Extended High Frequencies Provide both Spectral and Temporal Information to Improve Speech-in-Speech Listening

Allison Trine[1] and Brian B. Monson[1,2]
[1]Department of Speech and Hearing Science, [2]Neuroscience Program, University of Illinois at Urbana-Champaign

ILLINOIS

Introduction

Some studies indicate that extended high frequencies (EHF, defined as frequencies ≥ 8 kHz) are useful for some auditory tasks, but it is widely believed they play little to no role in speech perception. However, recent studies from our lab and others have investigated the utility of EHF for speech perception, particularly in speech-in-speech (the "cocktail party" problem) listening simulations.

Because the typical recording procedure for speech materials involves using a microphone located directly in front of a talker, most studies examining speech-in-speech listening simulate an unnatural scenario in which the target talker and maskers are all facing the listener (Fig. 1A). Our study design was more representative of realistic cocktail party listening, in which the target talker faced the listener while co-located maskers faced away from the listener (45° or 60° relative to the listener), as though talking to other listeners (Fig. 1B).

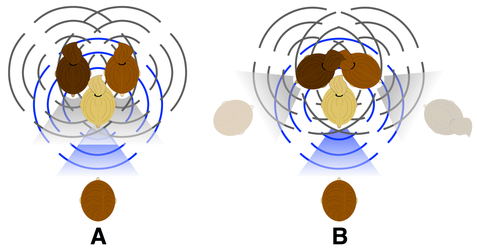


Figure 1. (A) The unnatural scenario typically simulated when evaluating cocktail party listening. This results in substantial masking at all frequencies. (B) The more ecologically valid scenario simulated in the present study. Due to the directionality of high-frequency radiation (shading) compared to low-frequency radiation (bars), this scenario results in substantial masking at low frequencies, but not at high frequencies. Note that maskers are **co-located** with target speech.

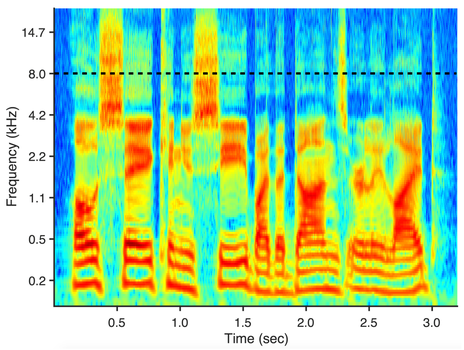
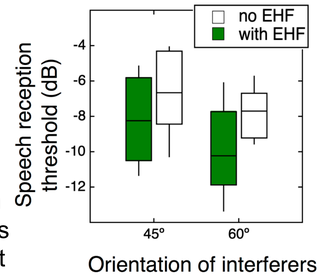


Figure 2. Cochleogram of the phrase "Oh say can you see by the dawn's early light" uttered by a male talker. There is considerable energy above 8,000 Hz.

Previous Study

- Effect of filtering condition
- Effect of masker head rotation
- Participants improved 1.7 dB on average with access to EHFs (Monson et al 2019)



Aim

- To determine whether EHF temporal information or spectral detail (or both) provide benefit for speech-in-speech performance.

Methods

Stimuli:

- Masker: two-female-talker babble stimulus created using previous recordings with microphones positioned at 45° and 60° relative to the talkers
  - To decrease predictability of the maskers, a semantically unpredictable speech signal was used for the maskers
- Target: female talker, recorded in a sound-treated booth at 0° relative to talker
  - BKB sentences
  - Type I microphone, 44.1-kHz sampling rate, 16-bit precision
- Low-pass filtered condition: all stimuli low-pass filtered, cutoff frequency of 8 kHz (**No EHF**)
- Full-band condition: all stimuli low-pass filtered, cutoff frequency of 20 kHz (**EHF Temporal + Spectral**)
- Full-band with EHF temporal information only: LP filter at 8 kHz; sum with EHF band "white" noise, amplitude modulated with the envelope of the speech EHF band (**EHF Temporal only**)

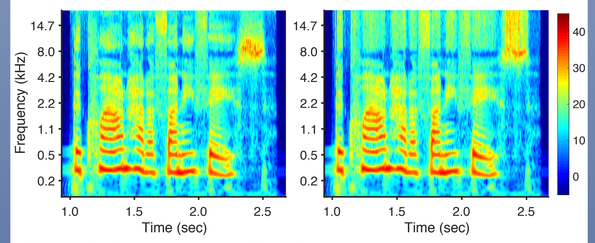


Figure 3. Cochleogram of the female target talker phrase, "The clown had a funny face." **Left** panel shows the full-band signal. **Right** panel shows the signal with EHF spectral detail removed, but EHF temporal information preserved.

Methods (continued)

Subjects:

- 20 female participants age 20-27 years with normal hearing (defined as thresholds ≤ 20 dB HL in at least one ear)

Procedure:

- Stimuli presented to listeners seated in a sound-treated booth at 1 m over a KRK Rokit 8 G3 loudspeaker with good high-frequency response
- Masker level set at 70 dB SPL at 1 m
- Target talker level (*i.e.*, signal-to-noise ratio; SNR) was adaptively varied
- One-down, one-up adaptive rule
- Both adaptive tracks started with a signal level of 4 dB SNR. SNR initially adjusted in steps of 4 dB, but switched to an adjustment of 2 dB after the first reversal
- Speech reception threshold (SRT; target-to-masker ratio necessary to achieve 50% accuracy of identification of words in a sentence)
- Brief training block consisting of 16 sentences
- Six conditions tested in separate blocks:
  - **EHF Temporal + Spectral** *vs.* **EHF Temporal only** *vs.* **No EHF**
  - Masker head rotation of 45° *vs.* 60°
- Block order randomized across participants

Results

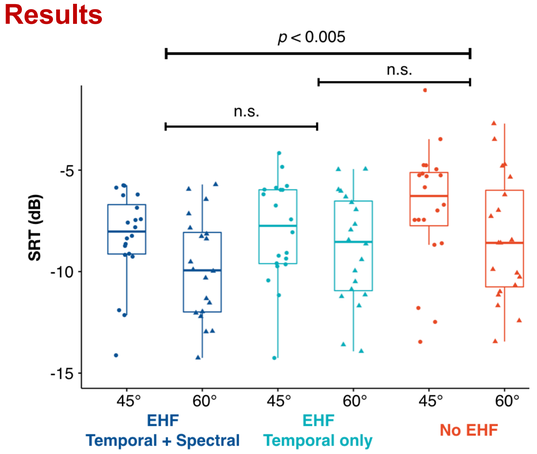


Figure 4. SRTs for three filtering conditions (color coded) and two masker head rotations (45° *vs.* 60°).

- Repeated-measures ANOVA
- Main effect of filtering condition (p=0.01)
- Main effect of masker head rotation (p=0.002)
- No interaction between filtering condition and masker head rotation (p=0.7)

Results (continued)

- Posthoc pairwise comparisons across filtering conditions revealed a significant difference only between **EHF Temporal + Spectral** and **No EHF** (p<0.005)
- Correlation observed between mean SRT for **EHF Temporal + Spectral** and 16-kHz pure tone threshold in the better ear (R=0.57, p=0.009), but not for EHF pure tone average (9-16 kHz) in the better ear (p=0.98)

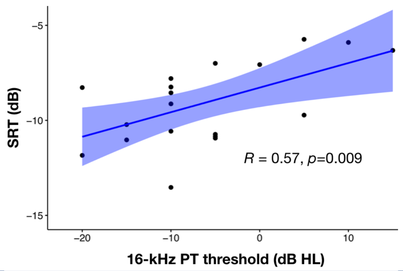


Figure 5. Mean SRTs for the EHF Temp + Spec condition plotted against pure tone threshold at 16 kHz for the better ear.

Conclusions

- Listeners performed better with access to EHFs, indicating that there is information present at EHFs that is utilized by young, normal-hearing listeners.
- Access to temporal cues at EHFs alone was insufficient to significantly improve speech-in-speech performance. Access to EHF spectral detail provided additional benefit.
- Masker head orientation impacts listener performance for speech-in-speech listening.
- On average, listeners performed 1.3 dB better for the 60° condition, signifying that a discrete change in masker head orientation (15°) led to a significant improvement in performance.
- For normal-hearing listeners, better EHF pure tone thresholds at 16 kHz predicted better performance on the speech-in-speech task.

References

Monson BB, Rock J, Schulz A, Hoffman E, and Buss E (2019). Ecological cocktail party listening reveals the utility of extended high-frequency hearing. *Hearing Research,* 381, 107773.
Monson BB and Caravello J (2019). The maximum audible low-pass cutoff frequency for speech. *J. Acoust. Soc. Am.* 146, EL496-501.
Monson BB, Hunter EJ, and Story BH (2012). Horizontal directivity of low- and high-frequency energy in speech and singing. *J. Acoust. Soc. Am,* 132, 433–441.
Zadeh LM, Silbert NH, Sternasty K, Swanepoel DW, Hunter LL, and Moore DR (2019). Extended high-frequency hearing enhances speech perception in noise. *Proc Nat Acad Sci, 116*(47), 23753–23759.
Hunter LL, Monson BB, Moore DR, Dhar S, Wright BA, Munro KJ, Motlagh Zadeh L, Blankenship CM, Stiepan SM, Siegel JH (in revision) Extended high-frequency hearing and speech perception implications in adults and children. *Hearing Research*