Published in final edited form as:

Ear Hear. 2022 ; 43(1): 90–100. doi:10.1097/AUD.0000000000001081.

Effect of Masker Head Orientation, Listener Age, and Extended High-Frequency Sensitivity on Speech Recognition in Spatially Separated Speech

Meredith D. Braza[1], **Nicole E. Corbin**[2], **Emily Buss**[3], **Brian B. Monson**[4,5]

[1]Division of Speech and Hearing Sciences, The University of North Carolina, Chapel Hill, North Carolina, USA

[2]Department of Communication Science and Disorders, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

[3]Department of Otolaryngology/Head and Neck Surgery, The University of North Carolina, Chapel Hill, North Carolina, USA

[4]Department of Speech and Hearing Science, University of Illinois, Urbana-Champaign, Urbana-Champaign, Illinois, USA

[5]Neuroscience Program, University of Illinois, Urbana-Champaign, Urbana-Champaign, Illinois, USA

Abstract

Objectives: Masked speech recognition is typically assessed as though the target and background talkers are all directly facing the listener. However, background speech in natural environments is often produced by talkers facing other directions, and talker head orientation affects the spectral content of speech, particularly at the extended high frequencies (EHFs; > 8 kHz). This study investigated the effect of masker head orientation and listeners' EHF sensitivity on speech-in-speech recognition and spatial release from masking in children and adults.

Design: Participants were 5- to 7-year-olds (*n* = 15) and adults (*n* = 34), all with normal hearing up to 8 kHz and a range of EHF hearing thresholds. Speech reception thresholds (SRTs) were measured for target sentences recorded from a microphone directly in front of the talker's mouth and presented from a loudspeaker directly in front of the listener, simulating a target directly in front of and facing the listener. The maskers were two streams of concatenated words recorded from a microphone located at either 0° or 60° azimuth, simulating masker talkers facing the listener or facing away from the listener, respectively. Maskers were presented in one of three spatial conditions: co-located with the target, symmetrically separated on either side or the target (+54° and −54° on the horizontal plane), or asymmetrically separated to the right of the target (both +54° on the horizontal plane).

Corresponding author: Emily Buss, 170 Manning Dr., G190 Houpt Bldg., CB7070, Chapel Hill, NC 27599, ebuss@med.unc.edu, (919) 843-9163.

Conflicts of interest: The authors have no conflicts of interest to disclose.

Results: Performance was poorer for the facing than for the non-facing masker head orientation. This benefit of the non-facing masker head orientation, or head orientation release from masking (HORM), was largest in the co-located condition, but it was also observed for the symmetric and asymmetric masker spatial separation conditions. SRTs were positively correlated with the mean 16-kHz threshold across ears in adults for the non-facing conditions but not the facing masker conditions. In adults with normal EHF thresholds, the HORM was comparable in magnitude to the benefit of a symmetric spatial separation of the target and maskers. Although children benefited from the non-facing masker head orientation, their HORM was reduced compared to adults with normal EHF thresholds. Spatial release from masking was comparable across age groups for symmetric masker placement, but it was larger in adults than children for the asymmetric masker.

Conclusions: Masker head orientation affects speech-in-speech recognition in children and adults, particularly those with normal EHF thresholds. This is important because masker talkers do not all face the listener under most natural listening conditions, and assuming a midline orientation would tend to overestimate the effect of spatial separation. The benefits associated with EHF audibility for speech-in-speech recognition may warrant clinical evaluation of thresholds above 8 kHz.

Keywords

auditory scene analysis; development; hearing loss; recording angle; spatial release from masking; speech-in-speech recognition; speech testing

Introduction

Speech communication often takes place in complex acoustic environments, such as busy restaurants and elementary school classrooms (Argus et al. 2009; McKellin et al. 2011). Listening to speech in the presence of background speech can be particularly challenging, relying on sound source segregation, selective attention, and linguistic knowledge (reviewed by Sobon et al. 2019). Collectively these listening challenges are often referred to as the "cocktail party problem" (Cherry 1953). Speech-in-speech recognition tends to be better in young adults than children (Wightman & Kistler 2005; Buss et al. 2017), and better in adults with normal hearing than those with hearing loss (Arbogast et al. 2005). It is well known that spatial separation of the target and masker on the horizontal plane can improve performance (e.g., Dirks & Wilson 1969; Noble & Perrett 2002), with smaller effects for young school-age children (Yuen & Yuan 2014; Corbin et al. 2016) and adults with hearing loss (Kidd et al. 2019) compared to adults with normal hearing. A recent study by Monson and colleagues (2019) suggested that differences in target and masker head orientation provide cues related to extended high-frequency (EHF; > 8 kHz) content that can also confer benefit for speech-in-speech recognition. This benefit presumably requires EHF audibility, which is reduced early in presbycusis (Matthews et al. 1997) and is commonly limited even in young adults (Green et al. 1987). It is unknown to what extent children rely on EHF cues. To better understand these effects, the present study evaluated speech-in-speech recognition for co-located and spatially separated stimuli as a function of masker head orientation, listener age, and EHF hearing sensitivity.

Spatial Separation and Talker Head Orientation

Speech recognition in the presence of background speech varies depending on the relative spatial positions of the target and background talkers. Performance is better when the target and masker are perceived as originating from different points on the horizontal plane, as opposed to co-located; perceptual differences in source location can be achieved by presenting the target and masker from different locations, or by manipulating perceived spatial location using the precedence effect (Freyman et al. 2001; King et al. 2019). The benefit of spatial separation, referred to as spatial release from masking (SRM), is thought to play an important role in functional hearing abilities (Vannson et al. 2015; Phatak et al. 2018), and several groups have developed tools to evaluate SRM clinically in both children and adults (Cameron & Dillon 2007; Jakien et al. 2017). Under natural listening conditions, SRM is believed to be the result of multiple factors including interaural differences that help to perceptually distinguish the target from the masker, improvements in signal-to-noise associated with the head shadow effect, and the combination of speech cues across ears (Brungart & Iyer 2012; Schoenmaker et al. 2016; Ellinger et al. 2017; Dieudonné & Francart 2018). The magnitude of SRM is larger for two-talker speech than speech-shaped noise in both children and adults with normal hearing (Freyman et al. 2001; Corbin et al. 2017), presumably due at least in part to the greater perceptual similarity between target and masker speech in the co-located condition.

Speech recognition in a speech masker also depends on the head orientation of the talkers relative to the listener. For example, Strelcyk and colleagues (2014) reported that digit recognition in a two-talker masker improved when stimuli were manipulated to simulate different head orientations for the target and masker talkers. Monson and colleagues (2019) also demonstrated an effect of masker head orientation and evaluated the role of EHF audibility in speech recognition under conditions of mismatched target and masker head orientation. In that study, target speech was recorded from 0°, and two masker talkers were recorded from either 45° or 60°. The target and masker speech were presented from a single loudspeaker (co-located) that was directly facing the listener, and speech reception thresholds (SRTs) were evaluated with and without an 8-kHz low-pass filter. In the full-band condition, a 15° change in masker head orientation (from 45° to 60°) resulted in approximately 2 dB of improvement in SRTs. Removing EHF content via low-pass filtering elevated SRTs by 1.4 dB and 2.0 dB for the 45° and 60° masker head orientations, respectively; effects of filtering were not significantly different for these two head orientations. These results provide evidence that EHF sensitivity plays a role in speech-in-speech recognition in natural multi-talker listening environments.

Information Provided at Extended High Frequencies

In addition to supporting speech-in-speech recognition, EHF audibility provides cues for judging the location and orientation of a sound source. It is well known that EHF content provides cues to elevation and front/back location for both speech (Best et al. 2005; Monson et al. 2014) and environmental sounds (Heffner & Heffner 2008). Human listeners can also judge orientation of a talker or a loudspeaker presenting speech based entirely on acoustic properties of sound (Kato et al. 2010; Imbery et al. 2019), with the greatest sensitivity around 0° (i.e., with the sound source facing the listener). The just noticeable difference in head orientation relative to 0° is better for full bandwidth stimuli than for stimuli that have been low-pass filtered at 8 kHz (Monson et al. 2019). The cues associated with orientation include overall level, interaural level differences, and spectral tilt; some of these cues are particularly pronounced at EHFs. In a sound-treated room, the EHF content associated with a talker's voice is maximized at the listener's ears when that talker is facing the listener (0°), and drops as the talker's head rotates off midline; attenuation in the octave band around 16 kHz is ~3 dB for 45° and ~30 dB for 180° head orientations (Monson et al. 2012; Kocon & Monson 2018).

Historically, hearing scientists believed that the spectral content required for accurate speech recognition extends only up to 4 or 8 kHz (Monson et al. 2014). However, current research indicates that EHF content also improves the quality and intelligibility of speech for children and adults with normal hearing (Monson et al. 2014; Hunter et al. 2020; Trine & Monson 2020). Listeners with high-frequency hearing loss rate sound quality as improved when target speech in these high-frequency regions (up to 10 kHz) is amplified (Moore et al. 2011; Arbogast et al. 2019). The speech spectrum contains a significant amount of EHF energy (Monson et al. 2012; Levy et al. 2015), which could convey speech information or serve as a non-spatial segregation cue for listeners in complex auditory environments (Monson et al. 2019; Trine & Monson 2020). If the benefits of EHF audibility in multi-source environments are due to the provision of a segregation cue, it is possible that this cue might be more helpful when the target and masker are co-located, where few other segregation cues are available, compared to the spatial separation condition, where segregation may be less of a challenge.

Listeners with clinically normal audiograms up to 8 kHz display substantial variability in speech-in-noise performance. One important question that remains unanswered is whether hearing loss at EHFs contributes to these observed difficulties with masked speech recognition (Hunter et al. 2020). Although ototoxicity monitoring involves measuring EHF thresholds (American Academy of Audiology 2009), standard audiologic procedures for both children and adults entail measuring hearing thresholds at octave frequencies only up to 8 kHz. If EHF hearing plays a role in speech-in-noise recognition, a relationship between EHF thresholds and speech-in-noise performance might be expected, but there are mixed findings on this topic. For instance, Badri et al. (2011) found that listeners who self-report and exhibit speech-in-noise difficulties have elevated EHF thresholds at 12.5 and 14 kHz relative to controls. Motlagh Zadeh et al. (2019) likewise reported a relationship between self-report of speech-in-noise difficulty and the severity of EHF loss at 10, 12.5, 14, and 16 kHz. They also found a correlation between average EHF thresholds and speech-in-noise performance measured using broadband speech and a broadband speech-shaped-noise masker. Yeend et al. (2019) found a correlation between average EHF thresholds (9, 10, 11.2, and 12.5 kHz) and a composite speech score that included both self-reported and measured speech-in-noise ability. Trine and Monson (2020) tested a group of young adults with normal hearing through 16 kHz; in that cohort, correlations between sentence recognition in a two-talker masker and pure-tone detection thresholds were observed to be stronger at 12.5 and 16 kHz than at lower frequencies.

On the other hand, Liberman et al. (2016) found no relationship between average EHF thresholds (9, 10, 11.2, 12.5, 14, and 16 kHz) and speech-in-noise performance, although the speech materials used in that study were low-pass filtered at 8.8 kHz (Noffsinger et al. 1994). Smith et al (2019) also failed to find a relationship between average EHF thresholds (10, 12.5, and 14 kHz) and speech-in-noise scores; however, all listeners in that study had relatively good EHF thresholds (averaging better than 10 dB HL at all frequencies), and it is unclear whether speech materials were bandlimited. Finally, Prendergast et al. (2019) reported that speech-in-noise performance was predicted by statistical models that included age, noise exposure, and 16-kHz thresholds as predictors; however, model predictions were improved when the 16-kHz threshold was replaced with pure-tone thresholds at standard audiometric frequencies.

Compared to adults, children require a higher signal-to-noise ratio (SNR) to recognize speech, and this effect is more pronounced for speech presented in a two-talker masker compared to speech-shaped noise (e.g., Wightman & Kistler 2005; Buss et al. 2017; Buss et al. 2019). Children also require greater stimulus bandwidths than adults to recognize speech (Stelmachowicz et al. 2001; Mlot et al. 2010; McCreery & Stelmachowicz 2011). Although young children appear to require a greater quality and quantity of cues than adults for correct speech recognition (e.g., McCreery & Stelmachowicz 2011), they tend to have good EHF hearing sensitivity. This raises the possibility that children may benefit from EHF content to a greater extent than adults.

Motivation and Predictions for the Present Study

The vast majority of research investigating SRM has been conducted with target and masker speech stimuli recorded from a microphone placed directly in front of the talker, simulating a listening environment in which all talkers are directly facing the listener. In the real world, masking speech is produced by talkers with a range of head orientations relative to the listener, a situation which introduces talker-specific differences in EHF content, among other cues. While Monson et al. (2019) confirmed that masker orientation significantly affects speech-in-speech performance in young adults with normal EHF thresholds, it is not clear whether this effect is also observed when the masker is spatially separated from the target speech, or whether this result generalizes to a broader population of listeners. As such, the motivation for the present study was to determine whether masker head orientation affects speech recognition under different spatial configurations in children and adults with normal hearing, and to determine the role of EHF sensitivity on performance. To that end, we measured SRTs in a two-talker masker for children and young adults with normal audiometric thresholds (≤ 8 kHz), with and without spatial separation on the horizontal plane, and with masker recordings made at 0° (like the target) and at 60°. These recordings simulate masker talkers facing and not facing the listener, respectively. The latter condition, with maskers facing away from the listener, simulates a situation that may more closely mimic a realistic cocktail party scenario than the case of all talkers facing the listener.

There were four primary predictions in this study. First, SRTs were expected to improve for the non-facing (60°) masker head orientation compared to the facing (0°) orientation, a benefit we will describe as head orientation release from masking (HORM). Second, SRM was expected to be smaller for the non-facing (60°) masker head orientation compared to the facing (0°) orientation due to HORM in the baseline, co-located condition. Third, a negative correlation was expected between EHF thresholds and the magnitude of HORM. Fourth, SRTs were expected to be poorer for children than for adults overall, as observed in previous speech-in-speech experiments, but HORM was expected to be greater for children than adults due to children's greater bandwidth requirements for speech recognition.

Materials and Methods

Participants

Participants were 15 children (5.1-7.8 yrs, mean 6.6 yrs, 9 female) and 34 adults (20.1-56.5 yrs, mean 32.4 yrs, 27 female). Pure-tone air conduction thresholds were measured at octave frequencies 0.25 – 8 kHz, as well as at EHFs of 11.2 and 16 kHz, using professionally calibrated Sennheiser HDA 200 circumaural headphones. Inclusion criteria were: (1) air conduction thresholds of 20 dB HL or less from 0.25 to 8 kHz, bilaterally (ANSI 2004), (2) typical development (for children, by parent report), (3) native American-English speaking, (4) normal middle ear function based on standard 226-Hz tympanometry, and (5) no history of chronic ear disease by self or parent report. All child participants had normal EHF thresholds (20 dB HL or less at 11.2, 16, and 20 kHz) bilaterally. Adults were separated into two groups based on EHF thresholds[1]. There were 17 adults with bilaterally normal EHF thresholds (20.1-34.2 yrs, mean 26.1 yrs, 13 female) and 17 adults with one or more elevated EHF threshold (23.1-56.5 yrs, mean 38.8 yrs, 14 female). Adults with normal EHF thresholds were significantly younger than those with elevated EHF thresholds when assessed with a Welch's t-test ($t_{22.3} = -4.58$, $p < 0.001$). Test procedures were approved by The University of North Carolina at Chapel Hill Institutional Review Board.

Stimuli and Conditions

Target speech was produced by a female native American-English speaker, with a mean F0 of 245 Hz. This talker read Revised Bamford-Kowal-Bench (BKB; Bench et al. 1979) sentences, which comprise 21 lists of 16 sentences, with 3-4 keywords in each sentence. Recordings were made in a double-walled sound treated room at a sampling rate of 44.1 kHz, with a ½-inch precision microphone (Brüel and Kjær 4189, Denmark) and windscreen placed 6 in directly in front of the talker's mouth (0°).

Two-talker masker speech was taken from a database of fully anechoic, multi-directional recordings made with ½-inch precision microphones (Larson Davis 2551, Provo, UT) positioned at 15° intervals, including at 0° (directly in front of the talker) and 60° (to the right of the talker), as described by Monson et al. (2012). Talkers were recorded reading non-sense sentences at a sampling rate of 44.1 kHz. The two masker talkers used for the present study were female and native American-English speakers, with mean F0s of 220 Hz and 224 Hz; these two talkers were selected because they were judged to be perceptually similar to the target talker. Recordings from each microphone condition were RMS

---

[1] Thresholds were not measured at 11.2 kHz for one adult. For this listener, thresholds were ≤ 10 dB HL at both 8 and 16 kHz. For the purposes of determining group membership, this adult was assumed to have normal thresholds at 11.2 kHz.

normalized. Words in these recordings, including articles, were isolated and concatenated in random order, separated by 50 ms of silence. The resulting masker recordings were truncated to 1.8 min in total duration. Target sentences were each approximately 1.5 to 2 secs in duration. The long-term magnitude spectra of the target, facing (0°) masker, and non-facing (60°) masker are shown in Figure 1.

Procedure

Testing was conducted in a 10 x 10, double-walled sound booth. Participants sat in an adjustable chair facing the middle loudspeaker of an arc of 11 loudspeakers (JBL LSR305, Los Angeles, CA), equally spaced in 18° increments from −90° to 90°. Each loudspeaker was connected via balanced line to one channel of a soundcard (MOTU 24i, MOTU, Cambridge, MA). Once the participant was seated, the chair was adjusted in height so the participant’s ear canals were on the same horizontal plane as the center of each speaker cone; in this position, every speaker was approximately 1 m from the center of the participant’s head. Target sentences were always presented from the front speaker (0° on the horizontal plane). There were three spatial conditions. In the co-located condition, both maskers were co-located with the target at 0° on the horizontal plane. In the asymmetrical condition, both maskers were played from +54° on the horizontal plane (to the right of the listener). In the symmetrical condition, one masker was played from +54° and the other from −54° on the horizontal plane (to the right and left of the listener, respectively). Each spatial condition was tested twice, once with the facing maskers (recorded at 0°) and once with the non-facing maskers (recorded at 60°), totaling to six stimulus conditions. These conditions are illustrated in Figure 2. The order of the stimulus conditions was randomized for each participant.

Simulating masker talker orientation using microphone recording angle and a fixed loudspeaker position has some advantages over other approaches used in previous studies. Those other approaches include: 1) playing speech recorded from 0° and rotating the loudspeaker relative to the talker (Moore & Popelka 2013; Imbery et al. 2019), and 2) low-pass filtering recordings made from 0° to match the off-axis long-term power spectrum (Strelcyk et al. 2014). Neither of these alternative procedures preserves the dynamic phoneme-level changes in directivity patterns associated with changes in talker head position (Monson et al 2012; Kocon & Monson, 2018). While recording speech from different angles preserves phoneme-level spectral cues, it fails to capture orientation-specific reverberation in the test environment and subtle interaural differences associated with frequency-specific acoustic propagation under natural listening conditions. RMS normalizing stimuli also removed the ~1 dB overall level reduction associated with increasing recording angle from 0° to 60° (Monson et al., 2012). However, these factors were all considered to be of secondary importance relative to differences in frequency content captured via recording angle.

The experiment was controlled using custom MATLAB scripts (MathWorks, Natick, MA). Stimuli were presented at a playback rate of 44.1 kHz. The target was temporarily centered in a masker sample that began 500 ms prior to target onset and ended 500 after target onset; masker samples were randomly chosen prior to each trial. Participants were instructed to repeat as many target words as they could after each target sentence was played. An experimenter scored each keyword as correct or incorrect. Participants were provided verbal encouragement following each response that varied by accuracy of the response (e.g., “good job” if the keywords were correct or nearly correct, and “nice try” if the keywords were completely incorrect). Participants were not otherwise given feedback about their responses.

The target plus masker level was 60 dB SPL irrespective of the SNR; this level was chosen because it is the average level of conversational speech (Olsen 1998). The SRT was defined as the SNR associated with 50% correct keyword recognition. Each estimate was based on two interleaved one-down, one-up adaptive tracks, each containing a list of 16 sentences. One track used a lax criterion for considering the response correct (≥ 1 keyword correct) and the other used a strict criterion (≤ 1 keyword incorrect). The initial step size for each track was 8 dB; this was reduced to 4 dB after the first reversal and 2 dB after the second reversal. Participant responses by keyword were combined across tracks and fitted with a logit function, and the midpoint of this function served as the final estimate of 50% correct (e.g., Sobon et al. 2019). The first sentence list used for the first adaptive track in a session was selected at random, and testing proceeded sequentially through the corpus. This procedure ensured that participants never heard a target sentence more than once. The initial SNR for an adaptive track depended on the spatial condition, with values of 10 dB SNR (co-located), 7-10 dB SNR (symmetrical), and 3 dB SNR (asymmetrical). Children and adults completed all testing in one visit of 1.5-2.5 hrs, including breaks between each track if needed.

Data Analysis

Data were analyzed in R (R Core Team 2016). Linear mixed models were implemented using nlme (Pinheiro et al. 2016), with selection of random intercepts based on the Akaike Information Criterion (AIC). All models included a random intercept for participant and spatial condition. Individual models included participant sex as a predictor variable, although no effect of sex was expected. In all other respects, the factors included in each model were determined by the associated research questions and hypotheses. A significance criterion of $\alpha = 0.05$ was adopted, and all tests were evaluated two-tailed unless otherwise specified.

Results

The quality of the logit function fits to participant responses by keyword were comparable for children and adults; across all data, the r^2 had a median value of 0.89 and interquartile range 0.81-0.94. An $r^2 < 0.5$ was obtained in five out of 360 fits, including data in one condition for a child participant and four conditions from adult participants. This represents < 2% of the total dataset. The associated SRTs did not appear to be outliers and were therefore included in the analysis reported below. Two adults did not detect the 16-kHz tone in their left ear at the maximum output of the audiometer (60 dB HL); a threshold of 65 dB HL was recorded in these cases.

Results are shown in Figures 3 and 4. In Figure 3, SRTs for adult participants are plotted as a function of the mean 16 kHz threshold across ears, with each panel showing data for a single spatial condition (by row) and masker head orientation condition (by column). The rationale for evaluating SRTs relative to the mean 16 kHz threshold was the contribution of hearing sensitivity from both ears in conditions with spatial separation, and the fact that 11.2-kHz thresholds were missing for one adult listener. However, selection of the mean 16 kHz threshold is not critical to the results reported below; the same pattern of results is observed when performance is evaluated relative to the poorer 16 kHz threshold or to 11.2 kHz thresholds. Filled symbols indicate data for adults with normal thresholds through 16 kHz, and open symbols indicate data for adults with one or more elevated EHF threshold. Figure 4 shows the distribution of SRTs for children and adults with normal EHF thresholds. Results in the three spatial conditions are shown in separate panels, and masker head orientation is indicated on the abscissa.

Effect of Masker Orientation on SRM in Adults with Normal EHF Thresholds

Data for adults with normal EHF thresholds indicate a reliable benefit of spatially separating the target and masker, and better performance for the non-facing masker orientation than the facing masker orientation; this is described as a positive HORM. These trends in the mean SRT are most evident for the adult data in Figure 4. For the facing masker orientation, mean SRTs improve from −4.9 dB in the co-located condition to −7.8 dB in the symmetrical condition and to −18.0 dB in the asymmetrical condition. SRTs are lower for the non-facing than the facing masker orientation, particularly for the co-located condition; mean SRTs for the non-facing masker orientation are −8.1 dB (co-located), −9.6 dB (symmetrical) and −19.5 dB (asymmetrical). Based on the median psychometric slope fitted to individual participants' data in the co-located spatial condition, the 3.3-dB difference between facing and non-facing masker orientations corresponds to a difference of 26 percentage points. Larger effects of masker orientation for the co-located than spatially separated masker conditions are reflected in the SRM. Mean values of SRM for the facing orientation conditions are 2.9 dB (symmetric) and 13.1 dB (asymmetric), compared to those for the non-facing masker orientation of 1.4 dB (symmetric) and 11.4 dB (asymmetric). That is, the SRM is 1.5-1.8 dB lower for the non-facing than the facing masker orientation conditions.

These observations are supported by a linear mixed model evaluating the effects of spatial condition and masker head orientation in the data of adults with normal EHF thresholds, with random intercepts for participant and spatial condition. This model indicates significant effects of spatial separation (symmetrical: $\beta = -2.95$, $t_{32} = -5.07$, $p < 0.001$; asymmetrical: $\beta = -13.15$, $t_{32} = -22.60$, $p < 0.001$) and a threshold reduction associated with the non-facing masker orientation ($\beta = -3.25$, $t_{48} = -7.55$, $p < 0.001$). There was also a significant interaction between masker orientation and spatial separation (symmetrical: $\beta = 1.50$, $t_{48} = 2.46$, $p = 0.018$; asymmetrical: $\beta = 1.79$, $t_{48} = 2.93$, $p = 0.005$). There was not a significant effect of participant sex ($\beta = -0.67$, $t_{15} = -0.79$, $p = 0.441$). A second model including just the data for the two spatial separation conditions produced a non-significant interaction between masker orientation and spatial separation ($\beta = 0.29$, $t_{32} = 0.57$, $p = 0.576$). These results indicate a larger HORM in the co-located condition than the spatial separation conditions, and a larger SRM for the facing than the non-facing masker orientation[2], but no difference in the effect of head orientation for the symmetrical or asymmetrical conditions.

Effect of EHF Sensitivity in Adults

As illustrated in Figure 3, there is a significant association between SRT and mean 16-kHz thresholds for the non-facing masker head orientation. Line fits in these conditions indicate that increases in threshold from −20 to 65 dB HL were associated with increases in SRT of approximately 4.4 dB (co-located), 2.6 dB (symmetrical) and 2.5 dB (asymmetrical). Comparable values for the facing masker orientation conditions were 1.3 dB (co-located), 0.3 dB (symmetrical), and 1.1 dB (asymmetrical); none of these associations for the facing masker orientation reached significance individually, as indicated by the *p*-values in each panel of Figure 3. These observations were confirmed with a linear mixed model evaluating the effects of spatial condition, masker head orientation, and mean 16 kHz threshold (across ears), as well as random intercepts for participant and spatial condition. The details of this model are reported in Table 1. As in the previous analysis, there were significant effects of spatial separation ($p < 0.001$ for both symmetrical and asymmetrical separation) and masker orientation ($p < 0.001$), and interactions between these factors (symmetrical: $p = 0.011$; asymmetrical: $p = 0.004$). There was not a significant main effect of EHF sensitivity ($p = 0.306$), but there was an interaction between EHF sensitivity and masker orientation ($p = 0.008$). These results are consistent with the conclusion that EHF audibility preferentially impacts performance in the non-facing masker orientation conditions. Although line fits in Figure 3 are consistent with a larger effect of EHF on SRTs in the non-facing co-located condition than the spatial separation conditions, the non-significant three-way interaction between spatial condition, masker orientation, and EHF thresholds failed to provide statistical support for this observation.

While there was no significant difference between the three-frequency pure-tone average (0.5, 1, and 2 kHz) for the two adult groups ($p = -0.310$ one-tailed), thresholds were significantly lower in the normal-EHF group at 4 kHz (mean difference of 4.1 dB, $p = 0.006$) and at 8 kHz (mean difference of 8.4 dB, $p < 0.001$). Further, thresholds tended to be correlated across frequency (e.g., 16 vs. 4 kHz: $r = 0.43$, $p = 0.011$; 16 vs. 8 kHz: $r = 0.60$, $p < 0.001$). This raises the possibility that sub-clinical threshold elevation at 4 or 8 kHz might be responsible for the associations between SRT and EHF thresholds reported above. Repeating the linear mixed model, replacing mean 16 kHz thresholds with mean thresholds at either 8 kHz or 4 kHz, increased the AIC from 863 to 877 and 881, respectively. Neither of these models resulted in a significant effect of -- or interaction with -- mean threshold ($p >= 0.151$).

As expected based on population-level data, EHF thresholds were positively correlated with participant age. For the mean 16-kHz threshold across ears, that correlation was $r = 0.78$ ($p < 0.001$). In fact, the strong correlation between EHF thresholds and SRTs for the co-located condition and non-facing masker head orientation is no longer significant after controlling for participant age via partial correlation ($r = 0.16$, $p = 0.386$). Repeating the analysis reported in Table 1 with age instead of EHF thresholds results in a nearly identical AIC value of (862.6 and 862.3, respectively). The effects of EHF sensitivity and

---

[2] The reduction in HORM with introduction of spatial separation is algebraically equivalent to the reduction in SRM when the masker orientation is increased from 0° to 60° (facing and non-facing conditions, respectively).

age are sufficiently correlated that we cannot differentiate statistically between these two alternatives. While it is theoretically possible that age could affect performance differently across spatial conditions and masker head orientations, there is no theoretical basis for a differential effect of age. In contrast, detrimental effects of elevated EHF thresholds in conditions containing EHF cues were predicted at the outset. The most parsimonious explanation of the results observed in Figure 3 are with respect to EHF thresholds, not participant age or thresholds ≤ 8 kHz.

Children vs. Adults with Normal EHF Thresholds

Figure 4 compares performance for children and adults with normal EHF thresholds. As commonly observed, children’s SRTs were elevated relative to adults’, an effect that ranged from 5.0 dB (co-located with facing masker orientation) to 8.3 dB (asymmetrical with non-facing masker orientation). Like the adults, children benefited from spatial separation of the target and masker, and they tended to perform better for the non-facing masker orientation than the facing masker orientation. However, both of these effects were smaller for children than adults. For example, the mean SRM observed with asymmetric maskers was 9.8 dB for children and 12.2 dB for adults, and the mean HORM in the co-located condition was 1.9 dB for children and 3.3 dB for adults.

These observations are supported by a linear mixed model evaluating the effects of spatial condition, masker head orientation, and age group; there were random intercepts for participant and spatial condition. The details of this model are reported in Table 2. As in the previous analysis, there were significant effects of spatial separation (both: $p < 0.001$) and masker orientation ($p < 0.001$), and significant interactions between spatial separation and masker orientation (symmetrical: $p = 0.010$; asymmetrical: $p = 0.002$). There was an effect of age group ($p < 0.001$), consistent with lower SRTs for adults than children. There was also a significant interaction between age group and spatial separation for the asymmetric condition ($p < 0.001$), consistent with reduced SRM for children in the asymmetric condition; this interaction did not reach significance for the symmetric condition ($p = 0.288$). There was a significant interaction between age group and masker orientation ($p = 0.022$), consistent with greater HORM for adults than children.

For data from children, there was evidence of improvement in SRT with increasing child age. This effect ranged from 1.2 dB per year of age ($p = 0.050$, one-tailed; asymmetric with facing masker orientation) to 2.3 dB per year ($p < 0.001$, one-tailed; symmetric with non-facing masker orientation). Visual inspection of these data did not reveal systematic differences in the age effect as a function of spatial condition or masker orientation. However, the present study was not designed to evaluate effects of child age, and the narrow range of child ages is not ideal for evaluating such effects. A figure showing SRTs for all participants (children and adults) as a function of age for each stimulus condition is included as supplemental data (See Graph, Supplemental Digital Content 1).

Discussion

This study investigated the role of listener age group (children vs. adults) and EHF hearing thresholds on speech-in-speech recognition as a function of the spatial separation between talkers and masker head orientation. All listener groups performed better when the target and masker were spatially separated (compared to co-located), and when the masker simulated talkers facing away from the listener (compared to facing towards the listener). These effects are described as spatial release from masking (SRM) and head orientation release from masking (HORM), respectively. While children and adults with normal EHF thresholds experienced both SRM and HORM, children performed more poorly than adults overall and experienced less masking release of both types. For data from adults, the magnitude of HORM was negatively correlated with EHF thresholds, with larger effects of EHF thresholds for the non-facing (60°) masker head orientation condition than the facing (0°) condition. This result is consistent with the conclusion that EHF content provides useful cues in realistic cocktail party scenarios, provided those cues are audible. It also suggests EHF hearing is necessary to take full advantage of head orientation cues. However, age and EHF sensitivity are highly correlated in the current dataset, and in the general population; as a result, possible effects of age cannot be ruled out conclusively.

Effect of Masker Orientation on SRM

The magnitude of SRM for adults differed for the two masker orientation conditions: it was larger in the facing condition than the non-facing condition. One possible explanation for this result is based on differences in informational masking in the baseline (co-located) condition. If a mismatch in target and masker head orientation introduces a segregation cue in the co-located condition, then it is possible that the additional benefit to segregation afforded by spatial separation would be limited. There is precedent in the literature for a negative association between informational masking at baseline and the magnitude of SRM. For example, Freyman et al. (2007) showed that SRTs differ across masker talkers more when the target and masker are co-located than when they are spatially separated condition. In other words, some masker talkers were more challenging to segregate from the target than others in the co-located condition, whereas they were all relatively easy to segregate in the spatially separated condition. In the case of the present dataset, the non-facing masker head orientation could provide a segregation cue in the baseline, co-located condition.

Effect of EHF Sensitivity on HORM

In adults, HORM is associated with EHF sensitivity. Specifically, there was a positive correlation between the SRT and the mean 16-kHz threshold across ears for the non-facing masker head orientation. In the non-facing masker orientation condition, the masker exerts little or no energetic masking of EHF target information, as displayed in Figure 1 (see also Monson et al. 2019). For participants with good EHF sensitivity, reduced EHF masking in the non-facing masker orientation condition could provide access to cues that support segregation, by helping to differentiate the target from the masker, or provide phonetic information about the target (Trine & Monson 2020). In contrast, the target and the masker in the facing masker head orientation condition have similar spectral content, including at EHFs. In this condition, reduced access to EHF content associated with threshold elevation is of less consequence to performance.

This result may help to resolve previous mixed findings regarding the relationship between EHF sensitivity and speech-in-noise performance. Although EHF sensitivity tends to correlate with subjective report of real-world speech-in-noise difficulty (Badri et al. 2011; Motlagh Zadeh et al. 2019; Yeend et al. 2019), correlations between EHF sensitivity and objective speech-in-noise measures are not consistently observed (Badri et al. 2011; Liberman et al. 2016; Motlagh Zadeh et al. 2019; Yeend et al. 2019; Smith et al. 2019; Prendergast et al. 2019). These inconsistencies may be due, in part, to limitations of testing speech-in-noise with maskers that are facing the listener. The present findings highlight the importance of incorporating more ecologically valid stimulus features into speech-in-noise testing, like differences in masker orientation, to obtain measures that more accurately reflect real-world experience.

As expected based on demographic data, participant age was positively correlated with EHF thresholds (Green et al. 1987), so theoretically associations between SRT and EHF threshold could be attributed to age. However, there is a theoretical rationale for expecting different associations between EHF thresholds and SRTs for the facing and non-facing masker orientation conditions, based on differential access to EHF cues, but no such rationale for expecting different associations between age and SRTs for the two orientation conditions. Further, Trine and Monson (2020) found correlations between SRTs with non-facing maskers and 16-kHz thresholds in young, normal-hearing listeners with EHF thresholds ≤ 20 dB HL in one or both ears. Therefore, an explanation based on EHF thresholds seems the more likely alternative until more data are available.

HORM and SRM in Children vs. Adults

Overall, children required a higher SNR than adults to recognize speech in the presence of the speech masker, replicating past research (e.g., Wightman & Kistler 2005; Buss et al. 2017; Buss et al. 2018). The SRM for asymmetric masker placement was smaller in children than adults, as observed previously (Yuen & Yuan 2014; Corbin et al. 2016). One novel finding of the present study is that children also received less HORM compared to adults. This result fails to support the initial hypothesis that children would derive *more* benefit from the introduction of EHF cues compared to adults, a prediction based on children’s greater bandwidth requirements for understanding speech (Stelmachowicz et al. 2001; Mlot et al. 2010; McCreery & Stelmachowicz 2011). However, there is precedent in the literature for children benefiting less than adults from cues known to reduce SRTs for speech in a speech masker. For example, children receive less benefit than adults from differences in target and masker voice F0 (Flaherty et al. 2018), clear speech modifications (Calandruccio et al. 2016), and semantic context (Buss et al. 2019). Children’s failure to derive adult-like HORM cue could be due to immature selective attention; for example, children are less adept than adults at listening selectively in frequency when the task calls for that strategy (Leibold & Buss 2016).

Implications

The results of this study have implications for how we think about speech recognition in complex multi-talker environments. There has been considerable interest in SRM in the past decade, with over 150 publications listed in PubMed[3]. In contrast, only a handful of studies have examined effects of talker head orientation (reviewed above). Despite this disparity in previous research, the present study indicates that the HORM is comparable in magnitude to the SRM under some conditions. For adults with normal EHF thresholds, the mean HORM observed for the co-located condition was 3.3 dB, and the mean SRM with symmetrical separation for the facing masker head orientation was 2.9 dB. Results also indicate that values of SRM are larger for facing than for non-facing masker head orientation; mean reductions in SRM for the non-facing compared to the facing masker orientation conditions are 1.5 dB and 1.8 dB for symmetrical and asymmetrical separation, respectively. These effects could be even more pronounced for more extreme differences in masker head orientation (e.g., 180°; Monson et al. 2012), given the larger effects on EHF and the introduction of lower-frequency effects. Considering the range of talker head positions in natural listening environments, sub-additive effects of SRM and HORM suggest that previous studies using recordings made with a microphone positioned directly in front of the target and masker talkers (0°) probably overestimate SRM. Clearly, more work is needed to document the impact of HORM on hearing in natural multi-talker listening environments.

The present results indicate a significant effect of masker head orientation, but one question is whether these effects are functionally significant for day-to-day listening, and whether this effect warrants including EHFs into the standard audiologic assessment. For the co-located spatial condition, where effects of masker orientation are the largest, the difference between SRTs for the facing and non-facing masker head orientations—the HORM—was 3.3 dB for adults with normal EHF thresholds. Among all adult participants, SRTs in the co-located non-facing condition rose by 4.4 dB as the mean threshold at 16 kHz increased from −20 to 65 dB HL. A 4.4-dB effect may not reach the 6-dB criterion that listeners appear to use when electing to change hearing aids (McShefferty et al. 2016), but it is nonetheless substantial. This suggests that EHF testing may be warranted to better understand speech recognition abilities in a realistic multi-talker environment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the National Institute of Deafness and Other Communication Disorders (R01 DC000397). Stephanie Berry, Stacey Kane, and Kathryn Sobon assisted with data collection. Preliminary data were presented at the 2019 meeting of the American Auditory Society.

Funding:

This study was funded by grant R01 DC000397 from the National Institutes of Health. The authors have no financial relationships relevant to this article to disclose.

References

Agus TR, Akeroyd MA, Noble W, & Bhullar N (2009). An analysis of the masking of speech by competing speech using self-report data. J Acoust Soc Am, 125(1), 23–26. doi:10.1121/1.3025915 [PubMed: 19173390]

---

[3] This search was run with the following search terms: (spatial release from masking) AND speech AND (("2011/01/01"[Date - Publication] : "2021/01/01"[Date - Publication]))

American Academy of Audiology. (2009). Position statement and clinical practice guidelines: Ototoxicity monitoring. Retrieved from https://audiology-web.s3.amazonaws.com/migrated/OtoMonGuidelines.pdf_539974c40999c1.58842217.pdf

ANSI. (2004). ANSI S3.6-2004, American National Standard Specification for Audiometers, New York: American National Standards Institute.

Arbogast TL, Moore BCJ, 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 Hear, 40(3), 741–756. doi: 10.1097/AUD.0000000000000661 [PubMed: 30300158]

Badri R, Siegel JH, & Wright BA (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. doi: 10.1121/1.3523476 [PubMed: 21361443]

Bench J, Kowal A, & Bamford J (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. Br J Audiol, 13, 108–112. doi: 10.3109/03005367909078884 [PubMed: 486816]

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. doi: 10.1121/1.1926107 [PubMed: 16119356]

Brungart DS, & Iyer N (2012). Better-ear glimpsing efficiency with symmetrically-placed interfering talkers. J Acoust Soc Am, 132(4), 2545–2556. doi: 10.1121/1.4747005 [PubMed: 23039448]

Buss E, Hodge SE, Calandruccio L, Leibold LJ, & Grose JH (2019). Masked sentence recognition in children, young adults, and older adults: Age-dependent effects of semantic context and masker type. Ear Hear, 1–10. doi: 10.1097/AUD.0000000000000692 [PubMed: 30557222]

Buss E, Leibold LJ, Porter HL, & Grose JH (2017). Speech recognition in one- and two-talker maskers in school-age children and adults: Development of perceptual masking and glimpsing. J Acoust Soc Am, 141(4), 2650–2660. doi: 10.1121/1.4979936 [PubMed: 28464682]

Calandruccio L, Buss E, & Leibold L (2016). Speaking clearly to young children: Can we reduce the child-adult deficit while listening to speech-in-speech? Hearing Across the Lifespan, Lake Como, Italy, June 2016.

Cameron S & Dillon H (2007). Development of the Listening in Spatialized Noise-Sentences Test (LISN-S). Ear Hear, 28, 196–211. doi: 10.1097/AUD.0b013e318031267f [PubMed: 17496671]

Cherry EC (1953). Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am, 25(5), 975–979. doi: 10.1121/1.1907229

Corbin N, Bonino AY, Buss E, & Leibold LJ (2016). Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers. Ear Hear, 37, 55–63. doi: 10.1097/AUD.0000000000000201 [PubMed: 26226605]

Corbin NE, Buss E, & Leibold LJ (2017). Spatial release from masking in children: Effects of simulated unilateral hearing loss. Ear Hear, 38(2), 223–235. doi: 10.1097/AUD.0000000000000376 [PubMed: 27787392]

Dieudonné B, & Francart T (2019). Redundant information is sometimes more beneficial than spatial information to understand speech in noise. Ear Hear, 40(3), 545–554. doi: 10.1097/AUD.0000000000000660 [PubMed: 30299342]

Dirks DD, & Wilson RH (1969). The effect of spatially separated sound sources on speech intelligibility. J Speech Lang Hear Res, 12, 508. doi: 10.1044/jshr.1201.05

Ellinger RL, Jakien KM, & Gallun FJ (2017). The role of interaural differences on speech intelligibility in complex multi-talker environments. J Acoust Soc Am, 141, EL170. doi: 10.1121/1.4976113 [PubMed: 28253635]

Flaherty MM, Buss E, & Leibold LJ (2018). Developmental effects in children's ability to benefit from F0 differences between target and masker speech. Ear Hear, 40, 927–937. doi: 10.1097/AUD.0000000000000673

Freyman RL, Balakrishnan U, & Helfer KS (2001). Spatial release from informational masking in speech recognition. J Acoust Soc Am, 109(5 pt 1), 2112–2122. doi: 10.1177/2331216518817464 [PubMed: 11386563]

Gallun FJ, Mason CR, & Kidd G Jr. (2005). Binaural release from informational masking in a speech identification task. J Acoust Soc Am, 118, 1614–1625. doi: 10.1121/1.1984876 [PubMed: 16240821]

Green DM, Kidd G Jr., Stevens KN (1987). High-frequency audiometric assessment of a young adult population. J Acoust Soc Am, 81, 485–494. doi: 10.1121/1.394914 [PubMed: 3558966]

Heffner HE, & Heffner RS (2008). High-frequency hearing. In Dallos P, Oertel D, & Hoy R (Eds.), Handbook of the Senses: Audition (pp. 55–60). Elsevier: New York.

Hunter LL, Monson BB, Moore DR, Dhar S, Wright BA, Munro KJ, … , & Siegel JH (2020). Extended high frequency hearing and speech perception implications in adults and children. Hear Res. doi: 10.1016/j.heares.2020.107922

Imbery C, Franz S, van de Par S, & Bitzer J (2019). Auditory facing angle perception: The effect of different source positions in a real and an anechoic environment. Acta Acust United Ac, 105(3), 492–505. doi: 10.3813/AAA.919331

Jakien KM, Kampel SD, Stansell MM, & Gallun FJ (2017). Validating a rapid, automated test of spatial release from masking. Am J Audiol, 26, 507–518. doi: 10.1044/2017_AJA-17-0013 [PubMed: 28973106]

Kato H, Takemoto H, Nishimura R, & Mokhtari P (2010). Spatial acoustic cues for the auditory perception of speaker’s facing direction. Proceedings of 20th International Congress on Acoustics, ICA 2010.

Kidd G Jr., Mason CR, Best V, Roverud E, Swaminathan J, Jennings T, Clayton K, & Steven Colburn H (2019). Determining the energetic and informational components of speech-on-speech masking in listeners with sensorineural hearing loss. J Acoust Soc Am, 145, 440. doi: 10.1121/1.5087555 [PubMed: 30710924]

King G, Corbin NE, Leibold LJ, & Buss E (2019). Spatial release from masking using clinical corpora: Sentence recognition in a colocated or spatially separated speech masker. J Am Acad Audiol, 1–7. doi: 10.3766/jaaa.19018

Leibold L (2012). Development of auditory scene analysis and auditory attention, human auditory development. Springer: New York.

Leibold LJ, & Buss E (2016) Factors responsible for remote-frequency masking in children and adults. J Acoust Soc Am, 140, 4367–4377. doi: 10.1121/1.4971780 [PubMed: 28040030]

Levy SC, Freed DJ, Nilsson M, Moore BCJ, & Puria S (2015). Extended high-frequency bandwidth improves speech reception in the presence of spatially separated masking speech. Ear Hear, 36, e214–e224. doi: 10.1097/AUD.0000000000000161 [PubMed: 25856543]

Liberman MC, Epstein MJ, Cleveland SS, Wang H, & Maison SF (2016). Toward a differential diagnosis of hidden hearing loss in humans. PLoS ONE, 11(9), e0162726. doi: 10.1371/journal.pone.0162726 [PubMed: 27618300]

Licklider JCR (1948). The influence of interaural phase relations upon the masking of speech by white noise. J Acoust Soc Am, 42, 601–608. doi: 10.1121/1.1906358

Litovsky RY (2005). Speech intelligibility and spatial release from masking in young children. J Acoust Soc Am, 117(5), 3091–3099. doi: 10.1121/1.1873913 [PubMed: 15957777]

Matthews LJ, Lee FS, Mills JH, & Dubno JR (1997). Extended high-frequency thresholds in older adults. J Speech Lang Hear Res, 41(1), 208–204. doi: 10.1044/jslhr.4001.208

McCreery RW, Stelmachowicz PG (2011). Audibility-based predictions of speech recognition for children and adults with normal hearing. J Acoust Soc Am, 130, 4070–4081. doi: 10.1121/1.3658476 [PubMed: 22225061]

McKellin WH, Shahin K, Hodgson M, Jamieson J, & Pichora-Fuller MK (2011). Noisy zones of proximal development: Conversation in noisy classrooms. J Socioling, 15, 65–93. doi: 10.1111/j.1467-9841.2010.00467.x

McShefferty D, Whitmer WM, & Akeroyd MA (2016). The just-meaningful difference in speech-to-noise ratio. Trends Hear, 20, 1–11. doi: 10.1177/2331216515626570

Mlot S, Buss E, Hall JW III. (2010). Spectral integration and bandwidth effects on speech recognition in school-aged children and adults. Ear Hear, 31, 56–62. doi: 10.1097/AUD.0b013e3181ba746b [PubMed: 19816182]

Moore BCJ, Füllgrabe C, & Stone MA (2011). Determination of preferred parameters for multichannel compression using individually fitted simulated hearing aids and paired comparisons. Ear Hear, 32, 556–568. doi: 10.1097/AUD.0b013e31820b5f4c [PubMed: 21285878]

Moore BCJ, & Popelka GR (2013). Preliminary comparison of bone-anchored hearing instruments and a dental device as treatments for unilateral hearing loss. Int J Audiol, 52(10), 678–686. doi: 10.3109/14992027.2013.809483 [PubMed: 23859058]

Moore BCJ, & Tan CT (2003). Perceived naturalness of spectrally distorted speech and music. J Acoust Soc Am, 114, 408–419. doi: 10.1121/1.1577552 [PubMed: 12880052]

Moore JK, & Linthicum FH Jr. (2007). The human auditory system: A timeline of development. Int J Audiol, 46, 460–478. doi: 10.1080/14992020701383019 [PubMed: 17828663]

Monson BB, Hunter EJ, Lotto AJ, & Story BH (2014). The perceptual significance of high-frequency energy in the human voice. Front Psychol, 5, 1–11. doi: 10.3389/fpsyg.2014.00587 [PubMed: 24474945]

Monson BB, Lotto AJ, & Story BH (2012). Analysis of high-frequency energy in long-term average spectra of singing, speech, and voiceless fricatives. J Acoust Soc Am, 132(3), 1754–64. doi: 10.1121/1.4742724 [PubMed: 22978902]

Monson BB, Hunter EJ, & Story BH (2012). Horizontal directivity of low- and high-frequency energy in speech and singing. J Acoust Soc Am, 132, 433–441. doi: 10.1121/1.4725963 [PubMed: 22779490]

Monson BB, Lotto AJ, & Ternström S (2011). Detection of high-frequency energy changes in sustained vowels produced by singers. J Acoust Soc Am, 129, 2263–2268. doi: 10.1121/1.3557033 [PubMed: 21476681]

Monson BB, 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. doi: 10.1016/j.heares.2019.107773

Motlagh Zadeh L, Silbert NH, Sternasty K, Swanepoel DW, Hunter LL, & Moore DR (2019). Extended high-frequency hearing enhances speech perception in noise. PNAS, 116(47), 23753–23759. doi: 10.1073/pnas.1903315116 [PubMed: 31685611]

Neuhoff JG (2003). Twist and shout: Audible facing angles and dynamic rotation. Ecol Psychol, 15(4), 335–351. doi: 10.1207/s15326969eco1504_7

Noble W, & Perrett S (2002). Hearing speech against spatially separate competing speech versus competing noise. Percept Psychophys, 64, 1325–1336. doi: 10.3758/bf03194775 [PubMed: 12519029]

Noffsinger D, Wilson RH, & Musiek FE (1994). Department of Veterans Affairs compact disc recording for auditory perceptual assessment: Background and introduction. JAAA, 5(4), 231–235. [PubMed: 7949294]

Olsen WO (1998). Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson, Bennett & Fidell (1997) report. Am J Audiol, 7, 1–5. doi: 10.1044/1059-0889(1998/012)

Phatak SA, Brungart DS, Zion DJ, & Grant KW (2018). Clinical assessment of functional hearing deficits: Speech-in-noise performance. Ear Hear, 40, 426–436. doi: 10.1097/AUD.0000000000000635

Pinheiro J, Bates D, DebRoy S, Sarkar D, and R Core Team (2016). nlme: Linear and Nonlinear Mixed Effects Models (R Foundation for Statistical Computing, Vienna, Austria).

Prendergast G, Couth S, Millman RE, Guest H, Kluk K, Munro KJ, & Plack CJ (2019). Effects of age and noise exposure on proxy measures of cochlear synaptopathy. Trend Hear, 23, 1–16. doi: 10.1177/2331216519877301

Prendergast G, Millman RE, Guest H, Munro KJ, Kluk K, Dewey RS, … Plack CJ (2017). Effects of noise exposure on young adults with normal audiograms II: Behavioral measures. Hear Res, 356, 74–86. doi: 10.1016/j.heares.2017.10.007 [PubMed: 29126651]

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Schoenmaker E, Brand T, & van de Par S (2016). The multiple contributions of interaural differences to improved speech intelligibility in multitalker scenarios. J Acoust Soc Am, 139, 2589–2603. doi: 10.1121/1.4948568 [PubMed: 27250153]

Smith SBB, 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. Hear Res, 371, 66–74. doi: 10.1016/j.heares.2018.11.008 [PubMed: 30504092]

Sobon KA, Taleb NM, Buss E, Grose JH, & Calandruccio L (2019). Psychometric function slope for speech-in-noise and speech-in-speech: Effects of development and aging. J Acoust Soc Am, 145(4), EL284. doi: 10.1121/1.5097377 [PubMed: 31046371]

Stelmachowicz PG, Pittman AL, Hoover BM, & Lewis DE (2001). Effect of stimulus bandwidth on the perception of /s/ in normal- and hearing-impaired children and adults. J Acoust Soc Am, 110(4), 2183–2190. doi: 10.1121/1.1400757 [PubMed: 11681394]

Strelcyk O, Pentony S, Kalluri S, & Edwards B (2014). Effects of interferer facing orientation on speech perception by normal-hearing and hearing-impaired listeners. J Acoust Soc Am, 135(3),1419–1432. doi: 10.1121/1.4864786 [PubMed: 24606279]

Trine A, & Monson BB (2020). Extended high frequencies provide both spectral and temporal information to improve speech-in-speech recognition. Trends Hear, 24, 1–8. doi: 10.1177/2331216520980299

Vannson N, James C, Fraysse B, et al. (2015). Quality of life and auditory performance in adults with asymmetric hearing loss. Audiol Neurotol, 20(Suppl. 1), 38–43. doi: 10.1159/000380746

Yeend I, Beach EF, & Sharma M (2019). Working memory and extended high-frequency hearing in adults. Ear Hear, 40(3), 458–467. doi: 10.1097/AUD.0000000000000640 [PubMed: 30052557]

Yuen KCP & Yuan M (2014). Development of spatial release from masking in mandarin-speaking children with normal hearing. J Speech Lang Hear Res, 57, 2005–2023. doi: 10.1044/2014_JSLHR_H_13_0060 [PubMed: 24950448]

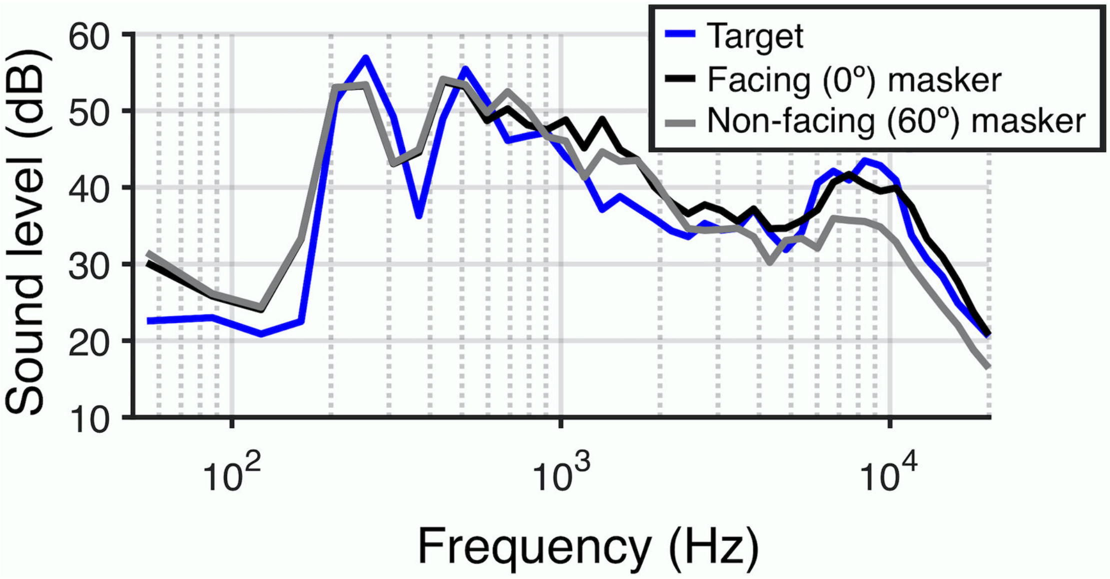

Figure 1:

Long-term magnitude spectrum for the target, facing (0°) masker, and the non-facing (60°) masker. Color reflects stimulus type, as defined in the legend.

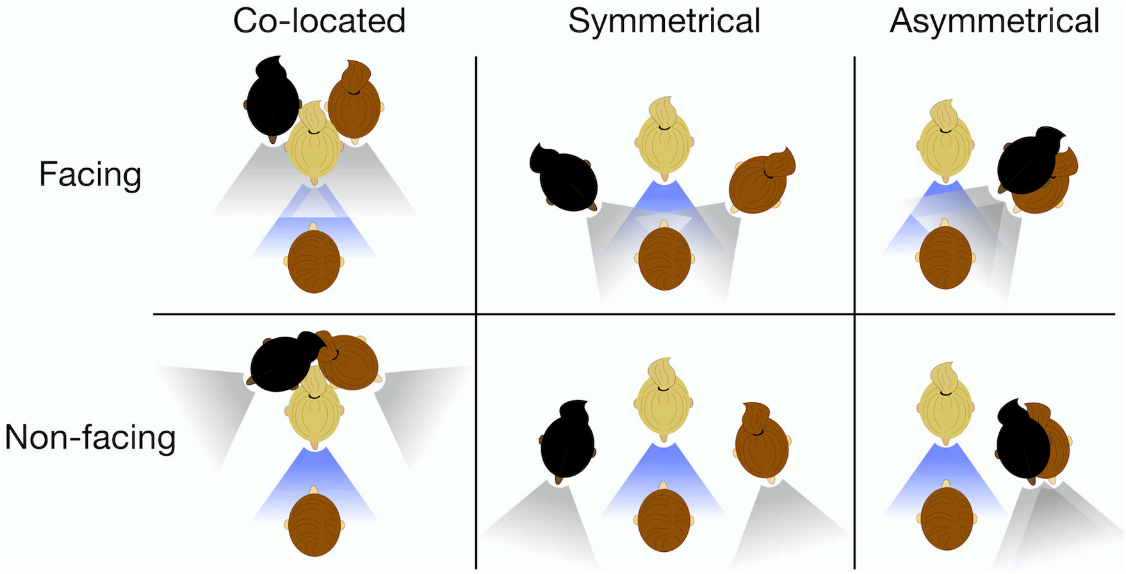

Figure 2:
Illustration of stimulus conditions. The target talker is always directly in front of the listener. The two masker talkers are either co-located with the target (left), symmetrically separated on either side of the listener (middle), or asymmetrically separated to the right of the listener (right). Masker head orientation was either facing (0°; top) or non-facing (60°; bottom). In this cartoon, shading indicates directional propogation of EHF energy, with blue indicating the target and gray indicating the masker.

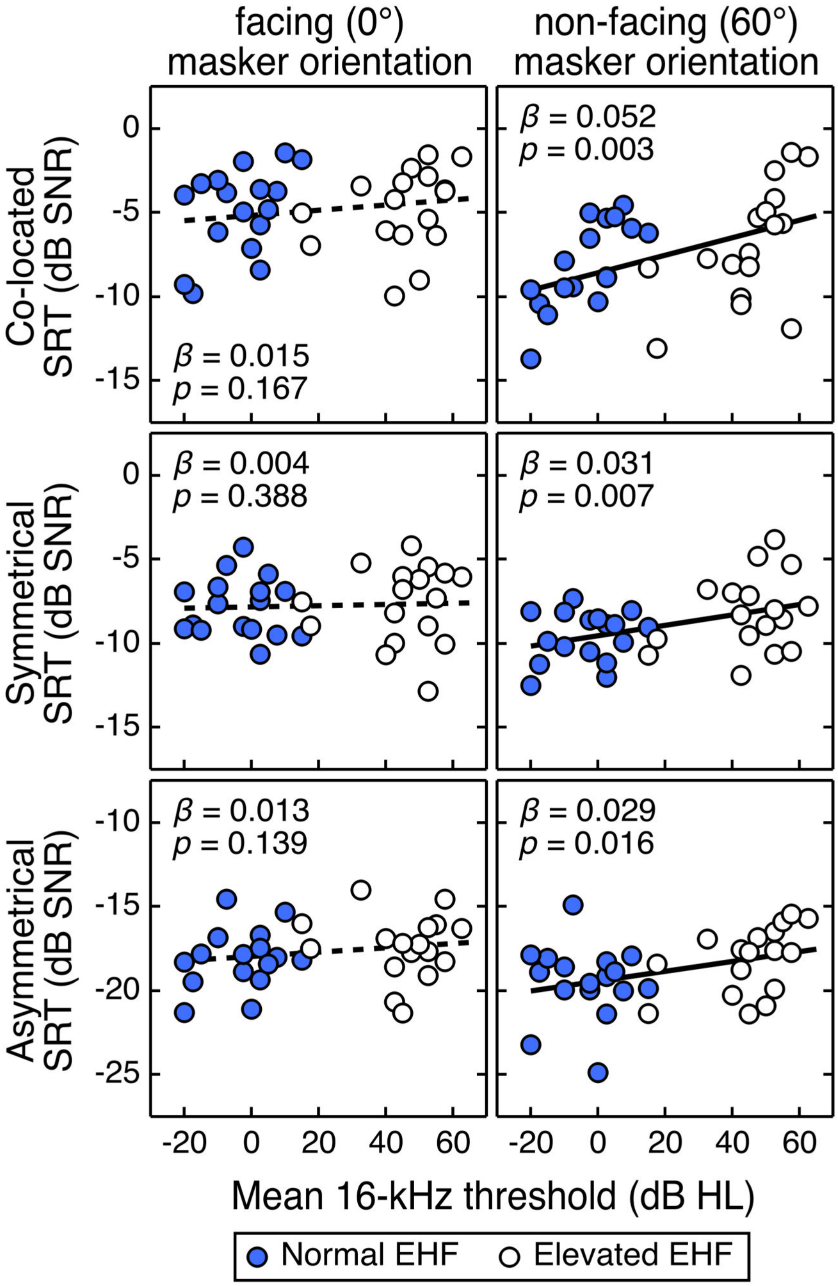

Figure 3:

SRTs in dB SNR for adult participants, plotted as a function of the mean 16 kHz threshold in the left and right ear, in dB HL. Symbol shading indicates participants with normal thresholds bilaterally through 16 kHz and those with one or more elevated EHF threshold, as defined in the legend at the bottom of the figure. Slopes and p-values associated with line fits to all adult data are indicated in each panel; solid lines indicate $p < 0.05$ (one-tailed), and dotted lines indicate $p \geq 0.05$.

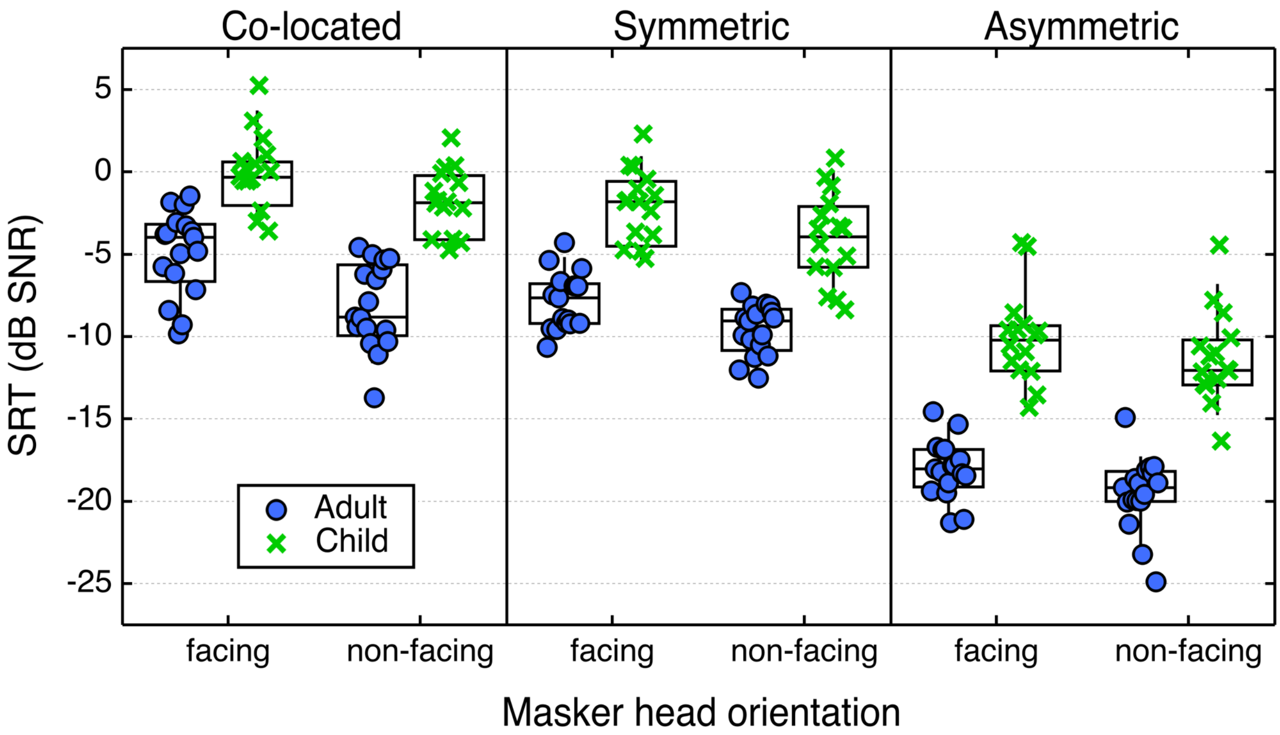

Figure 4:

Distribution of SRTs in dB SNR for the three spatial conditions, as indicated in the panel labels at the top of the figure. Masker head orientation is indicated on the abscissa, and participant age group is indicated by symbols, as defined in the legend. Horizontal lines indicate the median, boxes span the 25th-75th percentiles, and whiskers span the 10th-90th percentiles.

Table 1:

Results of a linear mixed model evaluating effects of spatial condition (co-located, symmetric, and asymmetric), masker talker head orientation (facing and non-facing), EHF (16 kHz) thresholds, and listener sex on SRTs for adult participants. Reference conditions were co-located spatial position, facing head orientation, and female sex.

	Coef.	**Std. Error**	***df***	***t*-value**	***p*-value**
(Intercept)	−4.980	0.490	96	−10.16	**< 0.001**
Spatial(sym)	−2.660	0.538	64	−4.95	**< 0.001**
Spatial(asym)	−12.808	0.538	64	−23.81	**< 0.001**
Orientation(non-facing)	−3.421	0.464	96	−7.37	**< 0.001**
EHF	0.014	0.014	31	1.04	0.306
Sex(male)	−0.777	0.636	31	−1.22	0.231
Spatial(sym):Orientation(non-facing)	1.709	0.657	96	2.60	**0.011**
Spatial(asym):Orientation(non-facing)	1.946	0.657	96	2.96	**0.004**
Spatial(sym):EHF	−0.012	0.016	64	−0.73	0.469
Spatial(asym):EHF	−0.002	0.016	64	−0.15	0.879
Orientation(non-facing):EHF	0.037	0.014	96	2.70	**0.008**
Spatial(sym):Orientation(non-facing):EHF	−0.010	0.019	96	−0.50	0.615
Spatial(asym):Orientation(non-facing):EHF	−0.020	0.019	96	−1.06	0.294

Table 2:

Results of a linear mixed model evaluating effects of spatial condition (co-located, symmetric, and asymmetric), masker talker head orientation (facing and non-facing), and sex on SRTs for two groups of participants with normal EHF detection thresholds (child and adult). This model includes a random intercept for participant and spatial condition. Reference conditions were co-located spatial position, facing head orientation, adult age group, and female sex.

	Coef.	**Std.Error**	***df***	***t*-value**	***p*-value**
(Intercept)	−5.047	0.575	90	−8.78	**< 0.001**
Spatial(sym)	−2.949	0.524	60	−5.63	**< 0.001**
Spatial(asym)	−13.145	0.524	60	−25.09	**< 0.001**
Orientation(non-facing)	−3.255	0.402	90	−8.11	**< 0.001**
Group(child)	4.910	0.811	29	6.05	**< 0.001**
Sex(male)	0.714	0.710	29	1.01	0.323
Spatial(sym):Orientation(non-facing)	1.500	0.568	90	2.64	**0.010**
Spatial(asym):Orientation(non-facing)	1.786	0.568	90	3.14	**0.002**
Spatial(sym):Group(child)	0.820	0.765	60	1.07	0.288
Spatial(asym):Group(child)	2.996	0.765	60	3.91	**< 0.001**
Orientation(non-facing):Group(adult)	1.368	0.587	90	2.33	**0.022**
Spatial(sym):Orientation(non-facing):Group(child)	−1.615	0.829	90	−1.95	0.055
Spatial(asym):Orientation(non-facing):Group(child)	−1.125	0.829	90	−1.36	0.178