# Band importance for speech-in-speech recognition in the presence of extended high-frequency cues

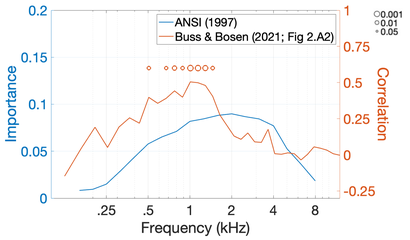
**Rohit M. Ananthanarayana[1], Emily Buss[2], Brian B. Monson[1]**

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

2 Department of Otolaryngology/HNS, University of North Carolina at Chapel Hill

## Introduction

- Band importance functions (BIFs) indicate the relative importance of spectral bands for speech understanding.
- Traditional methods for deriving BIFs (such as in the ANSI standard for speech intelligibility index) are used widely but have some key limitations:
  - The use of *successive low- and high-pass filtering* neglects the interactions between disjoint bands
  - The background masker is generally *steady or speech-shaped noise*, and not speech
  - Frequencies above *8-10 kHz* are considered to provide negligible benefit
- Issues with successive filtering have been addressed by employing correlation- and notch filtering-based methods.
- With regards to the masker and stimulus bandwidth, Buss and Bosen (2021) estimated BIFs up to 12 kHz for a speech-in-speech scenario.




*Band importance functions from literature. Circles on top mark significant correlations.*

- However, the remaining *extended high-frequencies (EHFs; 8-20 kHz) continue to be neglected,* in contrast with recent studies demonstrating the benefit of EHF cues for speech recognition in noisy backgrounds.
- EHFs have been shown to be useful particularly when the masker has reduced EHF levels relative to the target, which can occur in natural auditory scenes when the target talker is facing the listener and the masker talkers are not.
- Although EHF cues improve speech recognition, it is unclear how the magnitude of this benefit compares to that of other portions of the speech spectrum.

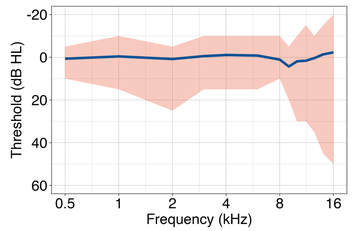
## Current study

- In this study, we estimated band importance functions (BIFs) for a *female target* and *two-talker masker* by *notch filtering* five contiguous bands from *40-20000 Hz.*
- With the target talker facing the listener, two masking conditions were tested: (1) masker talkers *facing* the listener; (2) maskers *facing 56°* (non-facing).
- We hypothesized a *significant interaction* between filtering and masker head orientation – i.e., higher importance for the EHF band in the condition with masker facing 56° compared to masker facing the listener.

## Methods

### A. Participants

- 37 native English speakers (31 F, 6 M), age 18-33 years (mean 21.14 years).
- Pure tone thresholds measured for standard frequencies (0.5-8 kHz) and EHFs (9, 10, 11.2, 12.5, 14, 16 kHz).
- All participants had thresholds <25 dB HL in at least one ear from 0.5 to 8 kHz.



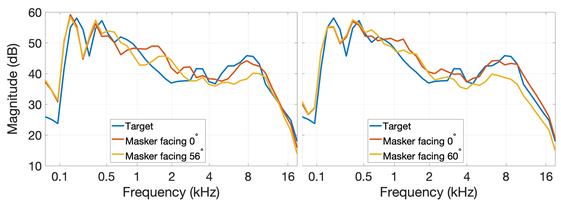

*Mean better-ear pure tone thresholds. Shaded region shows the range across participants.*

### B. Stimuli

- Target speech was the Bamford-Kowal-Bench (BKB) sentences spoken by a female talker.
- Masker was narrative speech by two female talkers.

### C. Conditions

- *6 filtering* conditions
  - Full-band (FB) and five notch-filtering: 40-400 Hz, 400-1k Hz, 1-3 kHz, 3-8 kHz, 8-20 kHz
  - Bands have same width on the equivalent rectangular bandwidth (ERB) scale
- *2 masker head orientations*:
  - Masker facing the listener, masker facing 56° away
- Total: 12 conditions




*ERB-scale long term average speech spectrum of the target and two-talker masker stimuli in the FB condition from current study (left) and Trine & Monson, 2019 (right). Note the difference in EHF levels between the two figures for non-facing masker.*

### D. Procedure

- Stimuli presented over a loudspeaker placed in front of the listener at a 1-m distance.
- Masker level set to 65 dB SPL, target level varied adaptively.
- Following a training block, the twelve conditions (six filtering × two masker head orientation) tested in separate blocks.

## Methods (continued)

### E. Analyses

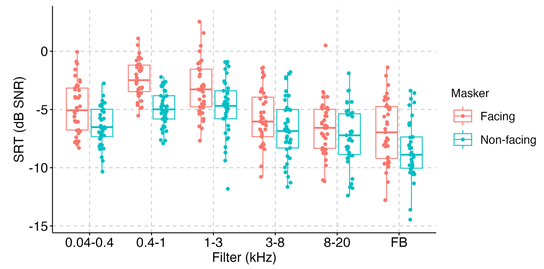
- Speech reception thresholds (SRTs) – SNR required for 50% correct performance – estimated for each condition.
- Linear mixed-effects models used to analyze effects of filtering and masker head orientation on SRT.
- Band importance for band '$i$' computed for each subject using the formula:

$$w_i = \frac{10^{(SRT_i - SRT_{FB})/10}}{\sum_j 10^{(SRT_j - SRT_{FB})/10}}$$

($SRT_{FB}$ is the SRT for the full-band condition)

- Effect of EHF pure tone thresholds on SRT examined.
- Exploratory logistic regression analysis of word-identification scores also conducted: sentence-level SNR, filtering, head orientation as fixed effects; subject, trial as random effects.

## Results



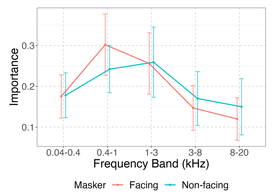

- SRTs in filtered conditions were *higher* (poorer) than for the FB (full-band; no filtering) condition, and *higher* when masker faced the listener.

**Linear mixed-effects model**:

- Model 1 (intercept=FB facing): main effects of all filtering conditions except 8-20 kHz.
- Model 2 (intercept=FB non-facing): main effects of all filtering conditions.
- main effect of masker head orientation in the FB condition.
- *no significant interaction terms*; however, model with interaction terms fit the data better (p=0.029).

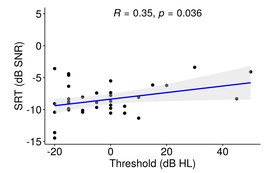
| | Intercept = FB, Facing | | Intercept = FB, Non-facing | |
|---|---|---|---|---|
| *Predictors* | *Estimates* | *p* | *Estimates* | *p* |
| (Intercept) | -6.87 | **<0.001** | -8.49 | **<0.001** |
| Filter[8-20k] | 0.26 | 0.520 | 1.21 | **0.003** |
| Filter[3-8k] | 1.17 | **0.004** | 1.79 | **<0.001** |
| Filter[1-3k] | 3.73 | **<0.001** | 3.68 | **<0.001** |
| Filter[0.4-1k] | 4.51 | **<0.001** | 3.57 | **<0.001** |
| Filter[0.04-0.4k] | 2.08 | **<0.001** | 2.15 | **<0.001** |
| Masker | -1.62 | **<0.001** | 1.62 | **<0.001** |
| 8-20k x Masker | 0.94 | 0.102 | -0.94 | 0.102 |
| 3-8k x Masker | 0.62 | 0.283 | -0.62 | 0.283 |
| 1-3k x Masker | -0.06 | 0.923 | 0.06 | 0.923 |
| 0.4-1k x Masker | -0.94 | 0.104 | 0.94 | 0.104 |
| 0.04-0.4k x Masker | 0.07 | 0.901 | -0.07 | 0.901 |

- Mean importance values of 3-8 kHz and 8-20 kHz bands *greater* when masker faced away.
- However, the standard deviation (error bar) was quite high in all bands.




## Results (continued)

- Better-ear 16-kHz thresholds were *significantly correlated with SRT* in the FB condition when masker faced away (p=0.036).
- However, *grouping listeners* based on 16-kHz thresholds (split at the median) did not significantly alter the filtering effects in the linear mixed effects model.




**Word-level analysis**:

- A generalized linear model indicated *significant interactions* between masker head orientation and each of 3-8 and 8-20 kHz bands, with *lower odds ratios when masker faced away.*

| *Predictors* | *Odds Ratios* | *p* |
|---|---|---|
| (Intercept) | 1.828 | **<0.001** |
| SNR | 1.363 | **<0.001** |
| Filter[8-20k] | 0.955 | 0.458 |
| Filter[3-8k] | 0.651 | **<0.001** |
| Filter[1-3k] | 0.311 | **<0.001** |
| Filter[0.4-1k] | 0.265 | **<0.001** |
| Filter[0.04-0.4k] | 0.561 | **<0.001** |
| Masker[Non-facing] | 1.769 | **<0.001** |
| 8-20k x Non-facing | 0.699 | **<0.001** |
| 3-8k x Non-facing | 0.830 | **0.038** |
| 1-3k x Non-facing | 0.875 | 0.132 |
| 0.4-1k x Non-facing | 1.054 | 0.549 |
| 0.04-0.4k x Non-facing | 0.878 | 0.139 |

## Conclusions

- While the pattern of SRT data suggested greater importance for EHFs when masker talkers faced away from the listener, the *effects of filtering on SRT were not significantly different* from the condition when maskers faced the listener.
- However, at the *word-level*, filtering EHFs did cause *significantly poorer identification scores* when the masker faced away.
- Despite a correlation between EHF pure-tone thresholds and $SRT_{FB}$, preliminary analyses do not indicate differences in the effects of filtering between listeners with better vs poorer thresholds.
- The effect of filtering EHFs on SRT when the masker faced 56° was about *1.2 dB*, which is smaller compared to the *1.8 dB* observed by Trine and Monson (2019) for 60°.
- Individual differences in masker *talker directionality* across stimuli could have influenced the effect of filtering EHFs.

## References

- ANSI. (1997). ANSI/ASA S3.5-1997 (R 2020). American National Standard Methods for Calculation of the Speech Intelligibility Index.
- Buss & Bosen (2021). Band importance for speech-in-speech recognition. JASA Express Letters, 1(8), 084402.
- Monson et al (2019). Ecological cocktail party listening reveals the utility of extended high-frequency hearing. Hearing Research, 381, 107773.
- Trine & Monson (2020). Extended High Frequencies Provide Both Spectral and Temporal Information to Improve Speech-in-Speech Recognition. Trends in Hearing, 24.

**This study was supported by NIH grant number R01-DC019745 to BBM.**

ILLINOIS