DOI: 10.61782/fa.2023.0081

EXTENDED HIGH FREQUENCY EFFECTS ON CHILDREN'S OPEN-SET SENTENCE RECOGNITION IN A TWO-TALKER MASKER

Mary Flaherty[1*] **Brian Monson**[1]
[1] Department of Speech and Hearing Science, University of Illinois Urbana-Champaign, USA

ABSTRACT

The goal of this study was to examine the extent to which children (5-17 yrs.) can take advantage of extended high frequency energy (EHF) when recognizing speech in a two-talker speech masker. Recent work demonstrated that EHF can be beneficial for adults' speech-in-speech recognition, but the role of EHF in children's speech recognition is not well understood. Given children's superior EHF hearing relative to adults, we hypothesized that EHF could be particularly useful to children in multitalker environments. The current study used an adaptive procedure to measure children's open-set sentence recognition in a two-talker masker. There were two filtering conditions: (1) full band stimuli, and (2) stimuli low-pass filtered at 8 kHz. In addition, because EHF energy in speech is dependent on talker head orientation, two masker head rotation conditions were tested: both maskers at 45 degrees or 60 degrees rotation, relative to the target talker. Results demonstrate children perform best when EHF was present, suggesting children can use EHF for speech-in-speech recognition. However, regardless of condition, overall performance was poorer for children compared to adults. This suggests that while EHF is a useful cue for children, increased EHF hearing sensitivity (relative to adults) did not increase their EHF benefit.

Keywords: *speech-in-speech recognition, children, extended high frequency hearing, head orientation*

[*]***Corresponding author**: maryflah@illinios.edu*

1. INTRODUCTION

Human hearing sensitivity to extended high-frequency (EHF) tones, which are typically above 8 kHz, is better in children compared to adults. However, age-related progressive loss of EHF hearing begins as early as late adolescence [1-5]. While EHF hearing has traditionally been considered to have little utility for speech perception, recent evidence contradicts this view [6]. Studies have shown that EHF hearing in adults contributes to speech localization, judgments of speech and voice quality, discrimination of a talker's head orientation, and speech recognition in noise [7-12]. Elevated EHF pure tone thresholds in otherwise normal-hearing adults have also been linked to poorer speech recognition in noise [13-15].
Research on EHF hearing in children is limited, with a focus on pure tone thresholds in quiet [3,16-18]. Some studies have shown that access to lower frequency bands that include some EHFs can improve speech perception in noise for children [19-20]. However, it remains unclear whether EHFs contribute to speech recognition in children. Children, who have enhanced EHF hearing relative to adults, may derive greater benefit from EHFs for speech perception.

EHF hearing may be especially important for when listening to speech in multitalker contexts. Children have greater difficulty than adults in recognizing speech in noisy environments, especially when the background is composed of competing talkers [21-24]. This is thought to be due, in part, to children's immature sound segregation and limitations in their ability to use spectro-temporally sparse cues for speech recognition speech [25-26].

In situations where multiple people are talking at the same time, listeners need to perceptually isolate the target speech from the competing masker and selectively attend to only those phonemic cues associated with the target. Acoustic differences, such as talker sex [27-28], language [29], and voice fundamental frequency [22,30], can facilitate this process by increasing perceptual dissimilarity between target and masker speech. Studies have shown that children's ability to use these acoustic differences for speech-in-speech recognition improves with age, but there are still developmental differences compared to adults.

The objective of the current study was to investigate how children use of EHFs and head orientation cues for speech-in-speech recognition during their school-age years. We hypothesized that children, with their generally heightened EHF hearing, would benefit more from EHFs compared to adults, and that children would also benefit from head orientation cues. Similar to natural listening environments where talkers face different directions, the head orientation of competing talkers was considered due to the directional nature of EHFs Furthermore, we aimed to explore whether child age would influence the degree of benefit from these cues. By determining the effectiveness of EHFs and head orientation cues in speech-in-speech recognition for children, we can gain insights into the role of EHF in speech perception during development and improve our understanding of the auditory mechanisms that children employ in complex listening environments. Additionally, if children rely more on EHF cues than adults in such contexts, it could have implications for children with high-frequency hearing loss.

2. METHODS

2.1 Participants

Thirty-nine children (ages 5.1-17.8 years, 19 females) were recruited from the local Central Illinois area. Criteria for participation include (1) native speaker of English, (2) normal hearing bilaterally, with thresholds ≤20 dB HL for octave frequencies between 250 and 16000 Hz (31), (3) no history of cognitive or language disorders. Data from adult listeners (n=18; ages 20-27 years, 14 female) with normal hearing were taken from (10) for comparison to mature performance. All parents provided written informed consent and children provided verbal assent prior to participation. Children were paid $15/hour for their participation.

2.2 Stimuli

The target speech stimuli were obtained from the Bamford-Kowal-Bench (BKB) sentences dataset [32], which were uttered by a single female talker and recorded in a sound-treated booth using a Class 1 precision microphone positioned at 0°. The original recordings had a sampling rate of 44.1 kHz and a precision of 16 bits. For the low-pass filtered condition, all stimuli, including the masker and target speech, were filtered using a 32-pole Butterworth filter with a cutoff frequency of 8 kHz. For the full-band condition, stimuli were filtered with a cutoff frequency of 20 kHz. The masker stimulus used in this study consisted of a babble created by combining recordings from a high-fidelity database of anechoic multi-channel recordings, as reported by [33]. Recordings from microphones positioned at 45° and 60° were used to create the masker stimuli. Semantically unpredictable speech signals were generated for each talker at each angle, and the speech signals from the two talkers were combined to create two versions of the masker stimulus, one with both talkers facing 45° and another with both talkers facing 60°.

2.3 Procedure

The experiment followed the methods conducted by Monson [33] with adult participants, but with modifications for child listeners. These modifications included: (1) increasing starting level (dB SNR) of the stimuli; 2) reducing the number of sentences per adaptive track; and (3) having an experimenter present inside the booth with the child. Further details regarding these changes can be found in the text below. All experimental procedures were approved by the Institutional Review Board at the University of Illinois at Urbana-Champaign.

Child participants were seated in a sound-treated booth and presented with co-located target and masker stimuli through a loudspeaker positioned 1 m directly in front of them. Children were instructed to repeat back the target sentence and were allowed to guess when unsure. An experimenter scored each key word as correct or incorrect.

Speech reception thresholds (SRT) were estimated using an adaptive procedure. The level of the two-talker masker was set at 70 dB SPL at 1 m, while the level of the target signal was adaptively varied. Two interleaved adaptive tracks were used, both employing a one-down, one-up adaptive rule. For one track, the signal-to-noise ratio (SNR) was reduced if the child correctly repeated one or more words; otherwise, it was increased. For the other track, the SNR was reduced if the child correctly repeated all words or all but one word; otherwise, it was increased. Only exact matches to the keyword were considered correct. The initial step size for adjusting the SNR was set at 4 dB, which was reduced to 2 dB after the first reversal.

Prior to testing, 16 practice trials were administered to familiarize the children with the task. During practice, both adaptive tracks started at 10 dB SNR. All children were able to successfully complete the practice. During testing, both adaptive tracks started at 7 dB SNR. Each of the two tracks comprised 20 sentences (modified from 32 sentences used in the adult experiment). Word level data from the two tracks were combined and fitted with a logit function with asymptotes at 0 and 100% correct. The SRT was defined as the SNR associated with 50% correct. Data fits were associated with r^2 values ranging from 0.64 to 0.99, with a median value of 0.89.

Two filtering conditions were tested: full band vs. all stimuli low-pass filtered at 8 kHz. Two masker head angle conditions were tested: both maskers facing 45° or both maskers facing 60° relative to the target talker. After the training, the four conditions (2 filtering conditions x 2 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 BKB sentence lists.

3. RESULTS

Children exhibited higher (poorer) SRTs compared to adults across all four conditions, consistent with prior findings indicating that children are more vulnerable than adults to the effects of competing speech maskers [26]. Additionally, there was a trend for lower (better) SRTs in conditions with EHF as well as conditions with a greater head angle, regardless of age. On average, children had higher SRTs in conditions without EHF (45° condition = -2.3 dB [SD = 2.9]; 60° condition = -3.3 dB [SD = 2.9]) compared to conditions with EHF (45° condition = -3.7 dB [SD = 3.0]; 60° condition = -5.0 dB [SD = 3.0]). The mean improvement in SRTs with EHF present was 1.5 dB across both conditions, which is similar to the mean improvement observed in adults (1.7 dB)[10]. Individual variability in children's SRTs was also evident, consistent with previous research on children's speech-in-speech recognition [22,26,28,34]. A mixed-effects model was fit using the afex command "mixed". P-values were estimated using a parametric bootstrapping method. There was a main effect of age group ($\chi2$ =30.40, $p < 0.001$), a main effect of EHF ($\chi2$ =61.95, $p < 0.001$), and a main effect of masker head orientation angle ($\chi2$ =46.86, $p < 0.001$). The two- and three-way interactions between these factors were not significant ($p > 0.243$). The lack of significant interactions indicates that the effects of EHF and of masker head orientation angle did not differ by age group.

A linear mixed model was used to evaluate SRTs within the child group. The model included the fixed effects of child age, filtering condition, and angle condition, as well as their interactions. There was a significant effect of EHF [β = 1.16, S.E. = 0.63, $t(111) = 1.86$, $p = 0.0325$], a significant effect of masker head orientation angle [β = -1.62, S.E. = 0.63, $t(111) = -2.58$, $p = 0.0055$],and a significant effect of age [β = -8.18, S.E. = 2.87, $t(37) = -2.85$, $p = 0.0035$], but no interactions ($p > 0.23$). The age effect indicates that mean SRTs decreased with increasing age for all conditions. Performance was best in the conditions with EHF cues present and in conditions when the masker head orientation angle was 60°. The absence of interactions suggests the magnitude of benefit did not differ by age. One-tailed bivariate correlation between the magnitude of EHF benefit and age was $r = 0.15$ ($p = 0.183$) in the 45° condition and was $r = 0.18$ ($p = 0.136$) in the 60° condition. The one-tailed correlation between the magnitude of head orientation angle benefit was $r = 0.25$ ($p = 0.072$) in the condition without EHFs and was $r = 0.14$ ($p = 0.197$) in the condition with EHFs.

4. DISCUSSION

The current study investigated the impact of limiting access to extended high frequency (EHF) energy in speech during children's multitalker listening situations. The results revealed that EHF energy above 8 kHz is beneficial for children, as its presence improved overall performance, even though speech recognition was still possible in the absence of EHFs. The head orientation of competing talkers also impacted performance as predicted. However, contrary to our expectations, there was no evidence to suggest that children take greater advantage of these cues compared to adults. Like adults, children's performance was significantly worse in conditions where EHFs were removed through low-pass filtering, as well as in conditions where the masker talkers were facing more towards the listener (10). Thus, there is no indication that children's heightened ability to detect EHFs translates into an enhanced ability to utilize EHFs in the presence of competing talkers. Additionally, there was no evidence of a relationship between age and the ability to utilize head orientation cues for speech-in-speech recognition within the child group. These findings highlight that EHFs play a role in speech-in-speech understanding for children as young as 5 years old, providing further support to the growing evidence of EHFs' importance in speech perception [6]. These findings also support previous work demonstrating children's increasing susceptibility to competing speech relative to adults [21-24]. Importantly, this study is the first to demonstrate the utility of EHFs for children's speech-in-speech recognition, as previous research has mainly focused on EHF pure-tone audibility. These findings have practical implications for real-world listening situations for children, such as in classrooms, where they need to understand speech from teachers and peers amidst competing noise. EHFs also have implications for the fitting of hearing aids or other amplification devices for children with hearing loss, as providing appropriate amplification in the higher frequency range can improve their ability to perceive speech.

In summary, EHFs can be used by children during speech recognition and appear to provide useful auditory cues in challenging listening conditions, contributing to their overall ability to understand speech in noisy environments.

5. ACKNOWLEDGMENTS

The authors thank Kelsey Libert for her contributions throughout the research process. We also thank Amber Dorn and Jasmine Thakkar for their assistance with data collection and Sa Shen for her assistance with statistical analyses. This work was supported the Department of Speech and Hearing Science at the University of Illinois at Urbana-Champaign.

6. REFERENCES

[1] Green, D.M., Kidd, G., Jr., Stevens, K.N., 1987. High-frequency audiometric assessment of a young adult population. J. Acoust. Soc. Am. 81, 485-494.

[2] Rodriguez Valiente, A., Trinidad, A., Garcia Berrocal, J.R., Górriz, C., Ramirez Camacho, R., 2014. Extended high-frequency (9-20 kHz) audiometry reference thresholds in 645 healthy subjects. Int. J. Audiol.53, 531-545.

[3] Schechter, M.A., Fausti, S.A., Rappaport, B.Z., Frey, R.H., 1986. Age categorization of high-frequency auditory threshold data. J. Acoust. Soc. Am. 79, 767-771. https://doi.org/10.1121/1.393466

[4] Stelmachowicz, P.G., Beauchaine, K.A., Kalberer, A., Jesteadt, W., 1989. Normative thresholds in the 8- to 20-kHz range as a function of age. J. Acoust. Soc. Am. 86, 1384-1391. https://doi.org/10.1121/1.398698

[5] Trehub, S.E., Schneider, B.A., Morrongiello, B.A., Thorpe, L.A., 1989. Developmental changes in high-frequency sensitivity. Int. J. Audiol. 28, 241-249. https://doi.org/10.3109/00206098909081629

[6] 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. Hear. Res. 397, 1-14. https://doi.org/10.1016/j.heares.2020.107922

[7] Best, V., Carlile, S., Jin, C., van Schaik, A., 2005. The role of high frequencies in speech localization. J. Acoust. Soc. Am. 118, 353-363. https://doi.org/10.1121/1.1926107

[8] Monson, B.B., Lotto, A.J., Story, B.H., 2014a. Detection of high-frequency energy level changes in speech and singing. J. Acoust. Soc. Am. 135, 400-406. https://doi.org/10.1121/1.4829525

[9] Moore, B.C.J., Tan, C.T., 2003. Perceived naturalness of spectrally distorted speech and music. J. Acoust. Soc. Am. 114, 408-419. https://doi.org/10.1121/1.1577552

[10] Monson, B.B., Rock, J., Schulz, A., Hoffman, E., Buss, E., 2019. Ecological cocktail party listening reveals the utility of extended high-frequency hearing. Hear. Res. 381, 1-7. https://doi.org/10.1016/j.heares.2019.107773

[11] 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. Proc. Natl. Acad. Sci. USA 116, 23753-23759. https://doi.org/10.1073/pnas.1903315116

[12] Trine, A., Monson, B.B. (2020), Extended high frequencies provide both spectral and temporal information to improve speech-in-speech recognition. Trends Hear. 24,1-8. https://doi.org/10.1177/2331216520980299

[13] 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. J. Acoust. Soc. Am. 129, 852-863. https://doi.org/10.1121/1.3523476

[14] Corbin, N., Buss, E., Monson, B., 2019. Effect of talker orientation on speech perception in spatially-separated speech, American Auditory Society, Scottsdale, AZ, February 2019.

[15] Yeend, I., Beach, E.F., Sharma, M., 2019. Working memory and extended high-frequency hearing in adults: Diagnostic predictors of speech-in-noise perception. Ear Hear. 40, 458-467.https://doi.org/ 10.1097/AUD.0000000000000640

[16] Schneider, B. A., Trehub. S. E., & Bull, D. (1980). High-frequency sensitivity in infants. Science, 207, 1003-1004.

[17] Trehub, S.E., Schneider, B.A., Morrongiello, B.A., Thorpe, L.A., 1989. Developmental changes in high-frequency sensitivity. Int. J. Audiol. 28, 241-249. https://doi.org/ 10.3109/00206098909081629

[18] Trehub, S.E., Schneider, B.A., Morrongiello, B.A., Thorpe, L.A., 1989. Developmental changes in high-frequency sensitivity. Int. J. Audiol. 28, 241-249. https://doi.org/ 10.3109/00206098909081629

[19] Stelmachowicz, P.G., Lewis, D.E., Choi, S., Hoover, B., 2007. Effect of stimulus bandwidth on auditory skills in normal-hearing and hearing-impaired children. Ear Hear. 28, 483-494. https://doi.org/10.1097/AUD.0b013e31806dc265

[20] Pittman, A.L., 2008. Short-term word-learning rate in children with normal hearing and children with hearing loss in limited and extended high-frequency bandwidths. J. Speech. Lang. Hear. Res. 51, 785-797. https://doi.org/10.1044/1092-4388(2008/056

[21] Corbin, N.E., Bonino, A.Y., Buss, E., Leibold, L.J., 2016. Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers. Ear Hear. 37, 55-63. https://doi.org/10.1097/AUD.0000000000000201

[22] Flaherty, M.M., Buss, E., Leibold, L.J., 2019. Developmental effects in children's ability to benefit from F0 differences between target and masker speech. Ear Hear. 40, 927-937. https://doi.org/ 10.1097/AUD.0000000000000673

[23] Hall, J.W., Grose, J.H., Buss, E., Dev, M.B., 2002. Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. Ear Hear. 23, 159-165. https://doi.org/ 10.1097/00003446-200204000-00008

[24] Leibold, L.J., Buss, E., 2013. Children's identification of consonants in a speech-shaped noise or a two-talker masker. J. Speech Lang. Hear. 56, 1144-1155. https://doi.org/10.1044/1092-4388(2012/12-0011

[25] Buss, E., Hodge, S.E., Calandruccio, L., Leibold, L.J., Grose, J.H., 2019. Masked sentence recognition in children, young adults, and older adults: Age-dependent effects of semantic context and masker type. Ear Hear. 40, 1117-112.

[26] Buss, E., Hodge, S.E., Calandruccio, L., Leibold, L.J., Grose, J.H., 2019. Masked sentence recognition in children, young adults, and older adults: Age-dependent effects of semantic context and masker type. Ear Hear. 40, 1117-1126. https://doi.org/ 10.1097/AUD.0000000000000692

[27] Leibold, L.J., Buss, E., Calandruccio, L., 2018. Developmental effects in masking release for speech-in-speech perception due to target/masker sex mismatch. Ear Hear. 39, 935-945. https://doi.org/10.1097/AUD.0000000000000554

[28] Wightman, F.L., Kistler, D.J., 2005. Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. J. Acoust. Soc. Am. 118, 3164-3176. https://doi.org/10.1121/1.2082567

[29] Calandruccio, L., Leibold, L.J., Buss, E., 2016. Linguistic masking release in school-age children and adults. Am. J. Audiol. 25, 34-40.

[30] Flaherty, M.M., Browning, J., Buss, E., Leibold, L.J., 2020. Effects of hearing loss on school-age children's ability to benefit from F0

differences between target and masker speech. Ear Hear, 1, 1084-1096.

[31] ANSI, 2018. American National Standard Specifications for Audiometers. ANSI S3.6-2018. New York

[32] Bench, J., Kowal, A., Bamford, J., 1979. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. Br. J. Audiol. 13, 108-112.

[33] Monson, B.B., Hunter, E.J., Story, B.H., 2012a. Horizontal directivity of low- and high-frequency energy in speech and singing. J. Acoust. Soc. Am. 132, 433-441. https://doi.org/10.1121/1.4725963

[34] Wightman, F., Kistler, D., Brungart, D., 2006. Informational masking of speech in children: Auditory-visual integration. J. Acoust. Soc. Am. 119, 3940-3949. https://doi.org/10.1121/1.2195121