameters consider many complexities of loudness besides spectral masking—such as temporal integration, outer/middle ear transfer functions, auditory filter banks and functions relating auditory excitation to specific loudness. While the statistical analyses indicate that  $EDT_N$  and  $T_N$  correlate well with the subjective data, care must be taken when using these parameters for estimating intelligibility because the parameters do not consider background noise and would yield a smaller value (which would be incorrect to interpret as increased intelligibility) as speech levels decrease below the hearing threshold.

To examine the performance of  $EDT_N$  and  $T_N$  as intelligibility parameters in the presence of background noise, an additional experiment was performed in the Great Hall of the University of Sydney—where the BRIR described in Section 3.2 was measured. The background noise at the time of experiment was 30 dB in  $L_{Aeq}$  averaged from 250 Hz to 8 kHz in octave band. The anechoic speech stimuli spoken by 'Lee' were presented at  $L_{Aeq}$  of 60 dB, 70 dB and 80 dB at receiver positions of 5 m and 20 m away from a source on stage. Like the laboratory experiment, subjects identified the target word (*i.e.* a randomly selected word from each set of 6 words having similar pronunciations) in the stimulus sentence from 6 words given on a paper. According to a correlation coefficient analysis,  $T_N$  and  $EDT_N$  correlate substantially less with the subjective data ( $r = -0.22$  and  $0.63$ , respectively) than do STI and SII ( $r = 0.86$  and  $0.89$ , respectively). Therefore,  $EDT_N$  and  $T_N$  seem to be inappropriate for estimating intelligibility when speech is mixed with background noise.

According to a one-way ANOVA test executed on the subjective data collected from the Great Hall experiment, the effect of speech level is significant at the two receiver positions, but this effect is stronger at 20 m, *i.e.*, ( $F = 2.76$ ,  $p < 0.1$ ) at 5 m and ( $F = 13.7$ ,  $p < 0.01$ ) at 20 m. One may raise a question as to why the significant effect of speech level is not observed at  $T_{20}$  of 1.9 s in the laboratory experiment, which simulates the acoustical condition of the Great Hall. This appears to be because a higher signal-to-noise ratio lessens the adverse effect of speech level on intelligibility [7] [23], leading to the stronger effect of speech level in the Great Hall experiment, as the laboratory experiment was conducted in a quiet listening condition. The additional gain of 5 dB in the Great Hall experiment does not seem to contribute towards the inconsistency, because the subjective data collected from this experiment are significantly different between speech level of 60 dB and 70 dB and between 60 dB and 80 dB, while as seen in **Figure 4(B)** the subjective data collected from the laboratory experiment do not have a significant mean difference between any speech level at  $T_{20}$  of 1.9 s—especially between 65 dB and 75 dB.

While the loudness-based parameters,  $EDT_N$  and  $T_N$ , show some promise in this study because of their consideration of temporal masking, they are not generally suitable as speech intelligibility parameters without substantial further development. Most obviously, they do not (on their own) consider the audibility of speech, and so will not be useful as predictors of intelligibility of quiet speech. Nevertheless, their usefulness in modeling



loud reverberant speech points to possible future refinements of speech intelligibility parameters for application in architectural acoustics contexts.

## 6. Conclusions

The effect of reverberation condition on speech intelligibility is significant for the three tested reverberation conditions, *i.e.*,  $T_{20}$  of 1.0 s, 1.9 s and 4.0 s. However, the effect of speech level is only significant between speech level of 55 dB and 75 dB at  $T_{20}$  of 4.0 s in a quiet listening condition. When there is background noise, the significant effect of speech level is observed between speech level of 60 dB and 70 dB and between 60 dB and 80 dB at  $T_{20}$  of 1.9 s. These findings lead to the conclusion that higher reverberation condition and background noise increase the adverse effect of speech level on intelligibility.

The loudness-based reverberation parameters,  $EDT_N$  and  $T_N$ , are successful in modeling the effect of speech level varying with a room's reverberation condition (for medium to loud reverberant conditions). As  $EDT_N$  and  $T_N$  incorporate many psychoacoustic phenomena (besides spectral masking that STI and SII mainly consider as psychoacoustic phenomenon important for speech intelligibility), comprehensive psychoacoustic approaches have some potential in the modeling of intelligibility at high speech level and in a high reverberation condition. However, the present study does not suggest using  $EDT_N$  and  $T_N$  as intelligibility parameters when there is background noise or speech level is very low (e.g. close to hearing threshold), because these parameters do not consider the effect of background noise and furthermore continuously yield a lower value (which represent high speech intelligibility) although speech level decreases below the hearing threshold.

## References

- [1] ISO 3382-1 (2009) Acoustics—Measurement of Room Acoustic Parameters—Part I: Performance Spaces. International Organization for Standardization.
- [2] IEC 60268-16 (2011) Sound System Equipment—Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index. International Electrotechnical Commission.
- [3] ANSI S3.5 (1997) Methods for Calculation of the Speech Intelligibility Index. American National Standard.
- [4] Fletcher, H. (1922) The Nature of Speech and Its Interpretation. *Bell Labs Technical Journal*, **1**, 129-144. <http://dx.doi.org/10.1002/j.1538-7305.1922.tb00384.x>
- [5] Kryter, K.D. (1946) Effects of Ear Protective Devices on the Intelligibility of Speech in Noise. *Journal of the Acoustical Society of America*, **18**, 413-417. <http://dx.doi.org/10.1121/1.1916380>
- [6] Pollack, I. and Pickett, J.M. (1958) Masking of Speech by Noise at High Sound Levels. *Journal of the Acoustical Society of America*, **30**, 127-130. <http://dx.doi.org/10.1121/1.1909503>
- [7] Hagerman, B. (1982) Sentences for Testing Speech Intelligibility in Noise. *Scandinavian Audiology*, **11**, 79-87. <http://dx.doi.org/10.3109/01050398209076203>
- [8] Lee, D. and Cabrera, D. (2010) Effect of Listening Level and Background Noise on the Subjective Decay Rate of Room Impulse Responses: Using Time-varying Loudness to Model Reverberance. *Applied Acoustics*, **71**, 801-811. <http://dx.doi.org/10.1016/j.apacoust.2010.04.005>
- [9] Lee, D., Cabrera, D. and Martens, W.L. (2012) The Effect of Loudness on the Reverberance of Music: Reverberance Prediction Using Loudness Models. *Journal of the Acoustical Society of America*, **131**, 1194-1205. <http://dx.doi.org/10.1121/1.3676602>
- [10] Glasberg, B.R. and Moore, B.C.J. (2002) A Model of loudness Applicable to Time-Varying Sound. *Journal of the Audio Engineering Society*, **50**, 331-342.
- [11] Chalupper, J. and Fastl, H. (2002) Dynamic Loudness Model (DLM) for Normal and Hearing-Impaired Listeners. *Acta Acustica United with Acustica*, **88**, 378-386.
- [12] Poulsen, T. (1981) Loudness of Tone Pulses in a Free Field. *Journal of the Acoustical Society of America*, **69**, 1786-1790. <http://dx.doi.org/10.1121/1.385915>
- [13] Sone, T., Suzuki, Y., Kumagai, M. and Takahashi, T. (1986) Loudness of a Single Burst of Impact Sound: Results of Round Robin Tests in Japan (I). *Journal of the Acoustical Society of Japan*, **7**, 173-182. <http://dx.doi.org/10.1250/ast.7.173>
- [14] Florentine, M., Buus, S. and Poulsen, T. (1996) Temporal Integration of Loudness as Function of Level. *Journal of the Acoustical Society of America*, **99**, 1633-1644. <http://dx.doi.org/10.1121/1.415236>
- [15] Stevens, S.S. (1955) The Measurement of Loudness. *Journal of the Acoustical Society of America*, **27**, 815-829.

- <http://dx.doi.org/10.1121/1.1908048>
- [16] Schroeder, M.R. (1965) New Method of Measuring Reverberation Time. *Journal of the Acoustical Society of America*, **37**, 409-412. <http://dx.doi.org/10.1121/1.1909343>
- [17] Moore, B J.C. and Glasberg, B.R. (2007) Modeling Binaural Loudness. *Journal of the Acoustical Society of America*, **121**, 1604-1612. <http://dx.doi.org/10.1121/1.2431331>
- [18] AS 2282 (1985) Acoustics—Methods of Assessing and Predicting Speech Privacy and Speech Intelligibility. Australian Standard.
- [19] Cabrera, D., Lee, D., Yadav, M. and Martens, W.L. (2011) Decay Envelope Manipulation of Room Impulse Responses: Techniques for Auralization and Sonification. *Proceedings of Acoustics*, Gold Coast, 2-4 November 2011, CD-ROM.
- [20] Cabrera, D., Jimenez, D. and Martens, W.L. (2014) Audio and Acoustical Response Analysis Environment (AARAE): a Tool to Support Education and Research in Acoustics. *Proceedings of Internoise*, Melbourne, 16-19 November 2014, CD-ROM.
- [21] Cabrera, D., Lee, D., Leembruggen, G. and Jimenez, D. (2014) Increasing Robustness in the Calculation of the Speech Transmission Index. *Building Acoustics*, **12**, 181-198. <http://dx.doi.org/10.1260/1351-010X.21.3.181>
- [22] Bradley, J.S., Reich, R. and Norcross, S.G. (1999) A Just Noticeable Difference in  $C_{50}$  for Speech. *Applied Acoustics*, **58**, 99-108. [http://dx.doi.org/10.1016/S0003-682X\(98\)00075-9](http://dx.doi.org/10.1016/S0003-682X(98)00075-9)
- [23] Studebaker, G.A., Sherbecoe, R.L., McDaniel, D.M. and Gwaltney, C.A. (1999) Monosyllabic Word Recognition at Higher-Than-Normal Speech and Noise Levels. *Journal of the Acoustical Society of America*, **105**, 2431-2444. <http://dx.doi.org/10.1121/1.426848>
- [24] Morimoto, M., Sato, H. and Kobayashi, M. (2004) Listening Difficulty as a Subjective Measure for Evaluation of Speech Transmission Performance in Public Spaces. *Journal of the Acoustical Society of America*, **116**, 1607-1613. <http://dx.doi.org/10.1121/1.1775276>