2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP) | 978-1-6654-7857-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSP54964.2022.9778684

Interaural time difference based spatial release from masking with asymmetric hearing over a video conference app

Jiachen Chen
Acoustics Laboratory, School of Physics and Optoelectronics,
South China University of Technology
Guangzhou, China
201920128443@mail.scut.edu.cn

Huali Zhou
College of Electronics and Information Engineering,
Shenzhen University
Shenzhen, China
zhouhuali2021@email.szu.edu.cn

Guangzheng Yu
Acoustics Laboratory, School of Physics and Optoelectronics,
South China University of Technology
Guangzhou, China
scgzyu@scut.edu.cn

Qinglin Meng
Acoustics Laboratory, School of Physics and Optoelectronics,
South China University of Technology
Guangzhou, China
mengqinglin@scut.edu.cn

***Abstract*—Due to the difficulty of conducting offline laboratory experiments during the coronavirus outbreak, remote experiments such as experiments over video conference apps have become an important method to collect data for hearing researchers. For remote testing using headphone presentations, compared to monaural (i.e., unilateral or diotic) audio stimuli, dichotic stimuli (i.e., sounds with differences between left and right ears) are used in relatively less studies. In this study, a binaural hearing task of spatial release from masking (SRM) was tested in laboratory and over a video conference app, i.e., Tencent Meeting. In the experiment, the effects of interaural time difference (ITD) on SRM were compared between symmetric and asymmetric hearing which were realized by using a fixed interaural level difference (ILD) of 0 dB and −15 dB respectively. Results showed that 1) SRM was observed in the remote test but it was >4 dB smaller than the laboratory test; 2) asymmetric hearing would lead to a ~2 dB significant decrease in the amount of masking release in both laboratory and remote conditions. The results indicate that binaural hearing could be measured remotely using the stereo sharing mode of video conference apps, but the effects of binaural cues especially ITDs may be degraded to some extent.**

Keywords—Interaural time difference, remote experiments, asymmetric hearing, spatial release from masking

I. INTRODUCTION

Subjective listening experiments typically require face-to-face measurements in the laboratory with measuring instruments. The outbreak of the coronavirus made it difficult to perform these experiments in laboratory. Many hearing researchers used online experiments over conference apps for data collection instead, such as Zoom and Tencent Meeting [1-2]. These methods facilitated the process of the experiments during the pandemic. Listeners can use headphone and stay in a quiet room to do the tasks. For remote testing with headphone presentations, monaural audio stimuli were widely used. For example, in [2] speech perception with cochlear implant users using different algorithms were measured through Tencent Meeting and in a sound booth. The study found that the relative differences between algorithms could be reliably measured but the absolute results may be inconsistent between testing modes for individual algorithm, because of influences from differences in software, hardware, and environment noise. Dichotic stimuli were used in relatively less studies. The online conference apps Zoom and Tencent Meeting both provide an optional function of stereo sound sharing, with monaural sound sharing as a default setting. Remote binaural hearing experiments have been carried out based on web [3] and downloaded software [4]. However, to the best of our knowledge, whether the video conference apps can reliably encode the binaural cues has not been reported in hearing research literatures.

In this study, a binaural hearing task of spatial release from masking (SRM) was carried out in laboratory and over Tencent Meeting. The target speech is sometimes masked by noise when they came from the same direction. If target speech and noise come from different directions, the listener can use binaural cues to separate the target from noise, behaving as a lower SRT or higher recognition accuracy than the co-located condition, resulting in SRM [5-9]. Many studies have demonstrated the importance of binaural cues, including interaural time difference (ITD) and interaural level difference (ILD), for improving the perception of target sounds in the presence of other interfering sounds [10-16]. Our recent study in laboratory showed that 1) when ILD is fixed at 0 dB, a ITD difference of 300 μs between target and noise could introduce significant SRM compared to a diotic condition; 2) dynamic modulation of ITD of target around 300 μs could further increase the SRM [17].

To examine the effects of remote testing with online conference apps on binaural hearing experiment, the ITD-SRM paradigm in [17] and Tencent Meeting was used in this study. An ILD condition of ILD = −15 dB was also added to simulate an asymmetric hearing threshold condition, which was

compared to the default ILD = 0 dB condition. The combinational conditions of ITD and ILD were used as a probe to study the strength and limitations of binaural hearing experiment over video conference apps. It was hypothesized that the app based stereo sound sharing function could be used to test binaural hearing remotely, but the binaural benefits may be negatively influenced by the remote setting as well as the asymmetric hearing thresholds.

II. METHOD

A. Participants

Two groups of participants each with ten (7 women and 13 men, range 19 to 26 years old, mean 23 years old) normal hearing subjects participated in these experiments. All subjects were native speakers of Mandarin Chinese, self-reported without any history of ear diseases, and confirmed that the 125-8000 Hz octave hearing threshold did not exceed 20 dB HL (confirmed prior to experiments by methods in [18]). One group participated in the online experiment and the other group participated in the offline experiment.

Offline experiment was carried out in a soundproof room, and the stimuli were presented by Roland OCTA-CAPTURE sound card and HD650 headphones, and the sound level was in a comfortable range, about 65-70 dBA. The online experiment was carried out over Tencent Meeting, in which the stereo audio was shared remotely but screen was not shared to the listeners. The listeners were asking to use wired headphones and the sound level was controlled in the same way as offline experiments. In both online and offline experiments, the same custom experimental software was used.

B. Stimuli

The target speech and noise materials used in these experiments were taken from the Mandarin Hearing In Noise Test (MHINT) corpus [19]. The corpus was recorded by a male speaker and contains 12 test sentence lists and 2 training sentence lists of 20 sentences each. These experiments used all 12 test sentence lists for testing, and 2 training sentence lists for training. The noise material was taken from the last four sentences of the second training sentence table, and in each trial, one of the sentences was randomly selected as the interfere noise.

Several ITD conditions for target speech were set, including three conditions of target speech ITD: S_0N_0, $S_{300}N_0$ and $S_{100\sim500}N_0$, where S represents the target speech, N represents noise, and the subscript represents the corresponding ITD value. The ITD of noise was fixed at 0 μs.

Condition of $S_{300}N_0$: The ITD adjustment method of the target speech was as followed: first, the single-channel original signal sampled at 16 kHz extracted from the audio file was used as the left-channel speech signal. Then it was up-sampled to 96 kHz sampling rate (the corresponding sampling interval is 1/96000 ≈ 10.4 μs) and then shifted by 29 samples (rounding of ITD×96000, for ITD = 300 μs). The shifted signal was down-sampled back to 16 kHz as the right channel signal..

Condition of $S_{100\sim500}N_0$: The dynamic ITD condition was realized by dynamically changing the ITD of the target speech with time (vertical axis) in a sinusoidal modulation mode within the range of 300 ± 200 μs at a modulation rate f_m of 0.5 Hz. The 0.5 Hz was selected because our previous study showed that a dynamic ITD condition of $S_{100\sim500}N_0$ with f_m = 0.5 Hz could introduce small but significant masking release compared to the fixed ITD condition of $S_{300}N_0$. The dynamic change of ITD was shown in equation (1),

$$ITD(t) = 300 + 200\sin(2\pi f_m t + \theta_0) \quad (1)$$

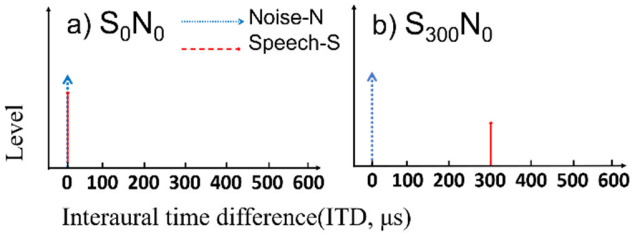

Fig.1. Two fixed ITD conditions. (S: Speech; N: Noise)

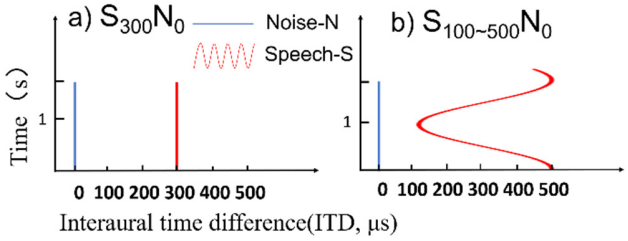

Fig.2. A fixed (left) and a dynamic (right) ITD condition (S: Speech; N: Noise)

where θ_0 is the initial phase (a uniformly distributed random value between 0 and 2π in each trial). The sampling frequency of the original speech was 16 kHz. A Hanning window of 200 sampling points is used. The corresponding window length was 200/16k = 12.5 ms, and the frame shift and inter-frame overlap were both 50% of the window length (i.e., 100 sampling point and 6.25 ms). Then the middle point time of the current frame was substituted as t to formula (1) to calculate an ITD value. The binaural signal of current frame was generated according to the procedure described in the paragraph of "*Condition of $S_{300}N_0$*". Finally, the binaural signals of all frames are superimposed at the corresponding sampling time [17].

For each of the three conditions, i.e. S_0N_0, $S_{300}N_0$ and $S_{100\sim500}N_0$, the final stimulus was obtained by superimposing the interfere noise and the speech signal of the left and right channels at a specific signal-to-noise ratio. The noise level was fixed, and the sound level of the target speech was adjusted to achieve different signal-to-noise ratios (SNRs). The calculation steps of the specific signal-to-noise ratio value were shown in the next section [17].

A asymmetric hearing was simulated by decreasing the level of the noisy speech at the right ear by 15 dB, so that the ILD between right and left ears was −15dB. Therefore, in total there were 6 conditions in the experiment, i.e., 3 ITD conditions (i.e., S_0N_0, $S_{300}N_0$ and $S_{100\sim500}N_0$)×2 ILD (0dB and −15dB) conditions.

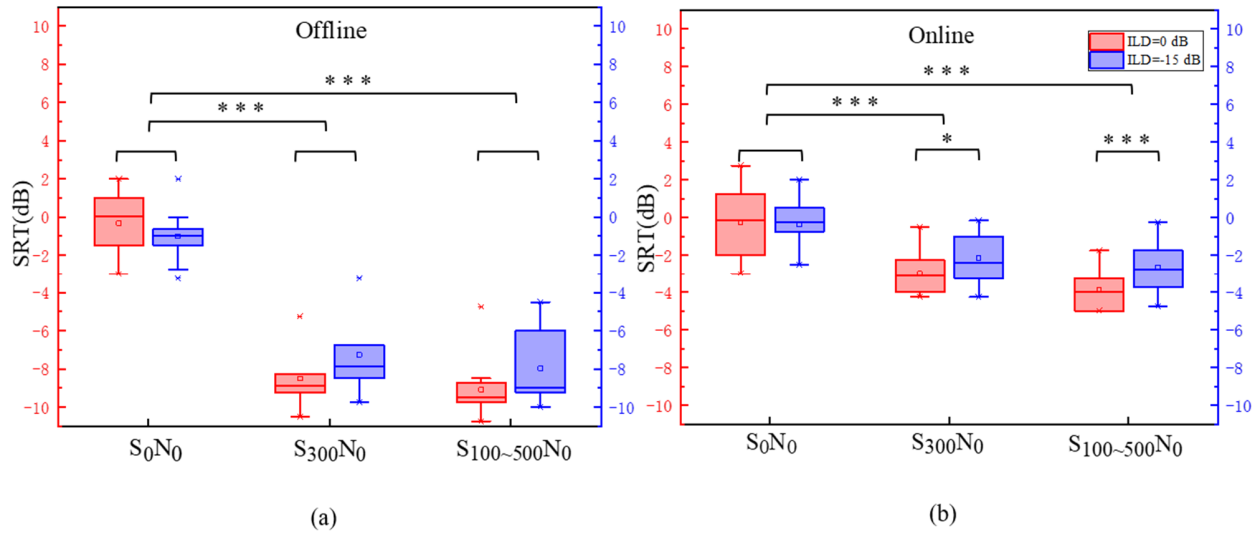

Fig.3. SRTs measured under different ITD conditions and different ILD conditions in the offline and (a) online test (b). The box lines represent the median and quartiles, the small squares represent the mean. Asterisks (*) indicate significant threshold differences: *p < 0.05, **p < 0.01, ***p < 0.001.

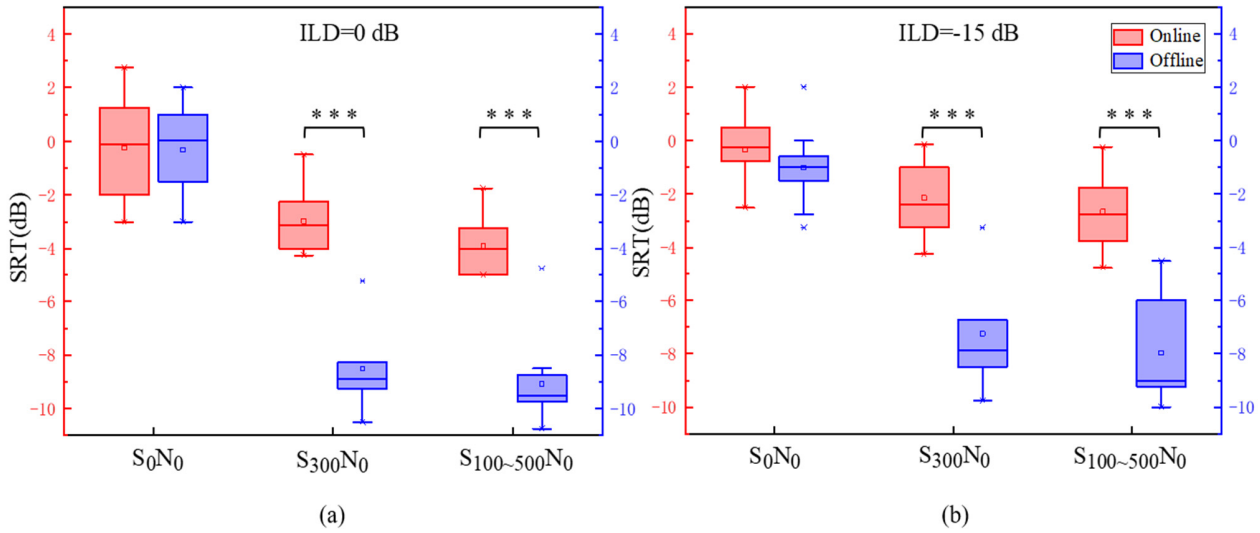

Fig.4. (a) SRTs measured under different ITD conditions in the offline and offline test ; (b) SRTs measured under different ITD conditions and 15 dB ILD condition in the online and offline test. The box lines represent the median and quartiles, the small squares represent the mean. Asterisks (*) indicate significant threshold differences: *p < 0.05, **p < 0.01, ***p < 0.001.

C. *Procedure*

A 1-down-1-up adaptive procedure was used to measure the speech reception threshold (SRT), which is the SNR with 80% recognition scores. The adaptive procedure used in this experiment is the same as that used in [17] and [20]. The initial signal-to-noise ratio is 10 dB, and the step size of the up-down change was 8 dB before the second reversal point, and then it became 4 dB. After 4 reversal points, it became 2 dB. The average signal-to-noise ratio of the last 8 sentences was taken as SRT. Before the presentation of each stimuli, a binaural complex tone containing 2 harmonics (1600 and 4800 Hz) with a length of 0.5 s was played as prompt tone, and the interval between prompt tone and stimuli was 0.2 s. After playing the stimuli, the subject was required to repeat the sentence, and if the correct word count exceeded 80% of the sentence, the sentence was recorded as intelligible. In each trial, the subject could request at most one chance to replay it.

Each condition was tested twice using different MHINT sentence lists, and the average of the two SRTs were taken as the final results of the condition. The six conditions (i.e, 3 ITD conditions × 2 ILD conditions) and the sentence lists were tested in random orders. In order to familiarize the subjects with the test process, before the formal test begins, two conditions, S_0N_0 and $S_{300}N_0$, were trained using the MHINT training lists.

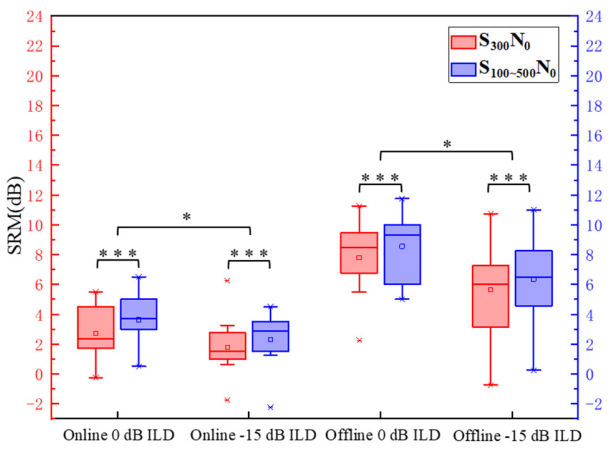

Fig.5. SRM produced in different conditions. (The box lines represent the median and quartiles, the small squares represent the mean. Asterisks (*) indicate significant threshold differences: *p < 0.05, **p < 0.01, ***p < 0.001.

III. RESULTS

Results are shown in in Fig.3. and Fig.4. Fig.3 shows SRTs with different ITD and ILD conditions in the offline test (a) and in the online test (b). In both the offline and online tests, participants had lower SRTs (better performance) when ITD was provided to target speech. To investigate the effects of ITD and ILD on SRT, a repeated measures two-way analysis of variance (ANOVA) was performed separately for the offline test and online test, using different ITD and ILD conditions as the independent variable and SRT as the dependent variable. A significant effect of ITD was found in the online test [$F(1,9) = 29.884$, $p < 0.001$], and also in the offline test [$F(1,9) = 128.207$, $p < 0.001$]. The effect of ILD was significant in online test [$F(1,9) = 6{,}683$, $p < 0.05$] but not in the offline test ($p = 0.176$). Post hoc comparisons with Bonferroni corrections showed that S_0N_0 had significantly higher SRTs than those of $S_{300}N_0$ ($p < 0.001$) and $S_{100\sim500}N_0$ ($p < 0.001$) in both online and offline tests. $S_{300}N_0$ with 15-dB ILD had significantly higher SRTs than those of $S_{300}N_0$ with 0-dB ILD ($p < 0.05$) in both online and offline tests. $S_{100\sim500}N_0$ with 15 dB ILD had significantly higher SRTs than those of $S_{100\sim500}N_0$ with 0-dB ILD in offline test ($p < 0.05$) and online test ($p < 0.001$).

To have a clear view of the difference between results of the offline and online tests, same results as in Fig.3 are rearranged in Fig.4. Online tests had significantly higher SRTs than those of the offline tests for both $S_{300}N_0$ and $S_{100\sim500}N_0$ ($p < 0.001$).

SRM, defined as the decrease in SRT when adding a non-zero ITD to the target speech, was shown in Fig. 5. SRMs in offline experiments were larger than those in online experiments. Asymmetric hearing thresholds would lead to a reduction of SRM. Using different ITD and ILD conditions as the independent variable and SRM as the dependent variable, mixed model ANOVA was performed. Results showed that different ILD conditions had a significant effect on SRM [$F(1,18)=13.399$, $p < 0.05$] and ITD conditions had a significant effect on SRM [$F(1,18)=25.430$, $p < 0.001$].

IV. DISCUSSION

The finding that that a simulation of hearing loss (15 dB attenuation in one channel) leads to a decrease of SRM is consistent with previous studies. For example, Bronkhorst et al. [16] found that SRM from both ITD and ILD was decreased when an overall 20 dB attenuation in one channel was applied.

The results showed in Fig.3. indicated that ITD difference between target and noise could derive significant SRM, which is in line with Kidd et al. and Culling et al. [21-22]. When asymmetric hearing was simulated, the SRM decreased. In [17], $S_{100\sim500}N_0$ derived significantly lower SRTs than $S_{300}N_0$ did, but in current study they did not show significant different performance. This may be because of the less training about the dynamic condition in this work than in [17].

As shown in Fig.4., there was little difference between the SRTs of S_0N_0 in both the online and offline experiments, which indicated that the effect of different hardware and software was small between the two test modes. However, there was noticeable difference between SRTs measured by the online and offline tests for $S_{300}N_0$ and $S_{100\sim500}N_0$, which may due to effects of the codecs of the app and transmission error of Internet [23]. There was still SRM produced in remote tests, but less than offline tests. The results indicate that the online video conference apps could encode ITD cues but in an imprecise manner. Further study about improving the precision of binaural cues coding may solve this problem.

V. CONCLUSSION

In this paper, an SRM (or binaural unmasking) experiment was carried out to examine whether binaural hearing could be measured remotely using the stereo sharing mode of video conference apps, the following preliminary conclusions were drawn:

- Asymmetric hearing may lead to a decrease in SRM in both online and offline experiments.
- Remote experiments showed less SRM produced, which may be attributed to the imprecise coding of ITD cues in the stereo sharing mode of video conference apps.

ACKNOWLEDGMENT

The authors thank all the participants. This work was supported by Fundamental and Applied Basic Research Fund of Guangdong Province (2020A1515010386), Guangzhou Science and Technology Program (202102020944), National Natural Science Foundation of China (11704129). Huali Zhou is the co-first author and Qinglin Meng is the corresponding author.

REFERENCES

[1] A.Yamamoto et al., "Comparison of remote experiments using crowdsourcing and laboratory experiments on speech intelligibility," 2021.

[2] Xi Chen et al., "Internet streaming audio based speech reception threshold measurement in cochlea implant users," ICASSP 2022. Accepted.

[3] A.L. Padilla-Ortiz, Felipe Orduña-Bustamante, ""Binaural speech intelligibility tests conducted remotely over the Internet compared with tests under controlled laboratory conditions," Applied Acoustics, 2021, pp, 172, 107574.

[4] Merchant, G. R., Dorey, C., Porter, H. L., Buss, E., & Leibold, L. J. "Feasibility of remote assessment of the binaural intelligibility level difference in school-age children," *JASA Express Letters*, 2021, pp, *1*(1), 014405.

[5] Y.Litovsky R., Spatial release from masking. Acoustic today, 2012, 8(2).

[6] S. A. Gelfand, L. Ross, and S. Miller, "Sentence reception in noise from one versus two sources: Effects of aging and hearing loss," J. Acoust. Soc., 1988, pp, Am. 83, 248–256.

[7] K. Allen, S. Carlile, and D. Alais, "Contributions of talker characteristics and spatial location to auditory streaming," J. Acoust. Soc. Am, 2008, pp, 1562–1570.

[8] E. C. Cherry, "Some experiments on the recognition of speech with one and two ears," J. Acoust. Soc. Am, 1953,pp, 25, 975–979.

[9] S. Cameron, H. Glyde, and H. Dillon, "Listening in spatialized noise–sentences test: Normative and retest reliability data for adolescents and adults up to 60 years of age," J. Am. Acad. Audiol.,2011, pp, 22, 697–709.

[10] Bronkhorst, A., "The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions,",Acustica, 2000, pp, 86, 117e128.

[11] Bernstein LR, Trahiotis C, Bernstein LR. "Binaural interference effects measured with masking-level difference and with ITD- and liD-discrimination paradigms.,"J Acoust Soc Am, 1995, pp, 98:155–163.

[12] Beutelmann R, Brand T. "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," J Acoust Soc Am, 2006, pp, 120:331–342

[13] Glyde, Helen, et al. "The importance of interaural time differences and level differences in spatial release from masking.". The Journal of the Acoustical Society of America, 2013, pp, 134.2: EL147-EL152.

[14] Festen JM, Plomp R, "Effects of fuctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing," J Acoust Soc Am, 1990, pp, 88:1725–1736.

[15] Freyman RL, Helfer KS, McCall DD, Clifton RK. "The role of perceived spatial separation in the unmasking of speech," J Acoust Soc Am, 1999, pp, 106:3578–3588.

[16] Bronkhorst, A. W. "The effect of head-induced interaural time and level differences on speech intelligibility in noise," The Journal of the Acoustical Society of America, 1988, pp, 83(4), 1508–1516.

[17] Jiachen Chen, Huali Zhou, Guangzheng Yu, Qinglin Meng. "Effects of interaural time difference on spatial release from masking," unpublished.

[18] Guo Z, Yu G, Zhou H, Wang X, Lu Y, Meng Q. "Utilizing True Wireless Stereo Earbuds in Automated Pure-Tone Audiometry," Trends Hear, 2021, pp, 25: 23312165211057367.

[19] Chen Y, Wong L L N. "Development of the mandarin hearing in noise test for children," Int J Audiol, 2020, pp, 59(9): 707—12.

[20] Meng Q, Zheng N, Li X. " Mandarin speech-in-noise and tone recognition using vocoder simulations of the temporal limits encoder for cochlear implants," J Acoust Soc Am, 2016, pp, 139(1): 301—10.M.

[21] J. F. Culling, M. L. Hawley, and R. Y. Litovsky, "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. , 2004, pp, 116(2), 1057–1065.

[22] G. Kidd, C. R. Mason, V. Best, and N. Marrone, "Stimulus factors influencing spatial release from speech-on-speech masking," J. Acoust. Soc. Am, 2010 , pp, 128(4) , 1965–1978.

[23] Pedro Mayorga, Laurent Besacier, Richard Lamy, and J-F Serignat, "Audio packet loss over IP and speech recognition," in 2003 IEEE Workshop on Automatic Speech Recognition and Understanding. IEEE, 2003, pp. 607–612. R.