Gender and speech material effects on extended high-frequency levels in the speech spectrum

Vahid Delaram[1], Rohit M. Ananthanarayana[1], Margaret K. Miller[2], 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 has audible acoustic energy at frequencies above 8 kHz. This extended high-frequency (EHF; >8 kHz) portion of the speech spectrum has been examined far less than lower frequencies, perhaps due to technological constraints and a longstanding belief that EHFs have a negligible contribution to speech perception.

Recent studies have revealed that EHFs play a role in speech perception. EHF cues in speech can contribute to speech recognition in noise, speech-in-speech recognition, talker head orientation identification, and speech localization in the vertical plane.

Audibility of EHF cues likely depends on speech spectral levels at EHFs, which may vary across talkers and different speech materials. Past studies have examined the long-term average speech spectrum (LTASS), but the reported levels for EHF bands have been inconsistent across studies, and factors that influence EHF levels have not been examined.

This study investigated the effects of gender and different speech materials on EHF levels for American English using a corpus of high-fidelity speech recordings.

Based on prior studies, we hypothesized EHF levels for female speech would be higher than for male speech. We also hypothesized that digits would exhibit higher EHF levels compared to BKB sentences due to a high concentration of voiceless fricatives in English digits.

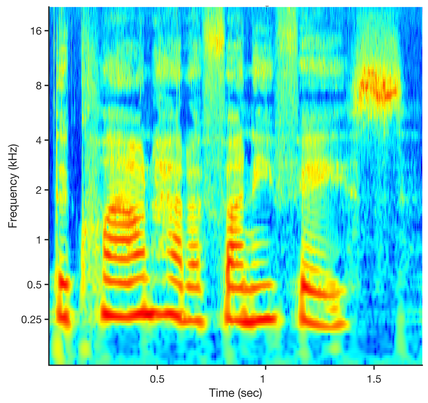

***FIG. 1**. Cochleogram of the sentence "The clown had a funny face," showing energy at extended high frequencies.*

Aims

- To create high-fidelity recordings of commonly used American English speech materials
- To investigate the effect of talker gender and different speech materials on EHF levels of speech

Methods

A. Subjects

- 30 subjects (15 female)
- Age range of 21.3 - 60.5 years old (mean 33.6)
- Native speakers of American English

B. Recordings

- Speech recordings were made in an anechoic chamber, with acoustical noise floor < 15 dB SPL at Boys Town National Research Hospital
- Class 1 precision microphone located 1m in front of talker mouth (0º)
- 48-kHz sampling rate and 24 bits/sample
- Speech material:
  - Unscripted narrative speech (~2.5 min)
  - Lists 1-4 of BKB sentences
  - Digits 0-10

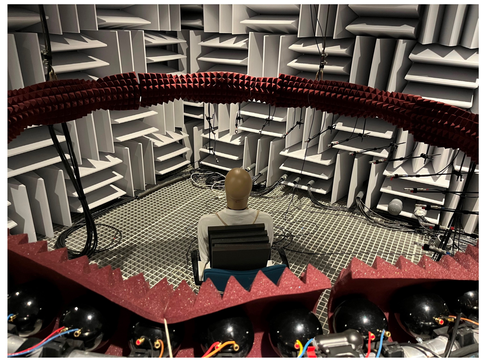

***FIG. 2.** Recording setup at Boys Town National Research Hospital anechoic chamber.*

C. Acoustical analysis

- Long-term average speech spectrum (LTASS) calculated using 2048-point FFT, resulting in a frame length of ~43 ms, with a Hamming window and 50% overlap
- LTASS converted to ERB scale using 1-ERB-wide analysis bands

D. Statistical analysis

- Narratives were excluded from statistical analysis due to phonetic variability across subjects' narratives
- Linear mixed effects models tested effects of gender, speech material (BKBs vs digits), and their interaction on EHF band levels, using subjects as random factor
- Pearson correlations were used to investigate the relationship between EHF levels across different speech materials

Results

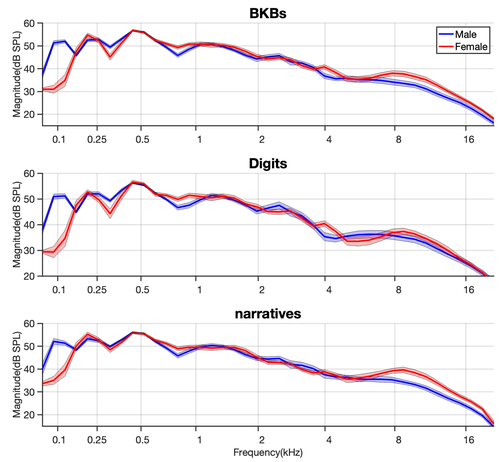

***FIG. 3**. LTASS of male (blue) and female (red) speech. Shaded region shows 95% confidence intervals.*

- Females had ~4 dB higher levels at EHFs compared to males for BKBs; this gap decreased to ~2 dB for digits
- EHF levels for female speech were ~18-20 dB lower than peak energy (at ~500 Hz)

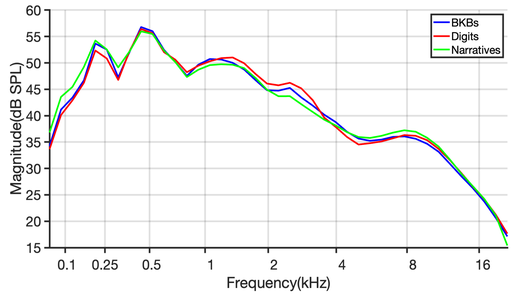

***FIG. 4.** Mean LTASS for all subjects for three sets of speech materials.*

	Estimate	SE	t-value	p
(Intercept)	37.71	0.89	41.93	<0.001
Speech material	-0.36	0.63	-0.56	0.57
Gender	-4.18	1.27	-3.28	**0.002**
Speech material x Gender	1.85	0.90	2.05	**0.04**

***Table 1:** Linear mixed effects model results for BKB and digit recordings.*

Results (continued)

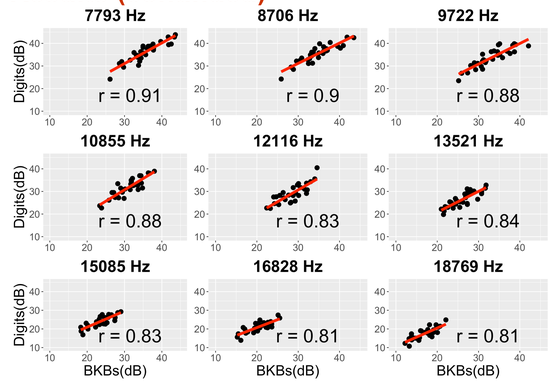

***FIG. 5**. Pearson correlations for EHF bands (7.8 -18.8 kHz) for BKBs vs. digits.*

- Similarly high correlations were observed in BKBs vs. narrative comparison

Conclusions

- Gender affects EHF levels in the LTASS, and this should be considered in studies that aim to investigate the role of EHFs in speech perception
- Although there was no main effect of speech material, there was an interaction between speech material and gender, suggesting speech material does influence gender differences at EHFs
- EHF levels are relatively stable across speech materials for a given talker
- Since EHF cues are useful for speech-in-noise recognition, it may be that higher EHF levels in female speech result in better audibility of EHF cues for female speech in complex auditory situations

References

- Monson, B. B., & Caravello, J. (2019). The maximum audible low-pass cutoff frequency for speech. The Journal of the Acoustical Society of America, 146(6), EL496–EL501. https://doi.org/10.1121/1.5140032
- Hunter, L. L., et al.,(2020) Extended high frequency hearing and speech perception implications in adults and children. Hearing Research, 397, 107922. https://doi.org/10.1016/j.heares.2020.107922

Acknowledgments

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

Scan to visit and download the corpus: