Tencent Ethereal Audio Lab

22-27 MAY 2022
ICASSP 2022
SINGAPORE

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

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

May 12th, 2022

01 Motivation

02 Experiment I

03 Experiment II

04 Conclusions

Background

- COVID-19 pandemic social distancing rules and regulations
- Rapid development of internet technology
- Widespread use of video conferencing applications
  - Pros: convenient, responsive, efficient
  - Cons: quality, environmental noise

Flexible and **reliable** remote assessment methodology

Recent work

- Tremendous growth in tele-audiology services
  - Low- or no-touch
  - Web- and app-based
- Remote platforms are available for speech intelligibility test in normal hearing (NH) listeners and cochlear implant (CI) users
  - Time-consuming：install standalone app, upload data to the cloud
  - direct audio input：exclusive，bypass the mic

[1] De Wet Swanepoel and James W Hall, “Making audiology work during covid-19 and beyond,” *The Hearing Journal*, vol. 73, no. 6, pp. 20–22, 2020..
[2] 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.
[3] 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.

Our Main Contribution

➢ Two remote speech reception threshold (SRT) assessments were conducted to evaluate the **feasibility** and **reliability** with CI users

- Characterizing speech intelligibility in remote and in-person settings
- Comparing the SRTs of the remote with conducted in-person

01

Motivation

02

Experiment I

03

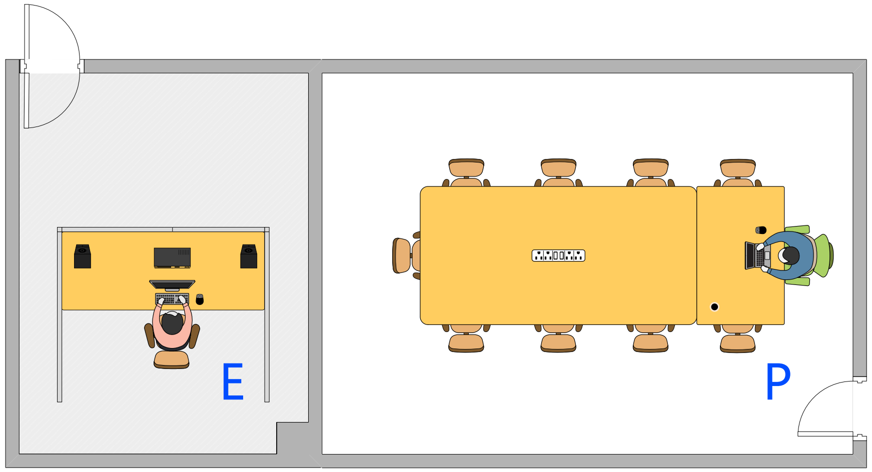
Experiment II

04

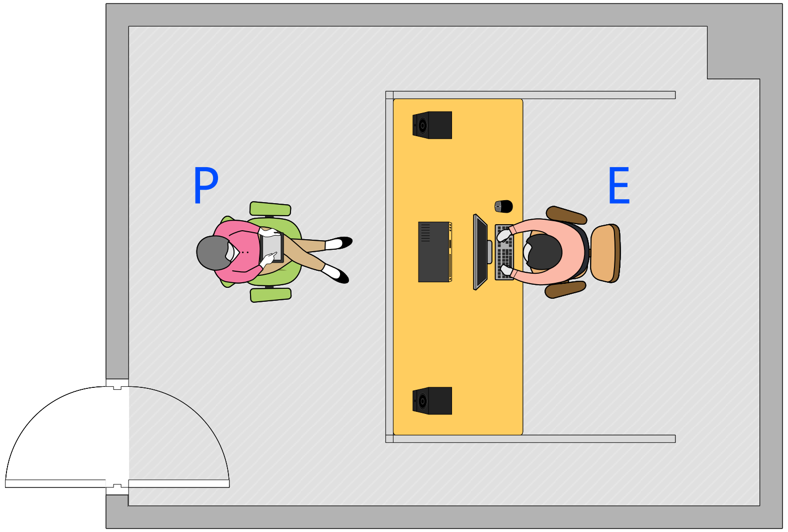
Conclusions

Conditions

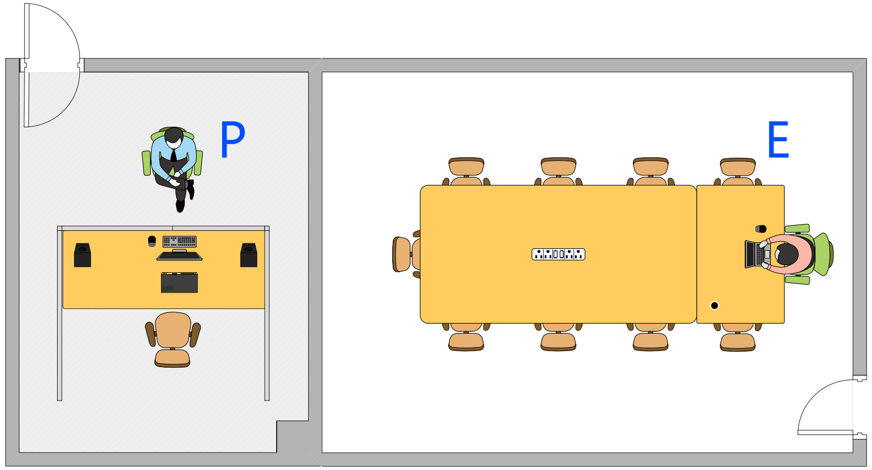
- Acoustic conditions
  - Noise-masking: DNN vs. Noisy
  - Noise type: Babble vs. SSN
- Scenes conditions:
  - Local vs. Remote 1 vs. Remote 2



Remote 1



Local



Remote 2

Procedure

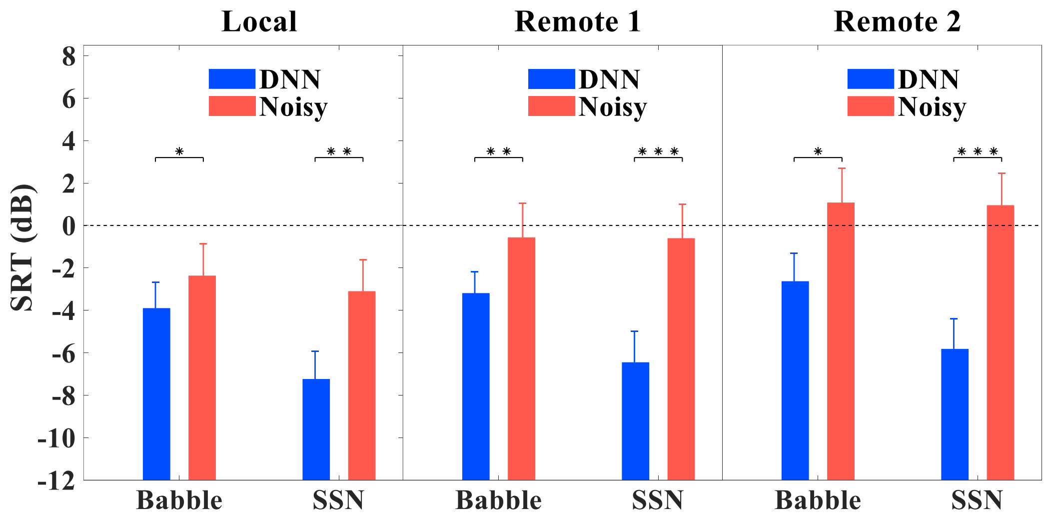
- Subjects: 7 CI users (aged 22 to 47), native Mandarin speaker
- Task: SRT assessment with adaptive staircase psychophysical procedure
- Material: Mandarin Chinese matrix corpus with randomize order
- SRT results under different conditions were measured and compared

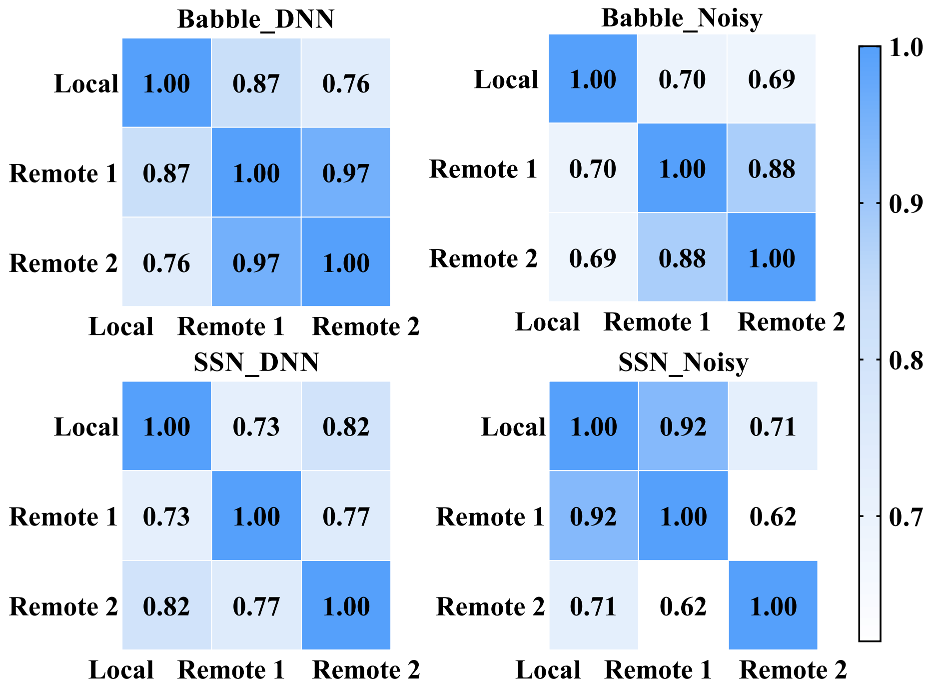
Scene	Noise Type	Noise Reduction	List	SRT (dB)
Remote 2	Babble	Noisy	list16	
	SSN	DNN	list19	
	Babble	DNN	list31	
	SSN	Noisy	list6	
Remote 2	Babble	DNN	list2	
	SSN	Noisy	list5	
	SSN	DNN	list15	
	Babble	Noisy	list30	
Local	Babble	Noisy	list37	
	Babble	DNN	list4	
	SSN	Noisy	list7	
	SSN	DNN	list9	
Remote 1	SSN	Noisy	list10	
	Babble	DNN	list14	
	Babble	Noisy	list8	
	SSN	DNN	list29	
Remote 1	SSN	DNN	list24	
	Babble	Noisy	list25	
	SSN	Noisy	list23	
	Babble	DNN	list17	
Local	Babble	Noisy	list34	
	SSN	DNN	list27	
	Babble	DNN	list33	
	SSN	Noisy	list11	

[4] Hongmei Hu, Xin Xi, Lena LN Wong, Sabine Hochmuth, Anna Warzybok, and Birger Kollmeier, "Construction and evaluation of the Mandarin Chinese matrix (CMNmatrix) sentence test for the assessment of speech recognition in noise," *International Journal of Audiology*, vol. 57, no. 11, pp. 838– 850, 2018.

Results and Discussions (1/2)

- Mean SRT: SSN<Babble, DNN<Noisy, Local < Remote 1 < Remote 2
- NR effects are **significant in all conditions**
- Remote assessments have **strong correlations** with local assessments regardless of the noise-related conditions





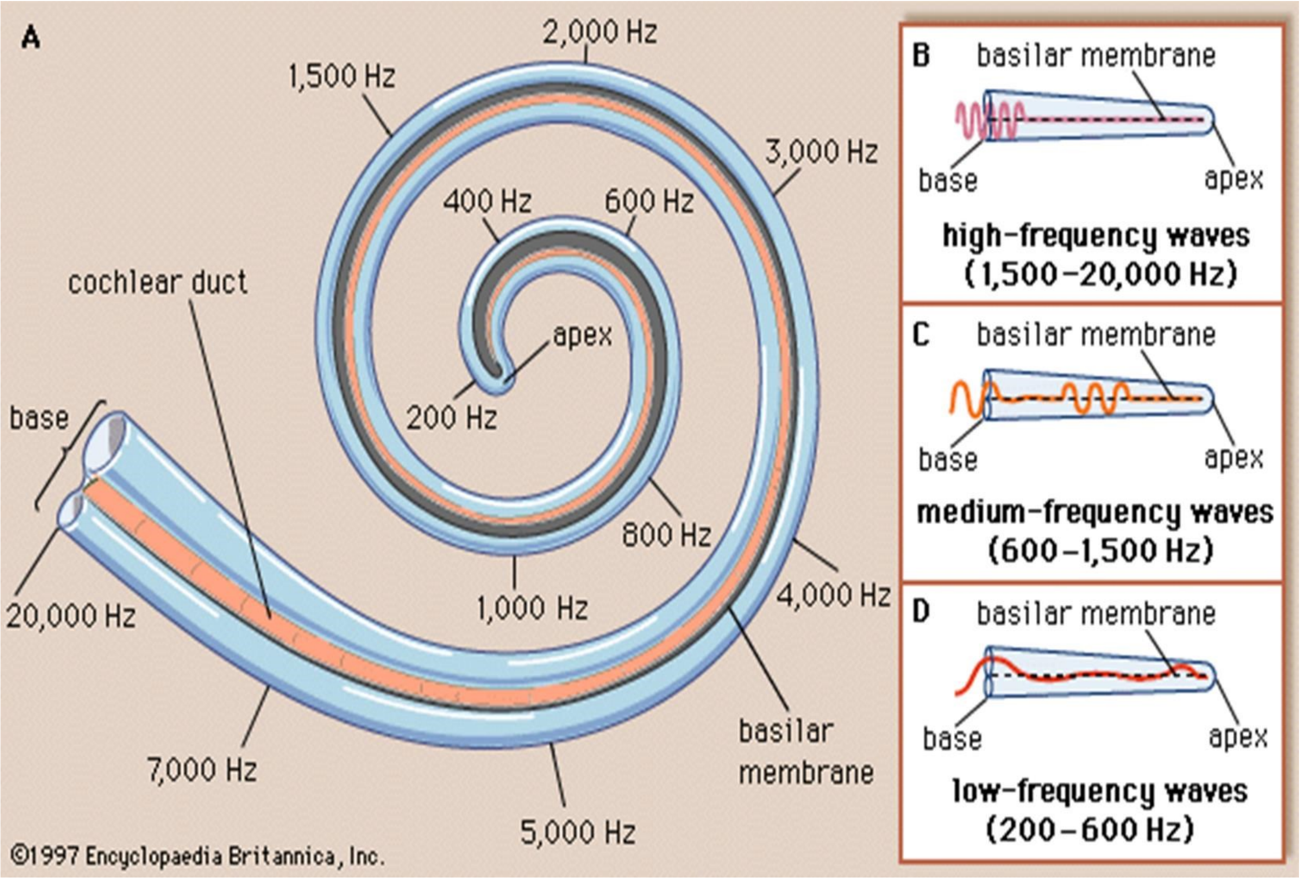
Results and Discussions (2/2)

- Scene/Noise type/Noise reduction effects were **statistically significant** on SRTs
- Interaction between scene and noise type was not significant
- Two significant interaction effects were observed
- NR effects differed across scenes and noise types
- No combined effects for these three factors

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

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

Rationale



[5] A. E. Vandali, L. A. Whitford, K. L. Plant, and G. M. Clark, "Speech perception as a function of electrical stimulation rate: Using the nucleus 24 cochlear implant system," *Ear and Hearing.*, vol. 21, no. 6, pp. 608–624, 2000.

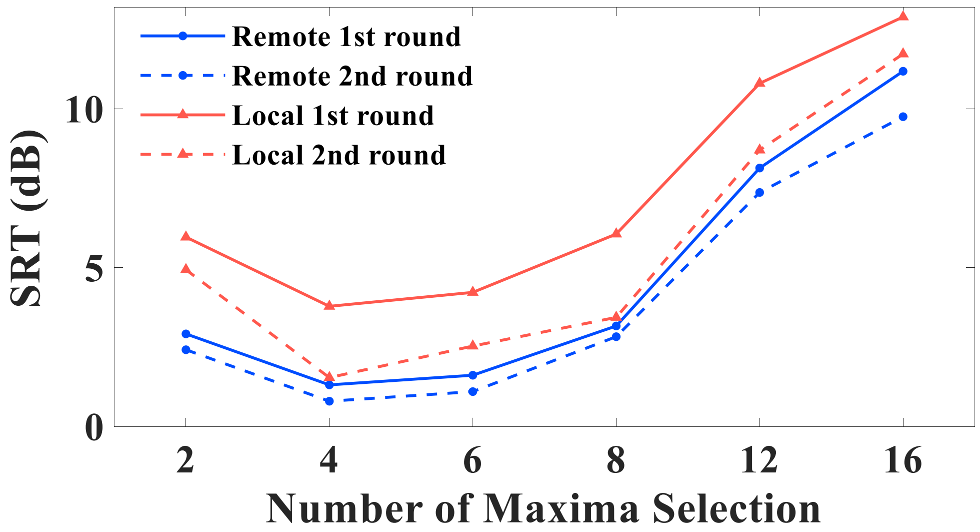
Conditions & Procedure

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

[5] A. E. Vandali, L. A. Whitford, K. L. Plant, and G. M. Clark, "Speech perception as a function of electrical stimulation rate: Using the nucleus 24 cochlear implant system," *Ear and Hearing.*, vol. 21, no. 6, pp. 608–624, 2000.

Results and Discussions

- Remote and local assessments gave similar trends of SRTs (r = 0.995, p < 0.001)
- Mean SRTs in Remote < Local due to the insufficient training
- No significant difference (p > 0.05 for all comparisons) between SRTs for the 2nd Local and the 1st Remote, suggesting that training is the main factor



01

Motivation

02

Experiment I

03

Experiment II

04

Conclusions

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
- The absolute values should be carefully compared and explained according to experimental conditions
- Future work will aim to address these noted issues

Thank you for your attention!