

Extended High Frequencies Provide Both Spectral and Temporal Information to Improve Speech-in-Speech Recognition

Allison Trine¹ and Brian B. Monson^{1,2} 

Trends in Hearing
Volume 24: 1–8
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2331216520980299
journals.sagepub.com/home/tia


Abstract

Several studies have demonstrated that extended high frequencies (EHFs; >8 kHz) in speech are not only audible but also have some utility for speech recognition, including for speech-in-speech recognition when maskers are facing away from the listener. However, the contribution of EHF spectral versus temporal information to speech recognition is unknown. Here, we show that access to EHF temporal information improved speech-in-speech recognition relative to speech bandlimited at 8 kHz but that additional access to EHF spectral detail provided an additional small but significant benefit. Results suggest that both EHF spectral structure and the temporal envelope contribute to the observed EHF benefit. Speech recognition performance was quite sensitive to masker head orientation, with a rotation of only 15° providing a highly significant benefit. An exploratory analysis indicated that pure-tone thresholds at EHF are better predictors of speech recognition performance than low-frequency pure-tone thresholds.

Keywords

speech perception, speech in noise, head orientation

Received 3 July 2020; Revised 12 November 2020; accepted 19 November 2020

The frequency range of human hearing extends up to approximately 20 kHz for young, healthy listeners. Speech perception research has generally focused on the frequency range below about 6–8 kHz, likely because key phonetic features of speech occur in this range (e.g., vowel formants), and it is therefore understood to have the greatest influence on speech perception. The prevailing viewpoint has been that extended high frequencies (EHFs; >8 kHz) provide little information useful for speech perception. Thus, whereas the audibility of speech frequencies below 8 kHz and corresponding effects on speech perception have been studied extensively over the past several decades (e.g., Ching et al., 1998; McCreery & Stelmachowicz, 2013), the audibility of higher frequency bands and corresponding effects have been studied far less (Monson et al., 2014a).

The EHF range in speech is audible and has some utility for speech perception (Hunter et al., 2020). For example, the average young, normal-hearing listener can detect the absence of speech energy beyond approximately 13 kHz, although listeners with better 16-kHz pure-tone thresholds can detect losses at even higher

frequencies (Monson & Caravello, 2019). It has also been demonstrated that EHF audibility contributes to speech localization (Best et al., 2005), speech quality (Monson et al., 2014b; Moore & Tan, 2003), talker head orientation discrimination (Monson et al., 2019), and speech recognition in the presence of background speech (Monson et al., 2019) and noise (Motlagh Zadeh et al., 2019).

In a previous study, we showed that access to EHF in speech supported speech-in-speech listening when the target talker was facing the listener while colocated maskers were facing away from the listener (Monson et al., 2019). This listening scenario departs from the traditional design but reflects a more realistic ‘cocktail

¹Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, Champaign, United States

²Neuroscience Program, University of Illinois at Urbana-Champaign, Champaign, United States

Corresponding Author:

Brian B. Monson, 901 S Sixth St, Champaign, IL 61820, United States.
Email: monson@illinois.edu



party' listening environment where the talker of interest is typically facing the listener and background talkers are typically facing other directions (Figure 1). Our approach followed from the hypothesis that rotated maskers result in less masking at the highest frequencies. This is because directivity patterns of speech radiation

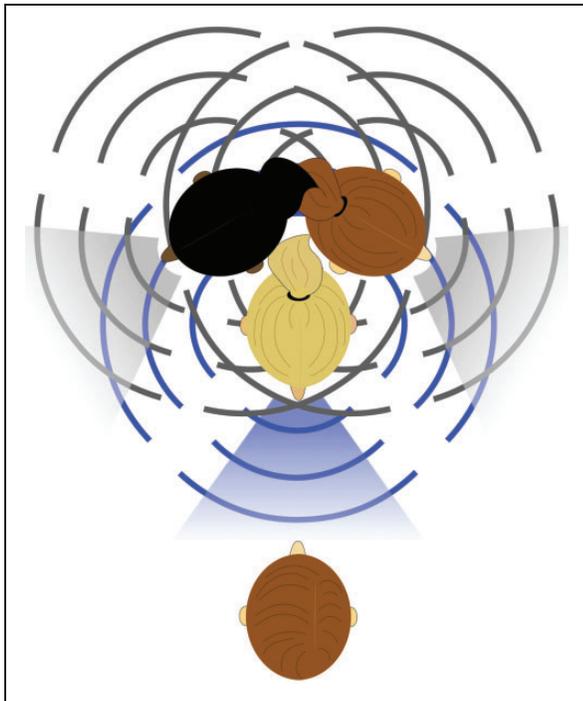


Figure 1. The Target (Blue) and Masker (Gray) Arrangement Simulated in this Study. Due to the directionality of extended high-frequency radiation (shading) compared with low-frequency radiation (bars), this scenario results in substantial masking at low frequencies, but not at extended high frequencies. Target and masker were presented from a single loudspeaker in front of the listener.

are frequency dependent, with low frequencies radiating more omnidirectionally around the talker's head and high frequencies radiating more directionally (i.e., with less horizontal spread away from the front of the talker; Chu & Warnock, 2002; Halkosaari et al., 2005; Kocon & Monson, 2018; Monson et al., 2012a). Because of the increasing directionality at higher frequencies, rotating a masker's head to face away from the listener effectively low-pass filters the masker speech signal received at the ear of the listener, providing potential spectral cues to the auditory system for detection and segregation of a target speech signal amidst masker speech signals. Under these conditions, we found that providing access to full-band speech (cutoff frequency of 20 kHz) improved normal-hearing listeners' speech-in-speech recognition performance relative to speech bandlimited at 8 kHz.

The results of that study suggest that EHF energy in speech conveys information regarding the speech signal. However, the type of information provided by EHF that benefits speech recognition remains unclear. One possibility is that EHF temporal information (e.g., the temporal envelope) serves as a segregation and grouping cue, facilitating segregation of low-frequency phonetic information. This is possible because high-frequency energy in speech is at least partially temporally coherent with low-frequency energy (Crouzet & Ainsworth, 2001; see Figure 2). Temporal coherence facilitates the grouping of sound features into a single stream, improving sound segregation for auditory scene analysis (Shamma et al., 2011), and it has been demonstrated that temporal (envelope) information becomes increasingly important for higher frequency bands in speech recognition in noise (Apoux & Bacon, 2004). Another possibility is that EHF spectral detail per se provides phonetic information. EHF spectral energy does provide information useful for phoneme identification when low-frequency

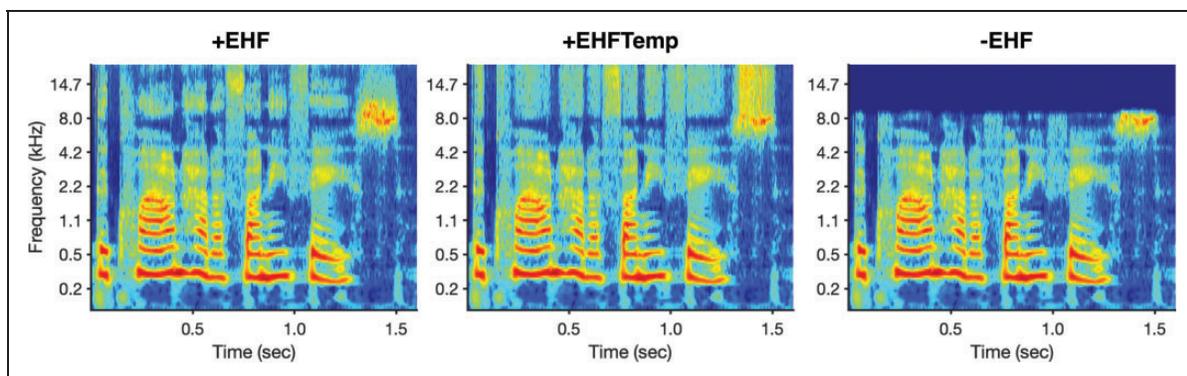


Figure 2. Cochleograms of the Female Target Talker Phrase, "The Clown Had a Funny Face." The three filtering conditions are shown: the full-band signal (+EHF; left), the signal with EHF spectral detail removed, but EHF temporal envelope preserved (+EHFTemp; middle), and the signal low-pass filtered at 8 kHz (-EHF; right). EHF = extended high frequency.

information is absent (Vitela et al., 2015) or severely degraded (Lippmann, 1996). However, it may be that phonetic information provided by EHF is redundant with phonetic information provided by lower frequencies and may not be useful when phonetic information at low frequencies is accessible. Indeed, the likelihood of this redundancy is supported by the history of speech intelligibility research, which resulted in models predicting negligible contribution from frequencies above 7 kHz for speech recognition when low and/or midrange frequencies are accessible (Monson et al., 2014a).

There is evidence that listeners with clinically normal audiograms but poorer pure-tone thresholds at EHF have diminished speech-in-noise abilities. Badri et al. (2011) showed that listeners who self-reported and exhibited speech-in-noise difficulties had elevated EHF thresholds at 12.5 and 14 kHz compared with a control group. Motlagh Zadeh et al. (2019) also found group-level differences in self-reported speech-in-noise difficulty, with greater likelihood of reporting difficulty for groups with more severe EHF hearing loss (measured at 10, 12.5, 14, and 16 kHz). They also reported a correlation between EHF pure-tone averages (PTAs) and speech-in-noise scores when the noise masker was a broadband speech-shaped noise, although, surprisingly, no such relationship was observed when the noise masker was bandlimited to 8 kHz. Yeend et al. (2019) found that EHF PTAs (measured from 9 to 12.5 kHz) correlated with a composite speech score derived from both self-reported difficulty and objective speech-in-noise assessments.

These studies suggest that EHF hearing loss could potentially be a diagnostic or predictive factor for speech-in-noise difficulty. Others studies, however, have failed to find a relationship between EHF thresholds and speech-in-noise performance. Liberman et al. (2016) found that, although group-level differences in EHF thresholds (measured at 9, 10, 11.2, 12.5, 14, and 16 kHz) were present between individuals at high risk versus low risk for cochlear synaptopathy, EHF PTAs did not predict speech-in-noise performance. However, that study used speech materials that were bandlimited at 8.8 kHz. Similarly, Smith et al. (2019) found no relationship between EHF PTAs (measured at 10, 12.5, and 14 kHz) and speech-in-noise scores, although listeners in that study all had relatively good EHF thresholds. Prendergast et al. (2019) reported that speech-in-noise performance was predicted by statistical models that included 16-kHz thresholds as predictors, along with age and noise exposure. However, replacing the 16-kHz threshold with pure-tone thresholds at standard audiometric frequencies as predictors resulted in improved model predictions. Thus, there are mixed findings on the relationship between EHF pure-tone thresholds and speech-in-noise difficulty.

Another study provided data that might help resolve these mixed findings. Taking into consideration the effects of directivity of speech radiation, Corbin et al. (2019) demonstrated that better 16-kHz thresholds were associated with better speech-in-noise scores when maskers were facing away from the listener while the target talker was facing the listener. However, there was *no* relationship between 16-kHz thresholds and speech-in-noise scores when maskers and the target talker were all facing the listener. As described earlier, the rotating of the maskers' heads introduces low-pass filtering effects, increasing the salience of EHF acoustic features for the target speech. This approach may have teased out the true relationship between EHF thresholds and speech-in-noise difficulty. Notably, listeners in that study had clinically normal audiograms but exhibited EHF pure-tone thresholds ranging from -20 to 60 dB HL.

To gain a better understanding of how spectral and temporal information contribute to the observed benefit of access to EHF in speech, the present study assessed whether access to temporal information alone in the EHF speech band provided a benefit for speech-in-speech listening, and whether access to spectral detail provided any additional benefit. We hypothesized that access to both temporal information and spectral detail at EHF would provide a benefit beyond that provided by temporal information alone. We assessed the effect of a change in masker head orientation, hypothesizing that maskers that were facing further away from the listener would lead to improved performance. In addition, we examined whether better pure-tone thresholds predicted better performance in our speech-in-speech task for a group of listeners who had normal hearing at both standard audiometric frequencies and EHF.

Methods

Participants

Forty-one participants (six male), ages 19–25 years (mean = 21.3 years), participated in this experiment. Participants had normal hearing across the frequency range of hearing, as indicated by pure-tone audiometric thresholds better than 25 dB HL in at least one ear for octave frequencies between 0.5 and 8 kHz and EHF of 9, 10, 11.2, 12.5, 14, and 16 kHz. All experimental procedures were approved by the Institutional Review Board at the University of Illinois at Urbana-Champaign.

Stimuli

The masker stimuli consisted of two-female-talker babble with both talkers facing 45° or both talkers

facing 60° relative to the listener. Masker stimuli were generated using recordings made at angles to the right of the talkers, taken from a database of high-fidelity (44.1-kHz sampling rate, 16-bit precision) anechoic multichannel recordings (Monson et al., 2012a). Left-right symmetry in speech radiation from the talker was assumed during the recording process. A semantically unpredictable speech babble signal was created for each angle. Target speech stimuli were the Bamford-Kowal-Bench sentences (Bench et al., 1979) recorded by a single female talker in a sound-treated booth using a class I precision microphone located at 0°, with 44.1-kHz sampling rate and 16-bit precision.

Three filtering schemes were used. For the low-pass filtered condition, all stimuli were low-pass filtered using a 32-pole Butterworth filter with a cutoff frequency of 8 kHz. For the full-band condition, all stimuli were low-pass filtered at 20 kHz. For the third condition, designed to preserve temporal EHF information while removing EHF spectral detail, the amplitude envelope of the EHF band of each target and masker stimulus was extracted by (a) high-pass filtering at 8 kHz using a Parks-McClellan equiripple finite impulse response (FIR) filter, (b) computing the Hilbert transform of the high-pass filtered signal, and (c) low-pass filtering the magnitude of the Hilbert transform at 100 Hz. Each 8-kHz low-pass filtered target and masker stimulus was then summed with a spectrally flat EHF noise band (8–20 kHz) that was amplitude modulated using the envelope of the EHF band (i.e., a single-channel vocoded EHF band) corresponding to that stimulus (Figure 2).

Procedure

Stimuli were presented to listeners using a KRK Rokit 8 G3 loudspeaker at 1 m directly in front of the listener seated in a sound-treated booth. The level of the two-talker masker was set at 70 dB sound pressure level at 1 m, while the level of the target was adaptively varied. Two interleaved adaptive tracks were used, each incorporating a one-down, one-up adaptive rule. For one track, the signal-to-noise ratio (SNR) was decreased if one or more words were correctly repeated; otherwise, the SNR was increased. For the second track, the SNR was decreased if all words or all but one word were correctly repeated; otherwise, the SNR was increased. Both tracks started at an SNR of 4 dB. The SNR was initially adjusted in steps of 4 dB and then by 2 dB after the first reversal. Each of the two tracks comprised 16 sentences. Word-level data from the two tracks were combined and fitted with a logit function with asymptotes at 0 and 100% correct. The speech reception threshold (SRT) was defined as the SNR associated with 50% correct. Data fits were associated with r^2 values ranging from .50 to .99, with a median value of .85.

Three filtering conditions were tested: full band (+EHF), full band with only EHF temporal information (+EHFTemp), and low-pass filtered at 8 kHz (–EHF). Two masker head orientation conditions were tested: both maskers facing 45° or both maskers facing 60° relative to the target talker. Following a single training block consisting of 16 sentences, the six conditions (three filtering conditions \times two masker head angles) were tested in separate blocks with block order randomized across participants. The starting sentence list number was randomized for each participant and continued in numerical order of the Bamford-Kowal-Bench sentence lists.

Analysis

Statistical analysis consisted of a two-way repeated-measures analysis of variance (ANOVA) to assess the effect of filtering condition and masker head angle. Univariate Pearson's correlation was used to assess the relationship between pure-tone thresholds and task performance. Statistical analyses were conducted using the ezANOVA and corr functions in R (R Core Team, 2018). Custom scripts written in MATLAB (MathWorks) were used for signal processing and experimental control. All recording materials and data for this study will be made available upon request.

Results

There was a main effect of filtering condition, with mean SRTs of -9.7 , -9.2 , and -8.3 dB (medians -9.9 , -9.4 , and -8.6 dB) for the +EHF, +EHFTemp, and –EHF conditions, respectively—two-way repeated-measures ANOVA, $F(80, 2) = 15.8$, $p < .001$ (Figure 3). Post hoc pairwise comparisons (Holm-Bonferroni corrected) revealed a

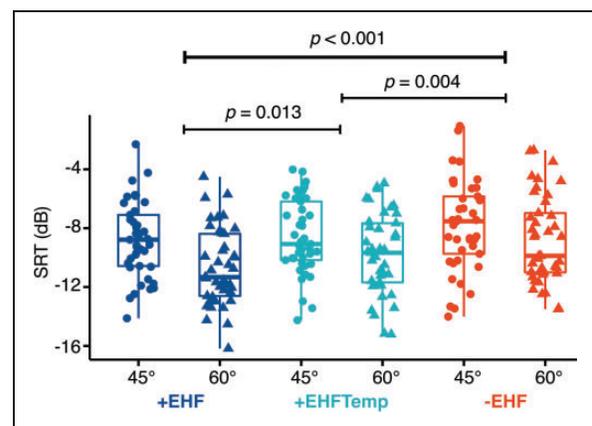


Figure 3. SRTs for the Three Filtering Conditions and Two Masker Head Orientations.

EHF = extended high frequency; SRTs = speech reception thresholds.

significant difference between all EHF conditions (corrected $p < .05$ for all comparisons; see Figure 3). There was a main effect of masker head orientation, with mean SRTs of -8.4 and -9.7 dB (medians -8.7 and -10.2 dB) for the 45° and 60° conditions, respectively, $F(40, 1) = 39.4$, $p < .001$, and no interaction between filtering condition and masker head orientation ($p = .2$).

We conducted an exploratory analysis to assess whether pure-tone thresholds across the frequency range of hearing predicted performance in our full-band task (Figure 4). The 12.5-kHz, 16-kHz, and EHF PTA (9–16 kHz) exhibited the highest correlation coefficients (Pearson's $r > .3$) between full-band (+EHF) task performance (averaged across masker head angles) and left-right-averaged pure-tone thresholds.

Discussion

We replicated our previous finding that access to EHF in speech improves normal-hearing listeners' speech-in-speech recognition performance relative to speech band-limited at 8 kHz (Monson et al., 2019). The improvements observed in the present study between the +EHF and -EHF conditions were of similar magnitudes to those reported previously. These findings continue to support the use of high-fidelity speech materials when testing and/or simulating speech-in-speech environments as

information at EHF is audible and useful for speech recognition for normal-hearing listeners.

We hypothesized that spectral detail at EHF provides benefit for listeners beyond that provided by EHF temporal information alone. Our results lend support for this hypothesis as a significant decrease in speech recognition was observed when spectral detail was removed and only temporal (i.e., envelope) information from the EHF band was provided to listeners. The size of this effect was small (0.5 dB on average), whereas EHF temporal information alone provided 0.9 dB of benefit, on average. Thus, our data suggest that EHF temporal information may account for a larger proportion of the EHF benefit, but the full complement of EHF benefit only occurs when additional spectral detail is also available. This finding highlights the exquisite sensitivity of the human auditory system to EHF in speech, despite poorer frequency discrimination ability (Moore & Ernst, 2012), poorer pure-tone audibility (International Organization for Standardization, 2003), and larger widths of auditory filters beyond 8 kHz (Glasberg & Moore, 1990).

Although it has not been shown definitively here, our findings lend credence to the idea that EHF provide phonetic information useful for speech-in-speech recognition rather than purely serving as a target speech segregation cue. This is possible because individual

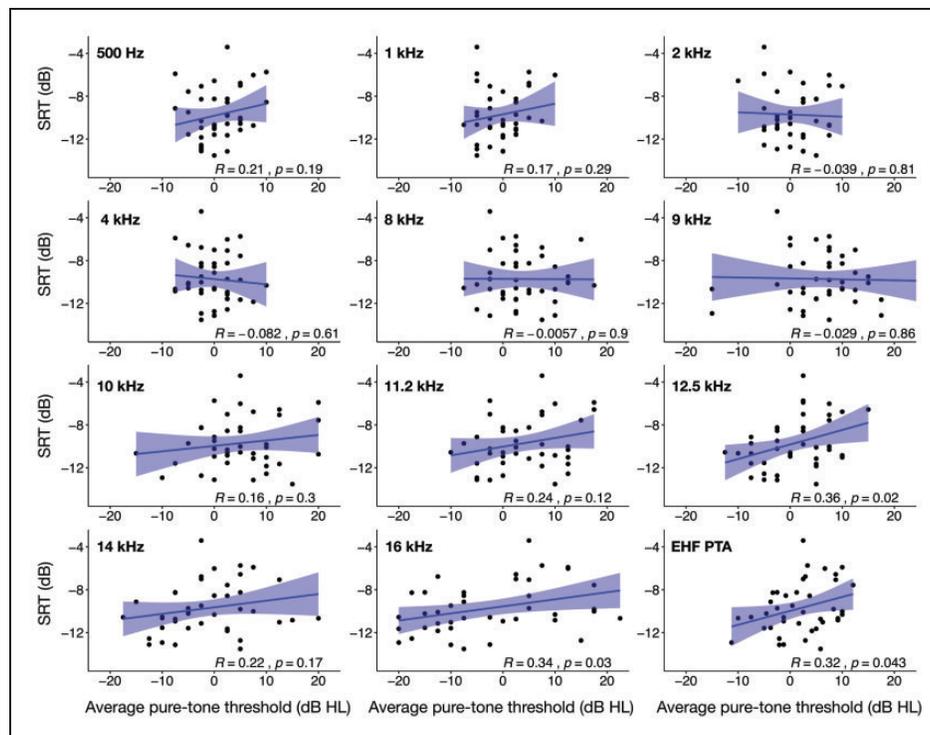


Figure 4. Mean SRTs for the +EHF Condition Plotted Against Pure-Tone Thresholds Averaged Across Both Ears. Shading represents 95% confidence intervals. Displayed p values are not corrected for multiple comparisons. SRT = speech reception threshold; EHF = extended high frequency; PTA = pure-tone average.

phonemes, such as voiceless fricatives, exhibit distinctive spectral features at EHF (e.g., energy peak loci, spectral slopes; Monson et al., 2012b) sufficient to facilitate phoneme recognition, especially for consonants (Lippmann, 1996; Vitela et al., 2015). This finding is of importance for potential amplification of EHF in hearing devices. For example, if EHF were to be represented in cochlear implants, our data suggest that devoting more than a single electrode channel to EHF may be useful to provide the intended EHF benefit.

The observed EHF benefit is also in line with previous reports that EHF hearing loss is correlated with both self-reported and objectively measured speech-in-noise difficulty (Badri et al., 2011; Barbee et al., 2018; Corbin et al., 2019; Motlagh Zadeh et al., 2019; Yeend et al., 2019). The inclusion of routine EHF examinations may help to identify listeners at risk of difficulties listening in noise with otherwise normal clinical audiograms. There are multiple reasons why EHF loss might lead to a speech-in-noise difficulty. As shown here, EHF contribute to speech-in-speech recognition when maskers are facing different directions, which is typical for real-world cocktail party environments. Similar to how visual cues of a social partner's head orientation and gaze can direct attention to that partner or other objects of interest (Frischen et al., 2007; Gamer & Hecht, 2007), highly directional EHF could serve to herald the potential importance of an interlocutor's speech signal, thereby drawing the listener's attention to that signal. That is, high-amplitude EHF energy will only be received from a talker that is directly facing a listener, which likely indicates that this listener is the intended recipient of the talker's utterance. In addition to this potential real-world cue, we have demonstrated here that spectral detail at EHF provides information useful for speech-in-speech recognition. EHF hearing loss might lead to the degradation of these multiple sources of information.

Our exploratory analysis revealed that relationships between full-band SRTs and pure-tone thresholds across the frequency range of hearing only emerged at EHF, in spite of our strict inclusion criterion for normal hearing (< 25 dB HL in at least one ear) at all frequencies, including EHF. This finding should inform future hypotheses regarding the relationship between EHF thresholds and speech-in-noise performance. The rotating of the maskers' heads in the present study introduces low-pass filtering effects, increasing the salience of EHF acoustic features for the target speech. This approach may elucidate the true relationship between EHF thresholds and speech-in-noise difficulty. We previously found that 16-kHz thresholds for normal-hearing listeners correlated with ability to *detect* EHF energy in speech (Monson & Caravello, 2019), and we found preliminary evidence here for a relationship with ability to *use* EHF for speech-in-speech recognition.

We observed approximately 2 dB of improvement in SRT when the maskers were rotated from 45° to 60° for full-band speech. That this consistent and highly significant improvement occurs with a change of only 15° in head orientation is striking and highlights the sensitivity of the auditory system to talker/masker head orientation, particularly as it pertains to speech recognition (Corbin et al., 2019; Strelcyk et al., 2014). We have shown previously that the minimum audible change in a talker's head orientation, relative to a 0° head orientation, is approximately 41° for the average normal-hearing listener (Monson et al., 2019). Although we have not tested the minimum audible change relative to a 45° head orientation, it is questionable whether this subtle 15° change in head orientation is detectable to the average listener. Nonetheless, it is clear that head orientation release from masking has a robust effect on speech-in-speech recognition for colocated maskers, although this effect may be reduced when maskers and target are spatially separated (Corbin et al., 2019).

In summary, we found that, despite the well-known decrease in sensitivity and acuity at EHF for the human auditory system, spectral detail at EHF conveys information useful for speech-in-speech recognition. EHF spectral detail provides additional gains beyond that provided by EHF temporal (i.e., envelope) information. Speech-in-speech performance is highly sensitive to masker head orientation, with a change of only 15° having a robust effect. We found evidence for a relationship between EHF pure-tone sensitivity and speech-in-noise scores when listeners have no substantial hearing loss at EHF. Implications include that the preservation of spectral detail at EHF may be beneficial in ongoing efforts to extend the bandwidth of hearing aids and other devices (Arbogast et al., 2019; Seeto & Searchfield, 2018; Van Eeckhoutte et al., 2020) or to restore audibility using frequency lowering or other amplification techniques (Alexander & Rallapalli, 2017; McCreery et al., 2014). Furthermore, the continued use of speech materials that are bandlimited by recording sampling rate and/or transducer frequency response for speech-in-noise testing in the clinic and the laboratory precludes the beneficial effects of EHF hearing. Finally, real-world speech signals include effects of talker head orientation, and incorporating these effects might improve the precision and predictive power of speech recognition measures.

Acknowledgments

The authors thank Andrew Oxenham for helpful discussions regarding the approach, Emily Buss for assistance with experimental software implementation, Sa Shen for biostatistical advice, and Chris Plack and one anonymous reviewer for helpful comments on the article. Melanie Flores, Raquel Rojas, Ana Sabic, and Paige Valente assisted with data collection.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Brian B. Monson  <https://orcid.org/0000-0002-0179-8179>

References

- Alexander, J. M., & Rallapalli, V. (2017). Acoustic and perceptual effects of amplitude and frequency compression on high-frequency speech. *The Journal of the Acoustical Society of America*, *142*(2), 908–923. <https://doi.org/10.1121/1.4997938>
- Apoux, F., & Bacon, S. P. (2004). Relative importance of temporal information in various frequency regions for consonant identification in quiet and in noise. *The Journal of the Acoustical Society of America*, *116*(3), 1671–1680. <https://doi.org/10.1121/1.1781329>
- Arbogast, T. L., Moore, B. C. J., Puria, S., Dundas, D., Brimacombe, J., Edwards, B., & Carr Levy, S. (2019). Achieved gain and subjective outcomes for a wide-bandwidth contact hearing aid fitted using CAM2. *Ear and Hearing*, *40*(3), 741–756. <https://doi.org/10.1097/AUD.0000000000000661>
- Badri, R., Siegel, J. H., & Wright, B. A. (2011). Auditory filter shapes and high-frequency hearing in adults who have impaired speech in noise performance despite clinically normal audiograms. *The Journal of the Acoustical Society of America*, *129*(2), 852–863. <https://doi.org/10.1121/1.3523476>
- Barbee, C., James, J., Park, J., Smith, E., Johnson, C., Clifton, S., & Danhauer, J. (2018). Effectiveness of auditory measures for detecting hidden hearing loss and/or cochlear synaptopathy: A systematic review. *Seminars in Hearing*, *39*(02), 172–209. <https://doi.org/10.1055/s-0038-1641743>
- Bench, J., Kowal, Å., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, *13*(3), 108–112. <https://doi.org/10.3109/03005367909078884>
- Best, V., Carlile, S., Jin, C., & van Schaik, A. (2005). The role of high frequencies in speech localization. *The Journal of the Acoustical Society of America*, *118*(1), 353–363. <https://doi.org/10.1121/1.1926107>
- Ching, T. Y. C., Dillon, H., & Byrne, D. (1998). Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification. *The Journal of the Acoustical Society of America*, *103*(2), 1128–1140. <https://doi.org/10.1121/1.421224>
- Chu, W., & Warnock, A. (2002). *Detailed directivity of sound fields around human talkers*. Institute for Research in Construction. <https://doi.org/10.4224/20378930>
- Corbin, N., Buss, E., & Monson, B. B. (2019). Effect of talker orientation on speech perception in spatially-separated speech. Annual Meeting of the American Auditory Society, Scottsdale, AZ, February 2019.
- Crouzet, O., & Ainsworth, W. A. (2001, September). On the various influences of envelope information on the perception of speech in adverse conditions: An analysis of between-channel envelope correlation [Paper presentation]. Workshop on Consistent and Reliable Acoustic Cues for Sound Analysis. Aalborg, Denmark.
- Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*(4), 694–724. <https://doi.org/10.1037/0033-2909.133.4.694>
- Gamer, M., & Hecht, H. (2007). Are you looking at me? Measuring the cone of gaze. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(3), 705–715. <https://doi.org/10.1037/0096-1523.33.3.705>
- Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1–2), 103–138. [https://doi.org/10.1016/0378-5955\(90\)90170-T](https://doi.org/10.1016/0378-5955(90)90170-T)
- Halkosaari, T., Vaalgamaa, M., & Karjalainen, M. (2005). Directivity of artificial and human speech. *Journal of the Audio Engineering Society*, *53*(7/8), 620–631.
- Hunter, L. L., Monson, B. B., Moore, D. R., Dhar, S., Wright, B. A., Munro, K. J., Motlagh Zadeh, L., Blankenship, C. M., Stiepan, S. M., & Siegel, J. H. (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>
- International Organization for Standardization. (2003). *ISO 226:2003 Acoustics—Normal equal-loudness-level contours*.
- Kocon, P., & Monson, B. B. (2018). Horizontal directivity patterns differ between vowels extracted from running speech. *The Journal of the Acoustical Society of America*, *144*(1), EL7–EL12. <https://doi.org/10.1121/1.5044508>
- Lieberman, M. C., Epstein, M. J., Cleveland, S. S., Wang, H., & Maison, S. F. (2016). Toward a differential diagnosis of hidden hearing loss in humans. *PLoS One*, *11*(9), e0162726. <https://doi.org/10.1371/journal.pone.0162726>
- Lippmann, R. P. (1996). Accurate consonant perception without mid-frequency speech energy. *IEEE Transactions on Speech and Audio Processing*, *4*(1), 66. <https://doi.org/10.1109/TSA.1996.481454>
- McCreery, R. W., Alexander, J., Brennan, M. A., Hoover, B., Kopun, J., & Stelmachowicz, P. G. (2014). The influence of audibility on speech recognition with nonlinear frequency compression for children and adults with hearing loss. *Ear and Hearing*, *35*(4), 440–447. <https://doi.org/10.1097/AUD.0000000000000027>
- McCreery, R. W., & Stelmachowicz, P. G. (2013). The effects of limited bandwidth and noise on verbal processing time and word recall in normal-hearing children. *Ear and Hearing*, *34*(5), 585–591. <https://doi.org/10.1097/AUD.0b013e31828576e2>
- 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>

- Monson, B. B., Hunter, E. J., Lotto, A. J., & Story, B. H. (2014a). The perceptual significance of high-frequency energy in the human voice. *Frontiers in Psychology*, 5(Jun), 1–11. <https://doi.org/10.3389/fpsyg.2014.00587>
- Monson, B. B., Hunter, E. J., & Story, B. H. (2012a). Horizontal directivity of low- and high-frequency energy in speech and singing. *The Journal of the Acoustical Society of America*, 132(1), 433–441. <https://doi.org/10.1121/1.4725963>
- Monson, B. B., Lotto, A. J., & Story, B. H. (2012b). Analysis of high-frequency energy in long-term average spectra of singing, speech, and voiceless fricatives. *The Journal of the Acoustical Society of America*, 132(3), 1754–1764. <https://doi.org/10.1121/1.4742724>
- Monson, B. B., Lotto, A. J., & Story, B. H. (2014b). Detection of high-frequency energy level changes in speech and singing. *The Journal of the Acoustical Society of America*, 135(1), 400–406. <https://doi.org/10.1121/1.4829525>
- Monson, B. B., Rock, J., Schulz, A., Hoffman, E., & Buss, E. (2019). Ecological cocktail party listening reveals the utility of extended high-frequency hearing. *Hearing Research*, 381, 107773. <https://doi.org/10.1016/j.heares.2019.107773>
- Moore, B. C. J., & Ernst, S. M. A. (2012). Frequency difference limens at high frequencies: Evidence for a transition from a temporal to a place code. *The Journal of the Acoustical Society of America*, 132(3), 1542–1547. <https://doi.org/10.1121/1.4739444>
- Moore, B. C. J., & Tan, C.-T. (2003). Perceived naturalness of spectrally distorted speech and music. *The Journal of the Acoustical Society of America*, 114(1), 408–419. <https://doi.org/10.1121/1.1577552>
- Motlagh Zadeh, L., Silbert, N. H., Sternasty, K., Swanepoel, D. W., Hunter, L. L., & Moore, D. R. (2019). Extended high-frequency hearing enhances speech perception in noise. *Proceedings of the National Academy of Sciences*, 116(47), 23753–23759. <https://doi.org/10.1073/pnas.1903315116>
- Prendergast, G., Couth, S., Millman, R. E., Guest, H., Kluk, K., Munro, K. J., & Plack, C. J. (2019). Effects of age and noise exposure on proxy measures of cochlear synaptopathy. *Trends in Hearing*, 23, 233121651987730. <https://doi.org/10.1177/2331216519877301>
- R Core Team (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Seeto, A., & Searchfield, G. D. (2018). Investigation of extended bandwidth hearing aid amplification on speech intelligibility and sound quality in adults with mild-to-moderate hearing loss. *Journal of the American Academy of Audiology*, 29(03), 243–254. <https://doi.org/10.3766/jaaa.16180>
- Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, 34(3), 114–123. <https://doi.org/10.1016/j.tins.2010.11.002>
- Smith, S. B., Krizman, J., Liu, C., White-Schwoch, T., Nicol, T., & Kraus, N. (2019). Investigating peripheral sources of speech-in-noise variability in listeners with normal audiograms. *Hearing Research*, 371, 66–74. <https://doi.org/10.1016/j.heares.2018.11.008>
- Strelcyk, O., Pentony, S., Kalluri, S., & Edwards, B. (2014). Effects of interferer facing orientation on speech perception by normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 135(3), 1419–1432. <https://doi.org/10.1121/1.4864786>
- Van Eeckhoutte, M., Folkeard, P., Glista, D., & Scollie, S. (2020). Speech recognition, loudness, and preference with extended bandwidth hearing aids for adult hearing aid users. *International Journal of Audiology*, 59, 780–791. <https://doi.org/10.1080/14992027.2020.1750718>
- Vitela, A. D., Monson, B. B., & Lotto, A. J. (2015). Phoneme categorization relying solely on high-frequency energy. *The Journal of the Acoustical Society of America*, 137(1), EL65–EL70. <https://doi.org/10.1121/1.4903917>
- Yeend, I., Beach, E. F., & Sharma, M. (2019). Working memory and extended high-frequency hearing in adults. *Ear and Hearing*, 40(3), 458–467. <https://doi.org/10.1097/AUD.0000000000000640>