School-age children benefit from voice gender cue differences for the perception of speech in competing speech

Leanne Nagels, Etienne Gaudrain, Deborah Vickers, Petra Hendriks, Deniz Başkent

School-age children benefit from voice gender cue differences for the perception of speech in competing speech

Leanne Nagels, Etienne Gaudrain, Deborah Vickers, et al.

Citation: The Journal of the Acoustical Society of America **149**, 3328 (2021); doi: 10.1121/10.0004791
View online: https://doi.org/10.1121/10.0004791
View Table of Contents: https://asa.scitation.org/toc/jas/149/5
Published by the Acoustical Society of America

JASA ARTICLE

School-age children benefit from voice gender cue differences for the perception of speech in competing speech

Leanne Nagels,[1,a)] Etienne Gaudrain,[2,b)] Deborah Vickers,[3,c)] Petra Hendriks,[1,d)] and Deniz Başkent[4,e)]

[1]*Center for Language and Cognition Groningen (CLCG), University of Groningen, Groningen 9712EK, Netherlands*

[2]*CNRS UMR 5292, Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics, Inserm UMRS 1028, Université Claude Bernard Lyon 1, Université de Lyon, Lyon, France*

[3]*Sound Lab, Cambridge Hearing Group, Clinical Neurosciences Department, University of Cambridge, Cambridge CB2 0SZ, United Kingdom*

[4]*Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen 9713GZ, Netherlands*

ABSTRACT:

Differences in speakers' voice characteristics, such as mean fundamental frequency (F0) and vocal-tract length (VTL), that primarily define speakers' so-called perceived voice gender facilitate the perception of speech in competing speech. Perceiving speech in competing speech is particularly challenging for children, which may relate to their lower sensitivity to differences in voice characteristics than adults. This study investigated the development of the benefit from F0 and VTL differences in school-age children (4–12 years) for separating two competing speakers while tasked with comprehending one of them and also the relationship between this benefit and their corresponding voice discrimination thresholds. Children benefited from differences in F0, VTL, or both cues at all ages tested. This benefit proportionally remained the same across age, although overall accuracy continued to differ from that of adults. Additionally, children's benefit from F0 and VTL differences and their overall accuracy were not related to their discrimination thresholds. Hence, although children's voice discrimination thresholds and speech in competing speech perception abilities develop throughout the school-age years, children already show a benefit from voice gender cue differences early on. Factors other than children's discrimination thresholds seem to relate more closely to their developing speech in competing speech perception abilities. © *2021 Acoustical Society of America*.

https://doi.org/10.1121/10.0004791

(Received 24 July 2020; revised 2 April 2021; accepted 8 April 2021; published online 18 May 2021)

[Editor: Jody Kreiman] Pages: 3328–3344

I. INTRODUCTION

In daily life, we are often presented with sounds originating from different sources and locations but overlapping in temporal and spectral characteristics. Although listening to one particular speech signal from this mixture may be demanding, adult listeners are generally able to effectively process the speech signal components that belong to the same source by segregating and grouping them, by principles of *auditory stream segregation* (Bregman, 1994). The perception of speech in competing background speech as opposed to competing noise differs in that the masking signal can interfere beyond pure perceptual obliteration of the target signal, also known as "energetic masking." For competing speech maskers, the masking signal can largely overlap with the target speech signal in their spectrotemporal properties, and the masking signal can cause lexical-semantic interference, also sometimes called "informational or perceptual masking" (Carhart *et al.*, 1969; Mattys *et al.*, 2009; Pollack, 1975). When listeners process the target and masking signals, they have to inhibit the information that is provided by the masker on a cognitive level to interpret the target signal correctly (Kidd *et al.*, 2008; Schneider *et al.*, 2007). Therefore, speech stream segregation in the presence of competing noise seems to be primarily a matter of peripheral perceptual processing, namely the correct grouping and interpretation of target speech components (Bronkhorst, 2015; Carhart *et al.*, 1969). In contrast, speech stream segregation in the presence of competing speech seems to rely additionally and heavily on central cognitive mechanisms, such as the allocation of attention to the target signal and the inhibition of masker interference. The similarities and differences between the effects of competing noise and competing speech maskers on listeners' ability to perceive speech and required cognitive resources involved have been studied extensively in adult listeners (Arbogast *et al.*, 2002;

[a)]Electronic mail: leanne.nagels@rug.nl. Also at: Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, Netherlands, ORCID: 0000-0003-4853-969X.

[b)]Also at: Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen 9713GZ, Netherlands, ORCID: 0000-0003-0490-0295.

[c)]ORCID: 0000-0002-7498-5637.

[d)]ORCID: 0000-0002-7584-4078.

[e)]ORCID: 0000-0002-6560-1451.

Brungart *et al.*, 2001; Brungart *et al.*, 2006; Cooke *et al.*, 2008; Evans *et al.*, 2016; Mattys *et al.*, 2009; Ruggles *et al.*, 2011; Scott *et al.*, 2004; Swaminathan *et al.*, 2015).

For children, there seems to be a discrepancy in the acquisition of adult-like speech perception for both masker types. It takes children considerably longer developmentally to reach an adult-like level for perceiving speech in competing speech compared to perceiving speech in competing noise, which is associated with their developing cognitive abilities, such as the inhibition of masker interference (Bonino *et al.*, 2013; Buss *et al.*, 2017b; Buss *et al.*, 2019; Corbin *et al.*, 2016; Hall *et al.*, 2002; Leibold and Buss, 2013). Also, the ability to perceive speech in competing speech often develops less linearly in children than their ability to perceive speech in competing noise (Corbin *et al.*, 2016). Listeners' susceptibility to the informational masking provided by a competing speech masker also largely varies across individuals, even among adult listeners (Swaminathan *et al.*, 2015). In addition, psychometric functions for perceiving speech in competing speech, i.e., accuracy as a function of the target-to-masker ratio (TMR), seem to be less steep for children compared to adults (MacPherson and Akeroyd, 2014; Sobon *et al.*, 2019). Bonino *et al.* (2013) found that 8–10-year-old children performed similarly to adults in recognizing disyllabic words in a speech-shaped noise masker but worse than adults in multi-talker babble or two-talker speech maskers. It seems that the underlying cause of this discrepancy specifically concerns informational masking and might be of a central cognitive nature (McCreery *et al.*, 2019; McCreery *et al.*, 2020; Sobon *et al.*, 2019). For instance, the ability to segregate different speech streams based on differences in speakers' voice characteristics may help with optimally using attentional mechanisms for understanding the target speech signal better. As the ability to discriminate subtle differences in voice cues seems to develop in children during the school-age years (Buss *et al.*, 2017a; Cleary *et al.*, 2005; Flaherty *et al.*, 2019; Nagels *et al.*, 2020a), their ability to benefit from voice differences between target and masker speakers for perceiving speech in competing speech may be limited, potentially affecting further processing stages.

Spatial differences and speech onset asynchrony are strong cues for speech stream segregation in adult listeners (Freyman *et al.*, 2001; Kidd *et al.*, 2005; Lee and Humes, 2012; Zobel *et al.*, 2019). Another cue that particularly improves speech stream segregation in competing speech maskers is differences in speakers' voice characteristics (Bird and Darwin, 1998; Broadbent, 1952; Brungart, 2001; Helfer and Freyman, 2009; Zekveld *et al.*, 2014). For instance, adults are better at perceiving speech in competing speech when the masker speakers are of a different sex than the target speaker (Brungart, 2001). Several studies have followed up on these findings by investigating the relative contribution of speakers' mean fundamental frequency (F0) and vocal-tract length (VTL) to the results (Başkent and Gaudrain, 2016; Darwin *et al.*, 2003). Speakers' mean F0 is defined by the vibration rate of speakers' vocal folds and affects the perceived voice pitch, while speakers' VTL strongly correlates with speakers' height (Fitch and Giedd, 1999) and affects their formant frequencies (Kreiman and Sidtis, 2011). Together, these two voice cues are the primary acoustic features that define speakers' perceived sex or so-called voice gender (Fitch and Giedd, 1999; Skuk and Schweinberger, 2014; Titze, 1989). In the studies by Başkent and Gaudrain (2016) and Darwin *et al.* (2003), a single-talker competing speech masker was created by taking the target speaker's voice and manipulating only the mean F0 or VTL voice parameters to keep all other speaker-specific acoustic features consistent. The results demonstrated that adult listeners also benefit when the masker speaker differs from the target speaker in either their mean F0 only or in their VTL only.

Children also benefit similarly to—or more than—adults when the masker speech is produced by speakers of a different sex than the target speaker (Leibold *et al.*, 2018; Wightman and Kistler, 2005). Even two-and-a-half-year-old toddlers already show this benefit from talker-sex differences for speech stream segregation (Newman and Morini, 2017). However, Flaherty *et al.* (2019) found that, when a two-talker speech masker only differed from the target speech in mean F0, children did not benefit from target-masker F0 differences of −3, −6, or −9 semitones (st) until 7 years of age. In addition, 8–12-year-old children showed a reduced benefit relative to 13–15-year-old children and adults. Therefore, Flaherty *et al.* (2019) have suggested that children may rely more on the combination of F0 and VTL differences to determine speakers' voice gender and hence do not benefit from differences in speakers' mean F0 only. More evidence for this hypothesis is provided by a follow-up study in which speakers' VTL was also manipulated (Flaherty *et al.*, 2021). The results showed that similarly to F0 differences, young children did not benefit from differences in speakers' VTL only, but they did benefit from a change in the two voice cues together, although this benefit was still lower than that observed in older children. This argument is also in line with the results of our previous study that indicate children weigh both F0 and VTL cues to categorize speakers' voice gender (Nagels *et al.*, 2020a). Children may also not be sensitive enough to the acoustic variations induced by only mean F0 or only VTL differences to benefit from these during speech stream segregation and could have been relying on additional acoustic differences for speech segregation in the aforementioned studies by Leibold *et al.* (2018) and Wightman and Kistler (2005). Hence, children's ability to perceive speech in competing speech, as opposed to competing noise, may relate to the development in how well they can discriminate differences in voice cues, which improves during the school-age years (Buss *et al.*, 2017a; Flaherty *et al.*, 2019; Nagels *et al.*, 2020a).

Children's ability to discriminate pitch differences based on pure tones (Jensen and Neff, 1993; Maxon and Hochberg, 1982), mean F0 based on voice stimuli (Buss *et al.*, 2017a; Flaherty *et al.*, 2019; Nagels *et al.*, 2020a), or VTL cues based on voice stimuli (Nagels *et al.*, 2020a) develops during the school-age years. Higher-order

cognitive mechanisms of voice perception also develop during this period, for instance, children's ability to recognize unfamiliar voices (Creel and Jimenez, 2012; Fecher *et al.*, 2019; Mann *et al.*, 1979) or their weighting of voice and speech cues for categorization tasks (Floccia *et al.*, 2009; Hazan and Barrett, 2000; Nagels *et al.*, 2020a; Nittrouer and Miller, 1997). However, the exact relationship between children's perception and discrimination of voice cue differences and their ability to use these differences for speech stream segregation is still not well understood. Flaherty *et al.* (2019) did not find a significant correlation between individual children's ability to discriminate F0 cues and their benefit from F0 differences for perceiving speech in competing speech. This finding indicates that voice discrimination and the functional usage of voice cue differences may, in fact, develop independently from each other.

An additional factor that may play a role in the development of children's ability to perceive speech in competing speech is their general language development (Klein *et al.*, 2017; McCreery *et al.*, 2017; McCreery *et al.*, 2020). When a target speech stimulus is presented simultaneously with a masker, some parts of the target are obliterated or inaccessible to the listener. This missing information can be *restored* by the listener by relying on the acoustic and linguistic redundancy inherent to speech and language. Speech information can sometimes be superfluous to listeners, as speech cues are coded in multiple ways and words can often be predicted based on their probability, i.e., word frequency or neighborhood density, or sentential context (Başkent *et al.*, 2016). The ability to restore missing segments depends on the listeners' language abilities, which are not fully developed yet for school-age children. Support for development in children's speech restoration abilities is provided by studies using a gating paradigm in which parts of words or sentences have been gated off, and the proportion of the word segment that listeners need for correct word recognition is examined. These studies have shown that young children require a greater amount of word segments for correct word recognition than older children and adults, who seem to use word probability and sentential context information more effectively (Craig *et al.*, 1993; Elliott *et al.*, 1987; Metsala, 1997). On the other hand, using a perceptual restoration paradigm (Samuel, 1996; Warren, 1970), Newman (2004) found that the perceptual restoration abilities of 5-year-old children were equal to those of adults. Nittrouer and Boothroyd (1990) also showed that 4–6-year-old children used lexical and syntactic constraints to the same extent as adults for perceiving speech in competing noise. Nevertheless, Buss *et al.* (2019) observed a discrepancy in young children's ability to benefit from sentential context for perceiving speech in competing noise compared to competing speech. Young children seem to benefit from sentential context equally to older children and adults for perceiving speech in competing noise but less so for perceiving speech in competing speech. Buss *et al.* (2019) suggested that the high cognitive demands associated with perceiving speech in competing speech may prevent benefit from sentential context in young children. In addition, a correlation between the ability to perceive speech in competing noise or competing speech seems to depend on the complexity of the linguistic stimuli that are used (Klein *et al.*, 2017; McCreery *et al.*, 2017; McCreery *et al.*, 2020). Deducing from such observations, it is not entirely evident yet how language development is related to concurrent speech perception and whether developmental effects are always present or depend on the specific task and materials that are used.

In the present study, we investigated how the benefit from differences in speakers' mean F0 and VTL for perceiving speech in competing speech develops in children during the school-age years (4–12 years of age). In addition, we examined whether children's benefit from F0 and VTL differences relates to their ability to discriminate these voice cues by using their F0 and VTL discrimination thresholds [taken from Nagels *et al.* (2020a)]. We used a child-friendly version of the coordinate response measure (CRM) [used earlier by, for instance, Bolia *et al.* (2000), Brungart (2001), Hazan *et al.* (2009), Moore (1981), Saleh *et al.* (2013), and Welch *et al.* (2015)] with a single-talker speech masker, which was created by manipulating the F0 and VTL parameters of the target speaker's voice and presented at three fixed TMRs. Based on previous research, we expected that children's ability to perceive speech in competing speech would improve as a function of age during the school-age years (Bonino *et al.*, 2013; Buss *et al.*, 2017b; Corbin *et al.*, 2016; Hall *et al.*, 2002; Leibold and Buss, 2013). Furthermore, if children only show substantial benefit from a combined change in F0 and VTL cues, as suggested by Flaherty *et al.* (2019, 2021), children's performance will improve when both F0 and VTL cues are manipulated, but not when only one individual voice cue is manipulated. Finally, if children's benefit from voice gender cue differences directly depends on their ability to discriminate differences in these cues, a significant correlation between these measures is expected.

We also collected vocabulary size scores as a marker of language development (Marchman and Fernald, 2008). Currently, the effects of language development on children's ability to perceive speech in competing speech remain unclear. While some studies have reported that children's perceptual restoration abilities and use of lexical and syntactic constraints are adult-like (Newman, 2004; Nittrouer and Boothroyd, 1990), other studies have reported that young children use word probability and sentential context information less effectively than older children and adults (Buss *et al.*, 2019; Craig *et al.*, 1993; Elliott *et al.*, 1987; Metsala, 1997). Vocabulary size could have some effects on children's ability to perceive speech in competing speech (Klein *et al.*, 2017; McCreery *et al.*, 2017, 2020), as vocabulary size is age-specific and increases during the school-age years. However, as we tested young children 4 years of age and older, our stimuli for the current study consisted of simple closed-set sentence materials, where the child only had to identify color terms and number words, words that children would be familiar with already and posing a closed-set of response options. As a result, the effects of vocabulary

https://doi.org/10.1121/10.0004791

size on school-age children's performance may be minimal in this study.

II. METHOD

The experiment was part of a larger project on the perception of indexical cues in kids and adults (PICKA) for which we collected data from the same population of children and adults for a number of studies on voice and speech perception (Nagels *et al.*, 2020a; Nagels *et al.*, 2020b).

A. Participants

Fifty-eight Dutch children between 4 and 12 years of age and 15 Dutch adults between 20 and 29 years of age took part in the experiment. The selected age range for children was based on the ages at which children attend primary school in the Netherlands (4–12 years) and would therefore be expected to be able to perform the experiments. However, three of the five 4-year-old children who completed the other PICKA experiments did not fully complete the current experiment due to attentional and motivational issues. All PICKA measures were done on the same day during one testing session of approximately 60–90 min. The speech in competing speech perception task was the last and longest task, and lasted approximately 15–20 min. The partial data of these children were not included in the data analysis, reducing the number of child participants to 55. Also, the vocabulary size of one 5-year-old participant was not measured, but their data were included in the analysis. The demographic characteristics of participants, categorized into five specified age groups used in parts of the data analysis, are summarized in Table I. The age groups each spanned 2 years. For instance, the 4–6-year-old age groups consisted of children who were 4 years of age or older but younger than 6 years of age ($\geq$4 years and $<$6 years). We primarily used age as a continuous variable for data analysis, but age groups were used to approximate the age at which children showed adult-like performance. We recruited child participants via local primary schools and after-school care centers, and adult participants via online advertisements. All participants were monolingually raised native speakers of Dutch and reported no hearing or language disorders.

To ensure that all participants had normal hearing, we used a portable Interacoustics (Middelfart, Denmark) AS608B screening audiometer to conduct a short 20 dB HL pure-tone audiometric screening at octave frequencies between 500 and 4000 Hz. The raw scores of children's vocabulary size were measured using the Dutch version of the Renfrew Word Finding Vocabulary Test (Renfrew, 1995), which had a maximum achievable score of 50 points. Before participants took part in the experiment, they were provided with detailed information about the study, and a written informed consent form was signed by the adult participants and by the parents or legal guardians of the child participants. Ethical approval of the study was given by the Medical Ethical Review Committee of the University Medical Center Groningen (METc 2016.689).

B. Stimuli and apparatus

We used a CRM task [first used by Moore (1981) and used by many others, e.g., Bolia *et al.* (2000), Brungart *et al.* (2001), Hazan *et al.* (2009), Saleh *et al.* (2013), and Welch *et al.* (2015)] with sentence stimuli adapted from the English stimuli used by Hazan *et al.* (2009) and Welch *et al.* (2015), translated into Dutch. The 48 target sentences consisted of a carrier phrase in which one of six colors and one of eight numbers were mentioned, e.g., *Laat de hond zien waar de rode* (color) *twee* (number) *is* [Show the dog where the red (color) two (number) is]. The six basic colors were all disyllabic words in Dutch (*rode, zwarte, groene, blauwe, witte*, and *gele*) [red, black, green, blue, white, and yellow], and the eight numbers were all monosyllabic words in Dutch (1–10; but excluding *zeven* [seven] and *negen* [nine], which are disyllabic words in Dutch). Using these closed-set sentence stimuli with words that are highly familiar and acquired early in life (Brysbaert *et al.*, 2014) makes the test-retest reliability of the CRM in general moderate to high (Saleh, 2013; Semeraro *et al.*, 2017). For the masker speech, we used a second set of 48 sentences with the same structure in which the call sign *hond* [dog] was replaced by *kat* [cat]. Sentence chunks ranging from 150 to 300 ms were then randomly selected from these sentences and concatenated after applying 50-ms raised cosine ramps, and avoiding sentences with the same color or number. We used sentence chunks instead of complete sentences for the masker speech, similar to the study of El Boghdady *et al.* (2019), to make the masker differ in structure from the target. We expected this would make the task easier to comprehend for children and help avoid potential confusion about which speaker they should attend to. To that purpose, the masker also started 750 ms before the target, and ended 250 ms after the target. The target and masker stimuli were produced by a female native speaker of Dutch with a standard Dutch accent, a mean F0 value of 242 Hz,

TABLE I. Demographic characteristics of participants divided into five specified age groups. Age is given in decimal years. Vocabulary corresponds to the raw scores on the Renfrew Word Finding Vocabulary Test (maximum score of 50 points). N/A, not applicable.

Participant group	Number of participants	Age (median; range)	Vocabulary (median; range)
4–6 years	10	5.42; 4.08–5.83	36; 29–41
6–8 years	13	7.17; 6.25–7.83	40; 37–46
8–10 years	16	9.00; 8.08–9.83	45; 40–49
10–12 years	16	11.08; 10.00–12.00	46; 41–48
Adults	15	24.42; 20.8–29.17	N/A

and an approximated VTL size of about 13.6 cm based on the speaker's height of 166 cm (Fitch and Giedd, 1999). The recordings were made in an anechoic room at a sampling rate of 44.1 kHz. The duration of the target stimuli ranged from 2.14 to 2.49 s with a mean duration of 2.27 s.

As differences in F0 are commonly measured in Hertz and differences in VTL in centimeters, we expressed the differences in F0 and VTL in semitones (st). This manipulation and comparison between cues were also done in prior studies from our research group (El Boghdady *et al.*, 2019; Fuller *et al.*, 2014; Gaudrain and Başkent, 2018; Nagels *et al.*, 2020a). There were four different voice conditions produced for the masker speech: (1) the same voice parameters as the target speaker, (2) a difference of −12 st in F0, (3) a difference of +3.8 st in VTL, or (4) a combined difference of −12 st in F0 and +3.8 st in VTL, relative to the female target speaker's voice. These voice differences correspond to mean F0 values of approximately 242 and 121 Hz and VTL sizes of approximately 13.6 and 16.7 cm. The specified differences of −12 st in F0 and +3.8 st in VTL are consistent with earlier findings by Smith *et al.* (2007), Smith and Patterson (2005), and Peterson and Barney (1952). Earlier studies from our research group confirmed that these differences in F0 and VTL, when processed in the same way as in this study, reliably change the perceived gender of a voice for normal-hearing adult listeners (Fuller *et al.*, 2014; Nagels *et al.*, 2020a).

All sentence stimuli were analyzed and resynthesized via STRAIGHT software (Kawahara and Irino, 2005) using MATLAB (MathWorks Inc., 2012). The mean F0 of the sentences was first normalized to a value of 242 Hz. Subsequently, the F0 contour and the spectral envelope of the sentences were extracted. We adjusted the masker speech stimuli to match the specified differences in F0 and VTL of one of the four masker speech voice conditions. We also resynthesized the masker speech stimuli with no differences in F0 and VTL to prevent giving any unfair advantage to this condition in case artifacts may have been arising from the resynthesis procedure itself. For the F0 manipulation, the original F0 fluctuations were preserved by multiplying the overall F0 contour by the specified change in mean F0 to only alter the mean F0 of the sentence. For the VTL manipulation, an overall spectral shift of the formant frequencies was produced by compressing the spectral envelope toward the lower frequencies to approximate a perceived change in speakers' VTL. The actual VTL size of the speaker could not be measured and hence had to be estimated. We then used STRAIGHT's pitch synchronous overlap-add (PSOLA) resynthesis method to recombine the modified F0 contour and spectral envelope.

C. Procedure

As mentioned before, the current speech in competing speech perception experiment was part of a larger project on children's perception of indexical cues. We had also examined children's ability to discriminate differences in F0 and VTL cues and their perceptual weighting of F0 and VTL cues for voice gender categorization prior to this experiment during the same test session (Nagels *et al.*, 2020a). For the discrimination experiment, we measured the just-noticeable differences (JNDs) in F0 and VTL that children could perceive via a three-interval, three-alternative forced-choice (3I-3AFC) adaptive procedure using CVCVCV nonwords, for instance, "ba-ki-mo." The F0 and VTL parameters of the stimuli were manipulated using the same STRAIGHT procedure as described in Sec. II B for the current experiment. The initial voice difference was 12 st in F0 or VTL values relative to the original female speaker's voice. After two consecutive correct responses, the voice difference decreased two step sizes, and after an incorrect response, the voice difference increased one step size (2-down, 1-up). The step size initially had a value of 2 st, but after 15 trials with the same step size or when the difference became smaller than 2 times the step size, the step size was divided by $\sqrt{2}$. The geometric mean of the voice difference values at the last six of eight reversals was calculated to determine participants' JND, which corresponds to the 70.7% correct discrimination point on the psychometric function (Levitt, 1971). Our results showed that the discrimination of differences in F0 was adult-like around the age of 8 for VTL, while the ability to discriminate differences in F0 was still not adult-like at the age of 12 for most children. More information about the methods and the procedure of these experiments is presented in Nagels *et al.* (2020a).

The speech in competing speech perception experiment started with a practice session consisting of eight trials to familiarize participants with the task. Participants heard the target sentences without the masker during the first three practice trials and a combination of the target and masker speech with different F0 and VTL parameters with a TMR of +6 dB for the remaining practice trials. We did not set a criterion level for performance before moving on to the experiment trials. The experiment session consisted of seven items per TMR (–6, 0, or +6 dB) for each of the four masker speech voice conditions, resulting in a total number of 84 trials (7 items × 3 TMRs × 4 voice conditions). The TMR values were chosen based on the adaptive TMR values of Flaherty *et al.* (2019), which were mostly between −6 and +6 dB, and pilot testing with four children between 6 and 12 years of age who were not included in the study. All items were presented in a randomized order in a single block. The total duration of the experiment was approximately 15 min, including two optional breaks.

The experiment was conducted on a Dell (Round Rock, TX) XPS 13 in. touchscreen laptop using a child-friendly experiment interface (Fig. 1) created in MATLAB. The stimuli were presented to participants via Sennheiser (Wedemark, Germany) HD 380 Pro headphones. Child participants were tested in a quiet room in their homes, and adult participants were tested in a quiet testing room at the University of Groningen. Before the experiment, young children were asked to name all six basic colors and eight numbers used to ensure they knew the correct words. We instructed participants to attend only to the target speech, which started 750 ms later than the masker speech and contained the

https://doi.org/10.1121/10.0004791

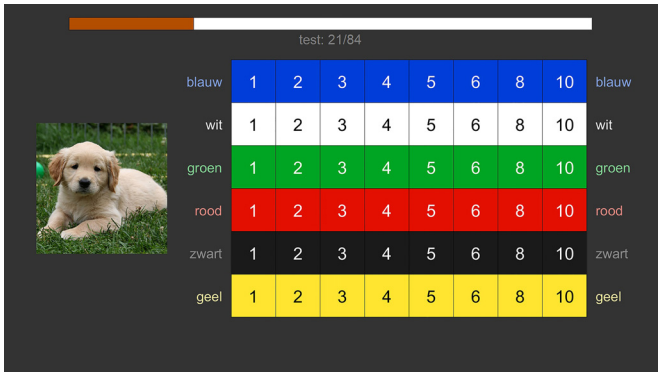

FIG. 1. (Color online) The experiment response interface.

carrier phrase, e.g., *Laat de hond zien waar de* (color) (number) *is* [Show the dog where the (color) (number) is]. The participants were told to press the color-number combination button that was mentioned in the target speech as fast as possible after hearing the stimulus. They could choose between 48 buttons (6 colors × 8 numbers) and would receive 1 point if they had both the color and number correct and 0 points when they had the color, number, or neither of them correct, similar to, for instance, Brungart *et al.* (2001). Thus, the probability of giving a correct response due to chance was 2.08%. After a response, the experiment would continue to the next trial. Participants did not receive any feedback on the accuracy of their responses during the practice and experiment sessions.

D. Data analysis

Children's accuracy scores were analyzed as a function of age, TMR, and target-masker differences in F0 and VTL. We normalized the target-masker differences in F0 and VTL in semitones by defining them as $\delta F0 = -\Delta F0/12 - 0.5$ and $\delta VTL = \Delta VTL/3.8 - 0.5$, to make the differences in F0 and VTL commensurate for model fitting. As a result, $\delta F0$ and δVTL values of −0.5 corresponded to the original female speaker's voice parameters, while $\delta F0$ and δVTL values of +0.5 corresponded to the manipulated male-sounding voice. Subsequently, we fitted a mixed-effects logistic regression model using the lme4 package (Bates *et al.*, 2015) in R (R Core Team, 2020) with random intercepts per participant. We only fitted the model on children's data without those of adults due to the non-continuous distribution of age. We used analysis of variance (ANOVA) chi-square tests to assess the improvement in models of children's accuracy scores due to the deletion of individual factors, starting from the full factorial model, in lme4 syntax: `correct ~ δF0*δVTL*TMR*log(age) + (1|participant)`. The outcome variable *correct* indicated whether the child had both the color and number correct (1 point) or not (0 points). The predictor variables $\delta F0$ and δVTL consisted of the normalized differences in F0 and VTL; *TMR* represented the different TMRs of −6, 0, and +6 dB; and *log(age)* indicated children's log-transformed age in decimal years.

Furthermore, we used a Dunnett's test using the DescTools package (Signorell *et al.*, 2018) to compare the mean benefit from target-masker differences in F0 and VTL of different child age groups to that of adults and approximate at what age children's benefit is adult-like. For this analysis, we converted participants' statistical benefit from target-masker differences in F0 and VTL into "Berkson" units per st (Bk/st) (Hilkhuysen *et al.*, 2012). The conversion into Bk/st makes the differences in participants' benefit from F0 and VTL changes easier to interpret, as an increase in 1 Bk/st corresponds to doubling the odds of getting a correct response for each semitone of voice difference. To obtain the voice-benefit values, we first computed two mixed-effects logistic regression models with random intercepts and slopes for $\delta F0$ per participant, in lme4 syntax: `correct ~ (δF0|participant)`, and for δVTL per participant, in lme4 syntax: `correct ~ (δVTL|participant)`. Subsequently, we extracted participants' model coefficients and scaled them to correspond to $\log_2$ odds per semitone, i.e., `δF0.coefficient/[12*log(2)]` and `δVTL.coefficient/[3.8*log(2)]`, because the logit is based on the natural log. In addition, we examined at what age children's accuracy scores were approximately adult-like by performing a Dunnett's test comparing the mean logit-transformed accuracy scores of the specified child age groups to that observed in adults. We applied a logit-transformation on the accuracy scores to take into account the near ceiling-level scores of adults and older children, primarily in the +6 dB TMR condition.

We performed an additional analysis in which we interpolated participants' accuracy scores across TMRs to the same performance level for all participants, including adults, as the proportion of the benefit from target-masker differences seems to depend on the overall accuracy scores. Particularly, because adult participants often demonstrated ceiling-level performance, except in the −6 dB TMR condition, there was less room for improvement in their performance, which makes the proportion of their benefit from target-masker differences relatively small compared to that observed in children, despite the fact that the logistic model would take the saturation into account. We calculated the mean logit-transformed accuracy scores of children and adults and interpolated these across TMRs to a logit score of 1.79 (equal to an accuracy score of 85.7% correct) in the same-voice condition with no target-masker differences in F0 or VTL. This specific score was chosen for interpolation, as it resulted in the fewest outliers. The scores of six children were all below 85.7%, and the scores of three children and three adults were all above 85.7% in the same-voice condition. We computed a linear mixed-effects model to examine participants' benefit of F0 and VTL differences once their accuracy scores were interpolated, in lme4 syntax: `interpolated logit score ~ δF0*δVTL*log(age) + (1|participant)`. Furthermore, we used a Dunnett's test to compare the mean benefit from target-masker differences in F0 and VTL of the specified child age groups to that of adults based on their interpolated logit-transformed accuracy scores.

Finally, we performed several correlation analyses to investigate if children's benefit from target-masker differences

in F0 and VTL and their overall ability to perceive speech in competing speech were related to their F0 and VTL discrimination thresholds (Nagels *et al.*, 2020a). To prevent a correlation coming forth from merely a general effect of age, we used the residuals of children's benefit from target-masker differences in F0 and VTL, their overall accuracy scores, and their F0 and VTL discrimination thresholds. To calculate the residuals, we used linear regression models with children's F0 or VTL voice-benefits (in Bk/st), their overall accuracy scores, and children's F0 or VTL JNDs as outcome variables and only age as a fixed effect, e.g., overall accuracy scores ∼ age. Subsequently, we calculated the Pearson's correlation coefficient for the correlations between the residuals of children's F0 and VTL voice-benefits and their corresponding JNDs and the correlations between the residuals of their overall accuracy scores and F0 and VTL JNDs. In addition, we examined if children's benefits from F0 and VTL differences were related and if children's vocabulary size potentially affected their ability to perceive speech in competing speech by calculating the Pearson's correlation coefficient between the residuals of children's vocabulary scores and their overall accuracy scores.

III. RESULTS

A. F0 and VTL benefit for perceiving speech in competing speech

Figure 2 shows the accuracy scores of participants in percentage points as a function of TMR, masker speech

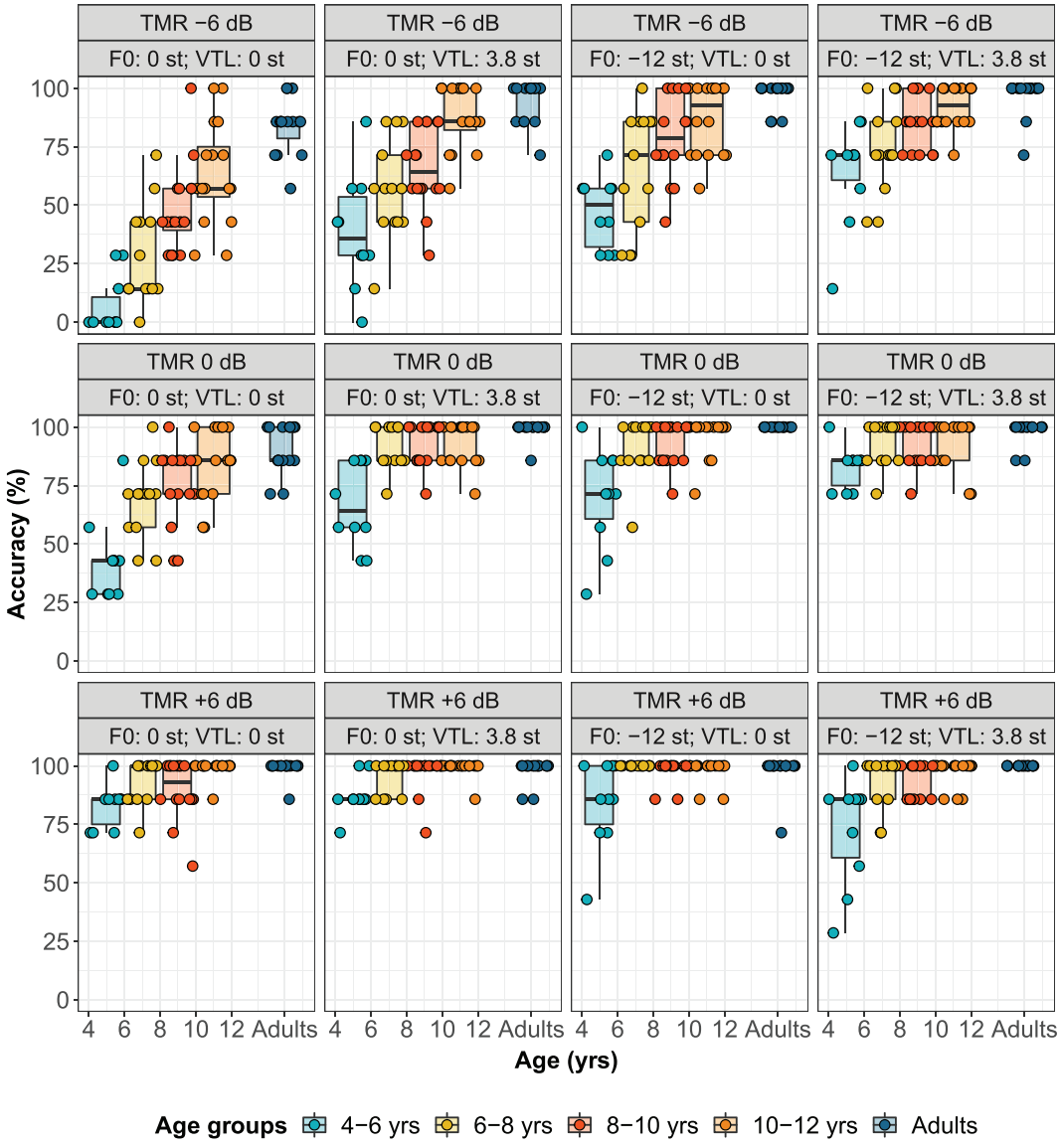

FIG. 2. (Color online) Accuracy scores of participants for perceiving speech in competing speech in percentage points as a function of age, TMR, and voice condition ($N_{children} = 55$, $N_{adults} = 15$). The panels from top to bottom show the accuracy scores in percentage points for the −6 dB TMR (upper panels), 0 dB TMR (middle panels), and +6 dB TMR (lower panels) conditions. Each row consists of four plots that show the accuracy scores per masker speech voice condition, arranged from the condition with target-masker differences of 0 st in F0 and VTL (left panels) to the condition with target-masker differences of −12 st in F0 and +3.8 st in VTL (right panels). The boxplots show the median accuracy scores of participants per age group and the lower and upper quartiles. The dots represent individual data points at participants' age, and the whiskers indicate the lowest and highest data points within ±1.5 times the interquartile range.

https://doi.org/10.1121/10.0004791

voice condition, and age. While adults demonstrated near ceiling-level performance in most conditions, most 4–6-year-olds could not perform the task in the condition with no target-masker differences in F0 or VTL at a TMR of −6 dB. Figure 3 shows the increase in participants' accuracy scores in percentage points as a function of the masker speech voice condition with respect to the same-voice condition, averaged across TMRs. We used backward stepwise model comparison using ANOVA chi-square tests to select the best fitting, most parsimonious model for children's total number of correct responses by deleting one factor at a time from the four-way interaction in the full model. The model comparison analysis indicated that the model with three-way interactions between *δF0*, *δVTL*, and *TMR* and between *δF0, TMR*, and *age* and a two-way interaction between *δVTL* and *age* was the best fitting model, in lme4 syntax: `correct ~ δF0*δVTL*TMR + δF0*TMR*log(age) + δVTL*log(age) + (1|participant)`.

The best model shows children's accuracy scores significantly increased as the TMR became more advantageous [$z = 2.87$, estimate = 0.20, standard error (SE) = 0.071, $p < 0.01$] and as a function of age ($z = 11.80$, estimate = 2.83 SE = 0.239, $p < 0.001$). Children's accuracy scores also improved as a result of target-masker differences in δVTL ($z = 2.50$, estimate = 1.59, SE = 0.637, $p < 0.05$). There were significant two-way interactions between TMR and δF0 ($z = -3.03$, estimate = −0.43, SE = 0.141, $p < 0.01$) and between TMR and δVTL ($z = -3.97$, estimate = −0.08, SE = 0.021, $p < 0.001$), demonstrating that the beneficial effect of target-masker differences in F0 or in VTL on accuracy scores became smaller as the TMR became more advantageous. Furthermore, there was a significant two-way interaction between δF0 and δVTL ($z = -5.65$, estimate = −1.16, SE = 0.205, $p < 0.001$), showing that a combined target-masker difference in F0 and VTL increased children's accuracy scores less than the additive effect of individual cues. Also, there was a significant three-way interaction between δF0, δVTL, and TMR ($z = -1.99$, estimate = −0.08, SE = 0.041, $p < 0.05$), indicating the beneficial effect of a combined target-masker difference in F0 and VTL on children's accuracy scores decreased as the TMR became more advantageous. Finally, the three-way interaction between δF0, TMR, and age was significant ($z = 2.33$, estimate = 0.17, SE = 0.071, $p < 0.05$), showing that the beneficial effect of target-masker differences in F0 on accuracy scores became smaller as the TMR became higher, but that this differed across ages. However, based on Fig. 2, this effect mainly seems to be caused by some 4–6-year-old children who showed a detrimental effect of F0 differences in the +6 dB TMR condition. To summarize, children's accuracy scores generally improved with age, higher TMRs, and target-masker differences in either or both F0 and VTL cues. The benefit from target-masker differences in F0 and VTL decreased for children as the TMR became more advantageous. We did not find any differences in the benefit that children derived from target-masker differences across age, meaning the size of the improvement in accuracy as a result of differences in F0 and VTL was the same for children of all ages.

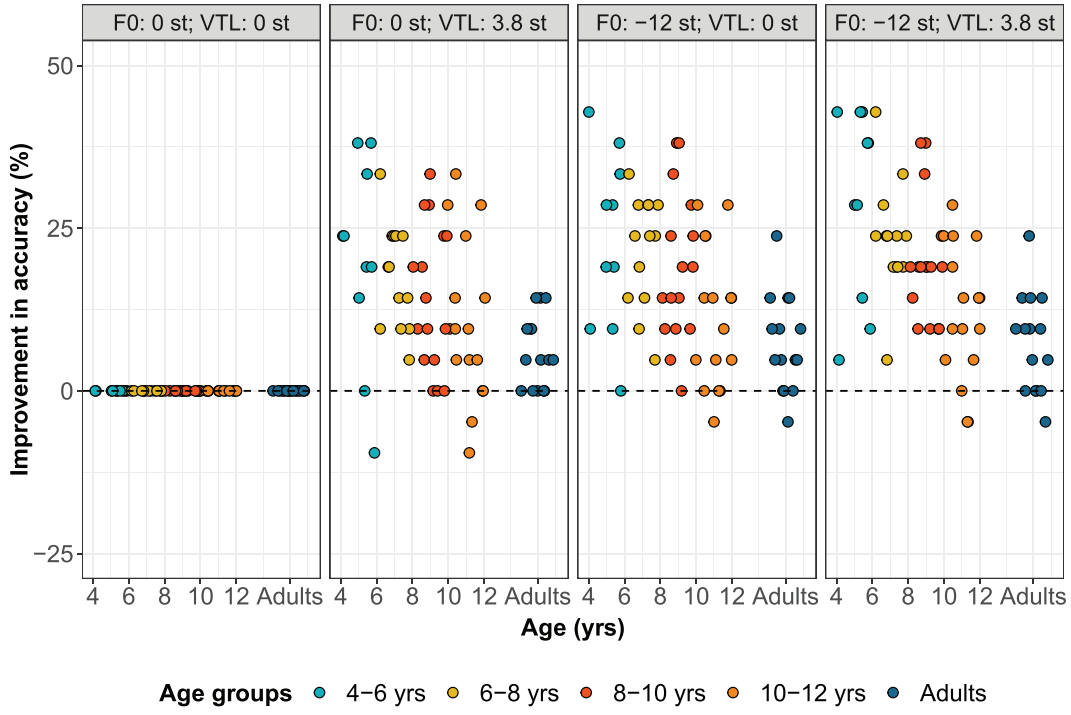

FIG. 3. (Color online) The improvement in the accuracy scores of participants, averaged across TMRs, as a function of the masker speech voice condition ($N_{children} = 55$, $N_{adults} = 15$) shown across the panels and as a function of age within each panel. Each panel shows the improvement in individual participants' accuracy scores with respect to their accuracy scores in the voice condition with no target-masker differences in F0 or VTL. The dots represent individual data points at participants' age.

B. Adult-like benefit from F0 and VTL differences and adult-like accuracy scores

Figures 4(A) and 4(B) show participants' benefit from target-masker differences in F0 and VTL cues in Bk/st for each TMR condition, and Fig. 4(C) shows participants' accuracy scores for each TMR condition presented in percentage points. Positive Bk/st values reflect a benefit from F0 or VTL differences, while negative Bk/st values reflect a detrimental effect of F0 or VTL differences. Based on Figs. 4(A) and 4(B), we observe that almost all children demonstrated a benefit of F0 and VTL differences in the −6 and 0 dB TMR conditions. The Dunnett's test results are presented in Table II. Since we compared the results of adults to those of different child age groups instead of using age as a continuous factor, this analysis only gives a rough approximation of when children's performance is adult-like rather than a precise estimate. For F0, children between 4 and

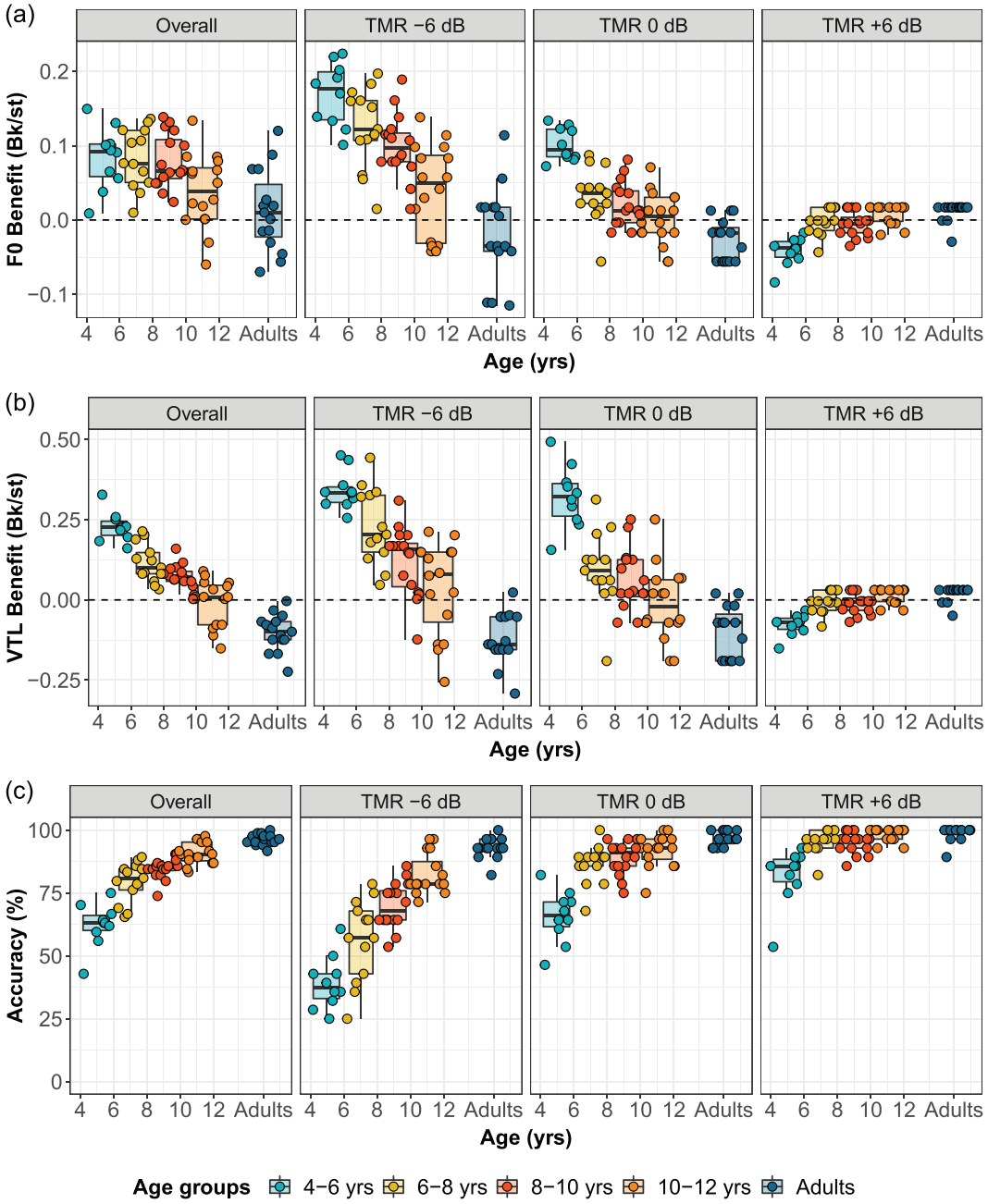

FIG. 4. (Color online) Participants' benefit from target-masker differences in F0 and VTL in Bk/st and their accuracy scores per TMR condition ($N_{children} = 55$, $N_{adults} = 15$). (A) Participants' benefit from target-masker differences in F0 for perceiving speech in competing speech in Bk/st overall and per TMR condition. (B) benefit from target-masker differences in VTL for perceiving speech in competing speech in Bk/st overall and per TMR condition. The boxplots show participants' median benefit from F0 or VTL in Bk/st per age group and the lower and upper quartiles of the values that were obtained. The dashed line at the value of 0 was included to indicate if participants benefited from voice difference, i.e., positive value, or not, i.e., negative value. (C) Participants' accuracy scores in percentage points per TMR condition. The boxplots show participants' median accuracy scores per age group. In all plots, the dots represent individual data points at participants' age, and the whiskers indicate the lowest and highest data points within ±1.5 times the interquartile range.

https://doi.org/10.1121/10.0004791

TABLE II. Dunnett's test analysis results for the differences between adults' and children's mean benefit from target-masker differences in F0 and VTL and their accuracy scores per age group, overall, and per TMR condition.

	Age group	Overall	TMR −6 dB	TMR 0 dB	TMR +6 dB
F0 benefit	4–6 years	0.07, $p < 0.01$**	0.19, $p < 0.001$***	0.13, $p < 0.001$***	−0.05, $p < 0.001$*** $p < 0.001$***
	6–8 years	0.05, $p < 0.001$***	0.14, $p < 0.001$***	0.06, $p < 0.001$***	−0.01, $p = 0.07$
	8–10 years	0.05, $p < 0.001$***	0.12, $p < 0.001$***	0.05, $p < 0.001$***	−0.02, $p < 0.05$*
	10–12 years	0.02, $p = 0.35$	0.06, $p < 0.05$*	0.03, $p < 0.05$*	−0.002, $p = 0.97$
VTL benefit	4–6 years	0.33, $p < 0.001$***	0.46, $p < 0.001$***	0.42, $p < 0.001$***	−0.10, $p < 0.001$***
	6–8 years	0.22, $p < 0.001$***	0.35, $p < 0.001$***	0.20, $p < 0.001$***	−0.03, $p = 0.06$
	8–10 years	0.17, $p < 0.001$***	0.24, $p < 0.001$***	0.16, $p < 0.001$***	−0.03, $p < 0.05$*
	10–12 years	0.09, $p < 0.001$***	0.16, $p < 0.001$***	0.09, $p = 0.06$	−0.004, $p = 0.98$
Overall logit *accuracy scores*	4–6 years	−2.93, $p < 0.001$***	−3.18, $p < 0.001$***	−2.70, $p < 0.001$***	−2.16, $p < 0.001$***
	6–8 years	−2.03, $p < 0.001$***	−2.45, $p < 0.001$***	−1.39, $p < 0.001$***	−0.58, $p = 0.07$
	8–10 years	−1.70, $p < 0.001$***	−1.85, $p < 0.001$***	−1.12, $p < 0.001$***	−0.69, $p < 0.05$*
	10–12 years	−0.99, $p < 0.001$***	−1.02, $p < 0.001$***	−0.72, $p < 0.05$*	−0.09, $p = 0.98$

10 years of age overall benefited more from F0 differences than adults, but the benefit from F0 differences derived from children between 10 and 12 years of age did not differ from that of adults. When we examined children's benefit from F0 differences in the different TMR conditions, we found that all children benefited more from F0 differences than adults in the −6 and 0 dB conditions. In the +6 dB TMR condition, 4–6-year-old and 8–10-year-old children differed from adults by, in fact, showing a small but significant detrimental effect of F0 differences on their accuracy scores. For VTL, children benefited more from VTL differences than adults at all tested ages. Children also showed a greater benefit from VTL differences than adults in the −6 dB TMR condition at all tested ages, and only 10–12-year-old children did not show a larger benefit than adults in the 0 dB TMR condition. Finally, similar to F0 differences, 4–6-year-old children and 8–10-year-old children showed a small but significant detrimental effect of VTL differences on their accuracy scores. In summary, children's overall benefit from F0 was significantly larger than adults' between 4 and 10 years of age but no longer differed from adults' after 10 years of age. The overall benefit that children had of VTL differences was significantly larger than that observed in adults for children at all tested ages. All children showed a larger benefit from F0 and VTL differences than adults in the −6 dB condition, 10–12-year-old children no longer showed a larger benefit from VTL differences in the 0 dB condition, and only 4–6-year-old and 8–10-year-old children differed from adults by showing a small but significant detrimental effect of F0 and VTL differences in the +6 dB condition.

For participants' overall logit-transformed accuracy scores, the Dunnett's test indicated that children's total number of correct responses was lower than adults' at all tested ages, even for the oldest children tested. In the −6 and 0 dB TMR conditions, all children's accuracy scores were lower than adults'. However, in the +6 dB TMR condition, 6–8-year-old and 10–12-year-old children's accuracy scores did not differ from those of adults. The accuracy scores of 4–6-year-old and 8–10-year-old children differed from adults' at all TMRs. To summarize, children's overall accuracy scores and their accuracy scores in the −6 and 0 dB conditions were generally lower compared to adults'. In the +6 dB condition, only 4–6-year-old and 8–10-year-old children had lower accuracy scores than adults.

C. Adult-like benefit from F0 and VTL differences for interpolated accuracy scores

Figure 5(A) shows participants' mean accuracy scores interpolated across TMRs to 85.7% correct in the same-voice condition, equal to the mean accuracy score of adults in the −6 dB TMR condition, and Fig. 5(B) shows the TMRs of individual participants, which were used for the interpolation. Note that 12 participants had scores that were either all below or above 85.7% correct, which explains why not all data points are centered to 85.7% correct in the upper left panel of Fig. 5(A). We fitted the model on the data of all participants, children and adults combined and used log-transformed age values to reduce the effects of the non-continuous distribution in age. The backward stepwise model comparison analysis showed that the model with *age* as a fixed factor and a two-way interaction between *δF0* and *δVTL* was the best fitting model, in lme4 syntax: `interpolated logit score ~ log(age) + δF0*δVTL + (1|participant)`.

There was a significant effect of δF0 ($t = 3.45$, estimate = 0.19, SE = 0.057, $p < 0.001$) and a significant two-way interaction between δF0 and δVTL ($t = -4.57$, estimate = −0.53, SE = 0.116, $p < 0.001$) that demonstrate participants' accuracy scores improved due to target-masker differences in F0 and a combined difference in F0 and VTL. We did not find a significant main effect of δVTL, although it was near-significant ($t = 1.80$, estimate = 0.10, SE = 0.058, $p = 0.07$). Also, we only found a significant main effect of age on participants' accuracy scores ($t = 4.92$, estimate = 0.37, SE = 0.075, $p < 0.001$). This lack of an interaction between age with δF0 and δVTL indicates that age did not affect the benefit participants derived from target-masker differences in F0 and VTL once their overall accuracy scores were interpolated to the same performance level in the same-voice condition.

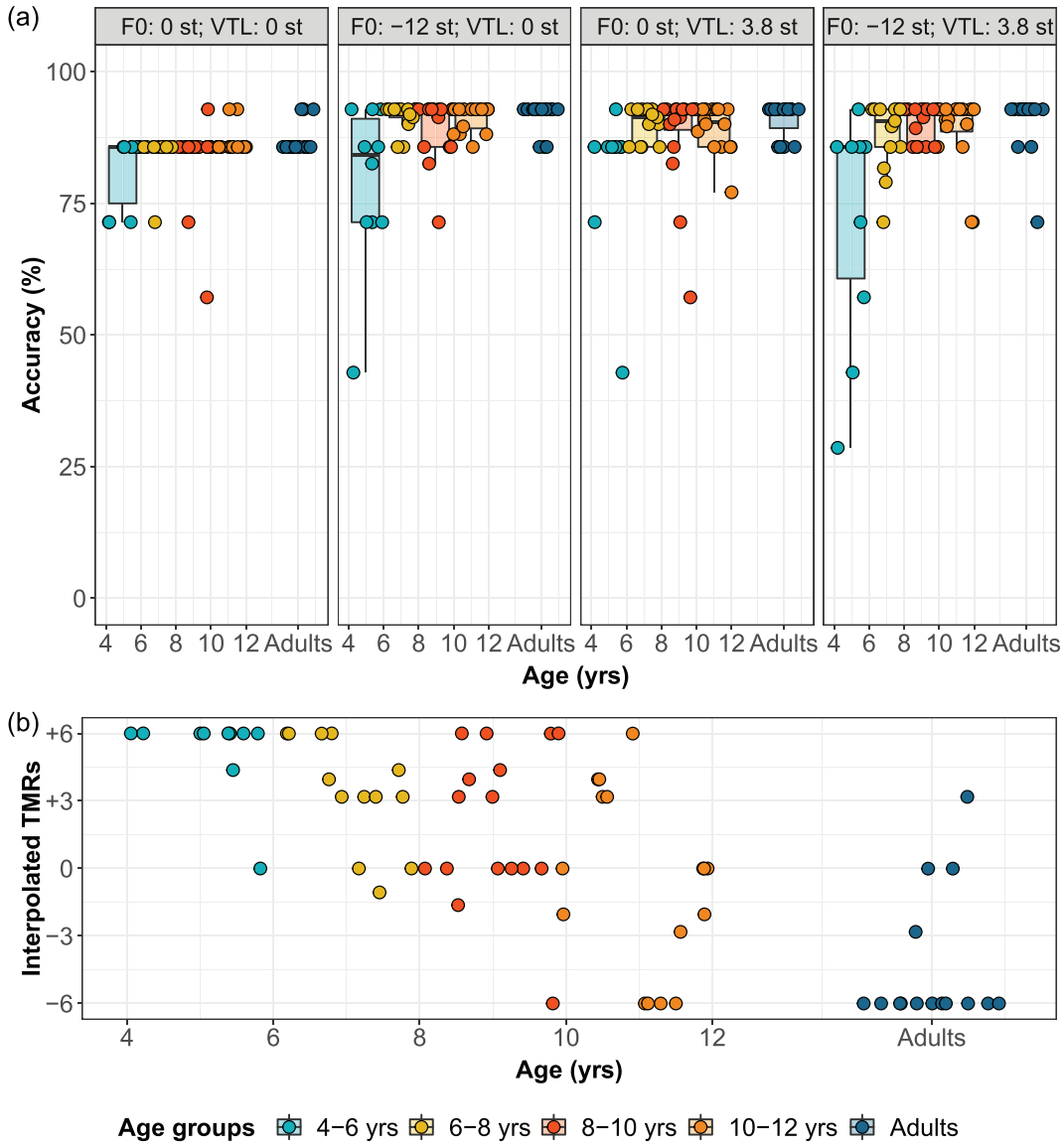

FIG. 5. (Color online) Mean accuracy scores of participants per age group and voice condition interpolated across TMRs ($N_{children} = 55$, $N_{adults} = 15$). (A) Participants' mean accuracy scores per age group and voice condition interpolated to 85.7% correct in the same-voice condition. The four panels show the accuracy score for each masker speech voice condition. (B) Participants' TMRs at 85.7% correct in the same-voice condition, which were used for the interpolation. The boxplots show the median accuracy scores of participants per age group and the lower and upper quartiles. The dots represent individual data points at participants' age, and the whiskers indicate the lowest and highest data points within ±1.5 times the interquartile range.

Furthermore, to evaluate at what age children benefited similarly from F0 and VTL differences as adults, we performed a Dunnett's test on participants' benefit, i.e., the difference between their interpolated logit-transformed accuracy scores in the masker speech voice condition with −12 st in F0 and 3.8 st in VTL minus the condition with no differences in F0 and VTL. These results indicated that the benefit only differed from adults for 4–6-year-old children [4–6 years: difference (diff) = −0.40, $p < 0.01$; 6–8 years: diff = 0.07, $p = 0.94$; 8–10 years: diff = 0.08, $p = 0.87$; 10–12 years: diff = −0.05, $p = 0.98$]. Intriguingly, as can be seen in Fig. 5(A), some 4–6-year-old children showed a detrimental effect of target-masker differences in F0. However, based on Fig. 2, this only seemed to be the case in the +6 dB TMR condition.

D. Correlations of F0 and VTL benefit and overall accuracy scores with JNDs

Figure 6(A) shows the correlations between the residuals of children's benefit from target-masker differences in F0 and VTL and the residuals for their corresponding JNDs. The correlation analysis between the residuals of children's F0-difference benefit (Bk/st) and their F0 JNDs, i.e., after partialling out the effect of age, indicated there was no significant correlation between the two measures (Pearson's $r = 0.14$, $p = 0.32$). For the residuals of participants' VTL-difference benefit and their VTL JNDs, we also did not find a significant correlation between the two measures (Pearson's $r = 0.19$, $p = 0.15$). Thus, children's benefit from target-masker differences in F0 and VTL was not directly related to their respective discrimination thresholds.

https://doi.org/10.1121/10.0004791

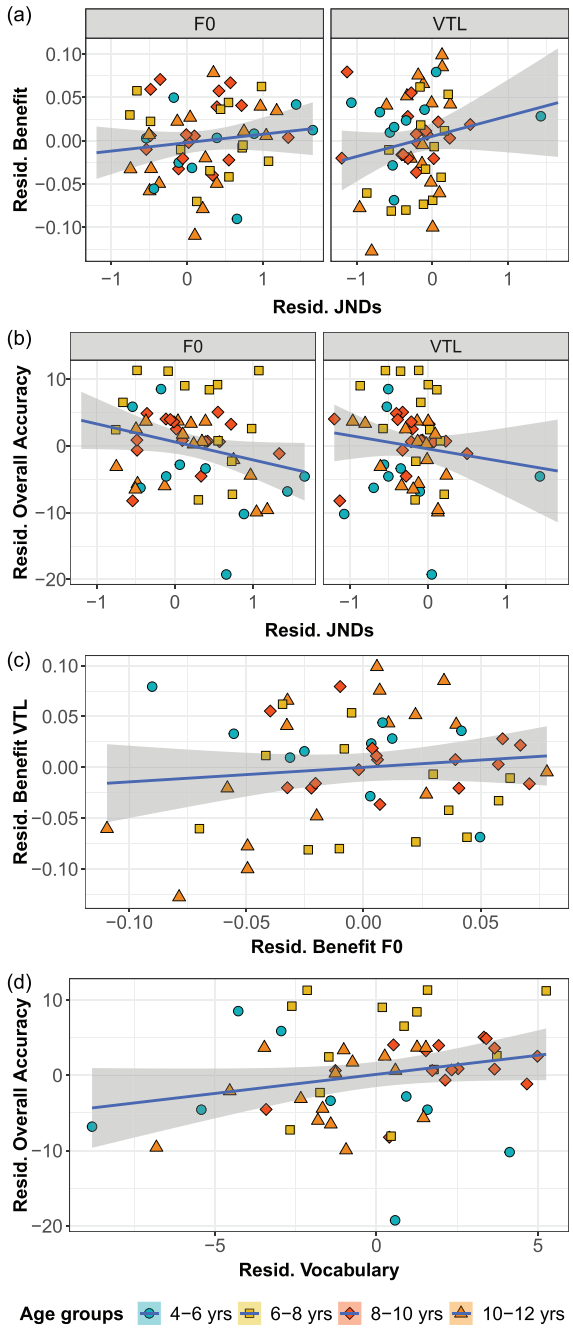

FIG. 6. (Color online) The correlations between the residuals of children's benefit from target-masker differences in F0 and VTL and their corresponding JNDs and the correlations between the residuals of children's overall accuracy scores and their F0 JNDs, VTL JNDs, and vocabulary scores ($N_{children} = 55$). (A) The correlations between the residuals of children's benefit from target-masker differences in F0 (left panel) and VTL (right panel) in Bk/st and their corresponding JNDs. (B) The correlations between the residuals of children's overall accuracy scores and their F0 (left panel) and VTL JNDs (right panel). (C) The correlation between the residuals of children's benefit from target-masker differences in F0 (left panel) and VTL (right panel) in Bk/st. (D) The correlation between the residuals of children's overall accuracy scores and the raw scores of their vocabulary size based on the Renfrew Word Finding Vocabulary Test. In all plots, the central line represents a linear regression line that indicates the relationship between both measures, and the surrounding area shows the 95% confidence intervals. The shape of the data points indicates the age group of participants.

Figure 6(B) shows the correlations between children's overall accuracy score residuals and their F0 and VTL JND residuals. The correlation analysis between the residuals of children's overall accuracy scores and their F0 JNDs indicated there was no significant correlation between the two measures (Pearson's $r = -0.26$, $p = 0.06$). There was also no significant correlation between the residuals of children's overall accuracy scores and their VTL JNDs (Pearson's $r = -0.14$, $p = 0.32$). Thus, children's overall accuracy scores were also not directly related to their F0 and VTL discrimination thresholds. In addition, Fig. 6(C) shows the correlation between children's benefit from target-masker differences in F0 and VTL. These two measures were also not significantly correlated (Pearson's $r = 0.05$, $p = 0.73$). Finally, Fig. 6(D) shows the correlation between the residuals of children's overall accuracy scores and the raw scores of their vocabulary size as measured by the Renfrew Word Finding Vocabulary Test. The correlation analysis indicated there was no significant correlation between the two measures (Pearson's $r = 0.24$, $p = 0.08$).

IV. DISCUSSION

In the current study, we investigated how children's benefit from target-masker differences in speakers' mean F0 and VTL for the perception of speech in competing speech develops during the school-age years (4–12 years of age). Our results show that the accuracy scores of children improved as a function of age and TMR as expected based on previous research (Bonino *et al.*, 2013; Buss *et al.*, 2017b; Buss *et al.*, 2019; Corbin *et al.*, 2016; Flaherty *et al.*, 2019, 2021; Hall *et al.*, 2002; Leibold *et al.*, 2018; Leibold and Buss, 2013; Wightman and Kistler, 2005). Children of all age groups showed a benefit from target-masker differences in either or both F0 and VTL. Also, the benefit from target-masker differences in F0 and VTL decreased as the TMR became more advantageous, likely due to the increase in children's overall accuracy scores. The size of the benefit from target-masker differences in F0 and VTL mainly seemed to be larger in children than in adults because of their overall lower accuracy scores, which were not adult-like at all tested ages and hence left more room for improvement. This explanation was confirmed by the lack of an effect of age on participants' benefit when their mean accuracy scores were interpolated across TMRs to the same performance level. Finally, we examined if the benefit from target-masker differences in F0 and VTL for children was related to their respective discrimination thresholds. Our results indicate that there were no significant correlations between children's benefit from F0 and VTL differences and their respective JNDs or between their benefit from F0 differences and their benefit from VTL differences. Moreover, we did not find any correlations for children's overall accuracy scores with their F0 JNDs, VTL JNDs, or vocabulary scores.

Our findings indicate that children of all age groups showed a benefit from target-masker differences in either or

both F0 and VTL for perceiving speech in competing speech. The benefit from voice differences was particularly large for young children at the −6 dB TMR, the highest masker speech level tested. While the performance of 4–6-year-old children was 7.14% on average in the same-voice condition, their performance increased substantially as a result of introducing target-masker differences in F0 (average performance of 47.1%), VTL (average performance of 38.6%), or both F0 and VTL (average performance of 64.3%). This considerable improvement in performance as a result of voice difference benefit is in agreement with previous research indicating that most children are able to distinguish different speakers when there are large differences between their voices (Cleary *et al.*, 2005; Creel and Jimenez, 2012). Children benefited from F0 and VTL differences at all tested ages, although their benefit only became as small as that observed in adults around 10 years of age. However, the observed differences in the benefit from F0 and VTL differences between children and adults mainly seemed to be caused by the differences in their overall accuracy scores. This explanation is also in line with our findings that children's overall accuracy scores did not reach an adult-like level for all tested ages. When we interpolated participants' accuracy scores across TMRs to a performance of 85.7% correct in the same-voice condition, we did not observe any significant differences across age. However, the results from the Dunnett's test indicated that 4–6-year-old children benefited less than adults, but this seems to be caused by a slightly detrimental effect that was observed in a small number of young children.

Counterintuitively, for three of ten 4–6-year-old children, there seemed to be a detrimental effect of target-masker differences in F0, although this was only observed in the +6 TMR condition. The underlying cause of this detrimental effect could be the specific combination of the lower presentation level of the masker compared to the target (6 dB lower) and the manipulation of the masker speech voice. While the favorable TMR should have given an advantage in selecting the target speech stream and inhibiting the masker speech stream, which is the case for older children and adults, it seems not to have worked in this manner for a small number of young children within our participant group. Some young children may not have mastered the effective use of selective attention mechanisms yet, which would allow them to focus on the target and ignore the masker like other children and adults. Given that the voice of the target speaker remained the same throughout the experiment, the exposure to these voice parameters was higher than the manipulated voice parameters of maskers, which may have caused a so-called novelty effect due to the change in voice gender (Darwin *et al.*, 2003). This effect could make the masker speech stream more distracting for children rather than facilitating the segregation and selection of the target speech stream. Supporting this idea, there are some studies on the mechanisms of selective attention in the processing of visual cues in children that suggest that the interference of distractors becomes larger when the perceptual load of the task becomes lower (Huang-Pollock *et al.*, 2002; Lavie, 2005). Tasks with a high perceptual load engage the full capacity of attentional resources for the processing of task-relevant stimuli, while tasks with a low perceptual load give more opportunity for the processing of also task-irrelevant stimuli. On the other hand, this detrimental effect of voice differences was mainly observed in a very small number of participants, three out of ten within the youngest group of 4–6-year-old children; hence, it could also have been a simple consequence of brief periods of inattention (Lutfi *et al.*, 2003; Sussman and Steinschneider, 2009). Wightman and Kistler (2005) also addressed some of the difficulties involved in estimating psychometric function data from young children, such as higher individual variability in their performance related to the rapid development that takes place at this age and the upper asymptote not reaching 100% correct due to general attention span.

Our findings partially contradict the earlier results from Flaherty *et al.* (2019, 2021), which showed that young children do not benefit from target-masker differences in F0 or VTL and benefit less than older children and adults from combined differences in F0 and VTL. The F0 differences tested in their studies, namely −3, −6, and −9 st, and VTL differences using scaling factors of 0.84 and 1.16 were smaller than the −12 st difference in F0 and 3.8 st difference in VTL that we have used. However, differences in speakers' mean F0 of −6 and −9 st seem to be well within the range of most school-age children's discrimination thresholds (Buss *et al.*, 2017a; Flaherty *et al.*, 2019; Nagels *et al.*, 2020a). The mean F0 discrimination threshold was 6.18 st for the youngest age group of 4–6-year-old children in the current study, as was previously reported by Nagels *et al.* (2020a). Also, participants' JNDs correspond to the 70.7% correct discrimination point on the psychometric function, so voice differences below their JND could still be expected to be somewhat audible. In addition, similarly to Flaherty *et al.* (2019), we did not find any significant correlation between children's discrimination thresholds and their benefit of target-masker differences in F0 or VTL or with their overall accuracy scores. The lack of a significant correlation also suggests that children's reduced sensitivity to differences in voice cues does not seem to be the primary factor explaining their poorer perception of speech in competing speech compared to adults. Instead, children's ability to segregate different speech streams seems to rely on different voice perception mechanisms than those involved in voice discrimination, such as selective auditory attention and inhibition. In agreement with this explanation, Sussman *et al.* (2007) and Sussman and Steinschneider (2009) reported a similar discrepancy between children's frequency discrimination abilities and their ability to use frequency differences to segregate two streams of pure tones. Nevertheless, it should be kept in mind that the target-masker voice differences of −12 st in F0 and 3.8 st in VTL in the current study were well above most children's discrimination thresholds. A closer relationship between discrimination thresholds and a benefit from target-masker voice differences might have

https://doi.org/10.1121/10.0004791

been present if more subtle voice differences nearer to children's discrimination thresholds had been used.

Another explanation for the discrepancy between our results and those of Flaherty *et al.* (2019, 2021) could be the use of a carrier phrase to mark the target sentences in the current study. Using the same consistent carrier phrase most likely helped children to segregate the different speech streams and use the voice differences more effectively for keeping track of the target speech stream (Bonino *et al.*, 2013; Freyman *et al.*, 2004; Sussman-Fort and Sussman, 2014). Bonino *et al.* (2013) reported a large improvement of 16.8% in the accuracy scores for children's perception of speech in competing speech when a carrier phrase was used. An idea for future research would be to make the task more difficult for adults, for example, by modulating the onset asynchrony between the target and masker sentences (Lee and Humes, 2012) and hence minimizing the chance for ceiling performance. Finally, our masker speech consisted of sentence chunks instead of full sentences and a single-talker speech masker instead of a two-talker speech masker. These design differences could have led to less effective overall masking in the current study (Buss *et al.*, 2017b; Litovsky, 2005; Rosen *et al.*, 2013). For instance, the single-talker masker may have provided more opportunity for "glimpsing" acoustic information due to temporary reductions in masking (Rosen *et al.*, 2013), although findings by Buss *et al.* (2017b) suggest that young children cannot use these low-level glimpses as efficiently as adults. Further research should examine the individual contributions of these design parameters to children's benefit from voice differences for perceiving speech in competing speech.

Furthermore, we did not find any significant correlation between children's overall ability to perceive speech in competing speech and their vocabulary scores in the current study. However, we chose to use simple sentences with the same carrier phrase and sentence structure as target sentences and words for basic colors and numbers up to 10 as target words in these sentences to ensure that children could perform the task at all tested ages. Such words are acquired very early in life, around 3 or 4 years of age (Brysbaert *et al.*, 2014). If there are effects of vocabulary size on the ability to perceive speech in competing speech, the use of simple closed-set sentence materials with very familiar words likely limited these effects in the current study, as such effects have been reported when more complex open-set sentence stimuli were used (Klein *et al.*, 2017; McCreery *et al.*, 2017, 2020). Another reason for choosing these simple sentence materials was that we are planning to follow up on these findings by testing children with cochlear implants, who often have smaller vocabulary sizes than their age-matched peers (Fagan and Pisoni, 2010; Lund, 2016), using the same testing materials and procedure.

Finally, children may have more difficulties with perceiving speech in competing speech due to their limited acoustic and language experience compared to adults, as was suggested earlier by Corbin *et al.* (2016). Children may need more time and experience to develop representations of speech that are robust enough to withstand the perceptual obliteration caused by masking. This idea is also supported by the fact that children seem to require a greater spectro-temporal resolution for speech recognition in degraded speech conditions (Eisenberg *et al.*, 2000; Mlot *et al.*, 2010) and use word probability and sentential context less effectively than older children and adults (Buss *et al.*, 2019; Craig *et al.*, 1993; Elliott *et al.*, 1987; Metsala, 1997). On the other hand, a reduced ability to perceive speech in competing speech may also negatively affect children's language development, as it limits the amount of language input they receive. This matter of causality has primarily been addressed with respect to the language development of children with hearing impairments, such as for children with cochlear implants (Fagan and Pisoni, 2010; Geers *et al.*, 2017), but has received little attention for children with normal hearing. Having a lower ability to perceive speech in competing speech may also negatively affect the language development of children with normal hearing.

In conclusion, children seem to benefit from target-masker differences in speakers' F0 or VTL at all tested ages in the current study. Their benefit from differences in F0 and VTL was larger than that observed in adults until the age of 10, but when we interpolated participants' accuracy scores to the same performance level, the benefit was proportionally the same across all tested ages. Also, children's overall accuracy scores did not become adult-like for all tested ages, which indicates that the ability to perceive speech in competing speech continues to develop even for the oldest tested age group of 10–12-year-old children. Moreover, we did not find any correlations between children's benefit from target-masker differences in F0 and VTL and their corresponding voice cue JNDs or between children's overall accuracy scores and F0 JNDs, VTL JNDs, or vocabulary scores. These findings suggest that children's ability to perceive speech in competing speech is not directly related to their voice discrimination abilities. Instead, other factors such as those related to selective auditory attention and inhibition seem to be more closely associated with the developing speech in competing speech perception abilities of school-age children.

ACKNOWLEDGMENTS

We are grateful to all children, parents, and students who participated, and particularly Basisschool De Brink in Ottersum, Basisschool de Petteflet, and BSO Huis de B in Groningen, which helped us recruit our child participants. We would also like to thank Dr. Paolo Toffanin, Iris van Bommel, Evelien Birza, Jacqueline Libert, Jemima Phillpot, Marta Matos Lopes, and Jop Luberti (illustrations) for their contribution to the development of the game interfaces and Dr. Stuart Rosen for providing the English stimuli used in the study by Hazan *et al.* (2009). This work was funded by the Center for Language Cognition Groningen (CLCG), a VICI Grant from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for

Health Research and Development (ZonMw) (Grant No. 918-17-603), the Medical Research Council (Senior Fellowship Grant No. S002537/1), a National Institute of Health Research Programme Grant (Grant No. NIHR201608), and the Heinsius Houbolt Foundation. This work was conducted in the framework of the LabEx CeLyA ("Centre Lyonnais d'Acoustique," ANR-10-LABX-0060/ANR-11-IDEX-0007) operated by the French National Research Agency and is also part of the research program of the University Medical Center Groningen (UMCG) Otorhinolaryngology Department: Healthy Aging and Communication. The raw data presented here can be accessed online at https://doi.org/10.34894/AKU9FR. The CCRM-NL corpus used in this study can be downloaded at https://doi.org/10.5281/zenodo.4700993.

Arbogast, T. L., Mason, C. R., and Kidd, G. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Başkent, D., Clarke, J., Pals, C., Benard, M. R., Bhargava, P., Saija, J., Sarampalis, A., Wagner, A., and Gaudrain, E. (**2016**). "Cognitive compensation of speech perception with hearing impairment, cochlear implants, and aging: How and to what degree can it be achieved?," Trends Hear. **20**, 1–16.

Başkent, D., and Gaudrain, E. (**2016**). "Musician advantage for speech-on-speech perception," J. Acoust. Soc. Am. **139**, EL51–EL56.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**, 1–51.

Bird, J., and Darwin, C. (**1998**). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr Ltd., London), pp. 263–269.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Bonino, A. Y., Leibold, L. J., and Buss, E. (**2013**). "Release from perceptual masking for children and adults: Benefit of a carrier phrase," Ear Hear. **34**, 3–14.

Bregman, A. S. (**1994**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Broadbent, D. E. (**1952**). "Listening to one of two synchronous messages," J. Exp. Psychol. **44**, 51–55.

Bronkhorst, A. W. (**2015**). "The cocktail-party problem revisited: Early processing and selection of multi-talker speech," Atten. Percept. Psychophys. **77**, 1465–1487.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

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**, 4007–4018.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., and Storms, G. (**2014**). "Norms of age of acquisition and concreteness for 30,000 Dutch words," Acta Psychol. **150**, 80–84.

Buss, E., Flaherty, M. M., and Leibold, L. J. (**2017a**). "Development of frequency discrimination at 250 Hz is similar for tone and /ba/ stimuli," J. Acoust. Soc. Am. **142**, EL150–EL154.

Buss, E., Hodge, S. E., Calandruccio, L., Leibold, L. J., and Grose, J. H. (**2019**). "Masked sentence recognition in children, young adults, and older adults: Age-dependent effects of semantic context and masker type," Ear Hear. **40**, 1117–1126.

Buss, E., Leibold, L. J., Porter, H. L., and Grose, J. H. (**2017b**). "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**, 2650–2660.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.

Cleary, M., Pisoni, D. B., and Kirk, K. I. (**2005**). "Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants," J. Speech Lang. Hear. Res. **48**, 204–223.

Cooke, M., Garcia Lecumberri, M. L., and Barker, J. (**2008**). "The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception," J. Acoust. Soc. Am. **123**, 414–427.

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**, 55–63.

Craig, C. H., Kim, B. W., Pecyna Rhyner, P. M., and Bowen Chirillo, T. K. (**1993**). "Effects of word predictability, child development, and aging on time-gated speech recognition performance," J. Speech Lang. Hear. Res. **36**, 832–841.

Creel, S. C., and Jimenez, S. R. (**2012**). "Differences in talker recognition by preschoolers and adults," J. Exp. Child Psychol. **113**, 487–509.

Darwin, C. J., Brungart, D. S., and Simpson, B. D. (**2003**). "Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," J. Acoust. Soc. Am. **114**, 2913–2922.

Eisenberg, L. S., Shannon, R. V., Schaefer Martinez, A., Wygonski, J., and Boothroyd, A. (**2000**). "Speech recognition with reduced spectral cues as a function of age," J. Acoust. Soc. Am. **107**, 2704–2710.

El Boghdady, N., Gaudrain, E., and Başkent, D. (**2019**). "Does good perception of vocal characteristics relate to better speech-on-speech intelligibility for cochlear implant users?," J. Acoust. Soc. Am. **145**, 417–439.

Elliott, L. L., Hammer, M. A., and Evan, K. E. (**1987**). "Perception of gated, highly familiar spoken monosyllabic nouns by children, teenagers, and older adults," Percept. Psychophys. **42**, 150–157.

Evans, S., McGettigan, C., Agnew, Z. K., Rosen, S., and Scott, S. K. (**2016**). "Getting the cocktail party started: Masking effects in speech perception," J. Cogn. Neurosci. **28**, 483–500.

Fagan, M. K., and Pisoni, D. B. (**2010**). "Hearing experience and receptive vocabulary development in deaf children with cochlear implants," J. Deaf Stud. Deaf Educ. **15**, 149–161.

Fecher, N., Paquette-Smith, M., and Johnson, E. K. (**2019**). "Resolving the (apparent) talker recognition paradox in developmental speech perception," Infancy **24**, 570–588.

Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.

Flaherty, M. M., Buss, E., and Leibold, L. J. (**2019**). "Developmental effects in children's ability to benefit from F0 differences between target and masker speech," Ear Hear. **40**, 927–937.

Flaherty, M. M., Buss, E., and Leibold, L. J. (**2021**). "Independent and combined effects of fundamental frequency and vocal tract length differences for school-age children's sentence recognition in a two-talker masker," J. Speech Lang. Hear. Res. **64**, 206–217.

Floccia, C., Butler, J., Girard, F., and Goslin, J. (**2009**). "Categorization of regional and foreign accent in 5- to 7-year-old British children," Int. J. Behav. Dev. **33**, 366–375.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.

Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q.-J., Free, R. H., and Başkent, D. (**2014**). "Gender categorization is abnormal in cochlear implant users," J. Assoc. Res. Otolaryngol. **15**, 1037–1048.

Gaudrain, E., and Başkent, D. (**2018**). "Discrimination of voice pitch and vocal-tract length in cochlear implant users," Ear Hear. **39**, 226–237.

Geers, A. E., Mitchell, C. M., Warner-Czyz, A., Wang, N.-Y., Eisenberg, L. S., and CDaCI Investigative Team (**2017**). "Early sign language exposure and cochlear implantation benefits," Pediatrics **140**(1), e20163489.

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**, 159–165.

Hazan, V., and Barrett, S. (**2000**). "The development of phonemic categorization in children aged 6–12," J. Phon. **28**, 377–396.

Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., and Shakespeare, B. (**2009**). "Speech perception abilities of adults with dyslexia: Is there

any evidence for a true deficit?," J. Speech Lang. Hear. Res. **52**, 1510–1529.
Helfer, K. S., and Freyman, R. L. (**2009**). "Lexical and indexical cues in masking by competing speech," J. Acoust. Soc. Am. **125**, 447–456.
Hilkhuysen, G., Gaubitch, N., Brookes, M., and Huckvale, M. (**2012**). "Effects of noise suppression on intelligibility: Dependency on signal-to-noise ratios," J. Acoust. Soc. Am. **131**, 531–539.
Huang-Pollock, C. L., Carr, T. H., and Nigg, J. T. (**2002**). "Development of selective attention: Perceptual load influences early versus late attentional selection in children and adults," Dev. Psychol. **38**, 363–375.
Jensen, J. K., and Neff, D. L. (**1993**). "Development of basic auditory discrimination in preschool children," Psychol. Sci. **4**, 104–107.
Kawahara, H., and Irino, T. (**2005**). "Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation," in *Speech Separation by Humans and Machines* (Springer, Boston), pp. 167–180.
Kidd, G., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (**2005**). "The advantage of knowing where to listen," J. Acoust. Soc. Am. **118**, 3804–3815.
Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (**2008**). "Informational masking," in *Audit. Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, Boston), pp. 143–189.
Klein, K. E., Walker, E. A., Kirby, B., and McCreery, R. W. (**2017**). "Vocabulary facilitates speech perception in children with hearing aids," J. Speech Lang. Hear. Res. **60**, 2281–2296.
Kreiman, J., and Sidtis, D. (**2011**). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception* (Wiley, New York).
Lavie, N. (**2005**). "Distracted and confused?: Selective attention under load," Trends Cogn. Sci. **9**, 75–82.
Lee, J. H., and Humes, L. E. (**2012**). "Effect of fundamental-frequency and sentence-onset differences on speech-identification performance of young and older adults in a competing-talker background," J. Acoust. Soc. Am. **132**, 1700–1717.
Leibold, L. J., and Buss, E. (**2013**). "Children's identification of consonants in a speech-shaped noise or a two-talker masker," J. Speech Lang. Hear. Res. **56**, 1144–1155.
Leibold, L. J., Buss, E., and Calandruccio, L. (**2018**). "Developmental effects in masking release for speech-in-speech perception due to a target/masker sex mismatch," Ear Hear. **39**, 935–945.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Litovsky, R. Y. (**2005**). "Speech intelligibility and spatial release from masking in young children," J. Acoust. Soc. Am. **117**, 3091–3099.
Lund, E. (**2016**). "Vocabulary knowledge of children with cochlear implants: A meta-analysis," J. Deaf Stud. Deaf Educ. **21**, 107–121.
Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (**2003**). "One factor underlies individual differences in auditory informational masking within and across age groups," Percept. Psychophys. **65**, 396–406.
MacPherson, A., and Akeroyd, M. A. (**2014**). "Variations in the slope of the psychometric functions for speech intelligibility: A systematic survey," Trends Hear. **18**, 1–26.
Mann, V. A., Diamond, R., and Carey, S. (**1979**). "Development of voice recognition: Parallels with face recognition," J. Exp. Child Psychol. **27**, 153–165.
Marchman, V. A., and Fernald, A. (**2008**). "Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood," Dev. Sci. **11**, F9–F16.
MathWorks Inc. (**2012**). "MATLAB: The language of technical computing," Desktop Tools and Development Environment, Version 2012a (MathWorks, Natick, MA).
Mattys, S. L., Brooks, J., and Cooke, M. (**2009**). "Recognizing speech under a processing load: Dissociating energetic from informational factors," Cognit. Psychol. **59**, 203–243.
Maxon, A. B., and Hochberg, I. (**1982**). "Development of psychoacoustic behavior: Sensitivity and discrimination," Ear Hear. **3**, 301–308.
McCreery, R. W., Miller, M. K., Buss, E., and Leibold, L. J. (**2020**). "Cognitive and linguistic contributions to masked speech recognition in children," J. Speech Lang. Hear. Res. **63**, 3525–3538.
McCreery, R. W., Spratford, M., Kirby, B., and Brennan, M. (**2017**). "Individual differences in language and working memory affect children's speech recognition in noise," Int. J. Audiol. **56**, 306–315.
McCreery, R. W., Walker, E. A., Spratford, M., Lewis, D., and Brennan, M. (**2019**). "Auditory, cognitive, and linguistic factors predict speech recognition in adverse listening conditions for children with hearing loss," Front. Neurosci. **13**, 1–11.
Metsala, J. L. (**1997**). "An examination of word frequency and neighborhood density in the development of spoken-word recognition," Mem. Cognit. **25**, 47–56.
Mlot, S., Buss, E., and Hall, J. W. III (**2010**). "Spectral integration and bandwidth effects on speech recognition in school-aged children and adults," Ear Hear. **31**, 56–62.
Moore, T. J. (**1981**). "Voice communications jamming research," in *AGARD Conference Proceedings 331 Aural Communication in Aviation*, February 1–6, Neuilly-Sur-Seine, France.
Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., and Başkent, D. (**2020a**). "Development of voice perception is dissociated across gender cues in school-age children," Sci. Rep. **10**, 1–11.
Nagels, L., Gaudrain, E., Vickers, D., Matos Lopes, M., Hendriks, P., and Başkent, D. (**2020b**). "Development of vocal emotion recognition in school-age children: The EmoHI test for hearing-impaired populations," PeerJ **8**, e8773-14.
Newman, R. S. (**2004**). "Perceptual restoration in children versus adults," Appl. Psycholinguist. **25**, 481–493.
Newman, R. S., and Morini, G. (**2017**). "Effect of the relationship between target and masker sex on infants' recognition of speech," J. Acoust. Soc. Am. **141**, EL164–EL169.
Nittrouer, S., and Boothroyd, A. (**1990**). "Context effects in phoneme and word recognition by young children and older adults," J. Acoust. Soc. Am. **87**, 2705–2715.
Nittrouer, S., and Miller, M. E. (**1997**). "Predicting developmental shifts in perceptual weighting schemes," J. Acoust. Soc. Am. **101**, 2253–2266.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Pollack, I. (**1975**). "Auditory informational masking," J. Acoust. Soc. Am. **57**, S5.
R Core Team (**2020**). R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria).
Renfrew, C. E. (**1995**). *Word Finding Vocabulary Test* (Speechmark Publishing, Bicester, UK).
Rosen, S., Souza, P., Ekelund, C., and Majeed, A. A. (**2013**). "Listening to speech in a background of other talkers: Effects of talker number and noise vocoding," J. Acoust. Soc. Am. **133**, 2431–2443.
Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. G. (**2011**). "Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication," Proc. Natl. Acad. Sci. U.S.A. **108**, 15516–15521.
Saleh, S. M., Saeed, S. R., Meerton, L., Moore, D. R., and Vickers, D. A. (**2013**). "Clinical use of electrode differentiation to enhance programming of cochlear implants," Cochlear Implants Int. **14**, 16–18.
Saleh, S. M. I. (**2013**). "The efficacy of fitting cochlear implants based on pitch perception," Ph.D. thesis, University College London, UK.
Samuel, A. G. (**1996**). "Does lexical information influence the perceptual restoration of phonemes?," J. Exp. Psychol. Gen. **125**, 28–51.
Schneider, B. A., Li, L., and Daneman, M. (**2007**). "How competing speech interferes with speech comprehension in everyday listening situations," J. Am. Acad. Audiol. **18**, 559–572.
Scott, S. K., Rosen, S., Wickham, L., and Wise, R. J. S. (**2004**). "A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception," J. Acoust. Soc. Am. **115**, 813–821.
Semeraro, H. D., Rowan, D., Besouw, R. M. v., and Allsopp, A. A. (**2017**). "Development and evaluation of the British English coordinate response measure speech-in-noise test as an occupational hearing assessment tool," Int. J. Audiol. **56**, 749–758.
Signorell, A., Aho, K., Alfons, A., Anderegg, N., Aragon, T., Arppe, A., Baddeley, A., Barton, K., Bolker, B., Borchers, H. W., Caeiro, F., Champely, S., Chessel, D., Chhay, L., Cooper, N., Cummins, C., Dewey, M., Doran, H. C., Dray, S., Dupont, C., Eddelbuettel, D., Ekstrom, C., Elff, M., Enos, J., Farebrother, R. W., Fox, J., Francois, R., Friendly, M., Galili, T., Gamer, M., Gastwirth, J. L., Gegzna, V., Gel, Y. R., Graber, S.,

Gross, J., Grothendieck, G., Harrell, F. E., Jr., Heiberger, R., Hoehle, M., Hoffmann, C. W., Hojsgaard, S., Hothorn, T., Huerzeler, M., Hui, W. W., Hurd, P., Hyndman, R. J., Jackson, C., Kohl, M., Korpela, M., Kuhn, M., Labes, D., Leisch, F., Lemon, J., Li, D., Maechler, M., Magnusson, A., Mainwaring, B., Malter, D., Marsaglia, G., Marsaglia, J., Matei, A., Meyer, D., Miao, W., Millo, G., Min, Y., Mitchell, D., Mueller, F., Naepflin, M., Navarro, D., Nilsson, H., Nordhausen, K., Ogle, D., Ooi, H., Parsons, N., Pavoine, S., Plate, T., Prendergast, L., Rapold, R., Revelle, W., Rinker, T., Ripley, B. D., Rodriguez, C., Russell, N., Sabbe, N., Scherer, R., Seshan, V. E., Smithson, M., Snow, G., Soetaert, K., Stahel, W. A., Stephenson, A., Stevenson, M., Stubner, R., Templ, M., Lang, D. T., Therneau, T., Tille, Y., Torgo, L., Trapletti, A., Ulrich, J., Ushey, K., VanDerWal, J., Venables, B., Verzani, J., Villacorta Iglesias, P. J., Warnes, G. R., Wellek, S., Wickham, H., Wilcox, R. R., Wolf, P., Wollschlaeger, D., Wood, J., Wu, Y., Yee, T., and Zeileis, A. (**2018**). "DescTools: Tools for descriptive statistics," R package version 0.99.41, https://cran.r-project.org/package=DescTools (Last viewed 11/1/2020).

Skuk, V. G., and Schweinberger, S. R. (**2014**). "Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender," J. Speech Lang. Hear. Res. **57**, 285–296.

Smith, D. R. R., and Patterson, R. D. (**2005**). "The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age," J. Acoust. Soc. Am. **118**, 3177–3186.

Smith, D. R. R., Walters, T. C., and Patterson, R. D. (**2007**). "Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled," J. Acoust. Soc. Am. **122**, 3628–3639.

Sobon, K. A., Taleb, N. M., Buss, E., Grose, J. H., and Calandruccio, L. (**2019**). "Psychometric function slope for speech-in-noise and speech-in-speech: Effects of development and aging," J. Acoust. Soc. Am. **145**, EL284–EL290.

Sussman, E., and Steinschneider, M. (**2009**). "Attention effects on auditory scene analysis in children," Neuropsychologia **47**, 771–785.

Sussman, E., Wong, R., Horváth, J., Winkler, I., and Wang, W. (**2007**). "The development of the perceptual organization of sound by frequency separation in 5–11-year-old children," Hear. Res. **225**, 117–127.

Sussman-Fort, J., and Sussman, E. (**2014**). "The effect of stimulus context on the buildup to stream segregation," Front. Neurosci. **8**, 1–8.

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, J. G., and Patel, A. D. (**2015**). "Musical training, individual differences and the cocktail party problem," Sci. Rep. **5**, 1–10.

Titze, I. R. (**1989**). "Physiologic and acoustic differences between male and female voices," J. Acoust. Soc. Am. **85**, 1699–1707.

Warren, R. M. (**1970**). "Perceptual restoration of missing speech sounds," Science **167**, 392–393.

Welch, G. F., Saunders, J., Edwards, S., Palmer, Z., Himonides, E., Knight, J., Mahon, M., Griffin, S., and Vickers, D. A. (**2015**). "Using singing to nurture children's hearing? A pilot study," Cochlear Implants Int. **16**, 63–70.

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**, 3164–3176.

Zekveld, A. A., Rudner, M., Kramer, S. E., Lyzenga, J., and Rönnberg, J. (**2014**). "Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech," Front. Neurosci. **8**, 88.

Zobel, B. H., Wagner, A., Sanders, L. D., and Başkent, D. (**2019**). "Spatial release from informational masking declines with age: Evidence from a detection task in a virtual separation paradigm," J. Acoust. Soc. Am. **146**, 548–566.