Spatial release of masking in children and adults in non-individualized virtual environments

Katharina Zenke and Stuart Rosen

Citation: The Journal of the Acoustical Society of America **152**, 3384 (2022); doi: 10.1121/10.0016360
View online: https://doi.org/10.1121/10.0016360
View Table of Contents: https://asa.scitation.org/toc/jas/152/6
Published by the Acoustical Society of America

JASA ARTICLE

Spatial release of masking in children and adults in non-individualized virtual environments

Katharina Zenke[a] and Stuart Rosen
Speech, Hearing and Phonetic Sciences, University College London, 2 Wakefield Street, London, WC1N 1PF, United Kingdom

ABSTRACT:
The spatial release of masking (SRM) is often measured in virtual auditory environments created from head-related transfer functions (HRTFs) of a standardized adult head. Adults and children, however, differ in head dimensions and mismatched HRTFs are known to affect some aspects of binaural hearing. So far, there has been little research on HRTFs in children and it is unclear whether a large mismatch of spatial cues can degrade speech perception in complex environments. In two studies, the effect of non-individualized virtual environments on SRM accuracy in adults and children was examined. The SRMs were measured in virtual environments created from individual and non-individualized HRTFs and the equivalent real anechoic environment. Speech reception thresholds (SRTs) were measured for frontal target sentences and symmetrical speech maskers at 0° or ±90° azimuth. No significant difference between environments was observed for adults. In 7 to 12-year-old children, SRTs and SRMs improved with age, with SRMs approaching adult levels. SRTs differed slightly between environments and were significantly worse in a virtual environment based on HRTFs from a spherical head. Adult HRTFs seem sufficient to accurately measure SRTs in children even in complex listening conditions.
 https://doi.org/10.1121/10.0016360
(Received 18 March 2022; revised 19 September 2022; accepted 14 November 2022; published online 12 December 2022)
[Editor: G. Christopher Stecker] Pages: 3384–3395

I. INTRODUCTION

Spatialized sound sources are often presented in a virtual auditory environment in experimental or clinical procedures [e.g., Best *et al.* (2017) and Glyde *et al.* (2013)]. The use of virtual environments over real ones has many advantages: it reduces the acoustic demands on the test room, reduces overall costs, is portable and makes it easier to ensure equal testing conditions across participants and repetitions. Tests in virtual environments can also be used in settings in which it is not possible to test the participant in the original environment, e.g., in an audiology clinic. Virtual auditory environments are commonly generated by computation from head-related transfer functions (HRTFs). HRTFs describe the free-field transmission of sound waves from specific directions in space to a person's ear canal entrances. They contain all acoustic features that result from the filtering of sound waves by a person's pinnae, head, and torso (Blauert, 1974). Generating environments from HRTFs is hardware efficient and the use of individual HRTFs results in a very realistic auditory display (Xie, 2013). HRTFs are typically measured acoustically in an anechoic room or calculated from a geometric model of a person's body (Huttunen *et al.*, 2014; Ziegelwanger *et al.*, 2015). However, they can also be approximated through anatomical features or selected from a suitable set of non-individualized HRTFs based on localization accuracy or subjective preference (Iida *et al.*, 2014; Kistler and Wightman, 1992; Middlebrooks and Green, 1992). Direct measurements are the most reliable method to obtain accurate HRTFs, but they are complex, time-consuming, and demanding in terms of equipment and test environment. Numerical calculations from geometric data also place high demands on the precision of the photogrammetric techniques used to generate a mesh of head and pinnae and often require manual post-processing.

Due to the complexity of obtaining individual HRTFs, sound sources are often spatialized with pre-existing sets of HRTFs from individuals with similar body dimensions and pinna shapes (Pelzer *et al.*, 2020) or from artificial heads with average adult dimensions. Whilst these generic HRTFs are easier to obtain than measuring HRTFs for each participant, the use of non-individualized or inaccurate HRTFs leads to a degradation of spatial auditory perception and a loss of immersion and realism of the auditory scene (Sunder and Gan, 2016). The extent of these effects depends on the similarity of the mismatched HRTFs to the listener's own HRTFs. Most noticeably, localization accuracy and acuity are decreased for non-individualized HRTFs, especially for elevated and frontal sources (Møller *et al.*, 1996; Wenzel *et al.*, 1993). The use of non-individualized HRTFs can also reduce the percept of externalisation in anechoic environments and lead to increased apparent source widths and a less vivid auditory display due to spectral colouring (Best *et al.*, 2020; Iida, 2019). Other aspects of spatial hearing, such as distance perception, appear to be more robust and are unaffected by the use of mismatched HRTFs (Yu and Wang, 2018). Only a few studies have been conducted on

[a]Electronic mail: k.zenke@ucl.ac.uk

https://doi.org/10.1121/10.0016360

speech perception, with mixed results. Rychtáriková *et al.* (2011) found very similar SRTs in an anechoic environment and a non-individualized virtual environment for sentence recognition in speech-shaped noise. Cuevas-Rodriguez *et al.* (2021), however, found significant differences in SRTs for word recognition in speech-shaped noise for environments based on seven sets of adult HRTFs and HRTFs from a spherical head model. These differences were mainly caused by increased SRTs in the spherical head environment, but some participants also displayed increased SRTs for one of the adult HRTF sets.

Compared to adults, very little research has been conducted on HRTFs in children mainly due to the difficulties in obtaining precise HRTFs for them. Conventional HRTF measurement techniques require participants to sit motionless for an extended duration, which is impossible for young children. Only recently, full sets of HRTFs have been measured in young children for the first time with a rapid measurement technique in a study by Braren and Fels (2022) but not yet evaluated in perceptual studies. Previous studies [for example, Fels *et al.* (2004)] calculated child HRTFs from geometric data. However, this approach is strongly dependent on the accuracy of the geometric mesh. For young children who are likely to move during data acquisition, mesh generation requires manual post-processing and is therefore only feasible for a small number of participants.

Instead of individual HRTFs, generic HRTFs of artificial heads with adult dimensions are often used in studies with children (Brown *et al.*, 2010; Cameron and Dillon, 2007). Due to children's smaller head sizes, the mismatch caused by the use of artificial HRTFs means that any effects on auditory perception are likely to be larger than in adults, although it is not straightforward to predict what those might be. On the one hand, larger head sizes lead to larger ILDs which should be beneficial. Kollmeier and Peissig (1990) did find improved performance with a single masker off to the side of a target using an interaural magnification algorithm, but not with two maskers, one to the right and one to the left of the target, as used here. Also, for speech maskers, a substantial part of the spatial release of masking (SRM) appears to arise from binaural cues allowing accurate spatial localisation of the auditory sources thus allowing the appropriate source to be attended to (Carlile and Corkhill, 2015). As higher-order perceptual processes are still in development in young children, this might increase any difficulties experienced with mismatched cues (Jones and Moore, 2015; Leibold *et al.*, 2019).

In this project, we aimed to investigate whether virtual acoustic environments based on non-individualised HRTFs are suitable to accurately measure speech perception in adults and children. Therefore, we measured SRTs in a real anechoic environment and in virtual acoustic environments generated from individual and non-individualized HRTFs. Two independent studies were conducted, one for adults and one for children. It was expected that SRTs in the real and individualized environments would be the same for all participants. For adults, we hypothesized that the HRTF mismatch for non-individualized HRTFs would be small and speech performance similar in individualized and non-individualized environments. In young children, however, the mismatch of spatial cues was expected to be larger in non-individualized environments based on HRTFs of an adult-sized artificial head due to larger deviations in anatomy. Since auditory perception is still developing in young children, they were also expected to be more susceptible to the potential degradation of sounds due to mismatched spatial cues. Therefore, we predicted that a potential effect of non-individualized cues would be larger in children, and lead to degraded speech perception, observable as an increase in speech reception thresholds (SRTs) in non-individualized environments.

II. STUDY 1: ADULTS

A. Methods

1. Participants

Seventeen young adults (14 females, 3 males) aged 18 to 35 years (mean 23.6 years ± 5.0 years) participated in this experiment. All participants were native British English speakers with no known listening difficulties or developmental disorders. They had normal hearing thresholds (below 20 dB HL at 250 to 8000 Hz), which was confirmed by a hearing screening at the beginning of the session. All participants gave informed consent and were financially compensated for their time. The study was approved by the UCL Research Ethics Committee.

2. Measurement of head-related transfer functions

At the beginning of the session, HRTFs were measured for −90°, 0°, and 90° azimuth at 0° elevation. The participant sat on a non-rotating chair in the centre of an anechoic chamber (Nevard and Fourcin, 1995) at a distance of 1.15 m from three Fostex 6301B loudspeakers positioned at −90°, 0°, and 90° azimuth. The height of the chair was adjusted to align the participant's ear entrances with the centre of the loudspeaker drivers. The impulse responses were recorded with Knowles FG 2332 omnidirectional electret condenser microphones with a diameter of 2.5 mm positioned at the entrances of the participant's ear canals. To secure them in place and block the ear canals, they were mounted on closed silicone domes manufactured for hearing aid receivers. Due to the short duration of the measurement of only three spatial positions and to avoid additional reflections, no head fixture was used. The participant was positioned before the measurement and instructed to remain motionless until the end of the measurement. Exponential sine sweeps between 20 and 22 500 Hz at 75 dB SPL were used as excitation signals. They had a duration of 2 s and a sampling rate of 44.1 kHz. Signal generation and playback were realized in Matlab using the ITA Toolbox for acoustic measurements developed at RWTH Aachen University (Berzborn *et al.*, 2017). Three transfer functions were measured for each direction and averaged to get a more precise measurement.

To account for the combined effect of all hardware components in the measurement chain, the averaged transfer functions for each participant were divided by the equivalent system transfer functions, measured for the same system without a participant but with the microphones mounted on a thin stand at the equivalent location of the centre of the head. The resulting head-related impulse responses were truncated to 23 ms to remove possible reflections from the metal grid floor. Multiple measurements were taken for each participant and the set of HRTFs with the fewest artefacts was selected for the following speech perception test.

3. Stimuli

For the SRT measurements, sentences from the Adaptive Sentence Lists (ASL) developed by Macleod and Summerfield (1990) were used as target stimuli. The sentence lists were constructed on similar principles to the widely used BKB sentences (Bench *et al.*, 1979) developed for children aged 8 or older. Unlike the BKB sentences, the ASL sentences have the advantage of being novel for all participants as they are not used in standard clinical tests. Both corpora consist of short sentences with three keywords and simple syntax and vocabulary suitable for young children's speech and language competence. The sentences are meaningful and partly predictable (e.g., "They're living by the sea," "Christmas is coming soon"). Due to the limited number of ASL sentences, BKB sentences were used for training purposes. ASL and BKB sentence lists were recorded by a male standard southern British English talker in the same anechoic chamber at 22.05 kHz sampling frequency and upsampled to 44.1 kHz. Three publicly available children's stories with similar linguistic complexity to the target sentences were used as maskers.[1] They were recorded for the same talker at 44.1 kHz and high-pass filtered at 50 Hz. Pauses between words were shortened to 100 ms and audio distortions were removed. The final masker signals had durations of 102, 142, and 217 s.

To create virtually spatialized sound sources, target and masker signals were independently filtered with the inverse HRTFs. The resulting three stereo signals were combined to form left and right headphone channels. The two channels were then multiplied by the inverse headphone transfer functions to remove spectral alterations caused by the Sennheiser HD 25 headphones. The headphone transfer functions were calculated from calibration measurements with four adults. The calibration was conducted with human participants instead of an artificial head to measure the transfer functions with realistic pressure and seal of the headphones. The inverse transfer functions were bandpass filtered between 70 Hz and 11 kHz. In conditions in the real environment, the stimuli were played from the three Fostex 6301B loudspeakers and filtered with the respective inverse loudspeaker transfer functions to correct for the influence of loudspeaker characteristics. Levels were calibrated using a sound level meter at the position of the head to ensure equal levels for loudspeaker and headphone presentation. Custom MATLAB scripts were used to create the stimuli and run the procedure. Stimuli were presented from loudspeakers or headphones via an RME Fireface UC audio interface.

4. Conditions

SRTs, the signal-to-noise ratios (SNRs) in dB between target and maskers at which 50% of speech is intelligible, were measured for two spatial configurations in four auditory environments. Target stimuli were always presented from the front at 0° azimuth and elevation. Two maskers were presented either from the same location (colocated) or from −90° and 90° azimuth (separated). The SRM was calculated as the difference between SRTs in these two conditions. The maskers were placed symmetrically to avoid better ear listening, only allowing the participant to make use of rapid better-ear glimpses available in both ears (Brungart and Iyer, 2012). Two speech maskers were used to obtain large effects of informational and energetic masking and a large SRM (Buss *et al.*, 2017; Hall *et al.*, 2002).

SRTs in adults were measured in four auditory environments: in virtual environments generated from (1) the individual HRTFs measured at the beginning of the session, (2) HRTFs from a large adult head, (3) HRTFs from a small adult head, and in (4) the equivalent real anechoic environment. The two non-individualized sets of HRTFs were taken from the publicly available ARI HRTF database of the Acoustics Research Institute Vienna which contains measured HRTFs and anthropometric data from a large number of adults (ARI, 2017). The largest and smallest head of the database were selected based on an approximation of the head volume from the product of head width, height and depth: The large (male) head had the dimensions 15.7 cm × 23.4 cm × 23.0 cm and a head circumference of 61 cm, the small (female) head 14.6 cm × 18.7 cm × 16.3 cm and 47 cm. Whilst these two heads were selected based on their large difference in head size and, thus, in binaural cues, differences in pinna shapes were not accounted for.

5. Procedure

The participant was seated in the anechoic chamber at the same position as for the HRTF measurements and instructed to listen to the target sentences and to repeat them. First, the participant completed a short training of 30 sentences under headphones to get familiar with the virtual environments and masker locations. Then, SRTs for the eight test conditions were measured in an adaptive 1-up 1-down procedure (Plomp and Mimpen, 1979). Each adaptive run consisted of 15 target sentences from an ASL list in random order. For each trial, random segments within the two masker stories were selected and started 1 s before the target presentation with a 100 ms ramp at the beginning and end to avoid onset and offset effects. The combined signal level of the two maskers was 65 dB SPL. The initial target level was +5 dB SNR in colocated and −10 dB SNR in separated conditions and was adjusted according to the participant's response (decreased if two or three keywords were

https://doi.org/10.1121/10.0016360

correct and increased otherwise) to track the SRT at 50% intelligibility. The step sizes decreased from 4 dB to 3 and 2 dB at the first two reversals. Due to the limited number of ASL sentences, the initial target level was set relatively low, corresponding to approximately 1.5 step sizes above the estimated average SRT based on findings from a pilot study and similar studies in the literature (Besser *et al.*, 2015; Cameron *et al.*, 2011). To ensure that participants were adapted to each condition at the beginning of the adaptive run and did not make mistakes due to initial inattention, three BKB practice sentences were presented at the initial SNR before each measurement. SRTs were calculated as the mean of the reversals, starting from the third reversal for an even number of total reversals or the fourth for an odd number.

The 16 sentence lists and 6 possible combinations of the three maskers were pseudo-randomized between conditions across participants using Latin square counterbalancing. The order of presentation was randomized across participants with the following restrictions: conditions alternated between colocated and separated maskers and between environments, with the exception of the two loudspeaker conditions which were presented sequentially to minimize interruptions due to switching between presentation forms. After a break, the conditions were repeated in reverse order. A total of 240 test sentences were presented. Including the hearing screening and HRTF measurement, the session took about 70 min.

B. Results

All statistical analyses were performed in R. For linear mixed-effects models, likelihood ratio tests were used to sequentially eliminate terms with a significance level greater than $p = 0.05$ from the initial model. Only the reduced models are reported.

SRTs for all eight conditions are shown in Fig. 1. The mean SRT was −7.6 dB SNR (SD = 2.9 dB) in colocated conditions and −16.4 dB SNR (SD = 2.5 dB) in spatially separated conditions. To investigate the effects of the test conditions and procedure on the SRTs, a mixed-effects model was fitted for the SRT. The model included the fixed effects of masker location, acoustic environment and repetition and their interactions as well as the random effects of participant, sentence list and order of the measurement within the entire set. After eliminating non-significant terms, the reduced model found significant main effects for masker location [$\chi^2(1) = 958$, $p < 0.001$] and repetition [$\chi^2(1) = 9.16$, $p = 0.003$]. Only the random effect of participant had a significant influence on the data [$\chi^2(1) = 111$, $p < 0.001$] accounting for 42% of the variance. Acoustic environment was neither significant as a main effect [$\chi^2(3) = 0.69$, $p = 0.559$] nor in any interaction suggesting that speech perception was the same in all four environments. The model showed a 0.9 dB reduction in SRT for the second repetition, possibly due to ongoing learning at the beginning of the procedure. Since no random effects of list and position were found, the ASL sentence lists appear to be sufficiently balanced to be used interchangeably. Presentation order, too, did not seem to influence the results.

The SRM was calculated as the difference of the averaged SRTs in the colocated and separated conditions of each environment (displayed in Fig. 2). The mean SRM was 8.8 dB (SD = 2.6 dB). As expected from the lack of interaction between masker location and environment in the SRT model, a mixed-effects model for the SRM with the fixed effect of environment and random effect of participant found no differences between environments [$\chi^2(3) = 1.96$, $p = 0.581$].

This study included environments based on a large and a small adult head to investigate whether participants perform better with HRTFs from a head that's more similar to their own in size than with others. Although participants as a group showed no differences between individualized and non-individualized conditions, inter-subject variability was high and individual results differed greatly between environments. As an indicator of head size, the maximum ITD was extracted from the measured HRTFs for each participant. It was calculated as the time shift of the maximum in the cross correlation function between the head-related impulse response of the left and right ear for the two lateral sound

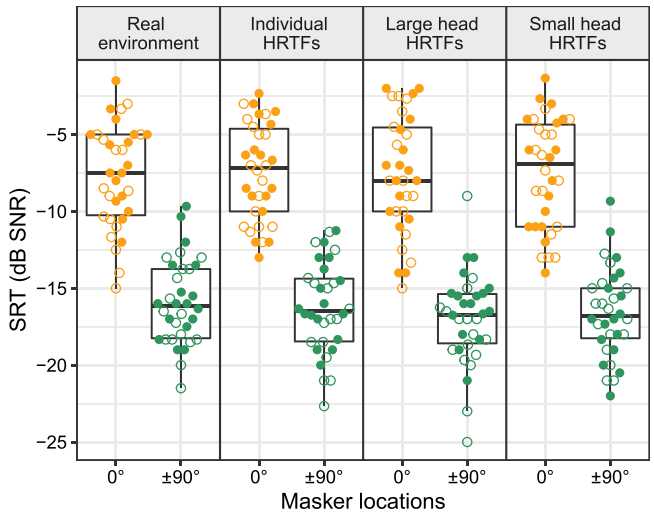

FIG. 1. (Color online) SRTs of adults for maskers at 0° and ±90° azimuth in the four environments for the first (filled symbols) and second repetition (open symbols).

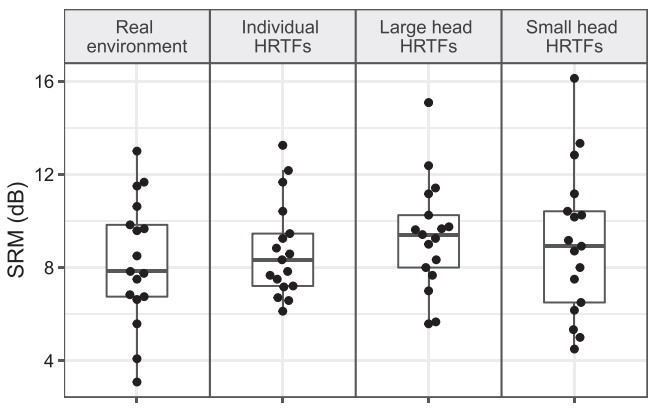

FIG. 2. SRM of adults in the four environments.

positions. To take into account only effects from the head but not the pinnae, the impulse responses were low-pass filtered at 3 kHz (Iida, 2019). A linear mixed-effects model was then fitted for the difference in SRT in the environments based on small and large head HRTFs with the fixed effects of maximum ITD and masker location and the random effect of participant. ITD did not have an effect on the relative SRT performance in the two conditions [$\chi^2(1) = 161.9$, $p = 0.305$] and there was no difference between the two spatial conditions [$\chi^2(1) = 160.9$, $p = 0.699$]. This suggests that SRT performance was not affected by differences in ITDs, and thus head size, between the participant and the heads used for the two non-individualized environments.

These results indicate that for normal-hearing adults, there is no reduction in speech intelligibility in non-individualized virtual environments, at least of the configuration studied here. Participants achieved the same SRM in virtual environments created from generic sets of HRTFs as for their own HRTFs. Two sets of HRTFs from very different sized heads were used in this study, suggesting that this finding is likely to hold for most sets of adult HRTFs, including commonly used artificial heads.

III. STUDY 2: CHILDREN

A. Methods

1. Participants

23 children (13 female, 10 male) aged 7 to 12 years (mean 9.9 years ± 1.7 years) took part in this study. All children had normal peripheral hearing (thresholds below 20 dB HL at 250 to 8000 Hz) as confirmed by an audiogram at the beginning of the session and no known listening difficulties or developmental disorders. All children were native English speakers. Most children were monolingual British English, but two children were bilingual and three children exposed to a second language by family members without speaking it. The children were recruited through online parent groups, local schools, and the university. Parents and children provided consent and the children received young scientist certificates and vouchers as rewards for their participation. Families were reimbursed for travel expenses and informed about the general findings. This study was approved by the UCL Research Ethics Committee.

2. HRTF measurement and head parameters

The same procedure as in the first study was used to measure HRTFs. Particular care was taken to ensure children felt comfortable remaining in the anechoic room by themselves. They were allowed to get accustomed to the room first and practice measurements were taken with a parent present in the room. During HRTF measurements, the child's position was monitored via a camera. At least three measurements were taken for each participant. For children who had difficulties remaining still, further instructions and assistance were provided and additional measurements were taken. In addition to measuring HRTFs, four parameters of head size were measured for each child to investigate differences in HRTFs: the circumference (at the level of the eyebrows), height (from below the chin to the highest point of the head), breadth (above the ears), and the depth (at the level of the eyebrows). The HRTF data set and anthropometric data are available online.[2]

3. Stimuli and norming study

Target and masker stimuli were the same as in the first study. To ensure equal intelligibility in young children for all sentences, a norming study was conducted prior to the experiment with 49 normal-hearing children (22 female, 27 male, native British English speakers) aged 8 to 10 years (mean 9.2 years ± 0.6 years). The ASL target sentences were presented simultaneously with the two speech maskers at multiple fixed levels around the estimated 50%-intelligibility level of each child. Based on the resulting psychometric functions for each sentence, sentence levels were adjusted for equal intelligibility. The sentences were then regrouped into 14 sentence lists (each containing 14 sentences) balanced in level adjustment, keywords, and phonemes.

4. Conditions

SRTs were measured in the real anechoic room and in virtual environments generated from three different sets of HRTFs: the individually measured HRTFs of each participant, non-individualized HRTFs of an artificial head and simplified HRTFs of a spherical head model. HRTFs from a KEMAR artificial head [Knowles Electronics Manikin for Acoustic Research, developed by Burkhard and Sachs (1975)] were used for the non-individualized environment. The KEMAR head was designed to have average adult dimensions of torso, head and pinna shapes based on anthropometric data of a large group of adults. The HRTF set used in this study is part of the publicly available CIPIC database from UC Davis (Algazi *et al.*, 2001). HRTFs recorded with the smaller pinna version were chosen because they are a better approximation for children and would typically be used in studies with them.

The results of the first study show no differences between the use of individual and non-individualized HRTFs in adults. To investigate whether realistic HRTFs are in fact necessary to measure accurate SRTs in children, or whether an approximation of ITDs and ILDs without pinna effects would be sufficient, an additional condition was included using a rigid sphere as a simplified model of a human head. The spherical head HRTFs were generated with a model from Brown and Duda (1998). It contains a head shadow model to simulate ILDs that calculates the diffraction of an acoustic wave at the sphere by Rayleigh's solution and a time delay model that approximates ITDs based on the ray-tracing formula by Woodworth and Schlosberg for high frequencies (Blauert, 1974). Similar to the measured HRTFs, each transfer function was normalized by dividing the pressure at the surface by the pressure at the centre of the sphere in free field. The sphere was modelled

https://doi.org/10.1121/10.0016360

to resemble the KEMAR manikin with a circumference of 57 cm and ear positions at ±100° azimuth (Blauert, 1974).

SRTs were measured for colocated and separated conditions in the real environment and virtual environments based on individual and KEMAR HRTFs. Due to limitations in time and sentence material, only the separated condition was included for the spherical head environment in which the potential difference in SRT was expected to be larger than for the colocated condition.

5. Procedure

The participant was seated in the anechoic room and instructed to repeat the target sentences. A short training session was conducted consisting of three adaptive runs with eight BKB sentences each to familiarize the child with the procedure and with the virtual environments and masker locations. In case a child needed more time to understand the task and to establish reliable responses, further training with BKB sentences was provided. Then, SRTs were measured adaptively for the seven test conditions: colocated and separated conditions for individual HRTFs, KEMAR HRTFs, and the real environment and the separated condition for the spherical head HRTFs. The 14 sentence lists and 6 possible masker combinations for each condition and participant were pseudo-randomized using Latin square counterbalancing. The conditions of the four different environments were presented in pseudo-random order, randomly starting with the separated or colocated condition. Since the previous study with adults showed no effect of presentation order, conditions in the same environment were always conducted consecutively to minimize interruptions caused by switching between headphone and loudspeaker presentation. All conditions were repeated in reverse order to increase measurement accuracy and assess test reliability. The children were monitored via a camera to ensure they faced to the front for conditions in the real environment.

As the adults showed an effect of training, more sentences were used per SRT measurement. Each adaptive run consisted of 22 sentences: 8 BKB sentences followed by an ASL list with 14 sentences in random order. The BKB sentences were used to approximate the threshold at the beginning of the run due to limited ASL sentence material. The maskers started at random points within the first 20 s of the audio files and were played in a loop throughout the run. Their combined level was 65 dB SPL. A 200 ms 1 kHz sine tone was presented as a prompt 500 ms before each target sentence. The initial target level was +10 dB SNR in the colocated and +5 dB SNR in the separated conditions. The target level was then adjusted to track 50% keywords correct with the step size decreasing from 4 dB to 3 and 2 dB at the first two reversals. The initial SNR was set high so that every child could clearly understand the first sentences, and thus remained engaged and motivated. All trials up to the second reversal were considered practice. SRTs were then calculated as the mean of the levels visited for all trials after the first reversal with ASL sentences (including the level following the last trial). Including breaks, the session lasted approximately 90 min.

B. Results

As in the previous study, statistical analyses were done in R and mixed-effects models were only reported in their reduced form, after elimination of non-significant terms and interactions.

Two SRTs were measured for each test condition. On average, the absolute difference between the repetitions was 1.8 dB (SD = 1.8 dB). For SRT pairs with more than three standard deviations difference between repetitions (2 of 161 cases), the higher threshold was considered an outlier and excluded from the analysis. For all other SRT pairs, the average of the two repetitions was used to calculate the SRM. SRTs as a function of age are displayed in Fig. 3. A strong effect of age and large differences between colocated and separated conditions are visible in all environments.

Since the experiment did not include the colocated condition for spherical HRTFs, two separate linear mixed-effects models for the SRT were fitted for the three colocated and the four separated conditions. They each contained the fixed effects of auditory environment, age and repetition and their interactions and a random effect of participant to control for variability between the children. The model for colocated conditions revealed large effects of age [$\chi^2(1) = 9.09$, $p = 0.003$] and repetition [$\chi^2(1) = 16.8$, $p < 0.001$] and a small but significant effect of environment [$\chi^2(2) = 6.88, p = 0.032$]. Performance improved with age by −0.6 dB per year. SRTs were also 1.1 dB lower in the second repetition, suggesting that there was a continuous learning process even after the initial training. Although a small effect of environment was detected (largest between the real and KEMAR environment with −0.8 dB), a *post hoc* Tukey HSD did not find any significant differences between pairs of environments (all $p > 0.39$). In the equivalent model for the separated conditions, significant differences were found for all three fixed factors [age: $\chi^2(1) = 15.7$, $p < 0.001$; environment: $\chi^2(3) = 77.3$, $p < 0.001$; repetition: $\chi^2(1) = 14.0$, $p = 0.003$] as well as the interaction of environment and repetition [$\chi^2(3) = 14.0$, $p = 0.003$]. Age accounted for an improvement of −1.0 dB per year. A *post hoc* Tukey HSD test showed that the differences between environments in this model were mainly due to large differences between the spherical head condition and the other two virtual conditions (both $p < 0.003$). No difference was found between the real, individualized and KEMAR conditions. SRTs in the spherical condition were on average 2.6 dB worse than in the other environments. SRTs in the second repetition were slightly lower in all virtual conditions (all less than 1 dB), similar to the colocated conditions. In the real environment, however, SRTs were 1.3 dB higher in the second repetition. It is unclear what caused this difference. Some participants might have needed more practice to become familiar with the virtual

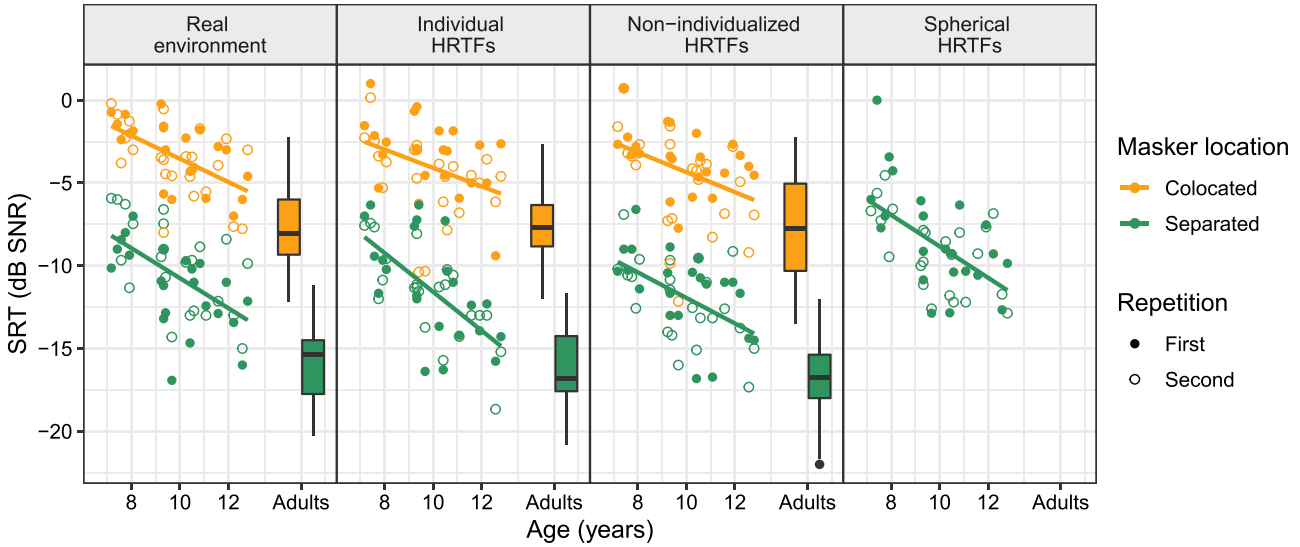

FIG. 3. (Color online) SRTs in the real anechoic environment and virtual environments based on HRTFs from individual participants, non-individualized heads (KEMAR head for children or large and small heads for adults) or a spherical head: results for children across age (points and regression lines) and adults (boxplots). Regression lines were fitted separately for each environment.

environments. In the real environment, SRTs might have been affected by small head movements especially towards the end of the procedure when children might have been more fatigued and less concentrated. However, this is only speculative, as head movements were not tracked.

Overall, and as shown in Fig. 4, large SRMs were reached in all three environments (mean 7.3 dB ± 2.1 dB). A linear mixed-effects model for SRM with the fixed effects of environment and age and random effect of participant showed only a weak trend for age [$\chi^2(2) = 3.491$, $p = 0.06$] resulting in an increase of 0.4 dB per year. No difference between environments was found [$\chi^2(2) = 0.193$, $p = 0.91$]. Since no colocated condition was measured for the spherical head environment, no SRM was obtained. The separated SRTs in this environment were significantly higher than in all other environments potentially due to a degradation of the auditory environment from mismatched binaural cues. As the colocated conditions only differ in the overall spectral filtering of all sound sources which is unlikely to lead to differences in SRTs, it is likely that the overall SRM would be smaller in the spherical head environment.

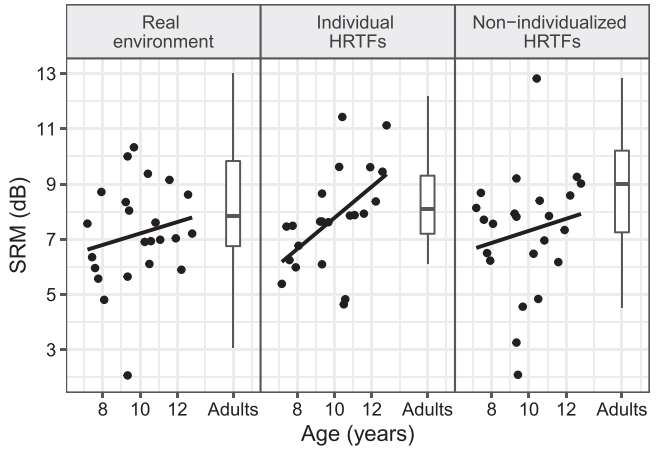

FIG. 4. SRMs in real and virtual environments as described in Fig. 3. Regression lines were fitted separately for each environment, so although it appears that the change of SRM with age depends upon environment, the interaction was not significant in the statistical model.

Whilst speech perception generally improves with age, large variations across participants are present in SRTs and SRM. In the KEMAR conditions, some of this variation may be related to the degree of similarity in head size between the child and the KEMAR head. Recall that four parameters of head size were measured for each child. Head circumference ranged from 52.0 to 56.7 cm (mean 53.7 cm), height from 18.0 to 21.7 cm (mean 19.4 cm), width from 12.6 to 16.2 cm (mean 14.3 cm), and depth from 16.5 to 19.3 cm (mean 18.0 cm). All parameters showed a slow increase with age, which, however, was small compared to the large individual differences between the participants (see supplementary material[3]). The largest head sizes among children approached adult dimensions. Regression models were fitted for all parameters, but the effect of age was only statistically significant for the head height [$F(1, 21) = 11.47$, $p = 0.003$].

To investigate the relationship between head size and performance in the KEMAR environment, age-normed SRT scores were obtained by fitting a linear regression model for the SRTs with age and conditions as main effects and their interaction. The model was then used to predict z-scores for SRTs for all participants. Circumference was used as the measure of head dimensions. As it slightly increased with age, from a mean value of 53.2 cm in 7-year old children to 54.6 cm in 12-year old children, the circumference was also normalized by age in a regression model. A mixed-effects model was then fitted for the normed SRTs in the KEMAR environment with the fixed factors of age-normed head circumference and the random factor of participant. No effect of circumference on normed SRTs was found [$\chi^2(1) = 1.868$, $p = 0.17$]. Children with relatively large heads were not performing better in this environment than others.

https://doi.org/10.1121/10.0016360

IV. GENERAL RESULTS

A. Speech perception in children and adults

In both studies, SRTs in colocated and separated conditions were measured in a real anechoic environment and virtual environments based on individual or non-individualized HRTFs. For children, the non-individualized HRTFs were taken from an artificial KEMAR head whereas for adults, HRTFs from two adults with very large and very small head dimensions were used. As the results for these two sets of HRTFs did not differ, they were combined for this analysis (and in the boxplots in Fig. 3).

SRTs improved strongly with age in children. The youngest children, aged 7 years, had mean SRTs of −2.1 dB SNR in colocated and −9.0 dB SNR in separated conditions. The oldest children, aged 12 years, had −5.6 dB SNR and −14.2 dB SNR, respectively. Thresholds changed on average by −0.6 dB per year in the colocated conditions and by 0.9 dB per year in the separated conditions. Adults had an average SRT of −7.6 dB SNR in colocated and −16.4 dB SNR in separated conditions, which was consistently lower than even the oldest children in all environments. To better understand the magnitude of these differences, a mixed-effects model for SRT was fitted with the fixed effects of participant group, age, auditory environment, and masker locations, their interactions and the random effect of participant. Group and age were included as separate factors to allow for an interaction to account for differences in the trends in SRT with age expected between children and adults. There was a main effect of group [$\chi^2(1) = 8.77$, $p = 0.003$] and age [$\chi^2(1) = 5.65$, $p = 0.018$] as well as an interaction between the two [$\chi^2(1) = 7.63$, $p = 0.006$], due to a change with age in the children but not in the adults. There was also a strong main effect of masker location [$\chi^2(1) = 136.7$, $p < 0.001$], thus SRM, and interaction of age with masker location [$\chi^2(1) = 20.67$, $p < 0.001$] due to the increase in SRM with age in children, albeit small. In addition, a small main effect of environment was observed [$\chi^2(2) = 7.74$, $p = 0.021$], mainly due to differences between the real and non-individualized environments, with SRTs being 0.7 dB lower in the latter.

SRMs in children also increased with age, but only by an average of 0.3 dB per year. Even the youngest children at age 7 had a large average SRM of 6.9 dB. The oldest children aged 12 years had a SRM of 8.6 dB, almost approaching adult performance which was 8.9 dB. No difference between the environments was found for the SRM.

B. Spatial cues in children and adults

Due to their smaller body dimensions, HRTFs from young children differ from those of adults. To understand the extent of these differences, binaural and spectral cues were extracted from the HRTFs measured in a larger pool of listeners. As part of this research project, HRTFs at −90°, 0°, and 90° azimuth and 0° elevation were measured in a total of 21 adults (the 17 participants of the first study plus 4 pilot participants) and 40 children aged 7 to 12 years (the 23 participants of the second study plus 17 children with listening difficulties). Multiple measurements were taken for each participant. The two measurements with the fewest artefacts were used to calculate binaural and monaural cues.

The maximum ITD for each participant was calculated as the time shift of the maximum in the cross correlation function between the head related impulse responses of the left and right ear for sound incidences of −90° and +90° azimuth. ITDs of the two repeated HRTF measurements were averaged to obtain a more stable measure. ITDs had a mean and standard deviation of 672 μs ± 30 μs in the 16 female adults and 723 μs ± 36 μs in the 5 males (see Fig. 5). ITDs in children were on average 645 μs ± 44 μs in the 21 girls and 666 μs ± 35 μs in the 19 boys. Statistical analyses revealed differences between genders and between children and adults as a group but the expected increase with age in children was far outweighed by the large variability in the children.

The maximum ILD was calculated as the averaged level difference between left and right ear HRTFs for sound incidences at −90° and +90°. ILDs were calculated for 1/3rd-octave bands (which roughly correspond to human auditory filters) with centre frequencies from 31.5 Hz to 20 kHz. ILDs are, of course, larger at high frequencies. To specifically investigate differences in ILDs relevant for speech perception and the results in the two studies, the ILDs were weighted with an importance function based on the speech intelligibility index (SII) for frequency bands between 160 and 8000 Hz (ANSI, 1997). Since this weighting removes

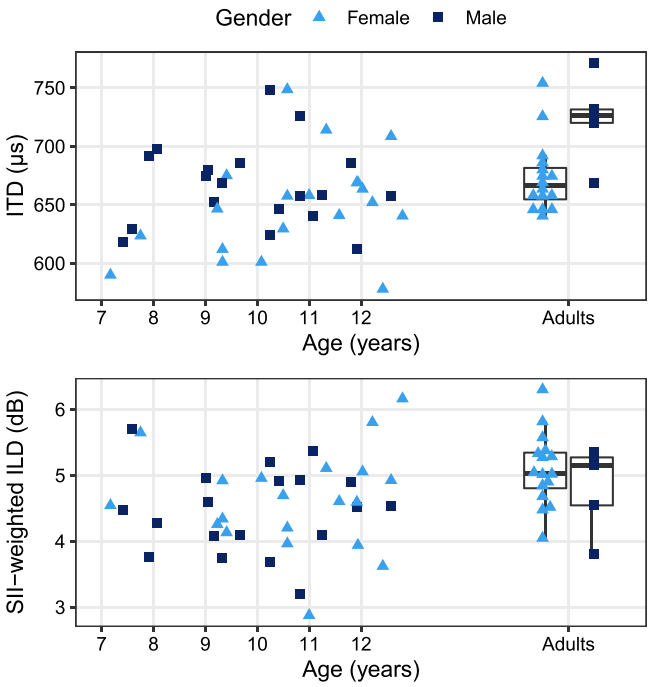

FIG. 5. (Color online) Binaural cues in children and adults: Interaural time differences (top) and interaural level differences weighted by the speech intelligibility index (bottom) for female (light triangles) and male participants (dark squares).

the contribution of frequencies over 8000 Hz, differences between the participants are smaller. The means and standard deviations of the SII-weighted ILDs were 5.0 dB ± 0.6 dB in adults and 4.5 dB ± 0.7 dB in children (see Fig. 5). No gender difference was found but ILDs were statistically smaller in the group of children than in the adults. Like ITDs, variability in children was much larger than in adults.

To investigate differences in spectral cues, the frequency and amplitude of the two lowest characteristic spectral peaks and notches in the HRTFs for 0° azimuth were extracted (see supplementary material[3]). No differences were found between children and adults. Systematic differences between the groups and other possible factors (such as gender differences) were far outweighed by the variability between individuals, which was again larger for children than adults.

V. DISCUSSION

Virtual acoustic environments offer a convenient and flexible framework for experimental and clinical procedures. Since obtaining individual HRTFs is time-consuming and costly, they are often generated from existing HRTFs of an adult head, which are a good approximation for most participants. Many aspects of hearing are robust enough to be accurately measured with non-individualised HRTFs. However, it is uncertain whether that is the case for speech-in-noise perception in challenging listening situations, particularly for young children whose auditory perception is still in development and who experience greater mismatch from non-individualised adult HRTFs. To address this question, two studies were conducted in which SRTs for colocated and spatially separated stimuli, and thus the SRM, were measured in individualized and non-individualised environments as well as in the underlying real anechoic environment in children and adults. Results showed that SRTs and SRMs increased with age and were not yet fully mature in the oldest children aged 12 years. For adults, no difference between the environments was found. In children, performance was also similar across the environments, except for the environment based on spherical HRTFs in which SRTs were much higher.

A. Maturation of speech perception in children

Speech-in-noise perception improves with age during childhood [reviewed by Leibold *et al.* (2019)]. However, large differences in the age of maturation are reported in the literature, likely due to differences in the test procedure, such as target complexity, masker type and spatial configuration of sources.

The current study found that SRTs in sentence recognition tasks with symmetrically placed intelligible maskers improved strongly with age and were not yet mature for the oldest children. The average predicted SRT of the oldest children aged 12 years was −5.5 dB in colocated conditions, 3.6 dB lower than that of the youngest children aged 7 years but still 2.1 dB higher than for adults. In the separated conditions, it was −13.1 dB, 4.2 dB below the 7-year-old children but 3.3 dB above adults. These findings are consistent with studies using similar procedures and stimuli that found improving SRTs up to adolescence for symmetric speech maskers [e.g., Brown *et al.* (2010), Cameron *et al.* (2011), and Misurelli and Litovsky (2012)]. SRT performance in children depends strongly on the type and number of the maskers. For a simple noise masker, Lovett *et al.* (2012) and Murphy *et al.* (2011) found mature SRTs in children aged 6 years. For a speech masker, SRTs were found to improve up to adolescence [e.g., Buss *et al.* (2017), Goldsworthy and Markle (2019), and Wightman and Kistler (2005)]. Speech maskers generally result in less energetic masking since they tend to be less spectrally and temporally dense than noise signals, but provide additional informational masking due to the strong similarity to the target stimuli (Brungart *et al.*, 2006). Due to the large combined effect of informational and energetic masking, SRTs of children and adults are highest for a two-talker masker. In addition to these higher SRTs, Corbin *et al.* (2016) also found longer developmental trajectories for a two-talker masker compared to speech-shaped noise. SRTs for the speech masker did not reach adult-like performance until 13 years of age whilst SRTs for the noise masker were mature at 10 years. Compared to adults, children appear to have particular difficulties with informational masking. Younger children seem to be less capable to allocate their attention to the target stimuli and to ignore information from the masking talkers because higher-order top-down processes such as selective auditory attention only develop late in childhood (Moore *et al.*, 2011; Wightman and Kistler, 2005).

Since the SRM is a derived measure from two SRT measurements, it is more robust against differences between participants, e.g., in their language abilities, their response behaviour but also in differences in their ability to perform the task related to their age. SRMs in the current study were already large for the youngest children and increased by a lesser amount than the SRTs. The youngest children aged 7 years had an average SRM of 6.9 dB in the three environments. In the oldest children aged 12 years the average SRM was 8.4 dB and approached adult levels. Studies also found that children are able to benefit from spatial separation of sound sources from a very young age. SRMs have been measured in children from the age of 2–3 years [e.g., Garadat and Litovsky (2007) and Hess *et al.* (2018)]. For single noise maskers, the SRM is smaller overall and appears to be mature by the age of 3 years (Lovett *et al.*, 2012). For speech maskers, the SRM is larger due to the strong informational masking in conditions with colocated stimuli and matures later in childhood (Cameron *et al.*, 2011). Whilst the SRM for noise maskers in children aged 8–10 years was the same as for adults in a study by Corbin *et al.* (2017), the SRM for two-talker maskers was reduced in children. Conditions with maskers presented only from one side of the head allow the participant to listen with the averted ear and take advantage of the consistently better SNR at that ear. For symmetric maskers, this better-ear

listening is greatly reduced as the participant can only take advantage of brief glimpses of better SNR in either ear (Brungart and Iyer, 2012). The SRM is generally smaller for symmetric maskers and matures later in childhood (Cameron *et al.*, 2011; Misurelli and Litovsky, 2012).

B. Differences between environments

SRTs and SRM in the virtual environments based on HRTFs from other adult heads or artificial heads were very similar to those measured in the real environment. For the adult participants, speech perception performance in environments based on the small or large adult head was not related to their own head size. Likewise in children, age-normed SRTs in the KEMAR head environment were not affected by the similarity of their head size to the KEMAR head. Binaural cues of children were smaller than those of adults, but no statistical differences were found for monaural cues. Overall, the children showed large variability in head size and spatial cues. Whilst most measures showed a small increase with age, individual differences in growth are likely to have a much larger influence. The largest children in this group were similar to adults in head size and spatial cues. The similarity of SRTs in individualized and non-individualized virtual environments suggest that HRTFs from artificial heads are sufficiently accurate to be used for speech perception tasks even in young children, at least for the spatial configurations that were used. In the current studies, large separation angles of $\pm$ 90° were used to measure the SRM. There is a possibility that test conditions with smaller separation angles that lead to smaller and potentially less robust effects of SRM (Marrone *et al.*, 2008; Srinivasan *et al.*, 2016), could be affected by a mismatch in HRTFs.

In children, SRTs were also measured in an environment based on a spherical head model. In this environment, SRTs were significantly higher than in the other three environments, by 2.6 dB on average. Similar results were also found for adults in a study by Cuevas-Rodriguez *et al.* (2021), where SRTs were measured in a word recognition task with symmetrically separated noise maskers in environments based on HRTFs of seven different adult heads and a spherical head. Whilst there was no difference between most human heads, SRTs were significantly higher in the spherical head environment. The spherical head model provides binaural cues similar to those of an adult head but no spectral cues which originate from the filtering of sound waves in the pinnae. Therefore, spherical HRTFs are very flat and do not contain the characteristic direction-dependent peaks and notches of a human head (Iida, 2019). Whilst most of these features are at high frequencies and unlikely to have a strong effect on speech perception, the lowest peak at 4 kHz produced by the first horizontal resonance mode in the pinna could potentially have an influence on speech perception in complex acoustic environments. Unfortunately, this environment was only examined in children and only in the condition with separated sound sources. Whilst the current findings from this condition suggest that a simple approximation of HRTFs with binaural cues is not sufficient to accurately measure SRTs in children, further research would be necessary to investigate whether this effect is the same in colocated conditions and for adults.

C. Voice differences between target and masker

The SRM obtained in this study was smaller than in comparable other studies with similar stimuli and spatial configurations. In a clinical test called the LiSN-S, which contains very similar stimuli and test conditions, both adults and children had larger SRMs (Brown *et al.*, 2010; Cameron *et al.*, 2009; Cameron and Dillon, 2007; Cameron *et al.*, 2011). For children aged 6 to 11 years, average SRMs were 11.8 dB in the Australian English version Cameron and Dillon (2007) and 9.3 dB in the American English version Brown *et al.* (2010) whilst in this study children had an average SRM of 7.1 dB in the equivalent KEMAR environment. This smaller SRM is mainly due to lower SRTs, and thus better performance, in the colocated conditions.

One possible explanation for this might be slight differences in voice characteristics between target and masker stimuli allowing easier auditory scene analysis. Both sets of stimuli were recorded by the same speaker, but with a gap of many years between the recording of targets and maskers. Many studies show that ageing can result in changes in voice characteristics [e.g., Stathopoulos *et al.* (2011)]. A direct comparison of the long-term average spectra of distractors and targets showed them to be quite similar up to 3.5 kHz, with maskers only having less energy at frequencies above 6 kHz. Such small changes are unlikely to be important.

A more likely possibility lies in differences in intonation. We found the mean fundamental frequency in the maskers to be lower than that of the targets (126 Hz vs 144 Hz), but more importantly, to be much less varied. The maskers thus reflect a more monotone reading style, whereas the targets are more highly inflected. These differences, again, might aid auditory scene analysis, and result in lower thresholds in the colocated conditions.

VI. CONCLUSION

To investigate whether individualization of virtual environments is necessary to measure speech perception accurately, SRTs and SRM for simple sentences in the presence of symmetric speech maskers were measured in different auditory environments for children and adults. The results were very similar for the real anechoic environment and virtual environments based on individual HRTFs measured for each participant and HRTFs from other adults or an artificial head. SRTs in children were worse in an environment based on spherical head HRTFs, suggesting that the spatial cues that this simplified model provides are not sufficiently close to those of a human head. The results showed that artificial head HRTFs, or more generally adult HRTFs, are suitable for speech perception tasks with intelligible maskers in children. This was found for target and masker stimuli that were

either colocated or had large spatial separations. It is still possible that for small separation angles the accuracy of spatial cues might be more important and speech-in-noise perception poorer for mismatched HRTFs. However, for the large separation angles typically used for SRM measurements, individual and non-individualized HRTFs resulted in the same SRTs even in children as young as 7 years old. Clinical tests like the LiSN-S are therefore suitable to measure accurate SRMs in children. Whilst this research also supports previous findings that HRTFs from children are more varied and differ from adult HRTFs, adult HRTFs generally provide a sufficient approximation for speech perception tasks.

ACKNOWLEDGMENTS

This research was funded by the Royal National Institute for Deaf People (Grant No. RNID-S49) and the Marie Curie Actions ITN iCARE (improving Children's Auditory Rehabilitation, Grant No. FPT7-607139) and supported by the National Institute of Health Research University College London Hospitals Biomedical Research Centre.

[1]Note that the maskers were recorded several years after the target sentences. The potential effect of age on vocal characteristics is described in the discussion.

[2]See https://doi.org/10.5281/zenodo.6985086 for the HRTF data set for children.

[3]See supplementary material at https://www.scitation.org/doi/suppl/10.1121/10.0016360 for figures of the children's HRTFs, head parameters, and spectral cues.

Algazi, V., Duda, R. O., Thompson, D., and Avendano, C. (**2001**). "The CIPIC HRTF database," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pp. 99–102.

ANSI (**1997**). ANSI-S3.5, *Methods for Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).

ARI (**2017**). "HRTF-database of the Acoustics Research Institute, Vienna," https://www.oeaw.ac.at/isf/das-institut/software/hrtf-database (Last viewed 9 March 2022).

Bench, J., Kowal, Å., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-bench) sentence lists for partially-hearing children," Br. J. Audiol **13**(3), 108–112.

Berzborn, M., Bomhardt, R., Klein, J., Richter, J. G., and Vorländer, M. (**2017**). "The ITA-Toolbox: An Open Source MATLAB Toolbox for Acoustic Measurements and Signal Processing."

Besser, J., Festen, J. M., Goverts, S. T., Kramer, S. E., and Pichora-Fuller, M. K. (**2015**). "Speech-in-speech listening on the LiSN-S test by older adults with good audiograms depends on cognition and hearing acuity at high frequencies," Ear Hear. **36**(1), 24–41.

Best, V., Baumgartner, R., Lavandier, M., Majdak, P., and Kopčo, N. (**2020**). "Sound externalization: A review of recent research," Trends Hear. **24**, 2331216520948390.

Best, V., Mason, C. R., Swaminathan, J., Roverud, E., and Kidd, G. (**2017**). "Use of a glimpsing model to understand the performance of listeners with and without hearing loss in spatialized speech mixtures," J. Acoust. Soc. Am. **141**(1), 81–91.

Blauert, J. (**1974**). *Räumliches Hören (Spatial Hearing)* (S. Hirzel Verlag, Stuttgart).

Braren, H. S., and Fels, J. (**2022**). "Towards child-appropriate virtual acoustic environments: A database of high-resolution HRTF measurements and 3D-scans of children," IJERPH **19**(1), 324.

Brown, C. P., and Duda, R. O. (**1998**). "A structural model for binaural sound synthesis," IEEE Trans. Speech Audio Process. **6**(5), 476–488.

Brown, D. K., Cameron, S., Martin, J. S., Watson, C., and Dillon, H. (**2010**). "The North American Listening in Spatialized Noise-Sentences Test (NA LiSN-S): Normative data and test-retest reliability studies for adolescents and young adults," J. Am. Acad. Audiol. **21**, 629–641.

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**(6), 4007–4018.

Brungart, D. S., and Iyer, N. (**2012**). "Better-ear glimpsing efficiency with symmetrically-placed interfering talkers," J. Acoust. Soc. Am. **132**(4), 2545–2556.

Burkhard, M. D., and Sachs, R. M. (**1975**). "Anthropometric manikin for acoustic research," J. Acoust. Soc. Am. **58**(1), 214–222.

Buss, E., Leibold, L. J., Porter, H. L., and Grose, J. H. (**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.

Cameron, S., Brown, D. K., Keith, R. W., Martin, J. S., Watson, C., and Dillon, H. (**2009**). "Development of the North American Listening in Spatialized Noise-Sentences Test (NA LiSN-S): Sentence equivalence, normative data, and test-retest reliability studies," J. Am. Acad. Audiol. **20**(2), 128–146.

Cameron, S., and Dillon, H. (**2007**). "Development of the Listening in Spatialized Noise-Sentences Test (LISN-S)," Ear Hear. **28**(2), 196–211.

Cameron, S., Glyde, H., and Dillon, H. (**2011**). "Listening in Spatialized Noise–Sentences Test (LiSN-S): Normative and retest reliability data for adolescents and adults up to 60 years of age," J. Am. Acad. Audiol. **22**(10), 697–709.

Carlile, S., and Corkhill, C. (**2015**). "Selective spatial attention modulates bottom-up informational masking of speech," Sci. Rep. **5**, 8662.

Corbin, N. E., Bonino, A. Y., Buss, E., and Leibold, L. J. (**2016**). "Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers," Ear Hear. **37**(1), 55–63.

Corbin, N. E., Buss, E., and Leibold, L. J. (**2017**). "Spatial release from masking in children: Effects of simulated unilateral hearing loss," Ear Hear. **38**(2), 223–235.

Cuevas-Rodriguez, M., Gonzalez-Toledo, D., Reyes-Lecuona, A., and Picinali, L. (**2021**). "Impact of non-individualised head related transfer functions on speech-in-noise performances within a synthesised virtual environment," J. Acoust. Soc. Am. **149**(4), 2573–2586.

Fels, J., Buthmann, P., and Vorländer, M. (**2004**). "Head-related transfer functions of children," Acta Acust. united Acust. **90**(5), 918–927.

Garadat, S. N., and Litovsky, R. Y. (**2007**). "Speech intelligibility in free field: Spatial unmasking in preschool children," J. Acoust. Soc. Am. **121**(2), 1047–1055.

Glyde, H., Buchholz, J. M., Dillon, H., Best, V., Hickson, L., and Cameron, S. (**2013**). "The effect of better-ear glimpsing on spatial release from masking," J. Acoust. Soc. Am. **134**(4), 2937–2945.

Goldsworthy, R. L., and Markle, K. L. (**2019**). "Pediatric hearing loss and speech recognition in quiet and in different types of background noise," J. Speech. Lang. Hear. Res. **62**(3), 758–767.

Hall, J. W., Grose, J. H., Buss, E., and Dev, M. B. (**2002**). "Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children," Ear Hear. **23**(2), 159–165.

Hess, C. L., Misurelli, S. M., and Litovsky, R. Y. (**2018**). "Spatial release from masking in 2-year-olds with normal hearing and with bilateral cochlear implants," Trends Hear. **22**, 1–13.

Huttunen, T., Vanne, A., Harder, S., Paulsen, R. R., King, S., Perry-Smith, L., and Kärkkäinen, L. (**2014**). "Rapid generation of personalized HRTFs," in *Audio Engineering Society—120th Convention Spring Preprints 2006*, pp. 1–6.

Iida, K. (**2019**). *Head-Related Transfer Function and Acoustic Virtual Reality* (Springer Nature, Singapore), pp. 1–234.

Iida, K., Ishii, Y., and Nishioka, S. (**2014**). "Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae," J. Acoust. Soc. Am. **136**(1), 317–333.

Jones, P. R., and Moore, D. R. (**2015**). "Development of auditory selective attention: Why children struggle to hear in noisy environments," Dev. Psychol. **51**(3), 353–369.

Kistler, D. J., and Wightman, F. L. (**1992**). "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," J. Acoust. Soc. Am. **91**(3), 1637–1647.

Kollmeier, B., and Peissig, J. (**1990**). "Speech intelligibility enhancement by interaural magnification," Acta Otolaryngol. **109**, 215–223.
Leibold, L. J., Buss, E., and Calandruccio, L. (**2019**). "Too young for the cocktail party? One reason why children and cocktail parties do not mix," Acoust. Today **15**(1), 37–43.
Lovett, R. E. S., Kitterick, P. T., Huang, S., and Summerfield, A. Q. (**2012**). "The developmental trajectory of spatial listening skills in normal-hearing children," J. Speech. Lang. Hear. Res. **55**(3), 865–878.
Macleod, A., and Summerfield, Q. (**1990**). "A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use," Br. J. Audiol. **24**(1), 29–43.
Marrone, N., Mason, C. R., and Kidd, G. (**2008**). "Tuning in the spatial dimension: Evidence from a masked speech identification task," J. Acoust. Soc. Am. **124**(2), 1146–1158.
Middlebrooks, J. C., and Green, D. M. (**1992**). "Observations on a principal components analysis of head-related transfer functions," J. Acoust. Soc. Am. **92**(1), 597–599.
Misurelli, S. M., and Litovsky, R. Y. (**2012**). "Spatial release from masking in children with normal hearing and with bilateral cochlear implants: Effect of interferer asymmetry," J. Acoust. Soc. Am. **132**(1), 380–391.
Møller, H., Sorensen, M., Jensen, C. B., and Hammershøi, D. (**1996**). "Binaural Technique: Do we need individual recordings?," J. Audio Eng. Soc. **44**(6), 451–469.
Moore, D. R., Cowan, J. A., Riley, A., Edmondson-Jones, A. M., and Ferguson, M. A. (**2011**). "Development of auditory processing in 6- to 11-year-old children," Ear Hear. **32**(3), 269–285.
Murphy, J., Summerfield, A. Q., O'Donoghue, G. M., and Moore, D. R. (**2011**). "Spatial hearing of normally hearing and cochlear implanted children," Int. J. Pediatr. Otorhinolaryngol. **75**(4), 489–494.
Nevard, S. P., and Fourcin, A. J. (**1995**). "The phonetics and linguistics anechoic room (UCL)," UCL Technical Report No. 8, pp. 1–10.
Pelzer, R., Dinakaran, M., Brinkmann, F., Lepa, S., Grosche, P., and Weinzierl, S. (**2020**). "Head-related transfer function recommendation based on perceptual similarities and anthropometric features," J. Acoust. Soc. Am. **148**(6), 3809–3817.
Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Int. J. Audiol. **18**(1), 43–52.
Rychtáriková, M., Bogaert, T. V. D., Vermeir, G., and Wouters, J. (**2011**). "Perceptual validation of virtual room acoustics: Sound localisation and speech understanding," Appl. Acoust. **72**(4), 196–204.
Srinivasan, N. K., Jakien, K. M., and Gallun, F. J. (**2016**). "Release from masking for small spatial separations: Effects of age and hearing loss," J. Acoust. Soc. Am. **140**(1), EL73–EL78.
Stathopoulos, E. T., Huber, J. E., and Sussman, J. E. (**2011**). "Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4–93 years of age," J. Speech. Lang. Hear. Res. **54**(4), 1011–1021.
Sunder, K., and Gan, W-s (**2016**). "Individualization of head-related transfer functions in the median plane using frontal projection headphones," J. Audio Eng. Soc. **64**(12), 1026–1041.
Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**(1), 111–123.
Wightman, F. L., and Kistler, D. J. (**2005**). "Informational masking of speech in children: Effects of ipsilateral and contralateral distracters," J. Acoust. Soc. Am. **118**(5), 3164–3176.
Xie, B. (**2013**). *Head-Related Transfer Function and Virtual Auditory Display*, 2nd ed. (J. Ross Publishing, Plantation, FL).
Yu, G., and Wang, L. (**2018**). "Effect of individualized head-related transfer functions on distance perception in virtual reproduction for a nearby source," in *2018 AES International Conference on Spatial Reproduction—Aesthetics and Science*, Vol. 44(2), pp. 3–7.
Ziegelwanger, H., Majdak, P., and Kreuzer, W. (**2015**). "Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization," J. Acoust. Soc. Am. **138**(1), 208–222.