EVALUATION OF SPEECH INTELLIGIBILITY IN TWO ACOUSTICALLY DIFFERENT SPACES USING LOGATOME TEST AND MEASURED IMPULSE RESPONSES

Andrea ANDRIJAŠEVIĆ – Bojan IVANČEVIĆ – Marko HORVAT

Abstract: *In this work, two closed spaces intended for speech will be compared on the basis of speech intelligibility level using two types of methods: subjective and objective ones. The first group of methods is represented with the logatome listening test. The results obtained with it (the percentages of correctly heard logatomes) are compared with the values of intelligibility correlated standardized acoustic parameters calculated from the acoustic impulse responses of the rooms. In the end the paper presents the comparison of the methods from the standpoints of time and money consumption and the conclusion regarding the suitability level of each room for speech.*

Keywords:
– listening test
– reverberation time
– speech transmission index
– acoustic impulse response
– speech intelligibility

1. INTRODUCTION

One of many ways by which closed spaces can be classified is, naturally, by their specific purpose (speech, listening or recording music, working, etc.), where the estimation of their suitability can be done by observing the values of parameters obtained using one or more standardized methods. For instance, if the room is to be used for speech (e.g. a lecture room), then the level of speech intelligibility in it should be checked by calculating the values of parameters highly correlated with it.

In this paper, speech intelligibility in two real closed spaces will be assessed using the results of objective measurements and of a subjective method of speech intelligibility evaluation. The subjective method is represented with the listening test of the series of logatomes, while the objective method is based upon the acoustic impulse response from which the values of objective parameters (clarity C_{50}, definition D_{50}, reverberation time T_{30}, *STI* and *RASTI*) are calculated.

The paper is organized as follows; The basics of subjective and objective methods for the estimation of intelligibility are presented in Chapter 2; The setup and the procedures for both methods are described in detail in Chapter 3 while Chapter 4 shows the results of both investigations; The advantages and disadvantages of both types of methods regarding time and money consumption are given in Chapter 5 and, finally, in Chapter 6 the conclusions are made.

2. BASICS OF THE METHODS

There are two distinct groups of methods for speech intelligibility assessment: objective and subjective. They will be described briefly in the following subchapters.

2.1. Subjective method

The listening tests are the most natural method for subjective estimation of speech intelligibility in a room. The basics of the method is that the subjects listen to some specific speech material mixed with disturbances (background noise) in the room under observation, and write down what they have heard. The type of material used can be very diverse, consisting of short words, syllables (with or without meaning) or sentences.

The final results, expressed as the percentages of correctly perceived components of speech material, are highly dependent on many factors, such as: the type of the text used (words or syllables require better listening conditions than the sentences), the familiarity of the listeners with the text, psychological and physiological state of both reader and the listeners, their experience in such kind of tests, etc. Therefore, the listening tests are a method for subjective relative comparison of listening conditions in different rooms using the same test setup and procedure; they do not provide an absolute value of speech intelligibility in a room.

2.2. Objective method

On the contrary, when objective methods for the assessment of speech intelligibility in different enclosures are used, absolute values of parameters correlated with it are obtained. Objective parameters used for describing the acoustical situation in a room, such as reverberation time, clarity, definition, *STI* and *RASTI*, can be obtained from the measured acoustic impulse response $h(t)$ in the manner described in the following section.

Every spatial position in a closed space can be addressed either as an acoustic monopole source point or as a receiving point. If $\boldsymbol{r}$ is the vector of position in the Cartesian coordinate system defined as:

$$\boldsymbol{r} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \qquad (1)$$

then the locations of the monopole source point and the receiving point can be written as:

$$\boldsymbol{r}_{\mathrm{S}} = \begin{bmatrix} x_{\mathrm{S}} \\ y_{\mathrm{S}} \\ z_{\mathrm{S}} \end{bmatrix}, \boldsymbol{r}_{R} = \begin{bmatrix} x_{\mathrm{R}} \\ y_{\mathrm{R}} \\ z_{\mathrm{R}} \end{bmatrix}, \qquad (2)$$

where $\boldsymbol{r}_{\mathrm{S}}$ is the spatial position of the source of acoustic signal, and $\boldsymbol{r}_{\mathrm{R}}$ is the position of the point of its reception. The closed space causes changes of the source (original) signal in every point of its reception, hence the path between every pair of source-receiver positions can be modeled with the corresponding acoustic impulse response $h(t,\, \boldsymbol{r}_{\mathrm{S}},\, \boldsymbol{r}_{\mathrm{R}})$. The received signal $s'(t,\, \boldsymbol{r}_{\mathrm{R}})$ can then be obtained by the convolution of the impulse response and original signal $s(t,\, \boldsymbol{r}_{\mathrm{S}})$:

$$s'(t,\, \boldsymbol{r}_{\mathrm{R}}) = h(t,\, \boldsymbol{r}_{\mathrm{S}},\, \boldsymbol{r}_{\mathrm{R}}) * s(t,\, \boldsymbol{r}_{\mathrm{S}}). \qquad (3)$$

In the frequency domain, the expression changes to multiplication of Fourier transforms and the source-receiver path can be viewed as a filter:

$$S'(f,\, \boldsymbol{r}_{\mathrm{R}}) = H(f,\, \boldsymbol{r}_{\mathrm{S}},\, \boldsymbol{r}_{\mathrm{R}}) \cdot S(f,\, \boldsymbol{r}_{\mathrm{S}}) \qquad (4)$$

For the purpose of simplicity, in the following text the variables $\boldsymbol{r}_{\mathrm{S}}$ and $\boldsymbol{r}_{\mathrm{R}}$ will be omitted and a generalized notation for acoustic impulse response of a room $h(t)$ will be used.

The reverberation time T_{60} is a traditional objective parameter strongly related with speech intelligibility in a sense that a higher value of reverberation time leads to lower speech intelligibility because of the overlap-masking effect where the low energy consonants are being masked by preceding high energy vowels [1]. It is defined as the time it takes for a sound pressure level to decrease by 60 dB after a continuous sound source has been shut off. To obtain its value from the impulse response, the energy decay curve (*EDC*) must be calculated first using the Schroeder's method of backwards integration of the squared impulse response, normalized by the total energy of the response:

$$EDC(t) = 10 \cdot \log_{10} \left[\frac{\int_t^{\infty} h(\tau)^2 \mathrm{d}\tau}{\int_0^{\infty} h(\tau)^2 \mathrm{d}\tau} \right]. \qquad (5)$$

After that, the best linear approximation of the *EDC* is obtained using the linear regression method and the reverberation time T_{60} can be calculated using the following expression, where A represents the time derivative of the approximation line [2, 3]:

$$T_{60} = \frac{60}{A} \text{ s}. \qquad (6)$$

Usually the 60 dB difference cannot be obtained because of the limitations of the measuring equipment and the level of ambient noise. The approximation line is then defined for the interval (-5, -15), (-5, -25) or (-5, -35) dB, where the *EDC* is above the level of ambient noise, providing parameters T_{10}, T_{20} and T_{30}, respectively. The procedures for obtaining these parameters are regulated by the ISO3382 standard.

Clarity C_{50} is an objective parameter describing the degree of clarity of the details of the original sound in the received sound [4]. The higher its value is, the more dominant the direct sound and early reflections are. For obtaining good speech intelligibility, its value should be -2 dB or higher [3].

$$C_{50} = 10 \cdot \log_{10} \left[\frac{\int_0^{50\text{ms}} h(t)^2 \text{d}t}{\int_{50\text{ms}}^{\infty} h(t)^2 \text{d}t} \right] \text{ dB.} \qquad (7)$$

Definition D_{50} is also based upon acoustic impulse response. It shows the contribution of early (useful) reflections in total energy of the impulse response. The higher its value is, the larger the share of the useful direct sound energy and, consequently, the speech intelligibility becomes better [5].

$$D_{50} = \frac{\int_0^{50\text{ms}} h(t)^2 \text{d}t}{\int_0^{\infty} h(t)^2 \text{d}t} \cdot 100 \text{ \%.} \qquad (8)$$

The most common way of assessing the degree of speech intelligibility today is by calculating the objective parameter called speech transmission index *STI* [6]. It was developed for the cases of both direct communication and amplified speech. The parameter called *RASTI* (Room Acoustics Speech Transmission Index or Rapid Speech Transmission Index) was developed later as the simplification of STI. Both parameters are based upon the assumption that the speech information is stored in the modulations of signal envelope and the reverberation causes the reduction of relative modulation depth thus affecting (degrading) the intelligibility.

In originally proposed methods for calculating *STI* and *RASTI*, bands of noise were modulated by a set of frequencies and the reductions of envelope depth were determined. Today, both parameters are calculated from the modulation transfer function given by the following expression, where *F* represents the modulation frequency:

$$m(F) = \left| \frac{\int_0^{\infty} h(t)^2 \cdot \text{e}^{-\text{j}2\pi F t} \text{d}t}{\int_0^{\infty} h(t)^2 \text{d}t} \right|. \qquad (9)$$

The range of the possible values of parameters is (0, 1), where zero expresses the complete lack of intelligibility, whereas unity represents excellent (perfect) speech intelligibility.

3. SETUP AND THE PROCEDURE

The acoustic properties regarding speech intelligibility of two rooms at the Department of electroacoustics at the Faculty of Electrical Engineering and Computing in Zagreb were examined. The first closed space, the listening room (10.2 m x 7.2 m x 3.2 m), was acoustically treated using absorbers while the second room, i.e. a lecture room of similar volume (12 m x 7.2 m x 3.2 m), was not treated, having three bare walls, and one covered with windows.

3.1. Subjective method

The following procedure was used in both rooms:

1. The position in the room for the female speaker was defined;
2. Three radii (2, 4 and 8 m) were measured from the position of the speaker;
3. One point on each of the radii was selected and the stand with the omnidirectional condenser microphone (Behringer ECM8000) was placed on it. The diaphragm of the microphone was set at the height of the speaker's mouth;
4. For each of the three microphone positions, a series of logatomes (meaningless words consisting of two consonants and a vowel in the form of CCV or CVC) were read by the same speaker and recorded on a computer.

Three different sets of logatomes were used: two of them consisting of 42 words, and the third having 38. The total number of logatomes recorded in each room was 122. The sets used at each specific distance in each room are noted in Table 1.

Table 1. The sets of logatomes used on each specific location

Radius from the speaker	**Listening room**	**Lecture room**
2 m	1st	2nd
4 m	2nd	1st
8 m	3rd	3rd

3.2. Objective method

As shown in the previous chapter, to be able to calculate the values of the objective parameters, acoustic impulse responses must be acquired first. Therefore, the following procedure for obtaining them was conducted in both rooms:

1. The position in the room for the dodecahedron loudspeaker was defined (the same one as for the female speaker in 3.1.);
2. The same three positions in each room as in chapter 3.1. were used for the placement of the microphone;
3. For each of the three positions, the impulse response was obtained in ARTA software using

the integrated impulse response method with pink noise as the excitation signal.

4. RESULTS

4.1. Subjective method

The group of seventeen employees of the Faculty of Engineering in Rijeka volunteered to participate in the listening test. Their age span was from 26 to 32, with the average value of 29 years. The volunteers listened with their headphones on and wrote down the logatomes as they heard them using a word processor. They were all informed before the beginning of the test that they would be listening to meaningless words consisting of three letters. The silences between the logatomes in the recordings were about 2 seconds.

After comparing the listeners' notes with the logatomes, the statistics were calculated and the graphs made. The percentages of correctly heard logatomes, are shown in Figure 1 for each of the 17 listeners. The differences in the percentages between the rooms are very diverse from one listener to the next. The percentage of the correctly written down logatomes is higher in the lecture room than in the listening room for the first, the fifth and the sixth listener.

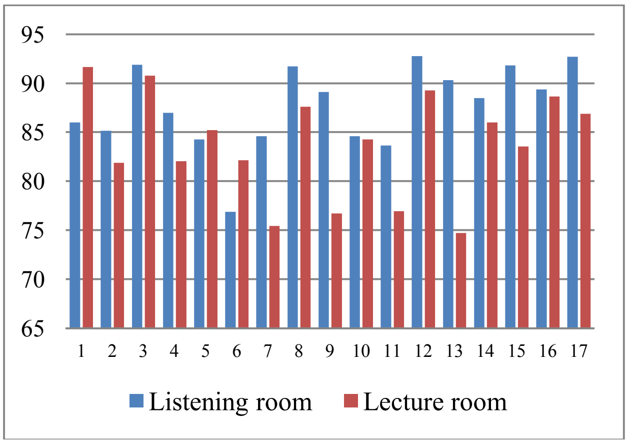
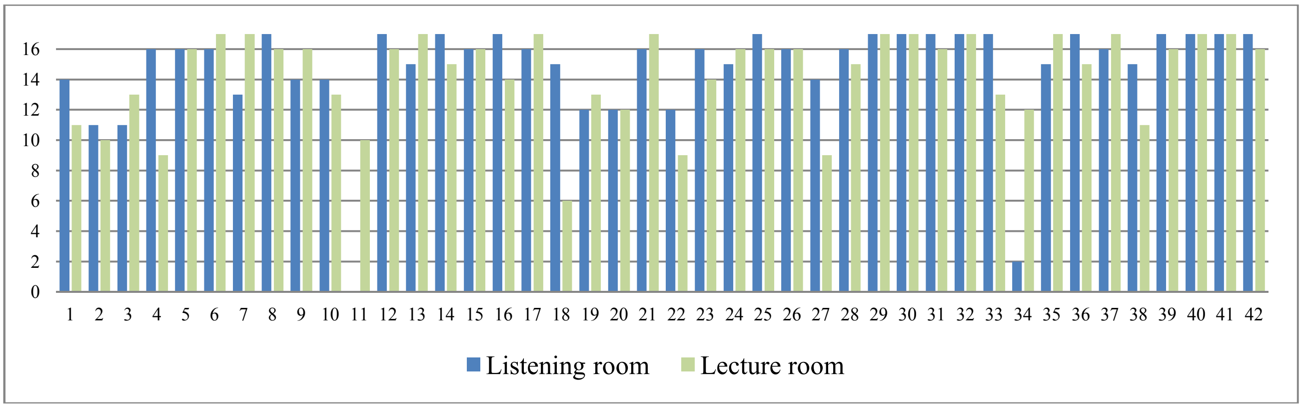
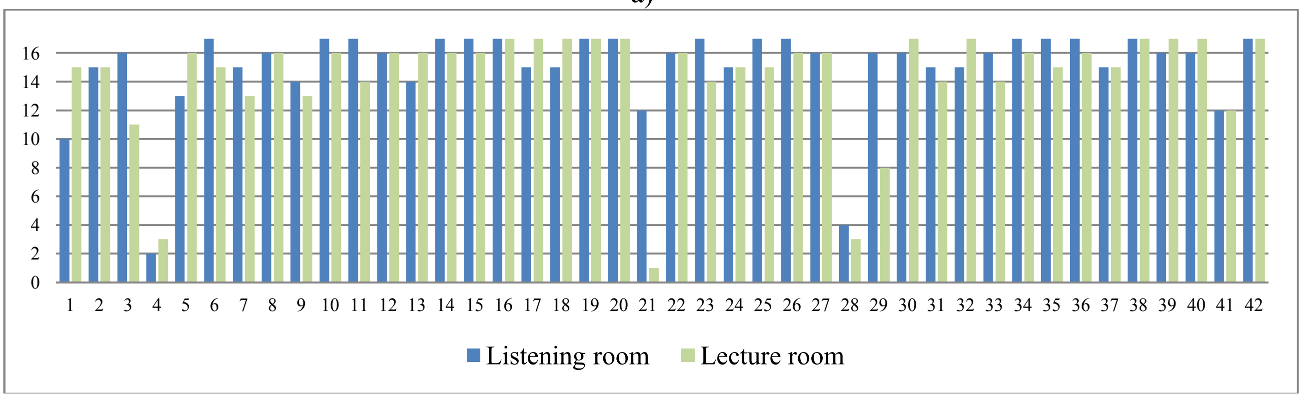


Figure 1. The percentages of correctly written down logatomes for the each listener



a)



b)

Figure 2. The number of listeners that wrote down correctly the logatomes of the: a) 1. series, b) 2. series

On the other hand, the ninth and the thirteenth listener recognized the logatomes significantly better in the listening room (89 % and 90 %, respectively) than in the lecture room (76 % and 74 %, respectively).
The number of the listeners was not a representative one, so the underlying reasons for these differences cannot be determined with certainty. Some of the reasons could be: the change in the acoustic quality or wrong speaker's articulation of a logatome, speaker's habitual way of pronunciation that is not understandable to the listeners, the degree of listeners' accuracy in typing and/or the change in the level of listeners' concentration.
It is visible in the Figure 2.a) that none of the listeners recognized the eleventh logatome recorded in the listening room while it was recognized by 10 listeners in the lecture room - it is highly probable that this is an example of speaker's error in pronunciation. The similar case is with the 34th logatome, while the opposite case is visible in Figure 2.b) where the 21st logatome was highly recognizable in the listening room but only one listener recognized it in the lecture room. The influence of the speaker's habitual way of articulation can be observed in the Figure 2.b) where the 4th and 28th logatomes were sparsely recognized correctly in both rooms.

Table 2. The average percentage of the correctly written down logatomes for each position and the average for the room

Radius from the speaker	**Listening room**	**Lecture room**
2 m	86.69 %	84.59 %
4 m	88.65 %	84.59 %
8 m	87.61 %	82.04 %
average	87.65 %	83.74 %

The statistics in Table 2 show the average percentage (for the group) of the correctly written down logatomes for each position, and average percentage value for the room. Only a small relative difference of 4 % between the rooms can be observed. It can be concluded that the listening room possesses slightly better acoustic properties regarding the level of speech intelligibility than the lecture room when assessing them with this specific subjective method.
Also, in order to examine the influence of acoustical properties of rooms on speech intelligibility, T-tests were made in which the results of the logatome tests made in the listening room and in the lecture room were tested against each other. The significance level was set to the common value of 0.05. The tests were made for data obtained at each of the three distances of the listener from the source, i.e. the speaker. The results of the tests show that a statistically significant difference in speech intelligibility between the two rooms clearly exists at the set significance level if the distance between the speaker and the listener is 8 meters. At the distance of 4 meters this difference barely reaches statistical significance, while at 2 meters no statistically significant difference in speech intelligibility can be observed between the rooms. These findings are consistent with the fact that speech intelligibility in a room is directly dependant on the ratio of direct and reverberant sound energy. At small distances from the source, the desired direct sound dominates over the reverberant one, thereby diminishing the influence of acoustical conditions in the room on speech intelligibility. As the distance from the source increases, the unwanted reverberant sound becomes dominant, thus degrading the speech intelligibility.

4.2. Objective method

Figures 3 and 4 show the acoustic impulse responses obtained for the 4 m distance between the loudspeaker and microphone.
By comparing them, it can be noticed that the duration of the impulse response of the lecture room (around 820 ms) is significantly longer than the one measured in the listening room (around 250 ms). Since the rooms are of similar volume, it follows from the Sabine's relation for the reverberant spaces [1] that the value of the average absorption coefficient in the listening room is higher than in the second room - it takes less time for the energy to be absorbed in the first room.
Energy time curves (*ETC*), representing the envelopes of the impulse responses, and energy decay curves are shown in the Figures 5 and 6. The axes were matched so that the curves could be more easily compared. It can be observed that the *EDC* for the listening room has a steeper slope, thus the value of reverberation time related to it must be smaller than the one related to the lecture room.
The values of objective parameters calculated from the impulse responses are presented in Table 3.

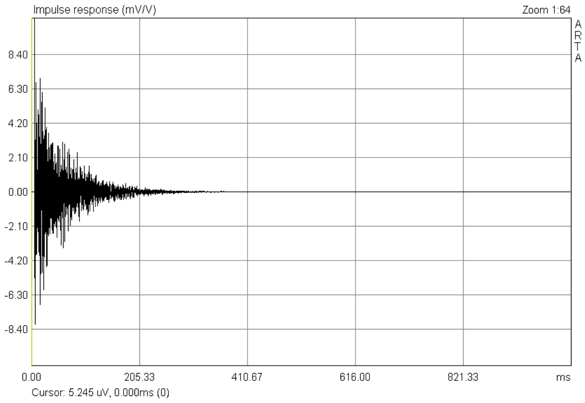


Figure 3. Impulse response of the listening room, distance: 4 m

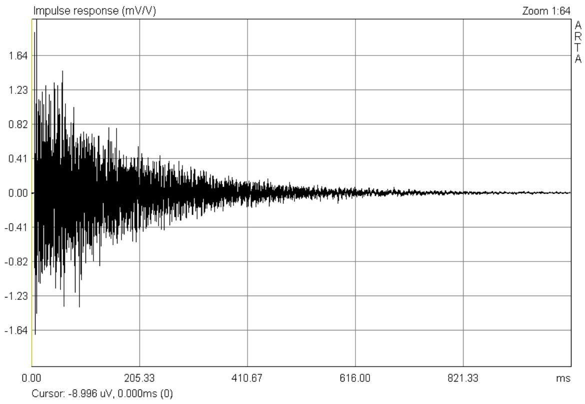


Figure 4. Impulse response of the lecture room, distance: 4 m

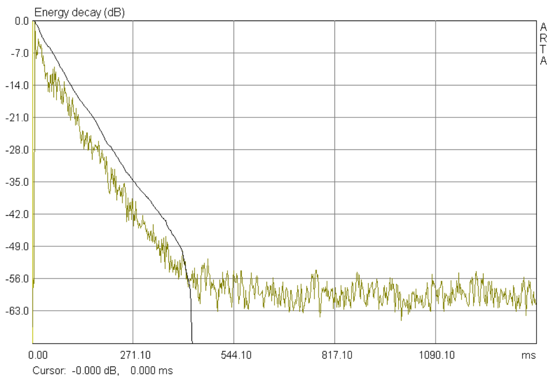


Figure 5. EDC and ETC for the listening room, distance: 4 m

The maximal values of energy decay curves in Figures 5 and 6 are more than 30 dB (50 dB and 40 dB, respectively) above the values of ambient noise, so the reverberation time parameter T_{30} can be calculated. The value of each reverberation time T_{30} in the columns "2 m", "4 m" and "8 m" is obtained

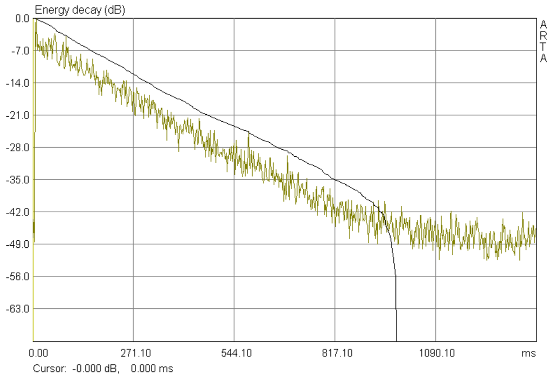


Figure 6. EDC and ETC for the lecture room, distance: 4 m

as an average from 250, 500, 1000 and 2000 Hz octave band values. The average value of T_{30} presented in the column "average" is significantly (three times) smaller for the listening room than for the lecture room. This quite a large ratio has been predicted from the durations of the impulse responses and the slopes of *EDC*s. For good speech intelligibility, it is necessary for the value of the reverberation time to be below one second [7]. This condition is satisfied for the listening room and intelligibility in it is excellent, while for the lecture room, this condition is not met and the intelligibility is fair.

The average value of parameter C_{50} for the lecture room is near the lower limit for good intelligibility (-1.3 dB) while it is much higher for the listening room (7.2 dB), confirming the abovementioned statement that the intelligibility in the lecture room is worse than in the listening room.

Above 80 % of the response energy in the listening room is the energy of direct sound and early reflections that causes the sound source to sound louder and clearer (more intelligible), while in the lecture room the value of D_{50} is only 40 %.

The values of these parameters reveal that the overall amount of reverberant energy is higher in the lecture room. As a consequence, the speech intelligibility is more degraded in the lecture room than in the listening room for the same source-microphone distances.

It can be seen from Table 3 that the values of *STI* and *RASTI* correlate well, which is normal because both parameters are obtained from the modulation transfer function of the acoustic impulse response. Their average value is 0.78 for the listening room, and 0.53 for the lecture room, confirming again the conclusion gained from values of other parameters

Table 3. Values of objective parameters

	The listening room				The lecture room			
	2 m	4 m	8 m	**average**	2 m	4 m	8 m	**Average**
T_{30} (s), 250 Hz – 2 kHz	0.465	0.469	0.484	0.473	1.367	1.382	1.408	1.386
C_{50} (dB), wideband	8.7	6.2	6.63	7.18	0.04	-1.96	-1.98	-1.3
D_{50} (%), wideband	88.1	80.6	82.1	83.6	50.2	38.9	38.8	42.6
STI – men	0.8055	0.7739	0.7769	0.7854	0.5630	0.5265	0.5274	0.5389
STI – women	0.8113	0.7750	0.7787	0.7883	0.5688	0.5283	0.5312	0.5427
Rating	Excellent	Excellent	Excellent	Excellent	Fair	Fair	Fair	Fair
RASTI	0.7969	0.7829	0.7659	0.7819	0.5586	0.5246	0.5187	0.5339

that the listening room is acoustically better suited for understanding of spoken words than the lecture room.

5. COMPARISON

In this section, the methods will be compared from the point of time and money consumption. Regarding the time consumption, the subjective method is much more demanding than the objective one - the listening material must be prepared, the representative number of listeners must be gathered to be able to obtain conclusive results, the speaker and the listeners should be trained by making the experimental tests first. Recording and listening processes are also very time-consuming, but the most consuming part of this method is the statistical processing of raw data acquired during the listening tests. From the standpoint of money consumption, the subjective method has the advantage over the objective one in which the accuracy of the results depends on the quality of the hardware and software used. Despite this major drawback, objective methods are more commonly used than the subjective ones because they enable the objective comparison of the values of parameters obtained from many standardized investigations of different enclosures.

6. CONCLUSION

In this paper, the subjective and objective methods for the estimation of speech intelligibility in two rooms were presented. When observing the results of the subjective method (listening test), regarding the human perception, no significant difference between the acoustic conditions in the listening and lecture room has been found. The average percentage of correctly written down logatomes was only 4 % higher for the first room (87.6 %) than for the second (83.7 %). The explanation for this small difference could lie in the fact that the test was made in the Croatian language whose lexical structure is the same as the phonetic one.

Using the method based on acoustic impulse response for obtaining the values of objective parameters correlated with the speech intelligibility level, larger differences in the acoustics properties were observed. Significantly better values of parameters were obtained for the listening room, arising from the higher sound absorption at its boundary surfaces, compared to the ones in the lecture room.

It can be concluded that, although the observed relative differences of intelligibility levels between the spaces differed significantly depending on the method used, both methods found the listening room to be acoustically better suited for intelligibility of the spoken language.

The other possible reason for the low level of correlation between the values obtained in subjective and objective methods could also lie in the realization of STI testing; the rooms were excited with an omnidirectional loudspeaker with pink noise excitation that excites more room reflections than a small loudspeaker with directional

characteristic similar to human head excited with noise signal with speech like spectrum that IEC 60268-16 recommends.
In the future work, the correlation between objective and logatome test methods for the Croatian language will be investigated in more detail. Also, the question whether CCV or CVC testing has higher correlation with the objective measures will be answered.

7. LIST OF SYMBOLS

acoustic impulse response	$h(t)$,	-
reverberation time	T_{60},	-
clarity	C_{50},	-
definition	D_{50},	-
energy decay curve	*EDC*	-
speech transmission index	*STI*,	-
rapid speech transmission index	*RASTI*,	-

REFERENCES

[1] Alton Everest, F.: *Master handbook of acoustics*, McGraw-Hill, 2009
[2] Kuttruff, H.: *Room acoustics*, Spoon press, 2009.
[3] ARTA version 1.6.1, *User manual*, 2010
[4] Rossing (ed.), T.D.: *Springer handbook of acoustics*, Springer, 2007
[5] Kuttruff, H.: *Acoustics – an introduction*, Taylor and Francis, 2007
[6] Larm, P., Hongisto, V.: *Experimental comparison between speech transmission index, rapid speech transmission index, and speech intelligibility index*, Journal of Acoustical Society of America, 119 (2), February 2006, p.1106.-1117.
[7] Long, M.: *Architectural acoustics*, Elsevier, 2006

Received: 21.04.2011. Accepted: 04.04.2012.

Author´s address:

Andrea Andrijašević
Faculty of engineering, University of Rijeka
Vukovarska 58, 51000 Rijeka
Bojan Ivančević
Marko Horvat
Faculty of electrical engineering and computing, University of Zagreb
Unska 3, 10000 Zagreb
andrea.andrijasevic@riteh.hr
bojan.ivancevic@fer.hr
marko.horvat@fer.hr