ARCHIVES OF ACOUSTICS
Vol. **41**, No. 2, pp. 255–264 (2016)

DOI: 10.1515/aoa-2016-0025

Speech Recognition in an Enclosure with a Long Reverberation Time

Jędrzej KOCIŃSKI, Edward OZIMEK

Institute of Acoustics, Faculty of Physics, Adam Mickiewicz University
Umultowska 85, 61-614 Poznań, Poland; e-mail: jedrzej.kocinski@amu.edu.pl

(*received September 30, 2015; accepted February 15, 2016*)

The aim of this work was to measure subjective speech intelligibility in an enclosure with a long reverberation time and comparison of these results with objective parameters. Impulse Responses (IRs) were first determined with a dummy head in different measurement points of the enclosure. The following objective parameters were calculated with Dirac 4.1 software: Reverberation Time (RT), Early Decay Time (EDT), weighted Clarity (C_{50}) and Speech Transmission Index (STI). For the chosen measurement points, a convolution of the IRs with the Polish Sentence Test (PST) and logatome tests was made. PST was presented at a background of a babble noise and speech reception threshold – SRT (i.e. SNR yielding 50% speech intelligibility) for those points were evaluated. A relationship of the sentence and logatome recognition vs. STI was determined. It was found that the final SRT data are well correlated with speech transmission index (STI), and can be expressed by a psychometric function. The difference between SRT determined in condition without reverberation and in reverberation conditions appeared to be a good measure of the effect of reverberation on speech intelligibility in a room. In addition, speech intelligibility, with and without use of the sound amplification system installed in the enclosure, was compared.

Keywords: speech intelligibility; speech recognition; sentence test; reverberation time; clarity; speech transmission index.

1. Introduction

Speech recognition (e.g. in terms of intelligibility) performance depends on many objective conditions such as acoustic parameters of an enclosure, signal-to-noise ratio (SNR), the spectro-temporal properties of the interfering noise, etc. Relationship between speech intelligibility and the acoustic parameters of an enclosure has been studied by many authors (Astolfi *et al.*, 2012; Brachmański, 2004; 2008; Bradley, 1986a; Bradley *et al.*, 2003; Houtgast, Steeneken, 1985; Houtgast *et al.*, 1980; Jacob *et al.*, 1991; Kang, 1998; Peng *et al.*, 2011; 2015; Steeneken, Houtgast, 1980; Yang, Bradley, 2009), however, it is still a challenging topic. Intelligibility in a room mainly depends on the reverberant conditions, which in turn depend on the listener's position in the room (Yang, Bradley, 2009). The reverberation effects speech intelligibility because of masking phenomena in which the reflected sounds which come later to the listener, mask the direct speech signal (Bradley *et al.*, 2003). Some studies demonstrated improvements in speech intelligibility due to early reflections (Bradley, 1986a; 1986b; 1998; Brandewie, Zahorik, 2010; Harvie-Clark *et al.*, 2014). Early reflection energy in real rooms is equivalent to increasing the level of the direct sound by up to 9 dB (Bradley *et al.*, 2003; Yang, Bradley, 2009). Later-arriving speech sounds are usually found to be detrimental to the intelligibility of speech (Bradley, 1998). Some measures of early reflections are used such as EDT obtained from the first 10 dB of decay (Standardization, 1998) and the energy index – Clarity (C_{50}) – as the ratio of early (from 0 to 50 ms) to late energy (over 50 ms) (Bradley, 1983; 1990; Marshall, 1994). A relative strength of the early reflection is expressed by various measures and their interdependence. These include the rise time, early decay times, various ratios of early- and late-arriving sounds (Bradley, 1983), and the Speech Transmission Index (STI) (Houtgast *et al.*, 1980; Steeneken, Houtgast, 1980). This index is a general measure of speech transmission quality and is often used to evaluate the influence of reverberation on speech intelligibility. Such influence results from the fact that reverberation decreases the envelope fluctuations in speech. A decrease in STI causes a reduction in sentence intelligibility (Houtgast, Steeneken, 1984).

Evaluation of speech intelligibility is mainly based on one-syllable words rhyme tests, words, logatomes or simple sentence tests (Brachmański, 2008; Hagerman, 1982; Kalikow *et al.*, 1977; Kollmeier, Wesselkamp, 1997a; Nilsson *et al.*, 1994; Ozimek *et al.*, 2009a; 2006; Peng *et al.*, 2011; 2015; Plomp, Mimpen, 1979a; Versfeld *et al.*, 2000). Word rhyme tests, e.g. (Bradley, 1986b, 1990; Bradley *et al.*, 2003; Prodi *et al.*, 2010) or logatomes (Brachmański, 2004; 2008; van Wijngaarden, Drullman, 2008) well correspond to the STI values and thus, can be used in real acoustical conditions (Houtgast, Steeneken, 1985). Those tests give the results which are monotonic functions of STI in a whole range of its values.

Interesting suggestion is taking into account a measure of intelligibility expressed as Speech Reception Threshold (SRT), defined as the SNR corresponding to 50% speech intelligibility. This measure is more phonemically representative for a given language and was proved to give more accurate speech intelligibility data than standard word tests. This is due to a relatively large slope of intelligibility functions at SRT point, i.e. S_{50}. It was shown that the larger S_{50}, the smaller spread of data (standard deviation) at SRT, i.e. the more accurate speech intelligibility measurement is possible (Kollmeier, Wesselkamp, 1997b; Nilsson *et al.*, 1994; Ozimek *et al.*, 2009b; 2010; Plomp, Mimpen, 1979b; Wagener, 2003). The adaptive procedure with the 1-up/1-down decision rule (Levitt, 1971) can be used to determine SRT. To adjust adaptive procedure to the reverberant conditions two ways can be chosen, namely speech intelligibility tests can be carried out both *in situ* or in the laboratory. The former means that there is a need to gather some subjects in the room and present them the tests. This method has some disadvantages, especially related to logistics and time consumption. The later solution seems to be more convenient since it can be carried out any time in the laboratory just by recording the signals in chosen places of the enclosure via a dummy head and presenting the recordings in the laboratory. The most flexible way, however, is available by recordings of IRs via a dummy head instead. The recorded IRs can be used to a so-called auralization, which is in fact their convolution with the test material. Then the listening session can be also carried out in the laboratory (Arai *et al.*, 2002; Brandewie, Zahorik, 2010; Culling, Lavandier, 2009; Jørgensen *et al.*, 1991; Longworth-Reed *et al.*, 2008; Peng, 2007; 2008; Peng *et al.*, 2011; Yang, 2006). It is worth noting that the SRT was recently used by George *et al.* (2010) in measurements of the effects of reverberation and noise on sentence intelligibility for hearing-impaired subjects.

The main purpose of the current study is to assess the speech intelligibility in normal-hearing subjects by measuring the SRT in the enclosure with a long reverberation time. The enclosure was the church characterized by place-dependant acoustic parameters. Two different sound sources were used in the study, namely an omnidirectional loudspeaker placed at the altar and the sound amplification system installed in the enclosure. A relationship of the sentence and logatome recognition vs. STI was determined. Speech intelligibility, with and without use of the sound amplification system installed in the enclosure, was compared. The experimental data showed that the reverberant listening environment was well reflected in the SRT data, which were correlated with speech transmission index (STI).

2. Method

2.1. Experimental set-up

A PC with B&K Dirac 4.1 software was used to record and collect impulse responses (IRs) of the enclosure. The software also allows calculation of the following objective parameters of the enclosure: RT, EDT, C_{50} and STI. To extract an IR, a Maximum Length Sequence (MLS) (Borish, Angell, 1983; Chu, 1990; Kuttruff, 2009) technique was used as a driving signal instead of an impulse burst. Two different types of receivers were used: an omnidirectional microphone (Svantek SV01A) and a dummy head (Neumann KU100). The former was used to get the objec-

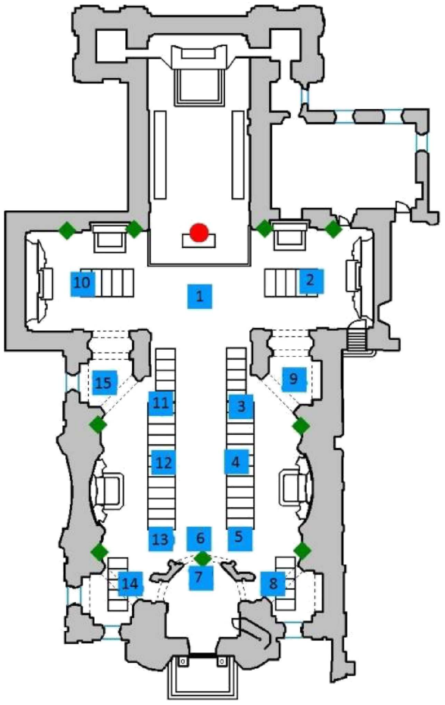

Fig. 1. A sketch of the tested church with the omnidirectional source (circle), sound amplification system loudspeakers (diamonds) and 15 measurement points (squares).

tive parameters, while the latter was used to collect the IRs via a head (with a HRTF) and subsequently convolved with the PST in the laboratory to measure the speech intelligibility. The MLS signal was generated by the software and fed via a D/A converter (ESI U2A) to the amplifier and then to the omnidirectional source placed at the altar or loudspeakers placed at the walls. All the recordings were carried out in 15 different places of the church to map the acoustical properties of the building.

Moreover, two different sound sources were used in the study: an omnidirectional loudspeaker placed at the altar and the sound amplification system installed in the church. A comparison of the speech intelligibility measurement for two sources might give an insight into the speech intelligibility improvement by the sound amplification system installed in the large volume enclosure.

Comparison of the acoustic parameter values in different places of the enclosure showed a high consistence of the symmetrical measurement points, namely 2 and 10, 3 and 1, 4 and 12, 5 and 13, 8 and 14, 9 and 15. Thus, in further analysis only data for one side of the enclosure (for 9 measurement points) were taken into account.

2.2. Recognition test and listening sessions

The Polish Sentence Tests (PST) presented against a masking noise (babble) was used in the present study (for details see (Ozimek *et al.*, 2006)). The so-called babble noise, made from the mixture of all sentences used in the test, was taken as a masker (for details see (Ozimek *et al.*, 2009b)). The power spectrum of the babble noise optimally matched the power spectra of the sentences. The precise spectral matching of masked speech and masker signal has been shown to be very important in getting a large steepness for the intelligibility function, i.e. for accurate SRT measurement. Thus, any statistically significant change in the SRT may be regarded as a measure of an effect of an external parameter (in our case a reverberation effect). The PST was composed of 25 lists each containing 10 sentences. The lists have been phonemically and statistically balanced. It was found that in anechoic conditions, the mean SRT (i.e. SNR yielding 50% speech intelligibility) was equal to −6.1 dB. This value was treated as one obtained in anechoic condition and was used in further study as reference value. Due to a relatively steep slope of the psychometric functions, the sentence test was shown to be accurate materials for speech intelligibility measurements. Additionally, the Polish Logatome Test (PLT) was also used (Brachmanski, Staroniewicz, 1999). Logatomes (non-sense words) are usually used to assess the distortions made by the path the signal has to go pass through (electrical, acoustical, etc.). These tests are based on the assumption that all the phonemes of a logatome should be heard out correctly to repeat the logatome. Thus, this kind of test is very robust, however, does not reflect a real communication process as sentences do.

The so-called auralization was used, i.e. the IRs recorded via dummy head in all the measurement points were convolved with the intelligibility test in the computer and presented to the subjects via Tucker-Davis Technology (TDT) RP2 (D/A converter) and Sennheiser HD580 headphones. The listening sessions were controlled using Matlab 6.5 software. The SNR was modified adaptively taking into consideration the most recent response of a subject. If the response was correct, the next sentence was presented at lower SNR. Conversely, if the response was incorrect, the SNR of the next utterance was increased. During the measurement, the SNR converged to the 50%-equilibrium point on the intelligibility function. SRT was computed as a mean of the adaptively changed SNR -values (excluding several initial values (Ozimek *et al.*, 2009b; Plomp, Mimpen, 1979b) or derived by fitting the model function to scores calculated for SNRs from the adaptive measurement (including also initial values) (Versfeld *et al.*, 2000).

Twenty normal hearing subjects took part in the listening sessions. Their age ranged from 23 to 28 years. They reported no problems with hearing or with speech reception. They had pure-tone hearing thresholds better than 10 dB HL at octave frequencies between 0.25 and 4.0 kHz. All of them listened to the test convolved with the IRs recorded in each measurement point for which both the omnidirectional source and sound amplification system were used. A particular list of the test was listened to only once to avoid the learning effect. Short training sessions were carried out to acquaint listeners with the task. The subjects were asked to write down the presented sentence. The subject was sited in a double-walled acoustically insulated booth. The so-called binary scoring was used in the assessment of speech intelligibility, namely only a correctly written logatome/sentence was counted as correctly understood and any mistake (except spelling mistakes) led to an incorrect note. The total level of the target signal presentation in the particular point was equal to the level measured in the enclosure during recordings, thus all the *in situ* conditions were preserved.

3. Results

3.1. Objective parameters

3.1.1. RT/EDT

Standard RTs/EDTs in six octave bands, computed with Dirac 4.1 software and their mean values

Table 1. RT/EDT values for six octave-bands and for nine measurement points.

RT/EDT [s]						
Measurement point	Octave-band frequency [Hz]					
	125	250	500	1000	2000	4000
1	5.1/3.8	4.4/3.6	4.6/3.8	3.9/3.2	2.7/2.5	2.0/1.9
2	4.3/4.6	5.1/3.6	4.4/3.5	3.5/3.0	2.5/2.5	1.6/1.7
3	5.1/6.0	4.9/4.9	4.8/4.4	4.0/4.0	3.0/3.0	2.1/2.2
4	4.5/5.7	4.8/5.3	4.9/4.9	4.1/4.4	3.1/3.3	2.4/2.2
5	4.7/5.7	5.6/5.7	4.9/5.2	4.0/4.7	3.2/3.5	2.5/2.6
6	4.5/4.8	5.0/5.2	4.9/5.1	4.2/4.5	3.2/3.5	2.2/2.6
7	4.9/5.6	4.9/5.9	4.8/5. 0	4.7/4.7	3.4/3.6	2.5/2.7
8	5.2/5.2	4.9/5.3	4.5/5.4	4.4/4.6	3.3/3.6	2.5/2.7
9	4.7/5.2	4.6/5.1	4.4/4.8	4.3/4.2	3.2/3.3	2.4/2.3

for nine measurement points are given in Table 1, and the mean RT/EDT values are presented in Fig. 2.

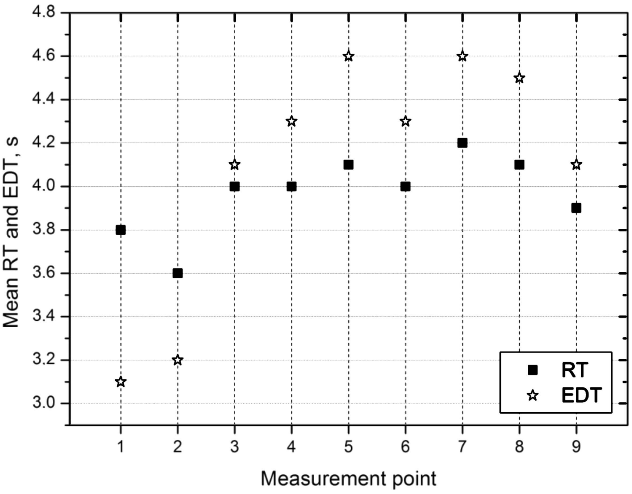

Fig. 2. Mean RTs and EDTs for nine measurement points.

Analysis of the RTs in different parts of the enclosure shows that the RT varies from 3.6 s (point 2) to 4.2 s (point 7). The general mean of RT is about 4.0 s. The EDT values are slightly longer than RT ones (except point 1 and 2). This suggests that the influence of reverberation on speech intelligibility might be also somewhat stronger, since early parts of reflection have a greater influence than it was suggested by RT, the intelligibility might be slightly higher.

3.1.2. Weighted Clarity (C_{50})

Since the speech intelligibility in an enclosure with a long RT is mainly related to the early reflection part of the sound energy (early part of an IR), the C_{50} parameter was also calculated. Figure 3 shows C_{50} values *versus* measurement points for omnidirectional source and sound amplification system. The prediction of speech intelligibility according to Marshall's rating (1994) is given in dashed lines. As this is energy ratio in time, the placement of the sound source as well as the distance between source and measurement point are crucial.

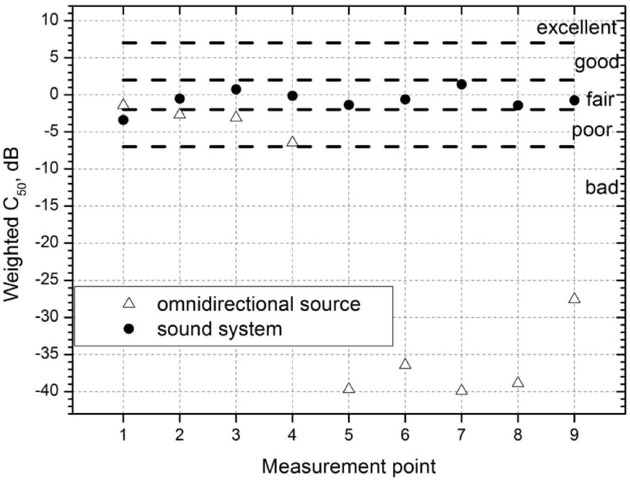

Fig. 3. Weighted C_{50} values for nine measurement points, for omnidirectional source and sound amplification system.

C_{50} data suggest that the use of the sound amplification system increases speech intelligibility especially in places located at the end of the church where the influence of early energy is minimal. In such a situation an increase in C_{50} caused by the sound amplification system leads to fair speech intelligibility (raise by two categories). Also in point 9, where there was almost no direct sound because of the columns of the arch between the target source and the measurement point, the speech intelligibility is bad for omnidirectional source and fair for the sound amplification system. In other measurement points the C_{50} values also suggest a speech intelligibility increment, however only by one category (from poor to fair).

3.1.3. Speech Transmission Index, STI

STI values calculated for both omnidirectional source and sound amplification system and for different measurement points are depicted in Fig. 4.

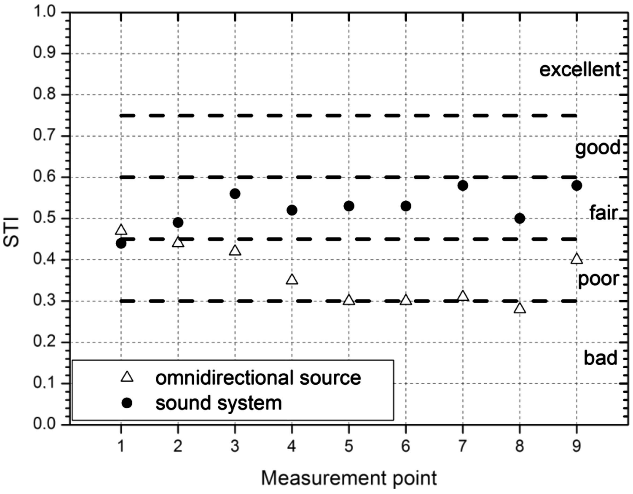

Fig. 4. Speech intelligibility prediction based on the STI values for omnidirectional source and sound amplification system for nine measurement points.

STI data are, generally, in agreement with C_{50} prediction. However, some differences can be found for points 5–8 where poor instead of bad speech intelligibility was predicted for the omnidirectional source. The use of the sound amplification system increases the speech intelligibility prediction to fair.

3.2. Speech recognition estimation

3.2.1. Logatome recognition

Figure 5 depicts the mean logatome recognition for omnidirectional source and for sound amplification system. Additionally, results from anechoic chamber as a reference value are depicted with an asterisks. As can be seen, logatome recognition is generally lower for omnidirectional source that for sound amplification system. Thus, the general statement that the sound amplification system makes the speech intelligibility much higher is confirmed. Since two-way-ANOVA has proven that both the way of presentation (omnidirectional source and sound system) and measurement point are statistically significant [$F = 407$, $p < 0.001$] and [$F = 3$, $p = 0.007$], respectively, the results were divided into two groups according to way of presentation. In both groups again the ANOVA was made to investigate whether the measurement points are statistically significant. For omnidirectional source the measurement point was proven to be statistically significant [$F = 7$, $p < 0.001$], however for sound system, the statistical significance of measurement point is on the border of significance [$F = 2$, $p = 0.07$], thus one can state that the sound system equalizes the conditions in the enclosure. Nevertheless, the results of ANOVA suggest that measurement points should be analyzed separately without any averaging.

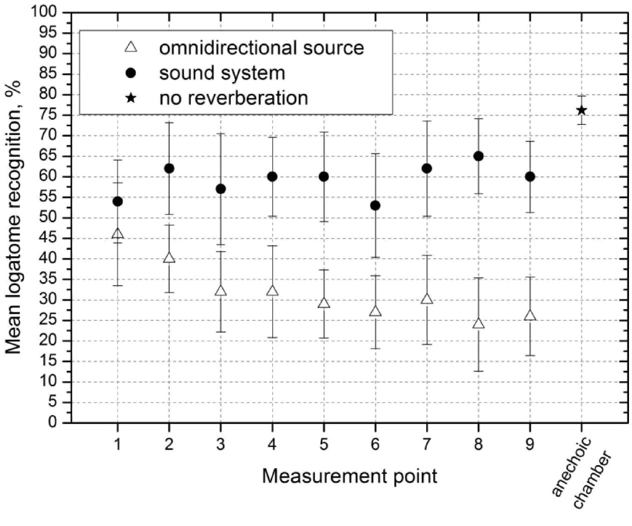

Fig. 5. Mean logatome recognition averaged across twenty subjects *versus* nine measurement points.

3.2.2. Sentence intelligibility based on SRT

A new approach to measure speech intelligibility in a room, consisted in measurements speech reception threshold (SRT) based on the Polish Sentence Test (PST) was undertaken.

First, the PST was recorded in an anechoic chamber with the dummy head placed in front of the signal source and the reference SRT values were measured. A standard 1-up/1-down adaptive procedure (Levitt, 1971) was used to determine SRT values. In this procedure, SNR was varied adaptively with respect to the most recent subject's response. The SNR was either increased or decreased by some value (step) when the most recent response was incorrect (1-up) or correct (1-down), respectively. SRT was determined as the mean of the last 8 (from 13) nominal SNRs. The mean SRT (across 20 subjects) obtained in the anechoic chamber was equal to −6.5 dB and is shown in Fig. 7 by the asterisk. In the next stage the PST was subjected to convolution with the IRs recorded via dummy head in the tested enclosure. Subsequently, sentence intelligibility measurement for this condition (with reverberation) was performed. Figure 6 depicts the mean SRTs obtained in reverberation condition, for omnidirectional source (open triangles) and sound amplification system (black circles). A high intelligibility increase (represented by a decrease of SRT values) can be observed when the sound amplification system was used. It can be also seen that the reverberation causes higher SD which suggests that the ability to understand the sentences in the reverberant conditions is more subject dependant. Moreover, the SRT values for the measurement points closer to the source are lower than those

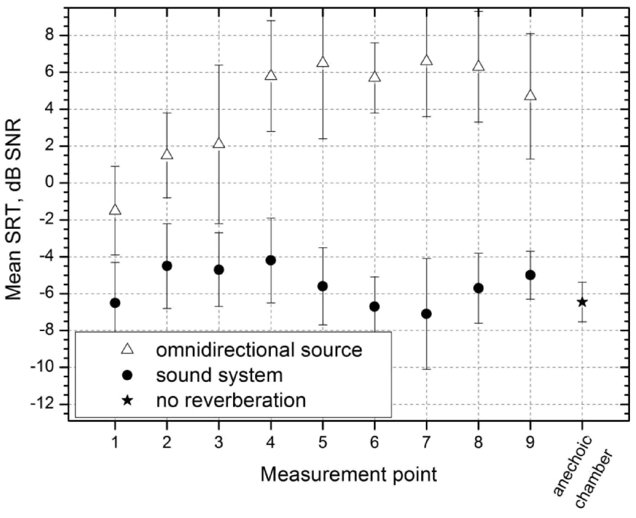

Fig. 6. Mean SRTs obtained, averaged across twenty listeners *versus* nine measurement points.

from the back of the church. It seems obvious since for the measurement points that are closer to the source, the early energy is higher and helps listener in speech intelligibility while for the most distant point, the late energy is higher causing the deterioration in speech intelligibility (increase in SRTs). This findings are in line with the STI results (see Fig. 4 for details).

Regarding logatome recognition, the same statistical analysis was made. Two-way-ANOVA has proven that way of presentation is statistically significant [$F = 618$, $p < 0.001$] as well as measurement point [$F = 5.5$, $p < 0.001$], thus the results were divided into two groups and another ANOVA was made. For both omnidirectional source and sound system, measurement point was proven to be statistically significant [$F = 13$, $p < 0.001$] and [$F = 3.5$, $p < 0.001$], respectively.

4. Discussion

First we will discuss the influence of late energy on the logatome intelligibility and SRT. As can be seen from Fig. 3, for omnidirectional source weighted C_{50} drastically decreases for points 5–9, which are far from the source placed at the altar. The use of sound system makes the sound source much closer, thus the early energy is higher. As a consequence, the C_{50} value is increased and the speech intelligibility should be higher. Comparing Fig. 3 and Fig. 5 it cannot be seen, however, that logatome recognition decreases for distant points (5–9) as it was predicted by C_{50} values. It may be caused by a low logatome intelligibility even for point which are close to the source. In such a situation the logatome test seems not to be sensitive enough. Moreover, comparing Fig. 3 with Fig. 6 (STI values for different points) it may be stated that for distant points the SRT values are higher (leading to lower speech intelligibility) than for close points, which suggests that the speech intelligibility for those points differ from each other. This results suggest that SRT is more sensitive for changes in early/late energy ratios. Nonetheless, when the signal was coming from sound system both measures of speech intelligibility show improvement which is in line with C_{50} measure.

With regard to STI, as shown by Houtgast and Steeneken (2002) the results of logatome recognition are well correlated with STI measure. Those results were also confirmed by Brachmański (2004; 2008). It can be assumed, according to the results showed by Houtgast and Steeneken (2002) and Brachmański (2004; 2008) that for obtained narrow range of STI values linear function is enough to model the relationship between logatome intelligibility and STI (see Fig. 7). The model was as follows:

$$\mathrm{LI}(\mathrm{STI}) = A * \mathrm{STI} + B, \tag{1}$$

where LI – logatome intelligibility, A and B, are the parameters to be estimated.

For clarity both groups in the Fig. 7 are depicted using different symbols, however they were analyzed together.

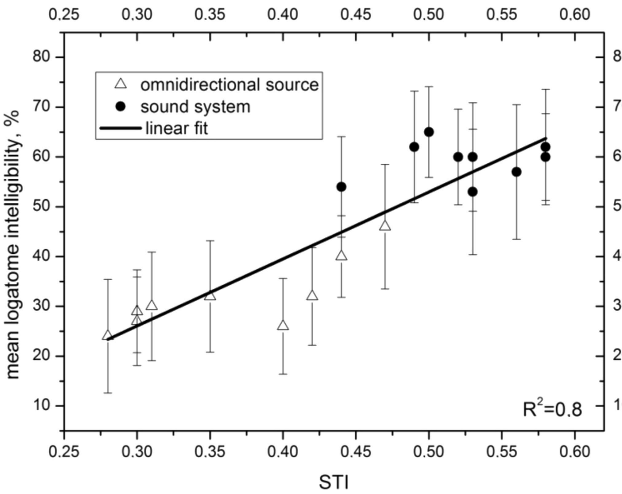

Fig. 7. Linear fit to the logatome recognition vs. STI function for both sources.

The fitting coefficient $R^2 = 0.8$ is high, and the parameters are as follows: $A = 134.5$; $B = -14.3$ in this range of STI values. It must be emphasized that for wide range of STI values the relationship is non-linear (Brachmański, 2004; 2008; Houtgast, Steeneken, 2002).

Nevertheless, the obtained results suggest smaller slope than that suggested by Houtgast and Steeneken (2002). This slight difference may be caused by the Polish language used here, which represents, in opposite to English, the group of languages based on fricatives which are more vulnerable by distortions. The results by Brachmański (2004;

2008), who also used Polish logatome test, confirm the data gathered here and have just slightly higher slope, but significantly lower than as suggested by Houtgast and Steeneken. Moreover, Houtgast and Steeneken suggest that for good conditions (high STI) logatome intelligibility aims to almost 100%, however it was not confirmed here: event in an anechoic chamber, the logatome intelligibility reaches just about 80%. The same results were obtained by Brachmański.

The ANOVA results for measurement points can be also analyzed in terms of STI (which characterizes each measurement point), thus it can be stated that there is no difference between intelligibility scores for different measurements point when sound system was used: [$F = 7$, $p = 0.07$]. This might be a result of narrow range of intelligibility scores and STI obtained for sound system. Nonetheless, for omnidirectional case the statistical significance of STI was noticed [$F = 7$, $p < 0.001$]. Thus it may be stated that the results of ANOVA have proven that the use of sound system equalizes the intelligibility among all points of the enclosure.

It is also worth mentioning that the logatome test does not reflect a real communication process, thus it is still not optimal solution for speech intelligibility testing. Sentence test and SRT (i.e. SNR yielding 50% speech intelligibility) seem to be more suitable measure here as they reflect the effect of distortions on real sentences, and, what is more, SRT in more sensitive for any change in conditions than classical speech intelligibility measured in percents. Therefore, the same analyzing procedure as for logatome was applied to SRT measure. According to previous finding of SRT (Ozimek *et al.*, 2009a; 2006) in such a range of SRT changes, a typical psychometric relation modeled by the logistic function can applied here to describe the SRT vs. STI relationship for both omnidirectional source and sound system at once (Fig. 8). Again for clarity, two different ways of presentation are depicted using different symbols, however to all points one curve was fitted. This function can be expressed by Eq. (2):

$$\mathrm{SRT(STI)} = A_2 + \frac{A_1 - A_2}{1 + \left(\frac{\mathrm{STI}}{x_o}\right)^p}, \tag{2}$$

where A_1, A_2, x_o and p are parameters to be estimated.

The fitting coefficient $R^2 = 0.98$ is very high which suggests that the relationship is of a psychometric type in the analyzed STI range. The parameters of the curve are as follows: $A_1 = 6.32$, $A_2 = -5.79$, $x_o = 0.44$, $p = 16.68$.

As can be seen from Fig. 8, sentence speech intelligibility determined by SRT is nonlinear function of STI. The most sensitive range of SRT relative to STI (the steepest slope of the psychometric function covers the range of STI changes from 0.38 to 0.48 (about 8 dB SRT/0.1 STI). Below and above this range, the SRT is less and less sensitive *versus* STI changes and much outside this range is practically independent of STI and reaches its minimum at about 0.55. This is because for this STI value SRT reaches the lowest possible value which is equal to the one obtained for anechoic conditions. This statement is the new and main finding resulting from the present study. Again it must be stated that only reverberant conditions (with no additional noise in the enclosure) were tested, thus the hearing-in-noise-test (HINT) used here gives the insight in the influence of reverberant conditions and amplification system on the intelligibility expressed in terms of SRT. Moreover, such a test is very sensitive to any condition change (like reverberation or amplification), thus gives very reliable date on the relationship between STI and SRT. However, for the STI values over 0.6 the dependency between these quantities will not be found because for anechoic conditions (which can be found as most "sterile" ones, with no convolutive distortions at all, only with additive distortion of masking noise) the SRT values reach about −6.5 dB which is obtained here for STI values of about 0.55.

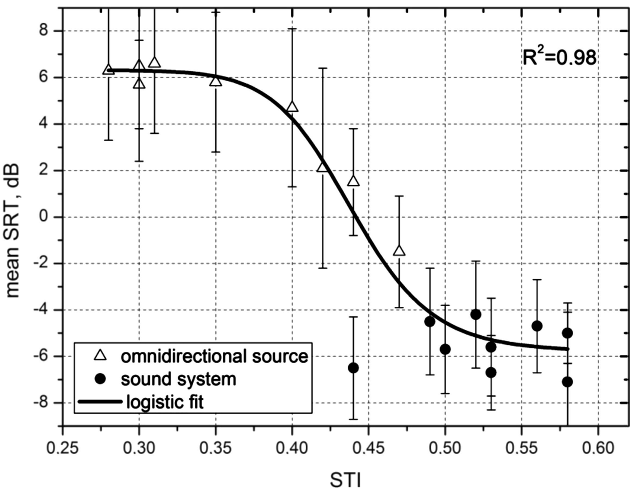

Fig. 8. SRT vs. STI and logistic curve fitting for averaged across subject SRT values.

To assess a relative change in SRT caused by reverberation, a differences ΔSRT between $\mathrm{SRT_{an}}$ obtained for non-reverberant condition and reverberant condition ($\mathrm{SRT_{rev}}$) were calculated for each measurement point. Zero value for such differences ($\Delta\mathrm{SRT} = \mathrm{SRT_{an}} - \mathrm{SRT_{rev}}$) means that speech intelligibility was not changed under reverberant condition, while positive values mean that speech intelligibility increased and negative values mean that the speech intelligibility decreased under reverberant condition (see Fig. 9). Calculation of the ΔSRT allows to estimate the effect of the sound amplification system on speech intelligibility in reverberant listening conditions. As shown in Fig. 9, for omnidirectional source a significant decrease in speech intelligibility can be noticed, while for sound

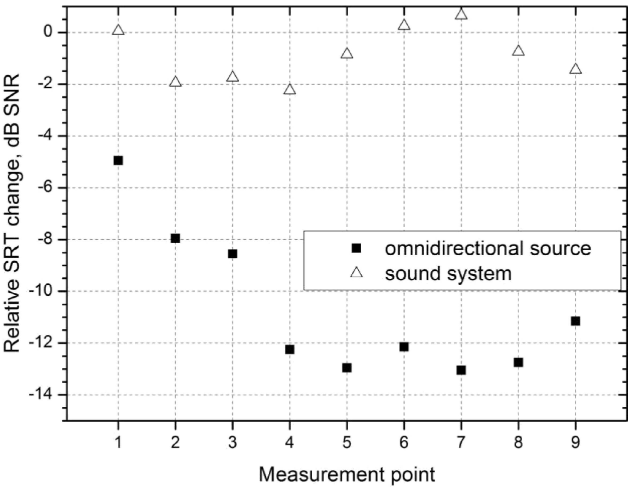

Fig. 9. Relative SRT changes for nine measurement points.

amplification system, speech intelligibility is generally less sensitive to a reverberant condition. The results of speech intelligibility obtained in the tested enclosure with a sound amplification system showed that it could be significantly improved (on average 30%). This is in line with known principle indicating that to improve speech intelligibility in a room with long reverberation time a well-designed sound amplification system should be used. It was shown in the current study that the SRT is a reasonable good indicator which well quantifies sentences speech intelligibility in reverberant conditions. The differences in SRT between measurement conditions (non-reverberant and reverberant) were statistically significant and were dependent on the location of the measurement point in an enclosure.

The significant decrease in speech intelligibility in reverberant conditions relative to intelligibility in non-reverberant condition (in noise only) indicates that listening to speech in reverberant environment is more difficult and requires higher cognitive abilities than listening to speech in noise which is mainly governed by the SNR and auditory profile. Thus, in the future research, the SRT may be applied to investigate the importance of cognitive and temporal processing in speech performance in reverberant listening conditions.

5. Conclusions

This study allows to draw the following conclusions:

- Presentation of the sentence test against babble noise in a room is more reliable method of speech intelligibility measurement than the logatome test, especially for high reverberation conditions.
- The SRT method used here seems to be more sensitive to changes in acoustic conditions of the room with a long reverberation time, especially for early/late energy ratio changes.
- For a room with a long reverberation time and rich architecture, the logatome vs. STI relationship can be modeled by linear function, but the sentence speech intelligibility expressed in terms of SRT vs. STI should be modeled using psychometric function.
- The most sensitive range of the SRT relative to STI changes corresponds to the middle range of STI values (around 0.35–0.5). Below and above this range, sentence intelligibility expressed in terms of SRT is much less dependent of STI changes (and significantly beyond this range is practically independent of STI changes).
- Difference ΔSRT between $\mathrm{SRT}_{\mathrm{anechoic}}$ and $\mathrm{SRT}_{\mathrm{reververant}}$ seems to be a good measure of the room reverberation effect on speech intelligibility.

Acknowledgments

Supported by a grant from Norway through the Norwegian Financial Mechanism (contract no. 7/2009) and a grant from National Science Centre: Project Number UMO-2011/03/B/HS6/03709. The authors thank M. Sołoducha and M. Benk for technical assistance.

References

1. Arai T., Kinoshita K., Hodoshima N., Kusumoto A., Kitamura T. (2002), *Effects of suppressing steady-state portions of speech on intelligibility in reverberant environments*, Acoustical Science and Technology, **23**, 4, 229–232.
2. Astolfi A., Bottalico P., Barbato G. (2012), *Subjective and objective speech intelligibility investigations in primary school classroom*, Journal of the Acoustical Society of America, **131**, 1, 247–258.
3. Borish J., Angell J.B. (1983), *An Efficient Algorithm for Measuring the Impulse Response Using Pseudorandom Noise*, Journal of the Audio Engineering Society, **31**, 7/8, 478–488.
4. Brachmanski S., Staroniewicz P. (1999), *Phonetic structure of a test material used in subjective measurements of speech quality* [in Polish], [Eds.], Speech and Language Technology, Poznan, pp. 71–80.
5. Brachmański S. (2004), *Estimation of logatom intelligibility with the STI method for polish speech transmitted via communication channels*, Archives of Acoustics, **29**, 4, 555–562.
6. Brachmański S. (2008), *Automation of subjective measurements of speech intelligibility in analogue telecommunication channels*, Archives of Acoustics, **33**, 3, 341–350.

7. Brachmański S. (2008), *Objective measure for assessment of speech quality in rooms*, Archives of Acoustics, **33**, 4 (Supplement), 177–182.
8. Bradley J.S. (1983), *Experience with new auditorium acoustic measurements*, Journal of Acoustical Society of America, **73**, 6, 2051–2058.
9. Bradley J.S. (1986a), *Predictors of speech intelligibility in rooms*, Journal of Acoustical Society of America, **80**, 837–845.
10. Bradley J.S. (1986b), *Speech intelligibility studies in classrooms*, Journal of Acoustical Society of America, **80**, 846–854.
11. Bradley J.S. (1990), *A comparison of three classical concert halls*, Journal of Acoustical Society of America, **89**, 3, 1176–1192.
12. Bradley J.S. (1998), *Relationships among measures of speech intelligibility in rooms*, Journal of the Audio Engineering Society, **46**, 396–405.
13. Bradley J.S., Sato H., Picard M. (2003), *On the importance of early reflections for speech in rooms*, Journal of Acoustical Society of America, **113**, 3233–3244.
14. Brandewie E., Zahorik P. (2010), Prior listening in rooms improves speech intelligibility, J. Acoust. Soc. Am., **128**, 1, 291–299.
15. Chu W.T. (1990), *Impulse-response and reverberation-decay measurements made by using a periodic pseudo-random sequence*, Applied Acoustics, **29**, 3, 193–205.
16. Culling J.F., Lavandier M. (2009), *Prediction of binaural speech intelligibility against noise in rooms*, J. Acoust. Soc. Am., **127**, 1, 387–399.
17. George E.L.J., Goverts S.T., Festen J.M., Houtgast T. (2010), *Measuring the Effects of Reverberation and Noise on Sentence Intelligibility for Hearing-Impaired Listeners*, Journal of Speech, Language, and Hearing Research, **53**, 1429–1439.
18. Hagerman B. (1982), *Sentences for testing speech intelligibility in noise*, Scand. Audiol., **11**, 79–87.
19. Harvie-Clark J., Dobinson N., Larrieu F. (2014), *Use of G and C50 for classroom design*, Forum Acousticum 2014, Krakow, Poland.
20. Houtgast T., Steeneken H.J. (1985), *A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria*, Journal of Acoustical Society of America, **77**, 1069–1077.
21. Houtgast T., Steeneken H.J. (2002), *Past, present, and future of the Speech Transmission Index*, TNO Human Factors, Soesterberg, The Netherlands.
22. Houtgast T., Steeneken H.J.M. (1984), *A multi-language evaluation of the RASTI-method for estimating speech intelligibility in auditoria*, Acustica, **54**, 185–199.
23. Houtgast T., Steeneken H.J.M., Plomp R. (1980), *Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics*, Acustica, **46**, 60–72.
24. Jacob K.D., Birkle T.K., Icker C.B. (1991), *Accurate Prediction of Speech Intelligibility without the Use of In-Room Measurements*, J. Audio Eng. Soc., **39**, 4, 232–242.
25. Jørgensen M., Ickler C.B., Jacob K.D. (1991), *Judging the Speech Intelligibility of Large Rooms via Computerized Audible Simulations*, 91st AES Convention, New York, NY.
26. Kalikow D.N., Stevens K.N., Elliot L.L. (1977), *Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability*, J. Acoust. Soc. Am., **61**, 1337–1351.
27. Kang J. (1998), *Comparison of speech intelligibility between English and Chinese*, Journal of the Acoustical Society of America, **103**, 2, 1213–1216.
28. Kollmeier B., Wesselkamp M. (1997), *Development and Evaluation of a German Sentence Test for objective and subjective Speech Intelligibility Assessment*, Journal of Acoustical Society of America, **102**, 4, 2412–2421.
29. Kuttruff M. (2009), *Room Acoustics*, 5th Edition, Spon Press, Abington, Oxon, UK.
30. Levitt H. (1971), *Transformed up-down methods in psychoacoustics*, J. Acoust. Soc. Am., **49**, 467–477.
31. Longworth-Reed L., Brandewie E., Zahorik P. (2008), *Time-forward speech intelligibility in time-reversed rooms*, JASA Express Letters, DOI: 10.1121/1.3040024,
32. Marshall L.G. (1994), *An acoustic measurement program for evaluating auditoriums based on the early/late sound energy ratio*, Journal of Acoustical Society of America, **96**, 4, 2251–2261.
33. Nilsson M., Soli S.D., Sullivan J.A. (1994), *Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise*, Journal of Acoustical Society of America, **95**, 1085–1099.
34. Ozimek E., Kutzner D., Libiszewski P., Warzybok A., Kocinski J. (2009a), *The new Polish tests for speech intelligibility measurements*, 13th IEEE SPA, Signal Processing: Algorithms, Architectures, Arrangements, and Applications, Poznań.
35. Ozimek E., Kutzner D., Sęk A., Wicher A. (2009b), *Polish sentence tests for measuring the intelligibility of speech in interfering noise*, International Journal of Audiology, **48**, 7, 433–443.
36. Ozimek E., Kutzner D., Sęk A.P., Wicher A., Szczepaniak O. (2006), *The Polish Sentence Test for Speech Intelligibility Evaluations Measurements*, Archives of Acoustics, **31**, 4, 431–438.
37. Ozimek E., Warzybok A., Kutzner D. (2010), *Polish sentence matrix test for speech intelligibility measurement in noise*, International Journal of Audiology, **49**, 6, 444–454.
38. Peng J. (2007), *Relationship between Chinese speech intelligibility and speech transmission index using diotic listening*, Speech Communication, **49**, 12, 933–936.

39. Peng J. (2008), *Relationship between Chinese speech intelligibility and speech transmission index in rooms using dichotic listening*, Chinese Science Bulletin, **53**, 18, 2748–2752.
40. Peng J., Bei C., Sun H. (2011), *Relationship between Chinese speech intelligibility and speech transmission index in rooms based on auralization*, Speech Communication, **53**, 7, 986–990.
41. Peng J., Yan N., Wang D. (2015), *Chinese speech intelligibility and its relationship with the speech transmission index for children in elementary school classrooms*, Journal of the Acoustical Society of America, **137**, 1, 85–93.
42. Plomp R., Mimpen A.M. (1979a), *Improving the reliability of testing the speech reception threshold for sentences*, Audiol., **18**, 43–53.
43. Plomp R., Mimpen A.M. (1979b), *Speech-reception threshold for sentences as a function of age and noise level*, J. Acoust. Soc. Am., **66**, 1333–1342.
44. Prodi N., Visentin C., Farnetani A. (2010), *Intelligibility, listening difficulty and listening efficiency in auralized classrooms*, J. Acoust. Soc. Am., **128**, 1, 172–181.
45. Standardization, I. O. f. (1998), *Acoustics – Measurement of the reverberation time of rooms with reference to other acoustical parameters*.
46. Steeneken H.J.M., Houtgast T. (1980), *A physical method for measuring speech-transmission quality*, Journal of the Acoustical Society of America, **69**, 318–326.
47. van Wijngaarden S.J., Drullman R. (2008), *Binaural intelligibility prediction based on the speech transmission index*, J. Acoust. Soc. Am., **123**, 6, 4514–4523.
48. Versfeld N.J., Daalder L., Festen J.M., Houtgast T. (2000), *Method for the selection of sentence material for efficient measurement of the speech reception threshold*, Journal of Acoustical Society of America, **107**, 1671–1684.
49. Wagener K. (2003), *Factors influencing sentence intelligibility in noise*.
50. Yang W. (2006), *Optimizing Acoustical Conditions for Speech Intelligibility in Classrooms*, Vancouver.
51. Yang W., Bradley J.S. (2009), *Effects of room acoustics on the intelligibility of speech in classrooms for young children*, Journal of Acoustical Society of America, **125**, 2, 922–933.