Gender and speech material effects on full-band speech LTASS

Vahid Delaram[1], Margaret K. Miller[2], Rohit M. Ananthanarayana[1], Allison Trine[1], Emily Buss[3], G. Christopher Stecker[2], 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

- Extended high-frequency (EHF; >8 kHz) cues play a role in speech recognition, talker head-orientation discrimination, speech localization, speech clarity, and word learning.
- The utility of these EHF cues in speech depends on the audibility of EHFs in speech, which in turn depends on speech spectral levels at EHFs.
- Gender and speech materials across the different languages are reported to influence speech spectral levels; however, accurate investigation is needed to analyze the effects of these factors on EHFs.
- Some prior data on gender effects were based on long-term average speech spectra (LTASS) up to 12.5 or 16 kHz, missing some portions of the EHF spectrum. [1,2] Others that used full-band, up to 20 kHz recordings did not test gender effects [3] or had a limited subject population.[4]
- Speech material effects on the EHF levels have been reported for different languages,[2] but this effect was not investigated for different speech materials within the same language.

Study	Bandwidth (kHz)	Talkers	Mic setup
Cox and Moore[1]	12.5	30 F, 30 M	30 cm at 0°
Byrne et al.[2]	16	Not specified	25 cm at 45°
Moore et al.[3]	20	8 F, 9 M	30 cm at 0°
Monson et al.[4]	20	8 F, 7 M	60 cm at 0°

Table1: Summary of methods of studies that have investigated the effects of gender and/or speech material on LTASS.

Aim

- In this study, we investigated the effects of gender and different speech materials on EHF levels for American English using high-fidelity recordings of BKB sentences, digits, and unscripted narratives.
- Based on past 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 and narratives due to a high concentration of voiceless fricatives in English digits.

Methods

A. Subjects

- 30 subjects (15 female, 15 male)
- Age range of 21.3 - 60.5 yr (mean 33.6)
- Native speakers of American English

B. Recordings

- Speech material from our corpus of high-fidelity multi-directional recordings
- Class 1 precision microphone located 1m in front of talker's mouth (0°)
- 48-kHz sampling rate and 24 bits/sample
- Speech material:
  - Unscripted narrative speech (~2.5 min)
  - Digits 0-10
  - Lists 1-4 of BKB sentences

Access to a subset of the recordings here

C. Acoustical analysis

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

D. Statistical analysis

- Linear mixed-effects models tested the effects of gender, speech material, and their interaction on EHF band levels, using subjects as a random factor
- Student t-tests were used to test gender differences in previously reported low-frequencies

Results

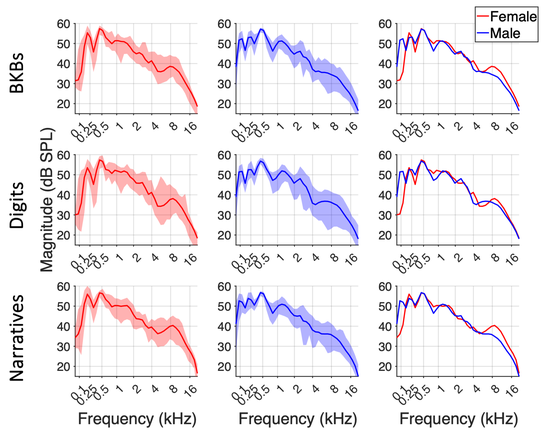

***FIG. 1**. ERB-scaled mean LTASS of female (red) and male (blue). The shaded regions show range.*

Results cont.

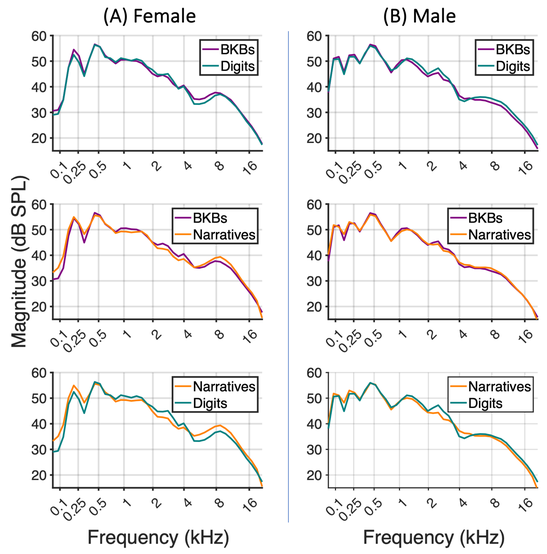

***FIG. 2.** Pairwise comparison of mean LTASS for speech materials for female (A) and male (B) talkers on the ERB-scale.*

Term	Estimate	S.E.	t-value	p-value
(Intercept)	38.75	0.8	48.53	<0.001 *
Band	-2.36	0.06	-37.41	<0.001 *
Male	-4.41	1.11	-3.98	<0.001 *
Digits	-0.31	0.37	-0.83	0.4
Narratives	1.78	0.38	4.74	<0.001 *
Band * Male	0.35	0.06	5.56	<0.001 *
Band * Digits	-0.02	0.08	-0.23	0.82
Band * Narratives	-0.14	0.08	-1.75	0.08
Male * Digits	1.85	0.35	5.24	<0.001 *
Male * Narratives	-0.73	0.36	-2	0.04 *

***Table 1:** Linear mixed-effects model results for EHF region. The intercept is female BKBs for ERB band 34 (centered at 8.7 kHz).*

- Females had higher levels than males by approximately 4 dB at EHFs and 2-6 dB at ERBs with center frequencies of 313, 806, and 924 Hz.
- Male levels were higher than females at ERBs with center frequencies ≤ 123 Hz

Results cont.

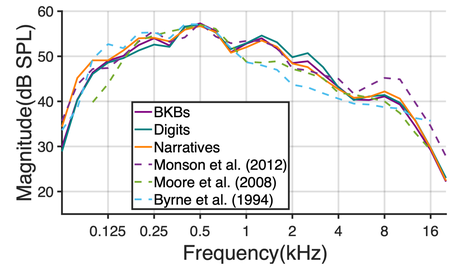

***FIG. 3.** Mean LTASS from present and previous studies plotted in third-octave bands on the log scale.*

Conclusion

- LTASS levels at EHFs are influenced mainly by gender rather than the phonetic content of speech and therefore, EHF cues may be more audible for female speech than male speech.
- EHF levels differ markedly across talkers within a gender, differing by up to 15 dB.
- Gender differences observed in this study may be related to anatomical differences in vocal tract length, size, and differences in fundamental frequency.
- Effects of speech material are potentially related to the representation of voiceless fricatives that tend to have spectral peaks at higher frequencies.
- Differences observed across studies could be due to factors including (1) differences in speech materials used, (2) recording setup and quality of recordings, and (3) high variability of the EHF levels across talkers.

References

[1] Cox, R. M., and Moore, J. N. (1988). "Composite Speech Spectrum for Hearing Aid Gain Prescriptions," J. Speech Lang. Hear. Res., 31, 102–107. doi:10.1044/jshr.3101.102
[2] Byrne, D., et al. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am., **96**, 2108–2120. doi:10.1121/1.410152
[3] Moore, B. C. J., et al. (**2008**). "Spectro-Temporal Characteristics of Speech at High Frequencies, and the Potential for Restoration of Audibility to People with Mild-to-Moderate Hearing Loss," Ear Hear., **29**, 907–922. doi:10.1097/AUD.0b013e31818246f6
[4] Monson, B. B., et al. (**2012**). "Analysis of high-frequency energy in long-term average spectra of singing, speech, and voiceless fricatives," J. Acoustic. Soc. Am., **132**, 1754–1764. doi:10.1121/1.4742724

Acknowledgments

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