

Extended-high-frequency cues for speech perception: talker head orientation and gender

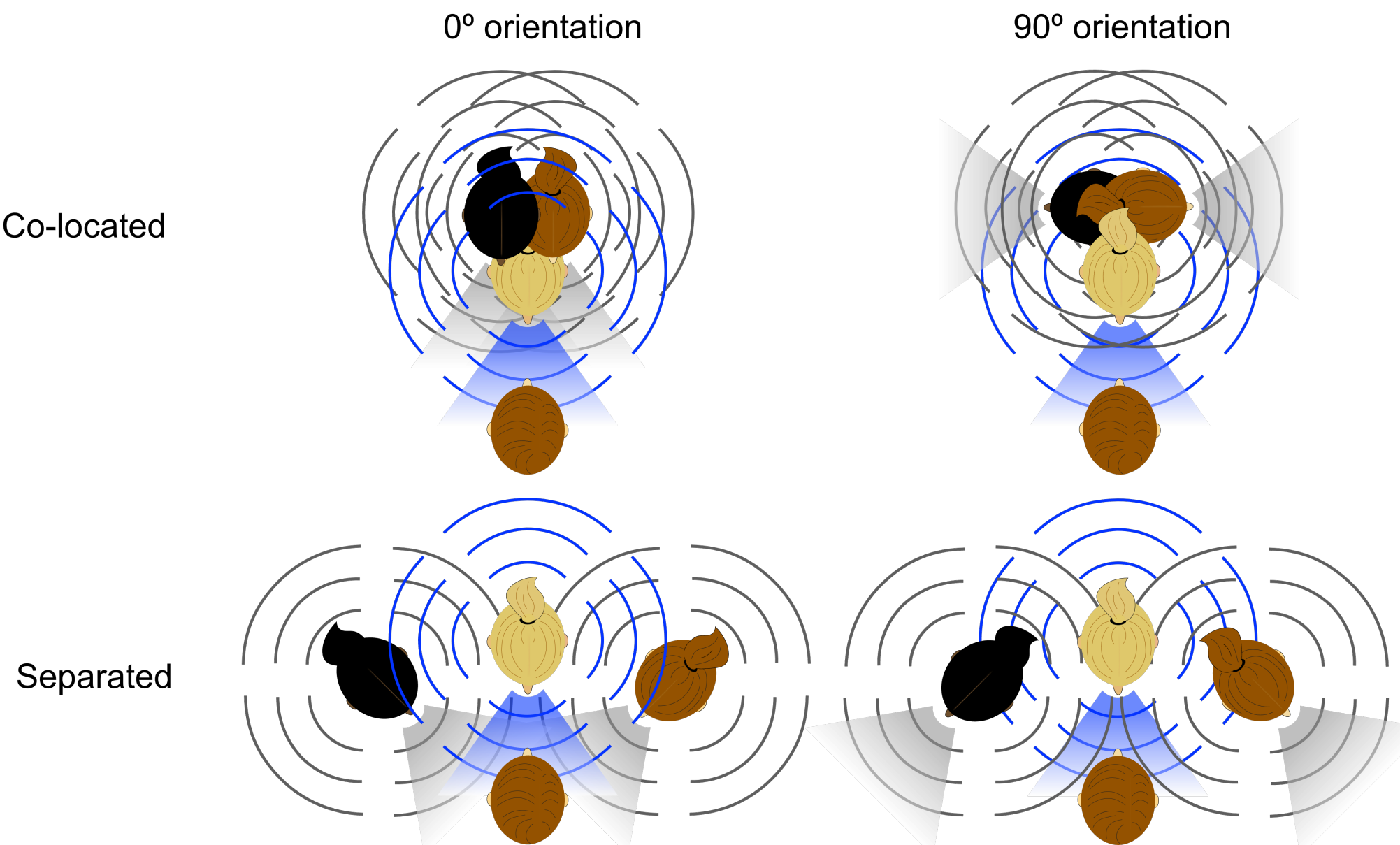
Allison Trine¹, Vahid Delaram¹, Rohit M. Ananthanarayana¹, Margaret K. Miller², G. Christopher Stecker², Emily Buss³, Brian B. Monson¹

1 Department of Speech and Hearing Science, College of Applied Health Sciences, University of Illinois Urbana-Champaign
2 Boys Town National Research Hospital, Omaha, Nebraska
3 Department of Otolaryngology/HNS, University of North Carolina at Chapel Hill



Introduction

Extended high frequencies (EHFs; >8 kHz) in speech provide beneficial cues for masked speech recognition, particularly when target talker EHF levels remain unmasked.¹ Due to the directionality of EHF levels in speech, there is less masking of EHF levels when masker talkers are facing away from the listener while the target talker faces the listener. This replicates a real-world scenario (see *90° orientation, Separated* below), where maskers are often rotated away from the listener, facing their own conversational partners. These effects have been demonstrated with female speech.² However, male speech has lower EHF levels than female speech. It is unclear whether EHF speech cues are audible and beneficial for male speech-in-speech recognition.



Background

- Previously, we compared the talker head-orientation related (THOR) benefit of non-facing masker talkers and talker spatial separation (spatial release from masking; SRM) for speech recognition with female speech.²
- Results indicated that spatial cues provided a greater benefit than head orientation cues, and that the EHF benefit was diminished when talkers were spatially separated. SRM was larger than the THOR benefit, in contrast to a previous study.³
- The largest EHF benefit occurred with spatially co-located talkers, with maskers facing away from the listener.
- SRTs were correlated with average-ear 16-kHz threshold in the spatially co-located, facing masker condition. The 16-kHz thresholds were also significantly correlated with age.

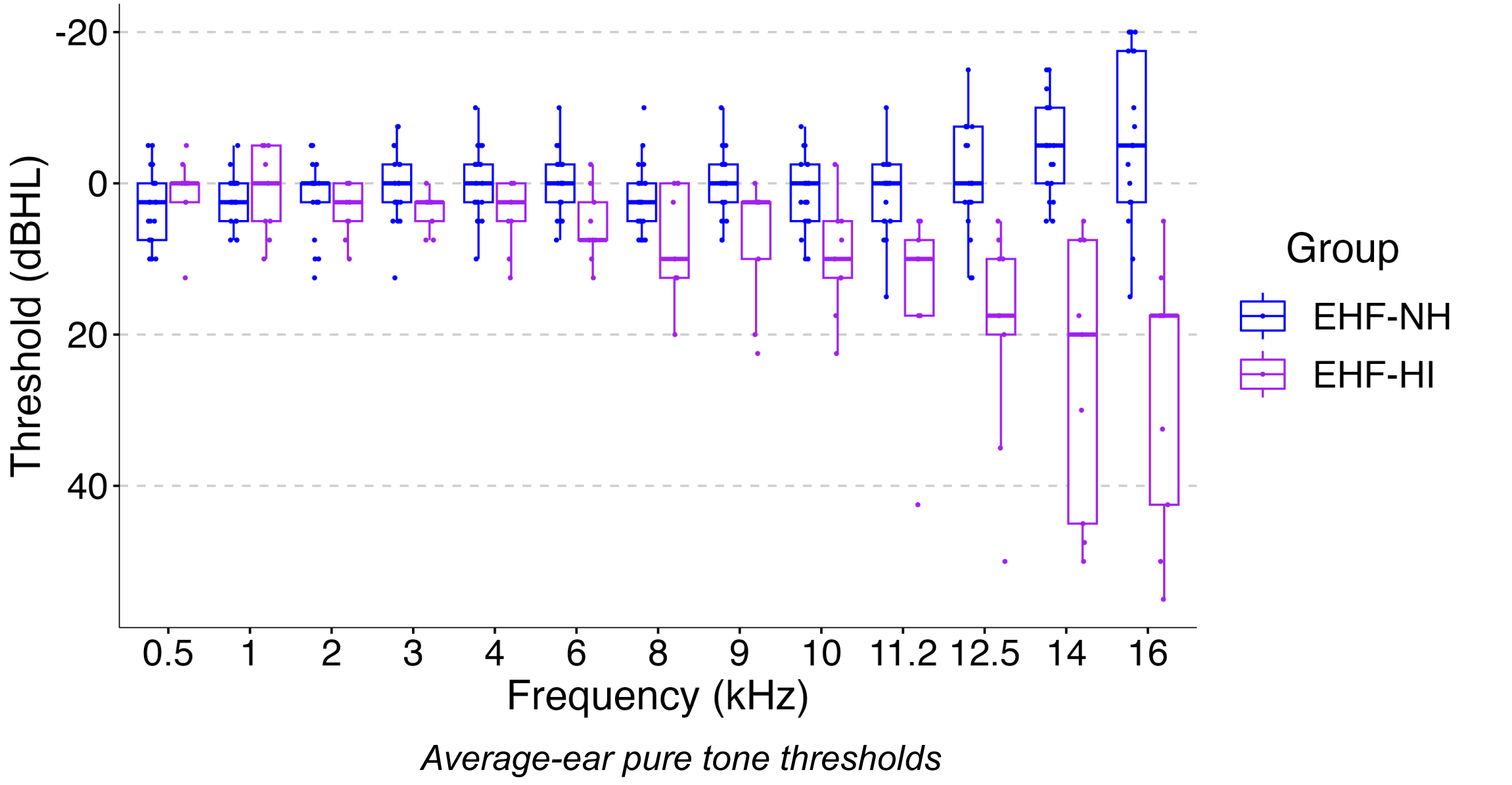
Current study

- EHF benefit was measured by comparing performance with full-band (20 kHz) speech vs. speech low-pass filtered at 8 kHz.
- We hypothesized that there would be an EHF benefit for speech-in-speech recognition for male speech.
- THOR benefit was measured by comparing performance with facing vs. non-facing maskers
- We hypothesized that there would be a THOR benefit for speech-in-speech recognition for male speech.

Methods

A. Participants

- 30 native American English speakers (21 F, 7 M, 2 Other), age 18-47 yr. (mean 23.5 yr.) with clinically normal hearing.
- 21 participants had thresholds <25 dB HL in both ears from 0.5-8 kHz and at EHF levels (9-16 kHz; **EHF-NH group**).
- 9 participants had thresholds <25 dB HL in both ears from 0.5-8 kHz but at least one elevated threshold at EHF levels (**EHF-HI group**).



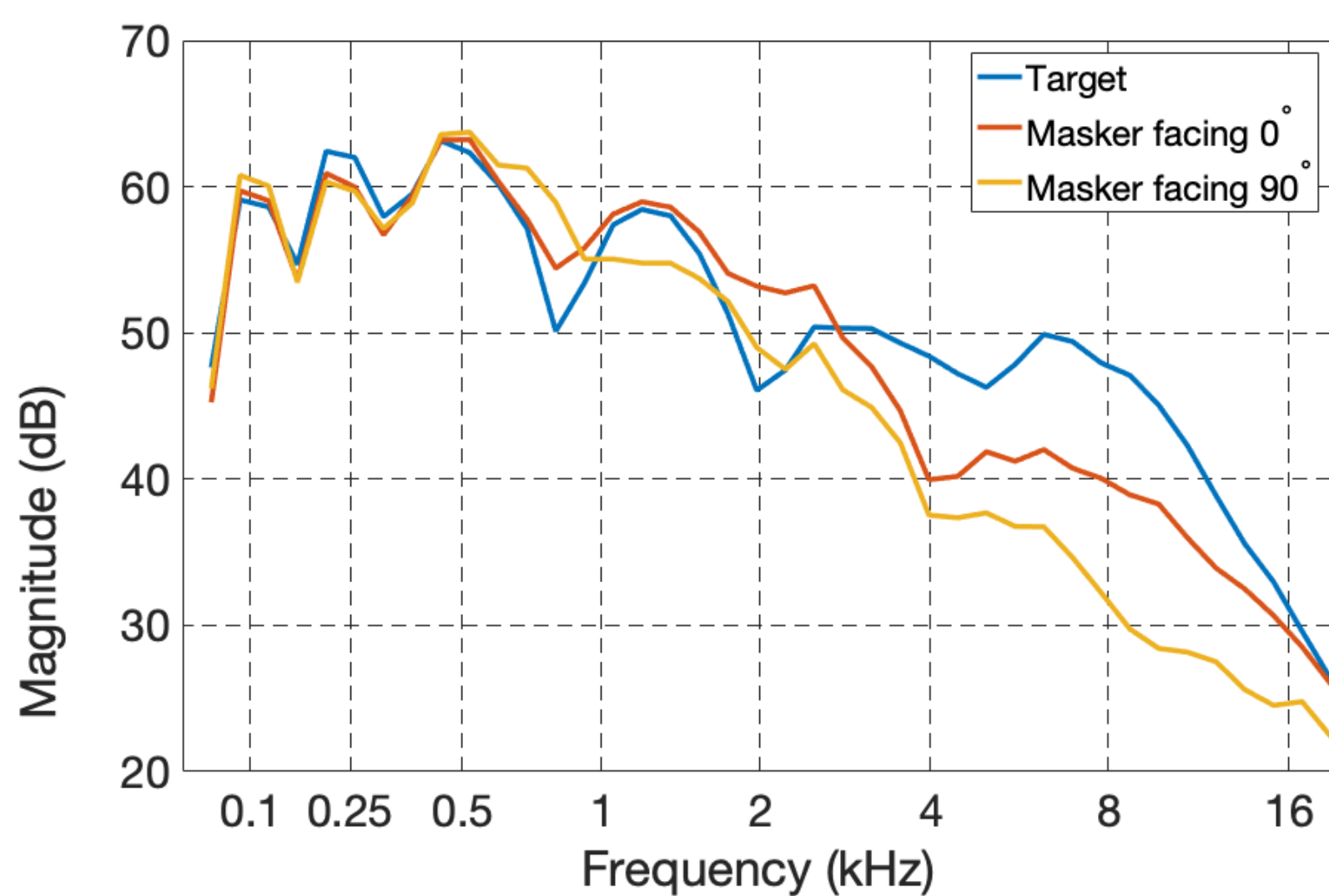
B. Stimuli

- Stimuli came from our publicly available corpus of anechoic recordings.
- Target speech: BKB sentences, male talker
- Masker speech: narratives, two male talkers



C. Conditions

- Spatial separation (Sep):
 - Target and masker co-located at 0° azimuth
 - Target at 0°, maskers at ±45° azimuth
- Masker head orientation (HO):
 - Facing the listener (0°)
 - Facing 90° away (non-facing)
- Filtering:
 - Full-band (FB)
 - Low-pass filtered at 8 kHz (LP8k)



ERB-scale long term average speech spectra of the target, facing masker and non-facing masker stimuli in the FB condition. Overall levels are set to 72 dB SPL.

D. Procedure

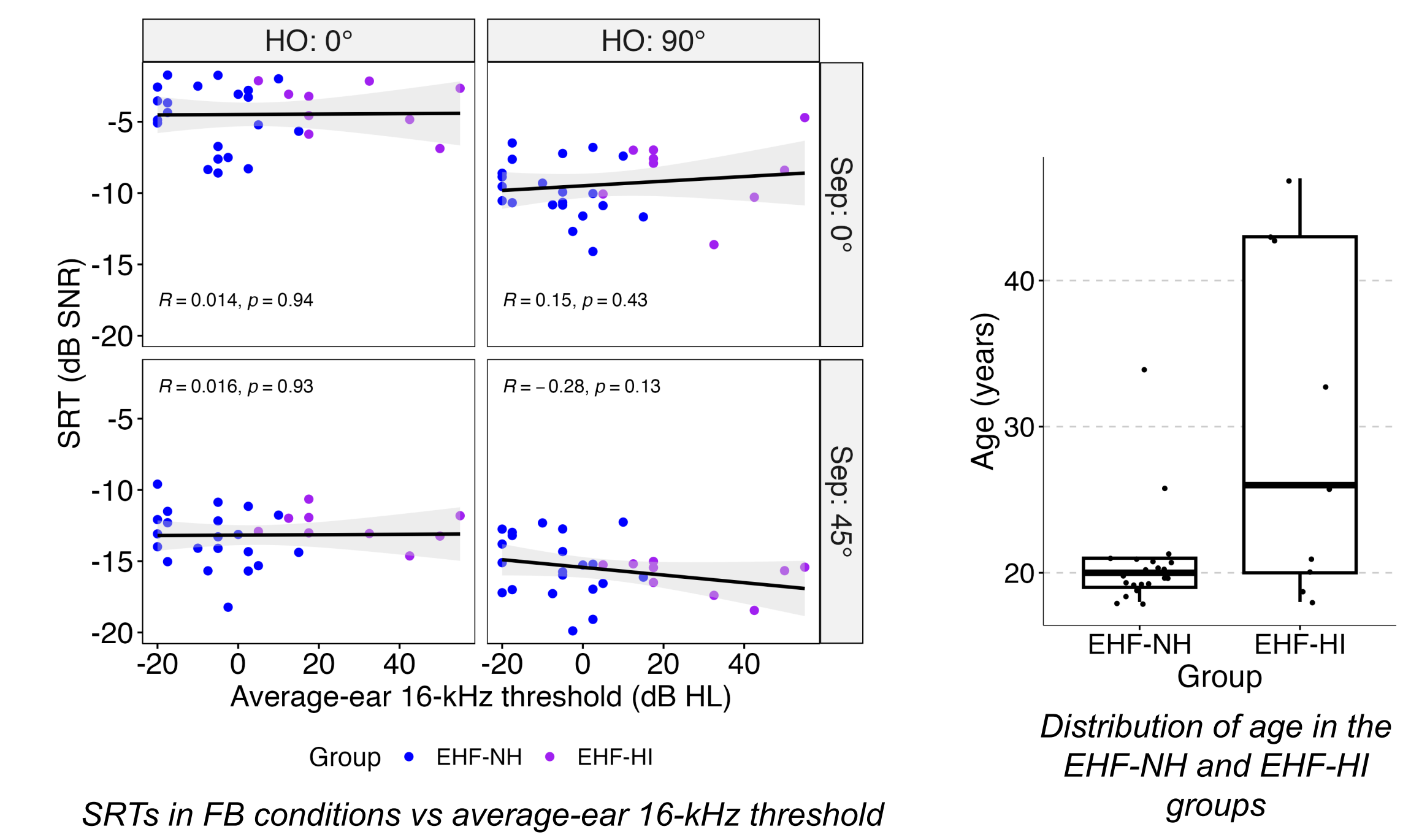
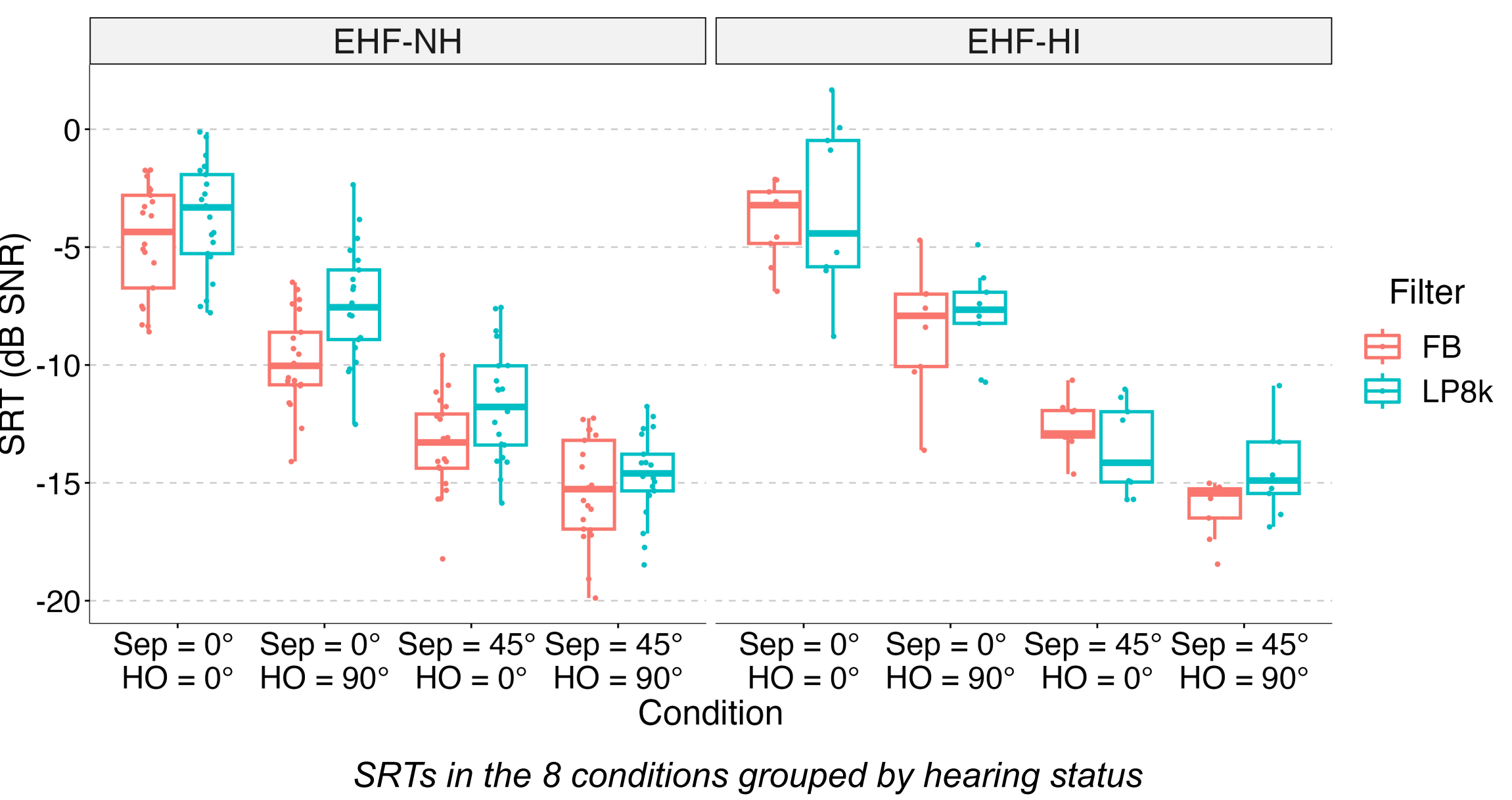
- Stimuli presented using loudspeaker array at 1-m radius.
- Each masker talker level set to 72 dB SPL, target level varied adaptively.
- Training block followed by eight experimental blocks (randomized order) with 32 trials each.

Methods (continued)

E. Analyses

- Speech reception threshold (SRT) – SNR required for 50% correct performance estimated for each condition by fitting psychometric functions.
- A linear mixed-effects model was used to analyze effects of different conditions and group on SRT.

Results



Male Talkers			Female Talkers		
Predictors	Estimates	p	Predictors	Estimates	p
(Intercept)	-9.826	<0.001	(Intercept)	-5.916	<0.001
Filter [LP8k]	2.421	<0.001	Filter [LP8k]	1.403	0.001
Sep [45]	-5.494	<0.001	Sep [45]	-7.848	<0.001
HO [0]	5.101	<0.001	HO [0]	4.204	<0.001
Grp [EHF-HI]	1.319	0.112	Grp [EHF-HI]	0.876	0.205
Filter [LP8k] × Sep [45]	-1.758	0.009	Filter [LP8k] × Sep [45]	-1.237	0.038
Filter [LP8k] × HO [0]	-1.441	0.031	Filter [LP8k] × HO [0]	-1.203	0.049
Sep [45] × HO [0]	-3.196	<0.001	Sep [45] × HO [0]	-2.845	<0.001
Filter [LP8k] × Grp [EHF-HI]	-1.770	0.040	Filter [LP8k] × Grp [EHF-HI]	-1.181	0.060
Sep [45] × Grp [EHF-HI]	-2.034	0.019	Sep [45] × Grp [EHF-HI]	-0.343	0.580
HO [0] × Grp [EHF-HI]	-0.527	0.540	HO [0] × Grp [EHF-HI]	-0.500	0.431
(Filter [LP8k] × Sep [45]) × HO [0]	2.523	0.008	(Filter [LP8k] × Sep [45]) × HO [0]	1.248	0.139
(Filter [LP8k] × Sep [45]) × Grp [EHF-HI]	2.605	0.033	(Filter [LP8k] × Sep [45]) × Grp [EHF-HI]	0.757	0.386
(Filter [LP8k] × HO [0]) × Grp [EHF-HI]	1.403	0.249	(Filter [LP8k] × HO [0]) × Grp [EHF-HI]	1.426	0.110
(Sep [45] × HO [0]) × Grp [EHF-HI]	2.079	0.088	(Sep [45] × HO [0]) × Grp [EHF-HI]	0.982	0.264
(Filter [LP8k] × Sep [45] × HO [0]) × Grp [EHF-HI]	-4.975	0.004	(Filter [LP8k] × Sep [45] × HO [0]) × Grp [EHF-HI]	-1.019	0.409

Linear mixed-effects model outputs with EHF hearing status represented by 'Grp'. 'Grp' compares EHF-NH vs EHF-HI. Results for this current study (left) and previous female talker (right) study are included.

Discussion

- For spatially co-located, non-facing maskers, the EHF benefit was 2.4 dB for male speech. The previous EHF benefit for female speech measured for this condition was 1.4 dB.
- This EHF benefit was reduced for spatially separated maskers, in both male and female speech. It was also reduced for facing maskers.
- The THOR benefit (for co-located talkers) was 5.1 dB for male speech. For female speech, the previous THOR benefit measured for this condition was 4.2 dB.
- SRM (for non-facing maskers) was 8.7 dB for male speech. Female speech SRM was previously measured at 10.6 dB.
- For both male and female speech, THOR benefit and SRM were reduced in the presence of the other cue.
- SRTs were not correlated with average-ear 16 kHz threshold in any condition for male speech. Conversely, a significant correlation was present with female speech in the spatially co-located, non-facing maskers.
- Similar to the female speech experiment, EHF-HI individuals were on average 9.4 years older than the EHF-NH (p = 0.002) and average-ear 16 kHz thresholds were significantly correlated with age (r = 0.81, p < 0.001).
- EHF-HI individuals may demonstrate larger SRM (for non-facing maskers) and reduced EHF benefit (spatially co-located, non-facing maskers).
- Comparison of the male and female speech experiments is limited due to a change in masker level (72 vs. 62 dB SPL per masker, respectively). Masker level was increased for the male speech experiment due to concerns that target talker EHF levels approached inaudibility for spatially separated conditions. With a masker level of 72 dB SPL, SRTs of -15 dB SNR result in a target talker level of 57 dB SPL, and target EHF spectral levels of approximately 33 dB SPL.

Acknowledgements

We thank our participating listeners. This study was supported by NIH grant number R01-DC019745.

References

- Monson et al (2019). Ecological cocktail party listening reveals the utility of extended high-frequency hearing. Hearing Research, 381, 107773.
- Ananthanarayana et al (2023). The interaction between talker head orientation, spatial separation, and extended high frequencies for speech-in-speech recognition. Acoustical Society of America 185th Meeting, Sydney, Australia.
- Braza et al (2022). Effect of masker head orientation, listener age, and extended high-frequency sensitivity on speech recognition in spatially separated speech. Ear and Hearing, 43(1), 90-100.