Differential benefits of unmasking extended high-frequency content of target vs. masker speech

Vahid Delaram[1], Rohit M. Ananthanarayana[1], Allison Trine[1], G. Christopher Stecker[2], Emily Buss[3], Brian B. Monson[1]

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

Speech energy at extended high-frequencies (EHFs; >8 kHz) provides acoustic cues useful for speech recognition in complex listening environments. These cues are particularly useful when maskers have attenuated levels of EHF energy relative to the target.

The EHF benefit in multi-talker environments has been demonstrated using natural EHF attenuation that occurs for real-world speech-in-speech (i.e., cocktail party) scenarios in which masker talkers are not facing the listener. This benefit can also be shown by low-pass filtering masker talkers at 8 kHz, leading to the unmasking of EHF content in the target talker.

At least three possible mechanisms for the EHF benefit exist: (1) EHFs provide useful phonetic information for the target; (2) EHFs provide segregation cues for target/masker talker separation; or (3) EHFs provide grouping cues that enhance a listener's ability to access target low-frequency information (e.g., temporal coherence). In this study, we examined possibility #2 by creating a substantial target-masker mismatch in EHF content.

The EHF benefit could also depend on the number of masker talkers in the speech-in-speech perception task. In contrast to one-talker speech, two-talker speech might be more effective as a masker because of increased informational masking and the limited number of opportunities to glimpse target speech.

In the present study, we tested speech recognition when co-located target and one- or two-talker maskers faced the listener, using four filtering combinations of full-band (FB; low-pass filtered at 20 kHz) and low-pass filtered (LP8k; low-pass filtered at 8 kHz) speech:

(1) FB target, LP8k masker (filter mismatched)

(2) FB target, FB masker (filter matched)

(3) LP8k target, FB masker (filter mismatched)

(4) LP8k target, LP8k masker (filter matched)

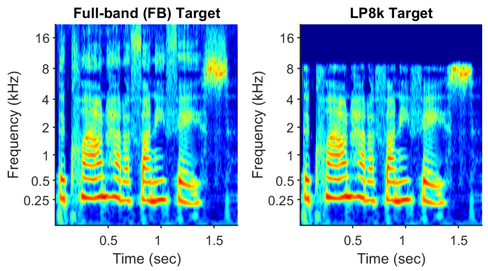

***FIG. 1**. Cochleogram of the sentence "The clown had a funny face," showing the spectral content of the FB target and the LP8k target.*

Aims

- To investigate the benefit derived from a mismatch in EHF content between target and masker talkers
- To study the impact of the number of masker talkers on the EHF benefit

Methods

A. Participants

- 39 native English speakers (27 female, 11 male, one other), age 18–26 years (mean 21.2 years)
- Pure tone audiometry conducted for standard audiometric frequencies (0.5-8 kHz) and the extended high frequencies of 9, 10, 11.2, 12.5, 14, and 16 kHz
- Participants' thresholds ≤25 dB HL in at least one ear from 0.5 to 8 kHz

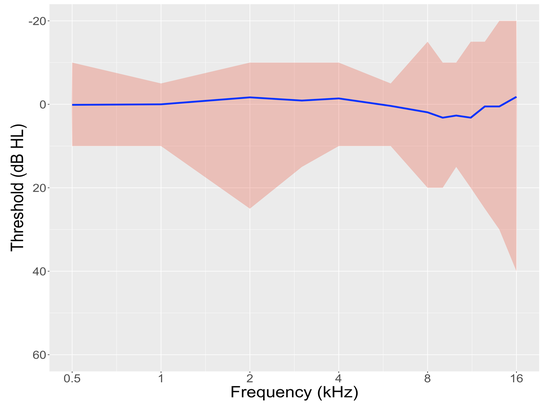

***FIG. 2.** Mean better ear pure tone thresholds. The shaded region depicts maximum and minimum responses across participants.*

B. Stimuli

- Masker was either one- or two-female-talker babble
- Target speech was the Bamford-Kowal-Bench (BKB) sentences spoken by a female talker
- Target and masker speech, independently low-pass filtered at either 20 kHz (FB) or 8 kHz (LP8k), resulting in four filtering conditions

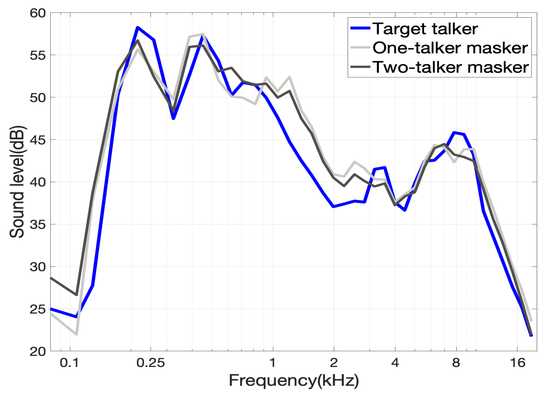

***FIG. 3.** Long term average speech spectrum (LTASS) of target talker, one- and two-talker maskers in FB condition*

Methods (continued)

C. Procedure

- Stimuli presented using a single loudspeaker in front of the listener at a 1m distance
- Masker level set to 65 dB SPL
- Target amplitude varied depending on the participant responses using two interleaved adaptive tracks
- Following a training block, the eight conditions (four filtering conditions × two masker conditions) tested in separate blocks
- Speech reception thresholds (SRTs), defined as the SNR required to produce 50% correct performance, estimated for each condition
- Linear mixed effects (LME) and regression models used to analyze the results
- Exploratory analysis used to investigate the association between the EHF content of each target word and the probability of getting that word correct

Results

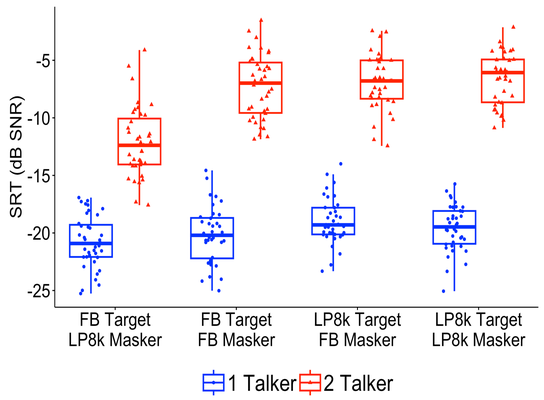

***FIG. 4**. Distributions of SRT values in dB SNR for each of the four conditions*

- Mean SRTs were consistently lower for the one-talker than two-talker masker
- Mean SRTs were similar for the one-talker masker but differed by up to 5.5 dB for the two-talker masker across four conditions
- Mismatch in EHFs only improved performance for the FB Target conditions, with larger effects for the two-talker masker
- EHF sensitivity was characterized as the better ear 16 kHz thresholds. No correlation between EHF benefit and 16kHz thresholds (R=0.0065, p=0.97)
- Results from a word-by-word analysis indicated higher recognition odds with increasing EHF energy level in the target word, particularly when the masker was LP8k (OR=1.03, p<0.001)

Results (continued)

	Estimate	SE	t-value	p
(Intercept)	-19.54	0.39	-49.79	<0.001
Mismatch	0.62	0.44	1.41	0.160
Two-talker Masker	12.89	0.44	29.17	<0.001
FB Target	-0.55	0.45	-1.24	0.220
Mismatch x Two-talker Masker	-0.82	0.65	-1.30	0.190
Mismatch x FB Target	-1.30	0.63	-2.08	0.038
Two-talker Masker x FB Target	-0.02	0.63	-0.03	0.970
Mismatch x Two-talker Masker x FB Target	-3.41	0.88	-3.86	<0.001

***Table 1:** Linear mixed effects model results*

Conclusions

- A large mismatch in EHF content improved speech recognition performance only when the target contained EHFs, especially for the two-talker masker
- Along with the word-level analysis, this finding could indicate that listeners take advantage of reduced energetic masking in the EHF band to make use of phonetic cues in EHFs
- Alternatively, target EHFs could provide cues that support segregation and selective attention to the target by preserving the natural temporal coherence of low and high-frequency bands
- Masker EHF energy does not confer any segregation benefit

References

- Monson, B. B., Rock, J., Schulz, A., Hoffman, E., and Buss, E. (2019). "Ecological cocktail party listening reveals the utility of extended high-frequency hearing," Hear. Res. 381, 107773.
- Motlagh Zadeh, L., Silbert, N. H., Sternasty, K., Swanepoel, W., Hunter, L. L., and Moore, D. R. (2019). "Extended high-frequency hearing enhances speech perception in noise," Proc. Natl. Acad. Sci. U. S. A. 116(47), 23753–23759.
- Hunter, L. L., Monson, B. B., Moore, D. R., Dhar, S., Wright, B. A., Munro, K. J., Zadeh, L. M., Blankenship, C. M., Stiepan, S. M., and Siegel, J. H. (2020). "Extended high frequency hearing and speech perception implications in adults and children," Hear. Res. 397, 107922.
- Trine, A. and Monson, B. B. (2020). "Extended high frequencies provide both spectral and temporal information to improve speech-in-speech recognition," Trends Hear. 24, 2331216520980299.

Acknowledgements

This study was supported by NIH grant number R01-DC019745.

ILLINOIS