Access to phonetic information at extended high frequencies improves speech-in-speech performance

Allison Trine and Brian B. Monson
Department of Speech and Hearing Science, College of Applied Health Sciences, University of Illinois at Urbana-Champaign

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).

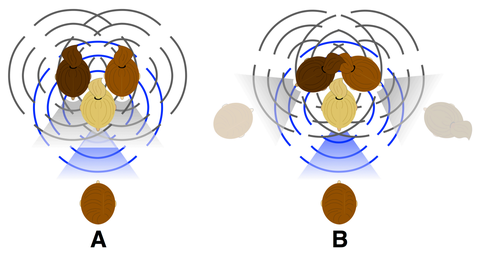

Figure 1. (A) The unnatural scenario typically simulated when evaluating cocktail party listening. This results in substantial masking at all frequencies. (B) The more ecologically valid scenario simulated in the present study. Due to the directionality of high-frequency radiation (shading) compared 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.

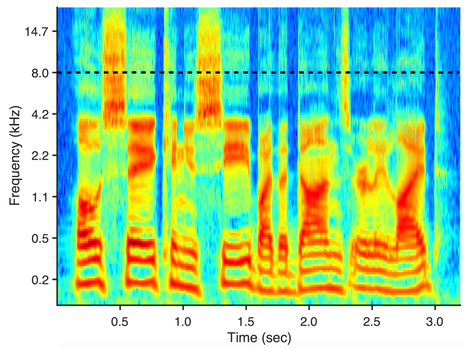

Figure 2. Cochleogram of the phrase "Oh say can you see by the dawn's early light" uttered by a male talker. There is considerable energy above 8,000 Hz.

Aim

- To determine whether EHF phonetic information provides benefit for speech-in-speech performance.

Method: Experiment 1

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 with 32-pole Butterworth filter, cutoff frequency of 8 kHz

Subjects:

- 20 (5 male) participants age 20-27 years with normal hearing (defined as thresholds better than 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) was measured
- Brief training block consisting of 16 sentences
- Four conditions tested in separate blocks:
  - With EHF *vs*. without EHF
  - Masker head rotation of 45° *vs*. head rotation of 60°
- Block order randomized across participants

Results: Experiment 1

- Two-way repeated measures ANOVA
- Main effect of filtering condition ($p<0.001$)
- Main effect of masker head rotation ($p<0.001$)
- Participants improved 1.7 dB on average with access to EHFs

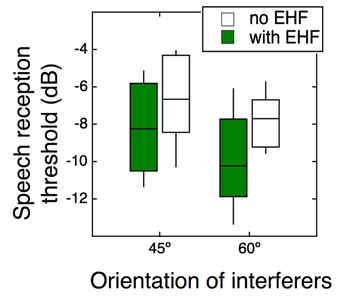

Figure 3. Experiment 1 results.

Method: Experiment 2 modifications

Stimuli:

- Full-band condition with temporal information only: EHF "white" noise, amplitude modulated with the envelope of the speech EHF band.

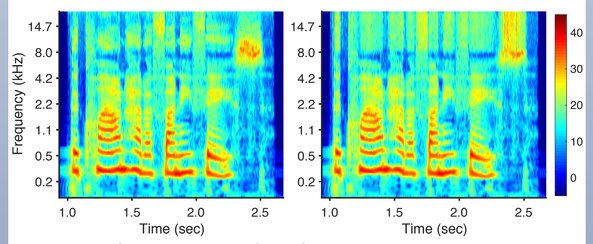

Figure 4. Cochleogram of the female target talker phrase, "The clown had a funny face." **Left** panel shows the full-band signal used in Experiment 1. **Right** panel shows the signal used in Experiment 2, with EHF phonetic detail removed, but EHF temporal information preserved.

Subjects:

- 25 participants (2 male) age 19-22 with normal hearing

Results: Experiment 2

- Mixed-effects ANOVA combining data from Exp. 1 (EHF temporal + phonetic information; **T+P**) and Exp. 2 (EHF temporal information only; **T**)
- Main effect of filtering condition ($p<0.001$)
- Main effect of masker head rotation ($p<0.001$)
- No main effect of group (T+P *vs*. T; $p=0.2$)
- Significant interaction between group and filtering condition ($p=0.049$)
- Group T participants improved 0.75 dB on average with access to only temporal information in EHF
- No correlation between EHF PTA (9-16 kHz) of the better ear and performance in the task

Results: Experiment 2 (continued)

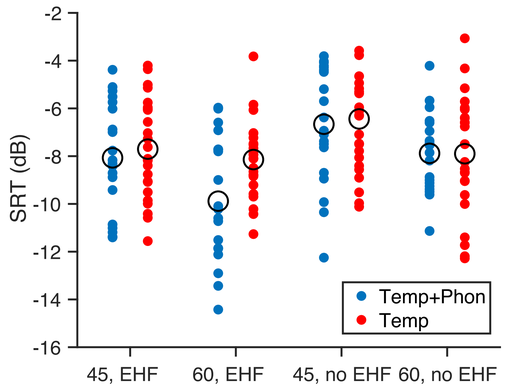

Figure 5. SRTs for T+P (blue) *vs* T (red) participants. Black circles indicate the mean for each condition.

Conclusions

- Masker head orientation impacts listener performance for speech-in-speech listening.
- Listeners in both groups performed 1.2 dB better, on average, for the 60° condition, signifying that a discrete change in masker head orientation (15°) led to a significant improvement in performance.
- Listeners performed better with access to EHFs, indicating that there is both phonetic and temporal information present in the EHFs and is utilized by young, normal-hearing listeners.
- Access to temporal cues in EHF alone was sufficient to improve speech-in-speech performance, however access to phonetic information provided additional gains.
- For normal-hearing listeners, lower EHF PTA thresholds were not indicative of performance in the task.

References

Monson, B. B., Rock, J., Schulz, A., Hoffman, E., and Buss, E. (submitted). Ecological cocktail party listening reveals the utility of extended high-frequency hearing.

Heffner, H.E., Heffner, R.S. (2008). High-frequency hearing. In: Dallos, P., Oertel, D., Hoy, R., (Eds.), Handbook of the senses: Audition. Elsevier, New York. pp. 55-60.

Monson, B. B., Hunter, E. J., and Story, B. H. (2012). Horizontal directivity of low- and high-frequency energy in speech and singing. *J. Acoust. Soc. Am,* 132, 433–441.

Vitela, A.D., Monson, B.B., Lotto, A.J. 2015. Phoneme categorization relying solely on high-frequency energy. *J. Acoust. Soc. Am.* 137, EL65-70.

ILLINOIS