DOI: 10.61782/fa.2023.0269

THE ECOLOGICAL UTILITY OF EXTENDED HIGH FREQUENCIES FOR SPEECH RECOGNITION

Brian B. Monson[1,2,3*]
[1] Department of Speech and Hearing Science, College of Applied Health Sciences
[2] Department of Biomedical and Translational Sciences, Carle Illinois College of Medicine
[3] Neuroscience Program
University of Illinois Urbana-Champaign, USA

ABSTRACT

Recent studies have demonstrated that extended high frequencies (EHFs; >8 kHz) in speech are audible and useful for speech recognition. In this paper I review recent and ongoing work in our lab examining audibility of EHF cues in speech and the conditions under which these cues are useful. Mismatches in head orientations between a target talker (facing the listener) and masker talkers (not facing the listener) influence the utility of EHFs. For example, EHFs provide a speech recognition benefit when masker talkers are facing away from the listener. Spatial separation of target and maskers may also influence this relationship. Our data indicate that EHF cues in speech are useful in real-world auditory scenes, suggesting the loss of EHF hearing, which typically begins in early adulthood, could contribute to speech-in-noise difficulties.

Keywords: *extended high frequency, speech perception, speech recognition*

1. INTRODUCTION

Extended high-frequencies (EHFs; >8 kHz) in speech have largely been unstudied until recently, likely due to early studies that concluded frequencies >7 kHz provide negligible benefit for speech intelligibility [1]. However, these early studies, driven by the desire to improve telecommunication, may not have considered other ecologically relevant scenarios in which EHFs could be beneficial for speech perception. The utility of EHF cues for speech perception has again been the topic of a number of recent studies [2-7]. Notably, EHFs in speech are audible [8-9], and audible EHF cues have been found to contribute to the detection and perception of speech, leading to improvements in: speech recognition for adults [2-6] and children [7]; speech localization [10]; discrimination of talker head orientation [2]; and speech and voice quality [9, 11].

*__Corresponding author__: monson@illinois.edu

2. METHODS

Investigations in our lab have included different experimental methods that have been published. To determine the benefit of EHFs, several of our studies have compared listener performance for perceptual tasks using full-band speech to performance using speech low-pass filtered at 8 kHz. It should be emphasized that critical methodology for examining EHFs includes using full-band speech signals that have been recorded on axis (i.e., directly in front of a talker) at a sampling rate of 44.1 kHz or higher, with microphones that have a flat frequency response to 20 kHz. Recording directly on axis is important because EHFs are highly directional toward the front of a talker [12-13], and common practices of recording slightly off axis or at other locations (e.g., using a lavalier microphone pinned to the chest) can disproportionately affect the EHF spectral levels recorded. For our studies, we have typically acquired and used anechoic speech recordings.

Transducers are important to consider for presentation of speech, as well. Headphone presentation can be especially problematic as most headphones deviate from a flat response at EHFs. For sound-field presentation, loudspeakers with a flat frequency response to 20 kHz are necessary, but are relatively easy to acquire.

Other aspects of the experimental setup must also be carefully considered, including the presence of reflective surfaces in the recording environment or listening environment. Given that wavelengths for sounds between 8 and 16 kHz range between ~2-4 cm, any surface or object with dimensions larger than this has the potential to affect EHF levels received at a microphone during recording, or at the ear the of the listener during stimulus presentation.

3. RESULTS AND DISCUSSION

A number of findings have resulted from our studies. We found that listeners demonstrated marked sensitivity to EHFs in speech. For example, the maximum audible low-pass filter cutoff frequency for speech was ~13 kHz for young adult listeners with typical hearing [8], indicating the 13-20-kHz band contains detectable information regarding the speech signal. Young, typical-hearing adult listeners could also detect speech spectral level changes of ~5-10 dB at EHFs [9]. In both studies, pure-tone thresholds at 16 kHz were correlated with audibility of EHFs in speech.

Given listeners' sensitivity to EHFs in speech, and considering that EHFs are highly directional, it seemed reasonable to expect that EHF cues would be useful in determining the head orientation of a talker. That is, the rotating of the head results in reduced EHF levels for the speech, providing a cue for detecting talker head orientation. Indeed, we found that the minimum audible change in head orientation (MACHO) was ~41º for full-band speech, whereas this sensitivity dropped to ~55º when speech signals were low-pass filtered at 8 kHz [2].

EHFs also contributed to speech recognition. Using only frequencies >6 kHz (i.e., high-pass filtered speech), adult listeners were able to discriminate consonants and vowels significantly better than chance [14]. For full-band speech, EHFs provided a speech recognition benefit when masker talkers were facing away from the listener for both adult listeners [2, 4] and for children [7]. This is likely because the rotating of the masker talkers' heads leads to reduced EHF levels, unmasking EHFs in the target speech. Interestingly, pure-tone thresholds at 16 kHz were correlated with full-band speech-in-speech recognition when masker talkers faced away from the listener, but not when masker talkers faced the listener, and this was true whether target and masker were co-located or were spatially separated [15].

4. CONCLUSIONS

In sum, EHFs in speech are audible and provide useful information for speech detection and recognition. We are continuing to investigate how the utility of EHF cues changes in real-world scenes that include mismatches in head orientations between a target talker and masker talkers, mismatches in spatial location, and reverberation.

5. ACKNOWLEDGMENTS

This work was supported by NIH Grant R01 DC019745.

6. REFERENCES

[1] B. B. Monson, E. J. Hunter, A. J. Lotto, and B. H. Story, "The perceptual significance of high-frequency energy in the human voice," *Frontiers in Psychology*, vol. 5, no. 587, pp. 1-11, 2014.

[2] B. B. Monson, J. Rock, A. Schulz, E. Hoffman, and E. Buss, "Ecological cocktail party listening reveals the utility of extended high-frequency hearing," *Hearing Research*, vol. 381, no. 107773, 2019.

[3] L. Motlagh Zadeh, N. H. Silbert, K. Sternasty, D. W. Swanepoel, L. L. Hunter, and D. R. Moore, "Extended high-frequency hearing enhances speech perception in noise," *Proceedings of the National Academy of Sciences*, vol. 116, no. 47, pp. 23753-23759, 2019.

[4] A. Trine and B. B. Monson, "Extended high frequencies provide both spectral and temporal information to improve speech-in-speech recognition" *Trends in Hearing*, vol. 24, no. 2331216520980299, 2020.

[5] L. L. Hunter, B. B. Monson, D. R. Moore, S. Dhar, B. A. Wright, K. J. Munro, L. Motlagh Zadeh, C. M. Blankenship, S. M. Stiepan, and J. H. Siegel "Extended high frequency hearing and speech perception implications in adults and children," *Hearing Research*, vol. 397, no. 107922, 2020.

[6] S. Polspoel, S. E. Kramer, B. van Dijk, and C. Smits "The importance of extended high-frequency speech information in the recognition of digits, words, and

sentences in quiet and noise," *Ear and Hearing*, vol. 43, no. 3, pp. 913-920, 2022.

[7] M. Flaherty, K. Libert, and B. B. Monson, "Extended high-frequency hearing and head orientation cues benefit children during speech-in-speech recognition," *Hearing Research*, vol. 406, no. 108230, 2021.

[8] B. B. Monson and J. Caravello, "The maximum audible low-pass cutoff frequency for speech," *The Journal of the Acoustical Society of America*, vol. 146, no. 6, pp. EL496–501, 2019.

[9] B. B. Monson, A. J. Lotto, and B. H. Story, "Detection of high-frequency energy level changes in speech and singing," *The Journal of the Acoustical Society of America*, vol. 135, no. 1, pp. 400-406, 2014.

[10] V. Best, S. Carlile, C. Jin, and A. van Schaik, "The role of high frequencies in speech localization," *The Journal of the Acoustical Society of America*, vol. 116, no. 1, pp. 353-363, 2005.

[11] B. C. J. Moore and C. T. Tan, "Perceived naturalness of spectrally distorted speech and music," *The Journal of the Acoustical Society of America*, vol. 114, no. 1, pp. 408-419, 2003.

[12] B. B. Monson, E. J. Hunter, and B. H. Story, "Horizontal directivity of low- and high-frequency energy in speech and singing," *The Journal of the Acoustical Society of America*, vol. 132, no. 1, pp. 433–441, 2012.

[13] P. Kocon and B. B. Monson, "Horizontal directivity patterns differ between vowels extracted from running speech," *The Journal of the Acoustical Society of America*, vol. 144, no. 1, pp. EL7–EL12, 2018.

[14] A. D. Vitela, B. B. Monson, and A. J. Lotto "Phoneme categorization relying solely on high-frequency energy," *The Journal of the Acoustical Society of America*, vol. 137, no. 1, pp. EL65–EL70, 2015.

[15] M. D. Braza, N. E. Corbin, E. Buss, and B. B. Monson, "Effect of Masker Head Orientation, Listener Age, and Extended High-Frequency Sensitivity on Speech Recognition in Spatially Separated Speech," *Ear & Hearing*, vol. 43, no. 1, pp. 90–100, 2022.