Internet Streaming Audio Based Speech Perception Threshold Measurement in Cochlear Implant Users

Xi Chen[1, 2], Yefei Mo[3], Kang Ouyang[1], Mingyue Shi[1], Huali Zhou[1],
Yupeng Shi[2], Wei Xiao[2], Shidong Shang[2], Qinglin Meng[3], Nengheng Zheng[1]

Email: nhzheng@szu.edu.cn, mengqinglin@scut.edu.cn

[1]The Guangdong Key Lab. of Intelligent Information Processing, College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China

[2]Tencent Ethereal Audio Lab, Shenzhen, China

[3]Acoustic Lab., School of Physics and Optoelectronics, South China University of Technology, Guangzhou, China

Paper ID: 5614

Problem Statement and our Contributions

- The COVID-19 pandemic has made traditional face-to-face listening test challenging due to social distancing rules and regulations
- Recent web-based platforms are available for speech intelligibility test in normal hearing (NH) listeners [1] and cochlear implant (CI) users [2]
  - [1]: installing standalone application, uploading data to the cloud
  - [2]: depending on direct audio input which bypassing the mic
- We conducted two **remote** speech reception threshold (SRT) assessments to evaluate the **feasibility and reliability** with CI users
  - Characterizing speech intelligibility in local and remote settings
  - Comparing the SRTs of the remote with conducted in-person

Experiment I: Experimental Conditions

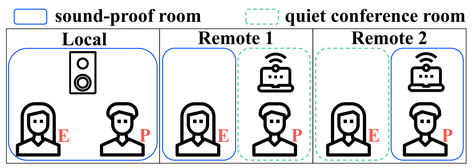

Fig. 1 Scenes schematic (E = experimenter; P = participant).

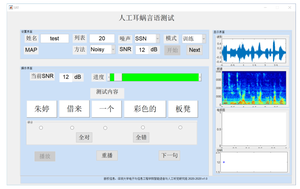

Fig. 2 SRT MATLAB Interface.

- Acoustic conditions
  - Noise-masking: DNN-based noise reduction (NR) (denoted by 'DNN') vs. without NR (denoted by 'Noisy')
  - Noise type: Babble vs. speech shaped noise (SSN)

姓名	动词	量词	形容词	名词
郭靓	带走	一个	彩色的	板凳
李锐	借来	两个	大号的	茶杯
沈悦	看见	三个	银田的	灯笼
王石	留下	四个	便宜的	钱盒
徐敏	买回	五个	漂亮的	花瓶
杨硕	拿起	六个	普通的	戒指
张伟	弄丢	七个	奇怪的	闹钟
郑贺	收好	八个	全新的	书包
周明	需要	九个	特别的	水壶
朱婷	找出	十个	用过的	玩具

每一句组成：姓名 + 动词 + 量词 + 形容词 + 名词

Fig. 3 CMNmatrix Table

- Scenes conditions: Local vs. Remote 1 vs. Remote 2 (Fig. 1)
  - Local: sound-proof room, high-quality monitor speaker
  - Remote 1: experimenter in sound-proof room, participant in quiet conference room using laptop built-in loudspeaker, communicate with each other via Tencent Meeting
  - Remote 2: participant in sound-proof room using high-quality monitor speaker, experimenter in quiet conference room , communicate with each other via Tencent Meeting

Experiment I: Procedure

- Subjects: 7 CI users (aged 22 to 47), native Mandarin speaker
- Task: SRT assessment (Fig. 2) with adaptive staircase psychophysical procedure
- Material: Mandarin Chinese matrix (CMNmatrix) corpus (Fig. 3) with randomize condition order
- SRT results under different conditions were measured and compared

Experiment I: Results and Discussions

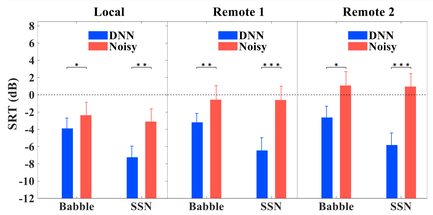

Fig. 4. Mean SRTs for different test conditions, error bars indicate the standard deviations., Asterisks above indicate the NR effect statistical significance (*p <0.05, **p <0.01, ***p <0.001).

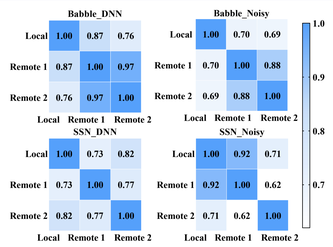

Fig. 5. Correlation matrices of the mean SRTs among different assessment scenes.

Table 1. ANOVA results of Experiment I.

Source	F value	p value
S	$F(2, 12) = 19.849$	$< 0.001^{***}$
NT	$F(1, 6) = 32.558$	0.001^{**}
NR	$F(1,6) = 48.669$	$< 0.001^{***}$
S × NT	$F(2,12) = 0.767$	0.486
S × NR	$F(2,12) = 9.258$	0.004^{**}
NT × NR	$F(1,6) = 19.493$	0.004^{**}
S × NT × NR	$F(2,12) = 0.483$	0.628

S represents scene, NT represents noise type, NR represents noise reduction.

- Mean SRT: SSN<Babble, DNN<Noisy, Local < Remote 1 < Remote 2
- Fig. 4 & Table 1: NR effects are significant in all conditions, but the effects differ in each conditions
- Fig. 5: Remote assessments have **strong correlations** with local assessments regardless of the noise-related conditions

Experiment II: Conditions & Procedure

- Noise type: Babble
- Scenes conditions: Local vs. Remote 1
- Processing conditions: vocoded speech based on Advanced Combination Encoder (ACE) strategy with 2, 4, 6, 8, 12 , or 16-of-22 channels selected
- Subjects: 10 NH listeners (aged 17 to 24), native Mandarin speaker
- Material and comparison are same as Experiment I

Experiment II: Results and Discussions

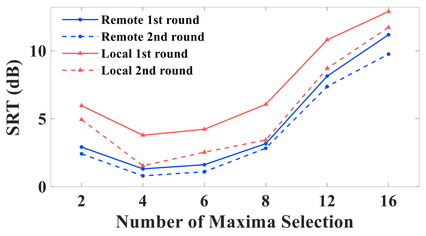

Fig. 6. Mean SRTs for different test conditions.

- Remote and local assessments gave **similar trends** of SRTs
- Mean SRTs in Remote < Local due to the insufficient training
- No significant difference between SRTs for the 2nd Local and the 1st Remote, also suggesting that training is the main factor

Conclusions

- Remote subjective assessments could be a reliable alternative to face-to-face assessments for CI research in the pandemic
- The relative variation of specific performance can be measured reliably, but the absolute values should be carefully compared and explained according to experimental conditions (e.g., internet transmission, background noise).

References

[1] Kevin M Chu, Leslie M Collins, and Boyla O Mainsah, "Assessing the intelligibility of vocoded speech using a remote testing framework," *arXiv preprint arXiv:2105.14120*, 2021.

[2] Joshua D Sevier, Sangsook Choi, and Michelle L Hughes, "Use of direct-connect for remote speech-perception testing in cochlear implants," *Ear and Hearing*, vol. 40, no. 5, pp. 1162–1173, 2019.