Evaluating speech intelligibility degradation under different orders of Ambisonics

Conference Paper · October 2022

CITATIONS
0

READS
21

3 authors, including:

Zhenyu Guo
South China University of Technology
11 PUBLICATIONS **9** CITATIONS
SEE PROFILE

Huali Zhou
Shenzhen University
18 PUBLICATIONS **22** CITATIONS
SEE PROFILE

PROCEEDINGS of the
24th International Congress on Acoustics
October 24 to 28, 2022, in Gyeongju, Korea

Extended abstract

ABS-0814

Evaluating speech intelligibility degradation under different orders of Ambisonics

Zhenyu GUO[1]; Huali ZHOU[2]; Yuezhe ZHAO[3]

[1] School of Architecture, South China University of Technology, Guangzhou 510640, China

[2] College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China

[3] School of Architecture, South China University of Technology, Guangzhou 510640, China

ABSTRACT

The state-of-the-art virtual audio display technologies reproducing complex sound scenes show considerable auditory research prospects. The authentic sound field recreated by multi-channel loudspeakers makes it possible to evaluate speech perception for both normal-hearing and hearing-impaired listeners under some critical scenarios. Among the loudspeaker-based sound reproduction system, the Ambisonics system is considered to have a good trade-off between authenticity and hardware cost. However, speech intelligibility could be damaged by the crosstalk of multiple loudspeakers, especially when the lower-order Ambisonics system with a large energy spread is used. To evaluate the degradation of speech intelligibility under various orders of Ambisonics, the speech reception thresholds (SRTs) of twelve participants under different orders of Ambisonics and the one-channel playback (reference) conditions were measured. Results show that SRTs under the 1st and 3rd order Ambisonics were significantly higher (worse intelligibility) than those under the reference condition. The SRTs of higher-order (7th and 9th) Ambisonics are slightly higher than those under the reference condition, but it is not statistically significant. The variation range of SRTs under first-order Ambisonics is more extensive than other conditions, which may be related to the limited sweet area of first-order Ambisonics.

Keywords: Speech intelligibility, Ambisonics, Sound reproduction

1. INTRODUCTION

Conventional auditory research predominantly adopts simple audio (e.g., free-field recording) stimuli in experiments. Albeit using simplified stimuli may be capable of revealing the intrinsic auditory mechanism, realistic sound scenes may fetch conclusions with higher ecological validity in some venues. In other cases, realistic sound scenes may be desired due to the research purpose, e.g., speech perception in a specific reverberation environment[1]. Hence, the state-of-the-art virtual audio display technologies reproducing the elaborate sound scenes confirm considerable prospects for auditory research.

Binaural virtual audio display technology with earphones meets its massive application in auditory research. It is still regarded as the most accurate method up to now, even when several apparent drawbacks, including front-back confusion and in-head localization, exist[2]. However, the inherent limitation of binaural reproduction due to the earphones restricts the usage for hearing-assisted individuals.

The authentic sound field recreated by multi-channel loudspeakers makes it possible to evaluate speech perception for normal-hearing and hearing-impaired listeners under critical scenarios[3]. Among the loudspeaker-based sound reproduction system, the Ambisonics system is considered to have a good trade-off between authenticity and hardware cost. However, the Ambisonics system cannot reproduce the ideal sound field without distortions due to the limitation of the actual system. Hence, understanding how these distortions impact specific aspects of auditory perception may be one of the essential premises for applications of the Ambisonics system.

[1] zhenyuguo404@qq.com
[2] tuobamao@qq.com
[3] arzhyzh@scut.edu.cn

The present study investigated the speech intelligibility damaged by the multiple-loudspeaker reproduction system, especially when the lower-order Ambisonics system with a large energy spread is used. The hypothesis of this study is that both the distortions from crosstalk of multiple loudspeakers and spatial masking release from the energy spread of Ambisonics may influence speech intelligibility. To evaluate the degradation of speech intelligibility, the speech reception thresholds (SRTs) under different orders of Ambisonics and the one-channel playback (reference) conditions were measured.

2. Methods

During the SRT measurement, the masking noise was speech shape noise, and the sentences from Mandarin Hearing In Noise Test (MHINT) corpus were used as the speech stimuli. The noise was played with a single front loudspeaker directly while the speech was played in different methods. Six types of speech stimuli were incorporated in the experiment, i.e., played with a single front loudspeaker (reference) and synthesized with 0-, 1-, 3-, 7-, and 9-order of Ambisonics. Note that the 0-order Ambisonics implies the signal was distributed to all loudspeakers equally; hence the speech had no definite direction. In other conditions, the locations of noise and speech were always spatially coincident. The sound level differences among 6 conditions were measured with a monaural microphone placed in the center of the loudspeaker array and compensated correspondingly.

The Ambisonics system used in the present study comprises 192 spherically distributed loudspeakers. The virtual stimuli were reproduction with the model-match methods.

Twelve participants (aged from 21 to 29) were recruited with payments. Each participant underwent 12 SRT measurements of 6 different conditions (1 reference and 5 different orders of Ambisonics) with 2 repetitions. The sequence of tests was balanced with a 12×12 Latin square. Before the formal experiment, a training phase including 2 measurements of SRT was assigned to participants.

3. Results and discussions

Totally 144 thresholds of speech recognition were gathered in the experiment. The SRTs of all conditions were illustrated in Fig. 1. It clearly shows that the SRTs decreased and approached the results of the reference conditions when the order was promoted except for the 0-order condition, which means higher speech intelligibility. SRTs of the 0-order condition were inferior to other conditions using Ambisonics.

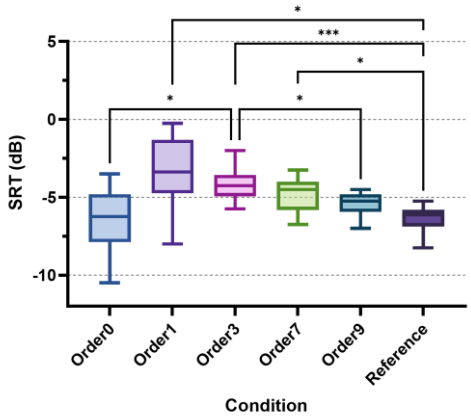

Fig 1. Tukey box and whisker plot of SRTs. The upper, middle, and lower lines of the box indicate the 75th, middle, and 25th of the SRTs, respectively. The * and ** caps denote significant differences at a level of 0.05 and 0.001, respectively.

The SRTs were statistically analyzed with the one-way repeated measures ANOVA, given the SRTs conform to the normality distribution. It reveals that conditions make a significant influence on the SRTs [$F(2.651, 29.16) = 7.652$, $p = 0.0009$]. According to the Tukey multiple comparisons, SRTs of the 1-, 3-, and 7-order conditions were significantly superior to those of the reference condition ($p = 0.0112$, 0.0001, and 0.0394, respectively). No significant difference between 9-order and reference conditions was observed. SRTs of 3-order were significantly higher than SRTs of 9-order condition ($p = 0.0409$). Besides, the variations of SRTs were also much larger for 0- and 1-order conditions. The

degraded speech intelligibility under lower order of Ambisonics probably comes from the crosstalk of multiple loudspeakers, which was verified for stereo[4] and Ambisonics system[5]. However, these results were disparate from those of Ahrens *et al*.[6], which show the SRTs barely not change with the order of Ambisonics when noise and speech are spatially coincident. Considering the differences between the two experimental configurations, further study may be needed.

While it seems SRTs of 0-order condition were below other conditions, a significant difference was only observed compared to 3-order condition (p = 0.0201). Given the noise was concentrated in the front loudspeaker in the present study, the spatial masking release may contribute to the speech intelligibility with the enlarged energy spread of speech in the lower order of Ambisonics. It may account for the phenomenon that the 0-order condition was completely contrary to the trend of other conditions.

4. Conclusion

The present study examined SRTs under different orders of Ambisonics. The results reveal that speech intelligibility was deteriorated by Ambisonics under the 9 order compared to playing with a single loudspeaker. Besides, the energy spread under the lower order of Ambisonics may induce perceivable spatial masking release. The study emphasizes the importance of considering the limitation of the Ambisonics system when using it to conduct auditory research.

ACKNOWLEDGEMENTS

Authors express gratitude to all participants involved in the experiment.

REFERENCES

1 Badajoz-Davila J, Buchholz JM, Van-Hoesel R. Effect of noise and reverberation on speech intelligibility for cochlear implant recipients in realistic sound environments. J Acoust Soc Am. 2020;147:3538.
2 Xie B. Head-related transfer function and virtual auditory display. USA: J. Ross Publishing; 2013.
3 Oreinos C, Buchholz JM. Evaluation of Loudspeaker-Based Virtual Sound Environments for Testing Directional Hearing Aids. J Am Acad Audiol. 2016;27:541–56.
4 Shirley B, Kendrick P, Churchill C. The effect of stereo crosstalk on intelligibility: comparison of a phantom stereo image and a central loudspeaker source. Journal of the Audio Engineering Society. 2007;55:852–63.
5 Solvang A. Spectral Impairment for Two-Dimensional Higher Order Ambisonics. Journal of the Audio Engineering Society. 2008;56:267–79.
6 Ahrens A, Marschall M, Dau T. Measuring and modeling speech intelligibility in real and loudspeaker-based virtual sound environments. Hear Res. 2019;377:307–17.