



Investigation on the effects of source directivity on Chinese speech intelligibility based on auralization

Jianxin Peng, Ting Wang, Shuoxian Wu

► To cite this version:

Jianxin Peng, Ting Wang, Shuoxian Wu. Investigation on the effects of source directivity on Chinese speech intelligibility based on auralization. Acoustics 2012, Apr 2012, Nantes, France. hal-00811128

HAL Id: hal-00811128

<https://hal.science/hal-00811128>

Submitted on 23 Apr 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Investigation on the effects of source directivity on Chinese speech intelligibility based on auralization

J. Peng^a, T. Wang^b and S. Wu^b

^aDepartment of Physics, School of Science, South China University of Technology, 381 Wushan Road, 510640 Guangzhou, China

^bState Key Laboratory of Subtropical Building Science, South China University of Technology, 381 Wushan Road, 510640 Guangzhou, China
arshxwu@scut.edu.cn

Subjective Chinese speech intelligibility was evaluated by using an omnidirectional source, i.e., a source with the approximate directivity of human speaker and a human speaker source in both real and simulated rectangular rooms with different reverberation time. The result shows that subjective Chinese speech intelligibility scores have statistical differences under different source conditions. Speech intelligibility scores obtained by using omnidirectional source are lower than those obtained by using the other two sources. Therefore, it will underestimate the speech intelligibility in rooms. Subject Chinese speech intelligibility obtained from auralization is basically the same as that obtained in actual rooms under different source directivity conditions. By using auralization technique one can properly evaluate the subjective Chinese speech intelligibility even with different directional sources.

1 Introduction

Room acoustical measurement is usually performed with omnidirectional sound sources according to ISO-3382[1]. However, in actual room, a real source (such as an instrument or human voice) is a directional source, which is obviously different from an omnidirectional source. Studies have indicated that the results of room acoustic parameters and subjective evaluation obtained from different directional sources show significant differences [2-8]. Dalenbäck et al. [3] investigated the subjective perception difference by changing the directivity of sound sources through auralization. They use two loudspeaker sources with different directional characteristics. Their results show that test subjects are able to recognize differences about clarity and reverberation between the two cases. Prince and Talaske [4] took measurements in a hall using both omnidirectional and directional sources. Large differences in the resulting clarity values are obtained. Giron [5] also studies the subjective effects of changing sound directivity by modeling a sound source using inverse spherical harmonics transforms. The subjects are able to differentiate the signals of different directivities. Otondo and Rindel have been further studied these differences by using instruments as directional sources [6]. The results show that source directivity has a direct effect on the distribution of objective parameters in a room, including sound pressure level (SPL), clarity index (C_{80}), lateral energy fraction (LF) and early decay time (EDT). Subjective testing also reveals that subjects can distinguish between the auralization signals made with a tone specific directivity and the averaged directivity, based on loudness and sometimes reverberance. Wang and Vigeant [7] extend the work about the effects of source directivity on room acoustic modeling prediction and auralization using three different source types with the same power: (a) an omnidirectional source; (b) sources with realistically-directional characteristics based on measurements from real instruments; and (c) an extremely directional artificial source. Their results show that directional sources induce both objective and subjective differences in room acoustical prediction and auralization.

The studies mentioned above mainly explore the effects of the directivity of instruments on the sound quality in rooms. Nevertheless, fewer investigations on the influence of source directivity on speech intelligibility are carried out. Speech intelligibility is a critical index for describing the acoustical quality in all speech - purpose rooms. The evaluation of speech intelligibility in these rooms includes objective room acoustical measurement and subjective listening test. The objective parameters related to speech intelligibility, such as speech transmission index (STI), early-late sound energy ratio (C_{50}), can be obtained through

measurement or prediction. Wang and Peng[8] compare the difference of objective parameters values in a classroom using sound field simulation by ODEON and find that source directivity has no significant effects on EDT and T_{30} , but it has significant effects on C_{50} in mid-high octave bands and on STI. Differences among subjective Chinese speech intelligibility scores under different directional source conditions need to be further investigated.

With the developing of signal processing technology, a new approach is provided to evaluate subjective speech intelligibility, which is called auralization[9, 10]. Listeners can assess the subjective acoustical quality in a given listening position when its binaural impulse responses can be obtained through acoustical simulation software. In this paper, the main works are the investigation into the effects of source directivity on subjective speech intelligibility in two rectangular rooms with different RT values through listening in actual room and simulated room with three different source types: (a) an omnidirectional source; (b) a source with the approximate directivity of human speaker and (c) a human speaker source. The goal is to explore the feasibility of investigating the subjective speech intelligibility difference due to different sources by using auralization technique.

2 Experimental method

2.1 Room model and acoustical simulation

A classroom and a multimedia lecture hall are chosen. Both rooms are rectangular and their dimensions are 15.82m×8.22m×4.90m and 10.56m×5.92m×3.25m respectively. B&K4296 dodecahedral loudspeaker is used to measure objective room acoustical indices for both rooms. During the testing, sound source is located at the platform and listening positions are distributed in seating area as showed in Figure 1. The corresponding source and listening positions were set according to measurement in real rooms in Odeon models[11]. First, an omnidirectional source is used and the acoustical properties of some surfaces in both models are adjusted to make the discrepancy between simulated objective acoustical parameters and measured ones for different listening positions is within 1 JND range [12]. Table 1 show the average values of EDT, T_{30} , C_{50} and STI in the two rooms at 500Hz octave band when omnidirectional source is applied. Then, a source with the approximate directivity of human speaker (JBL-LSR6325P loudspeaker) and a human voice source defined by BB93[13] are used and other sets remain the same in Odeon simulation. BB93 human voice source substituted for human talker source in real rooms during the simulation.

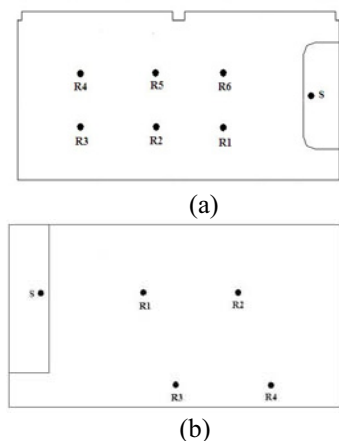


Figure 1 Sound source and listening positions in (a) University classroom and (b) Multimedia lecture hall

Table 1 Comparison of objective acoustical parameters from measurement and simulation

Acoustical parameters	University classroom		Multimedia lecture hall	
	measurement	simulation	measurement	simulation
EDT(500Hz) /s	0.49	0.51	2.17	2.19
$T_{30}(500\text{Hz})$ /s	0.44	0.47	2.11	2.17
$C_{50}(500\text{Hz})$ /dB	4.9	5.2	-4.24	-4.54
STI*	0.73	0.73	0.46	0.45

* The effects of background noise are ignored.

Figure 2 shows the directivity patterns of JBL-LSR6325P loudspeaker and a human voice. The simulated binaural room impulse responses (BRIRs) are obtained under three source conditions after the sound power of the sources is adjusted as 75dBA (sound pressure level) at 1 meter in front of the sources which is according to the measured value in real room. Total of 30 BRIRs are obtained at 10 different listening positions (see figure 1) for three different sources.

2.2 Speech intelligibility evaluation

Mandarin Chinese phonetically balanced word lists specified by GB/T 15508-1995 [14] are used for speech intelligibility test. Every list consists of 25 three-syllables rows. A total of 75 syllables are used and a balance between the difficulty level and phonemic characteristic is maintained[15]. The three syllables in each row are randomly selected. The test words are embedded in the carrier phrase as “The - row is $\times\times\times$ ”, where “-” stands for the row number and “ $\times\times\times$ ” for three syllables selected from one list randomly and without repetition. All test words were recorded at a rate of 4.0 words per second in an anechoic chamber. A mute interval of 9-10 seconds between two adjacent rows recordings is added by Cool Edit Pro [16] for subjects to write down the testing words.

The subjective speech intelligibility test was conducted at different positions using different sources in both real and simulated rooms. In actual rooms, the speech signals of PB test word list recorded in an anechoic chamber are reproduced through B&K 4296 loudspeaker and the source JBL-LSR6325P loudspeaker. The test words are also read by a man and a woman speaker at the same source position. The source and listening positions in both rooms are shown

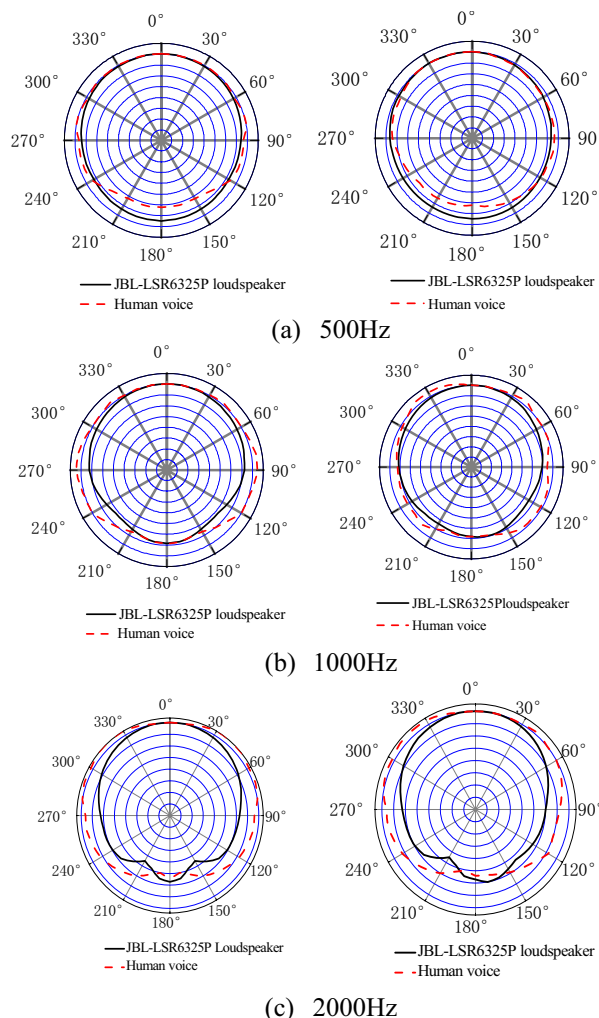


Figure 2 The directivity of the JBL-LSR6325P loudspeaker and human voice. Left: Horizontal Right: Vertical

in Figure 1. For auralization, the simulated BRIRs after headphone equalization at the corresponding listening position from different directional sources are convolved with the PB word list signals to obtain the speech test signals. Then the speech test signals are reproduced by headphone (Sennheiser HD580) at a certain sound pressure level which is set according to the result from ODEON simulation. The level estimation is based on the overall A-weighted root mean square (RMS) value and is corrected for the effect of silent periods by applying a threshold [17, 18]. The effect of background noise is ignored as the speech signal-to-noise ratio is more than 20dBA.

2.3 Subjects

The subjects were chosen from undergraduate students aged 19-24 years old. They have normal hearing ability. All subjects are trained and have passed a pretest which requires them to recognize test words in a clean condition with an identification rate of at least 95%. Two lists and 4-8 subjects are used under each listening condition. One of the lists has an odd number and the other an even number. The subjects are asked to write down the spellings of the key words which they heard. Only if the vowel, consonant and tone are all correct, the response is regarded as true, irrespective of the grapheme. The average of the subjective Chinese Mandarin intelligibility scores across all eight lists was then taken.

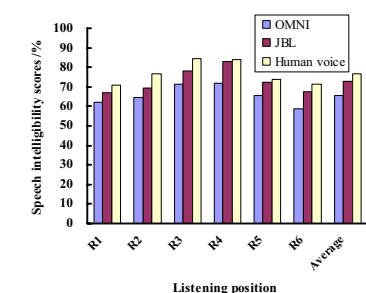
3 Result and Analysis

3.1 University classroom

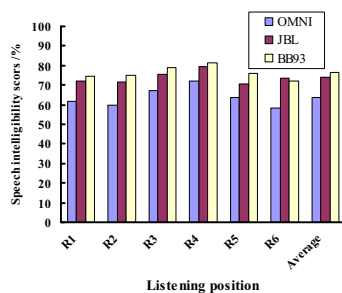
Figure 3 shows Chinese speech intelligibility scores in real classroom and in simulated one respectively with three different sources. It can be seen from Figure 3 that the discrepancies in both real and simulated classrooms exist between subjective speech intelligibility scores obtained from B&K 4296 loudspeaker and those obtained from JBL-LSR6325P loudspeaker and human voice. Figure 3(a) indicates that speech intelligibility scores from JBL-LSR6325P loudspeaker are higher than those from B&K 4296 loudspeaker in real classroom. The maximum difference is more than 10% at listening position 4, which the differences at other listening positions are between 4% and 10%. Similarly, Chinese speech intelligibility scores from human voice are higher than that which comes from B&K 4296 loudspeaker with differences at all listening positions ranging from 8% to a maximum of close to 13%. While discrepancies in speech intelligibility scores between JBL-LSR6325P loudspeaker and human voice are relatively smaller, subjects get more scores with human voice.

Figure 3(b) shows that speech intelligibility scores obtained from the simulated BRIRs with omnidirectional source are lower than those obtained with JBL-LSR6325P loudspeaker and BB93 based on auralization. At position 3, 4, 5, the scores from JBL-LSR6325P loudspeaker are 7% higher than those from omnidirectional source; at position 1, 2, 6, the discrepancies in scores are 10% more or less between JBL-LSR6325P loudspeaker and omnidirectional source. The average scores obtained from BB93 are slightly higher than those from JBL-LSR6325P loudspeaker by auralization. The scores obtained from BB93 are 3% higher than those from JBL-LSR6325P loudspeaker at other four listening positions except for at position 5 where the difference is almost 6% and at position 6 where the scores are slightly lower.

In order to investigate discrepancies of Chinese speech intelligibility scores obtained from different sound sources,



(a) Measurement



(b) Simulation

Figure 3 Chinese speech intelligibility scores from measurement and simulation in classroom

Least significant difference (LSD) method is applied to analyze subjective Chinese speech intelligibility scores obtained from real classroom and simulation model ($\alpha=0.05$).

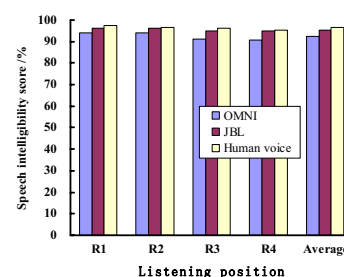
There is statistical significant difference ($p \leq 0.01$) between Chinese speech intelligibility scores from B&K 4296 loudspeaker and human voice in real room at all listening positions. There is statistical significant difference ($p < 0.01$) between scores from JBL-LSR6325P loudspeaker and B&K 4296 loudspeaker except at listening position 1 and 2. No statistical significant difference is found between those from human voice and JBL-LSR6325P loudspeaker except at listening position 2 and 3 ($p > 0.05$).

There is statistical significant difference ($p \leq 0.01$) in Chinese speech intelligibility scores from the simulated BRIRs between omnidirectional source and JBL-LSR6325P loudspeaker, and between omnidirectional source and BB 93 human voice. However, no statistical significant difference ($p > 0.05$) is found between JBL-LSR6325P loudspeaker and BB 93 source except at listening position 5.

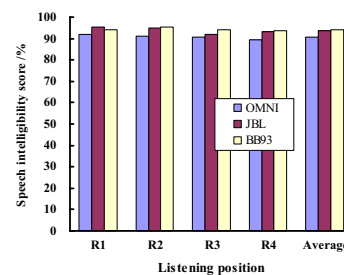
3.2 Multimedia lecture hall

Figure 4 shows Chinese speech intelligibility scores in multimedia lecture hall and in simulation model respectively with three different sources. It can be seen from Figure 4 that the discrepancies in both real hall and simulation model exist between subjective speech intelligibility scores obtained from B&K 4296 loudspeaker and the ones from JBL-LSR6325P loudspeaker and human voice. Figure 4(a) shows that speech intelligibility scores from JBL-LSR6325P loudspeaker are higher than those from B&K 4296 loudspeaker in real hall and the difference is less than 4%; Chinese speech intelligibility scores from human voice are also higher than those from B&K 4296 loudspeaker in real hall and the difference is less than 5%; Chinese speech intelligibility scores from human voice are also higher than those from JBL-LSR6325P loudspeaker in real hall but the difference is less than 2%.

It can be seen from figure 4(b) that the intelligibility scores obtained from the simulated BRIRs with



(a) Measurement



(b) Simulation

Figure 4 Chinese speech intelligibility scores from measurement and simulation in multimedia lecture hall

omnidirectional source in the lecture hall are lower than those with the other two sources ($p \leq 0.05$) and the differences are below 5%. The intelligibility scores obtained from the simulated BRIRs with BB93 are slightly higher than those with JBL-LSR6325P loudspeaker except at listening position 1.

LSD method is also applied to analyze subjective Chinese speech intelligibility scores obtained from real classroom and simulation model with three different sources in the lecture hall ($\alpha=0.05$). The results show that there is statistical significant difference between Chinese speech intelligibility scores from B&K 4296 loudspeaker and from JBL-LSR6325P loudspeaker except at listening position 1 ($p \leq 0.05$) and between those from B&K 4296 loudspeaker and from human voice at all listening positions. No statistical significant difference between those from human voice and from JBL-LSR6325P loudspeaker.

Except for R3 there is statistical significant difference ($P < 0.05$) when omnidirectional source and JBL-LSR6325P loudspeaker are used, so is it when BB93 human voice is used. But no statistical significant difference ($P > 0.05$) is found when human voice and JBL-LSR6325P loudspeaker are used.

There is statistical significant difference ($p < 0.05$) in Chinese speech intelligibility scores from the simulated BRIRs between omnidirectional source and JBL-LSR6325P loudspeaker source except at listening position 3, and between omnidirectional source and BB93 source except at listening position 1 based on auralization. However, no statistical significant difference ($p > 0.05$) is found between JBL-LSR6325P loudspeaker and BB 93 source.

4 Discussion

From the analysis mentioned above, we can find that there is statistical significant difference in Chinese speech intelligibility scores with, regardless of measuring in real room or using auralization technique, different directional sound sources under the same RT condition. The more reverberance is, the lower intelligibility score will be, and the differences in scores are more relatively under various source conditions.

There are no obvious effects of source directivity on RT and EDT in rooms when no significant directivity difference occurs between variant sources [7, 8]. RT and EDT represent energy decay in a certain time interval in a room, as a result, different directional sources have no obvious influence on the measures of RT and EDT. However, the curves of room impulse responses are different at the same listening position under different source conditions (see Figure 5). The change of late reflection and reverberant sound are different with time, which is important in calculating speech intelligibility parameter, such as sound energy ratio. The measured C50 value with directional source is higher than that with omnidirectional source. The difference of late reverberation sound in amplitude leads to different masking effects on speech signals. The more the late reverberation sound is, the stronger the masking of the late reverberation sound energy on speech signals tends to be, the lower the speech intelligibility will be. As can be seen from figure 5, there is more late reverberation sound energy with omnidirectional source than that with JBL-LSR6325P loudspeaker at the same time. This leads to that the speech intelligibility scores from omnidirectional source are lower than that

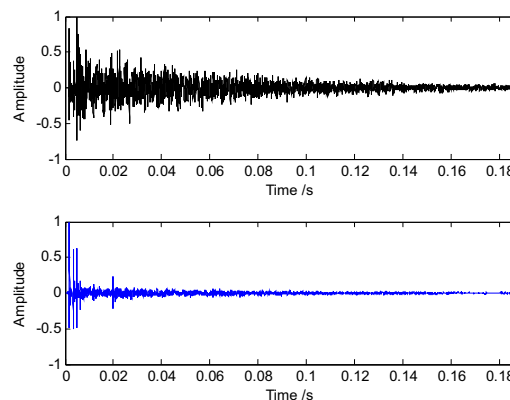


Figure 5 The measured room impulse response at the listening position 2 with different directional source. Up: B&K 4296 omnidirectional loudspeaker; Down: JBL-6325p loudspeaker.

obtained from JBL-LSR6325P loudspeaker.

By comparing Figure 3(a) and 3(b) and, Figure 4(a) and 4(b), the results of subjective Chinese speech intelligibility obtained from auralization simulation can be found. We used t-test to analyze the difference of subjective Chinese speech intelligibility in both real rooms and models. The results show that no significant difference exists between speech intelligibility scores from measurement and from auralization with omnidirectional source at each listening position in the classroom. So is the same with JBL-LSR6325P loudspeaker except at listening position 1 and 6, and with human voice except at listening position 3. In the hall, there is no significant difference between speech intelligibility scores from measurement and from auralization with omnidirectional source and with JBL-LSR6325P loudspeaker at each listening position. So as with human voice except at listening position 1.

Many factors would impact the results of subjective speech intelligibility evaluation by using auralization method. They include (1) the accuracy of room model and the absorption coefficients of surface materials[9]; (2) the mismatching of SPL and its calibration method; (3) the reproduction accuracy of speech signals in headphone[9].

Chinese speech intelligibility scores obtained from auralization are generally higher than that in real classroom. However, the difference in speech intelligibility scores is relatively smaller in multimedia lecture hall. This is mainly due to the method of the SPL calculation and calibration for auralization. RMS method is often used to calculate SPL of test signals for auralization in present studies. Length of speech test signals under a long reverberation time condition is longer than that under a short reverberation time condition. This may result that the SPL under a long reverberation time condition is higher than that in real classroom and thus obtaining a higher speech intelligibility scores.

From the analysis, it can be seen that speech intelligibility scores from an omnidirectional source are lower than those from the source with the approximate directivity of human talker and human talker source. In fact, it usually employs omnidirectional source to conduct the measurement of objective acoustical parameters and the simulation of sound field. This would underestimate speech intelligibility in rooms. Moreover, there is significant difference in directivity mode between omnidirectional

source and real sources. In order to reflect acoustical characteristics in real rooms, sound source that are close to actual one in directivity characteristic should be chosen for simulating and measuring.

5 Conclusion

Two rectangular rooms with different RTs are simulated by using ODEON. The objective acoustical parameters are also measured in both rooms using omnidirectional source. The speech intelligibility scores are investigated using an omnidirectional source, a source with the approximate directivity of human speaker and a human speaker source in real rooms and in simulated rooms based on auralization. The results show that there is statistical significant difference in subjective speech intelligibility scores obtained from different directional sources at the same listening positions. The scores from omnidirectional source are lower than those from the source with the approximate directivity of human speaker and a human speaker source. Speech intelligibility will be underestimated when an omnidirectional source is used. The speech intelligibility scores obtained from auralization agree with that in the real room. The subjective Chinese speech intelligibility under directional source condition can be evaluated by using auralization technique. In order to reflect acoustical characteristics in the real room, the similar directional sound source should be chosen for simulating and for taking measurement.

Acknowledgments

The authors extend their gratitude to the students who participated in subjective evaluation of Chinese speech intelligibility. This work is supported by National Natural Science Foundation of China (Grant No. 10774048 and 50938003), Science and Technology Planning Project of Guangdong Province, China (Grant No. 2008B080701020) and Project of the State Key Laboratory of Subtropical Building Science, South China University of Technology, China (Grant No. 2008KB32).

References

- [1] ISO 3382:1997. Acoustics – measurements of reverberation time of rooms with reference to other acoustical parameters. International Organization for Standardization(1997).
- [2] D.Rao, B.S.Xie, “Influence of sound source directivity on binaural auralization quality”, *Technical Acoustics*, 26, 899-903(2007).
- [3] BI Dalenback, M Kleiner, P. Svensson, “Audibility of changes in geometric shape, source directivity, and absorptive treatment–experiments in auralization”. *J Audio Eng Soc.* 41, 905–13 (1993).
- [4] D. Prince, R. Talaske, “Variation of room acoustic measurements as a function of source location and directivity”, *Wallace Clement Sabine centennial symposium*, 211–214 (1994).
- [5] F. Giron, *Investigations about the directivity of sound sources*, Ph.D. Dissertation. Ruhr-Universita’t Bochum, (1996).
- [6] F. Otondo, J. H. Rindel, “The influence of the directivity of musical instruments in a room”, *Acta Acustica united with Acustica*, 90, 1178–1184 (2004).
- [7] L.M.Wang, M.C.Vigeant, “Evaluations of output from room acoustic computer modeling and auralization due to different sound source directionalities”, *Applied Acoustics*, 69, 1281–1293 (2008).
- [8] T. Wang, J. X. Peng, H. T. Sun, “Influence of the Directivity of Source on the Objective Speech Intelligibility Parameters in a Classroom”, *Audio Engineering*, 34, 4-6 (2010).
- [9] W. Y. Yang, M. Hodgson, “Validation of the Auralization Technique: Comparative Speech Intelligibility Tests in Real and Virtual Classrooms”, *Acta Acustica united with Acustica*, 93, 991-999 (2007).
- [10] J. Peng, Feasibility of subjective speech intelligibility assessment based on auralization, *Appl. Acoustics*, 66, 591–601 (2005).
- [11] C. L. Christensen, *Odeon Room Acoustics Program User Manual*. Odeon A/S, Lyngby, Denmark (2009).
- [12] Oakley, Sebastian Alex Dalgas, *Comparison of Speech Intelligibility Measurements with ODEON Model Simulations*, Master thesis in Technical University of Denmark (2008).
- [13] GB 15508-1995. *Acoustics – speech articulation testing method. Standard of PR China* (1995).
- [14] Peng Jianxin, “Chinese speech intelligibility at different speech sound pressure levels and signal-to-noise ratios in simulated classrooms”, *Applied Acoustics*, 71, 386-390 (2010).
- [15] Syntrillium Software Corporation, *Cool Edit Pro Users Guide*(2003). Available: <http://www.syntrillium.com/>.
- [16] Steeneken H.J.M., Houtgast T, “Mutual dependence of the octave-band weights in predicting speech intelligibility”, *Speech Commun*, 28,109–23 (1999).
- [17] Steeneken H.J.M., Houtgast T. *Comparison of some methods for measuring speech levels*. Report IZF 1986-20, TNO-HFRI Human Factors Research Institute, Soesterberg, The Netherlands (1986).