applied sciences

MDPI

Article

The Effect of Noise on the Utilization of Fundamental Frequency and Formants for Voice Discrimination in Children and Adults

Liat Kishon-Rabin [1] and Yael Zaltz [1,2,*]

[1] Department of Communication Disorders, The Stanley Steyer School of Health Professions, Faculty of Medicine, Tel Aviv University, Tel Aviv 6997801, Israel; lrabin@tauex.tau.ac.il
[2] Sagol School of Neuroscience, Tel Aviv University, Tel Aviv 6997801, Israel
* Correspondence: yaelzalt@tauex.tau.ac.il

Abstract: The ability to discriminate between talkers based on their fundamental (F0) and formant frequencies can facilitate speech comprehension in multi-talker environments. To date, voice discrimination (VD) of children and adults has only been tested in quiet conditions. This study examines the effect of speech-shaped noise on the use of F0 only, formants only, and the combined F0 + formant cues for VD. A total of 24 adults (18–35 years) and 16 children (7–10 years) underwent VD threshold assessments in quiet and noisy environments with the tested cues. Thresholds were obtained using a three-interval, three-alternative, two-down, one-up adaptive procedure. The results demonstrated that noise negatively impacted the utilization of formants for VD. Consequently, F0 became the lead cue for VD for the adults in noisy environments, whereas the formants were the more accessible cue for VD in quiet environments. For children, however, both cues were poorly utilized in noisy environments. The finding that robust cues such as formants are not readily available for VD in noisy conditions has significant clinical implications. Specifically, the reliance on F0 in noisy environments highlights the difficulties that children encounter in multi-talker environments due to their poor F0 discrimination and emphasizes the importance of maintaining F0 cues in speech-processing strategies tailored for hearing devices.

Keywords: voice discrimination; background noise; speech perception; F0; formants; temporal processing; spectral processing; school-age children

Citation: Kishon-Rabin, L.; Zaltz, Y. The Effect of Noise on the Utilization of Fundamental Frequency and Formants for Voice Discrimination in Children and Adults. *Appl. Sci.* **2023**, *13*, 10752. https://doi.org/10.3390/app131910752

Academic Editor: John S. Allen

Received: 30 August 2023
Revised: 25 September 2023
Accepted: 26 September 2023
Published: 27 September 2023

1. Introduction

Recognizing speech in noisy environments is essential for successful communication in professional, educational, and social activities because it often occurs with background noise. One of the strategies that assist in listening in noisy environments is the ability to identify and follow the voice characteristics of a specific talker of interest [1,2]. The two fundamental voice characteristics that were found to be significant for voice identification are the speaker's fundamental frequency (F0) and formant frequencies [3–5]. F0 is the glottal pulse rate, which is influenced primarily by the length and mass of the vocal cords [6]. Formant frequencies reflect the resonance frequencies of the vocal tract and, as such, they provide cues regarding the speaker's vocal tract length (VTL) in association with his or her physical and perceived body size [6]. In quiet listening conditions, adults and children with normal hearing (NH) were shown to rely on formant frequency cues for voice discrimination (VD) [7–9] and for talker gender categorization [10–13] and considerably less on F0 cues [7–9]. Listening in noisy environments, however, introduces different challenges for discriminating between speakers because noise tends to mask spectral information, such as formants [14,15] and temporal information, including pitch [16,17]. To date, no study has been conducted to explore the utilization of voice cues for discriminating between speakers when in noisy environments. Thus, it remains unclear whether the outcomes

observed in noisy conditions will be similar to those observed in quiet conditions, with the formants being the most prominent cues. Moreover, children are known to be more vulnerable to the effects of noise [18–20] and differ from adults in their utilization of voice cues during speaker-discrimination tasks conducted in quiet conditions [9,21]. Therefore, the influence of noise on their reliance on both cues for VD is unknown. The purpose of the present study was to examine the effect of background noise on the utilization of F0 and formant cues for VD in young adults and school-age children with NH.

The perception of formants has been shown to rely on the frequency resolution capabilities of the cochlea (e.g., [22,23]). Formants are considered robust cues for vowels, mainly because they provide stable acoustic information for more than 100 msec [6] and are therefore readily available to listeners. They can be accompanied by visible articulatory cues that assist in the early acquisition of vowels for hearing and hearing-impaired children [24]. In contrast, the perception of F0 is known to depend on both spectral resolution and temporal processing of the changes in intensity over time in the speech envelope [25,26]. F0 cues are typically not accompanied by visible articulatory information and are considered subtle and therefore less available to listeners, especially in difficult listening conditions. In quiet conditions, children and adults were observed to yield greater advantages from formants compared to F0 cues for VD, supporting the notion that formant frequencies are reliable cues for recognizing a specific speaker [7,8]. The few studies that investigated the effect of background noise on the perception of these cues showed that noise distorted the spectral information conveyed in the speech signal [14,15] and thus negatively affected discrimination between formants [27,28]. Furthermore, the presence of background noise has been documented to impede tasks involving pitch perception, including identifying lexical tones [16] and recognizing emotional tones in speech [17], thereby indicating its potentially detrimental effect on F0 perception as well. These findings give rise to the hypothesis that both acoustic cues may be susceptible to noise and therefore may be differently utilized in noisy conditions compared to quiet conditions.

Children are considered a special group because they are particularly susceptible to the effects of background noise [29,30], with higher SNRs required to achieve similar performance levels as young adults on a wide range of auditory tasks [18,31,32]. These difficulties were explained by a combination of central ('top-down') and peripheral ('bottom-up') factors. Immature 'top-down' processes in children are influenced by the developing cognitive and language capabilities. With particular relevance to listening in noisy conditions, children show deficits in their ability to attribute incoming sounds to their corresponding sources or to selectively attend to an auditory object for further processing while disregarding irrelevant auditory information [19,33–36]. Despite the fact that the cochlea is developed at birth [37], studies showed age-related increase in sensitivity to frequency differences with pure tones or F0 [38–40], as well as improved recognition of spectrally reduced speech with age [41,42], supporting the notion of immature peripheral factors, such as spectral and temporal processing [18,43,44] and/or less efficient top-down cognitive processing, including attention and short-term memory [45,46]. Moreover, in quiet conditions, VD was observed to gradually improve with age in 4–12-year-old children, with the children's formant discrimination becoming adult-like earlier than their F0 discrimination [47]. The presence of background noise may therefore have a different effect on VD for children compared to adults, considering the immaturity of central and auditory processing of the former.

The present study aimed to examine the use of voice cues (F0 and formants and a combination of both) for discriminating between speakers with background noise in NH young adult and school-age children. It was hypothesized that in quiet conditions, formant cues will yield better VD compared to F0 cues for both children and adults, whereas in noisy conditions, the adults and children will exhibit distinct outcome patterns, with the adults better utilizing F0 for VD.

2. Materials and Methods

2.1. Participants

A total of 40 participants took part in the study: 24 adults (18–35 years old; Mean = 23.02, SD = 2.53 years) and 16 children (7–10 years old; Mean = 8.93, SD = 0.86 years). All participants enrolled in the study were proficient in the Hebrew language from birth and adhered to the following inclusive criteria: (1) hearing thresholds within the standard auditory range in both ears; that is, pure-tone air-conduction thresholds not exceeding 20 dB HL at octave frequencies ranging from 500 to 4000 Hz [48]; (2) no prior history of language or learning disorders; (3) no documented attention deficit disorders; (4) minimal musical training experience (less than a year) or no musical training at all; (5) no prior engagement in psychoacoustic evaluations. The background information of participants was established through either self-reporting or through information provided by their parents or guardians. Prior to their involvement, all adult participants and the parents or guardians of the children participants provided informed consent.

2.2. Stimuli

The stimuli utilized in the experiment were composed of three sentences extracted from the Hebrew adaptation of the Matrix sentence [49]. These sentences were verbally recorded by a female speaker who was a native Hebrew speaker. The selection of sentences followed the approach employed in prior investigations [7,8,50]. To mitigate the cognitive load on working memory, the sentences were shortened to three words: subject, predicate, and object. The manipulation of the sentences was conducted by means of a 13-point stimulus continuum. This continuum was designed to follow an exponential progression of $\sqrt{2}$ intervals, spanning from a reduction of -0.18 semitone to a reduction of -8 semitones, as described in previous studies [7,8,50]. This manipulation process created three distinct dimensions of change: (1) Fundamental Frequency (F0); (2) formants, wherein all formants were uniformly shifted in a downward direction based on the 13-point stimulus continuum; and (3) a combined variation involving both F0 and formants, where adjustments mirrored each other. For instance, in the case of the F0-manipulated sentences, the average F0 exhibited variations of 0, -0.18, -0.26, -0.36, -0.51, -0.72, -1.02, -1.44, -2.02, -2.86, -4.02, -5.67, and -8 semitones relative to the original mean F0 of the sentence. Accordingly, for the first sentence, the mean F0 was 175.62 Hz and the comparison sentences underwent exponential alterations at $\sqrt{2}$ increments, ranging from 174 Hz to 110.35 Hz. This was accomplished by utilizing the Pitch Synchronous Overlap and Add (PSOLA) algorithm, which facilitated pitch extraction and manipulation [51]. For formant frequency modifications, a comparable exponential approach was adopted, with adjustments occurring in $\sqrt{2}$ steps. This ranged from a value of 0.99 (representing the smallest ratio between the initial formant frequencies and the adjusted formant frequencies) to 0.63 (reflecting the highest ratio). This manipulation necessitated resampling of the stimuli to compress the frequency axis by a range of factors analogous to those employed for F0 adjustments. Subsequently, the PSOLA algorithm was applied to reinstate the original pitch and duration. Considering that the typical vocal tract length (VTL) for an adult female is approximately 140 mm [52] and that formant frequencies are inversely correlated with VTL [53], the range of formants corresponded to a hypothetical change in VTL from 2 mm to 88 mm. It should be noted that this VTL range was intentionally broader than the physiological variability seen in humans. This decision was initially made to prevent potential lower-bound effects for participants with cochlear implants (CIs) who encountered challenges in distinguishing normal differences in VTL [7]. All manipulations were executed using PRAAT software, version 5.4.17 (copyright © 1992–2015 by Boersma & Weenink).

2.3. Background Noise

Three steady-state speech-shaped noise segments approximately 4 min long were used in a semi-random order. The noise segments were generated by superimposing all synthesized sentences from the Hebrew version of the Matrix test multiple times [49]. As a result,

the long-term spectra of the noise segments matched that of the sentences [54]. Steady-state speech-shaped noise (SN) was used because it creates energetic masking, that is, loss of information representation at the peripheral level [14], and does not involve confounding factors, such as top-down language processes (i.e., informational masking) [14], that may differ between children and adults [55–57].

2.4. Voice Discrimination Threshold (VDT)

A three-interval, three-alternative forced-choice methodology was employed to assess the difference limens (DL) for F0 cues, formant cues, and the combined F0 + formant cues, following previous approaches [7,8,50]. In this procedure, each trial consisted of two reference sentences and one comparison sentence. Upon presentation of a sentence, a selection square on the computer monitor was illuminated, corresponding to the respective sentence. Participants were instructed to indicate the sentence that exhibited a 'different sound' by clicking the corresponding square using a mouse. Consistency was maintained in using the same sentence for each measurement of the VD threshold, with alterations being limited to the tested vocal cue (F0, formants, or F0 + formants). No immediate feedback was provided, and participants were not constrained by time for their responses. To ascertain the thresholds at which a detection rate of 70.7% was achieved on the psychometric function [58], a two-down, one-up adaptive tracking technique was adopted. Initially, the difference between stimuli was halved until the first reversal point, subsequently being incrementally decreased, or augmented by a factor of $\sqrt{2}$ until the sixth reversal point. The calculation of the DLs was based on the geometric mean of the last four reversal points.

2.5. Voice Discrimination Threshold in Noisy Conditions (VDTn)

The SNR at which the VDTn was determined was different for adults and children. Specifically, for the young adults, VDTn was determined at the same SNR for all participants and was set at -5 dB. This was based on previous studies on the Hebrew Matrix in young NH adults, which showed that their mean speech reception threshold in noisy conditions (that is, the SNR that yields 50% correct recognition of speech in noisy conditions, or the SRTn) is -8 dB, with a low amount of variability [32,49]. To test VD in SNRs that yield adequate speech recognition on the psychometric function, based on a slope of approximately 1.8 [59], we doubled the intensity of the sentences (i.e., we increased the sentences that were originally extracted from the Matrix test by 3 dB, resulting in the VD test being conducted at SNR of 5 dB) [60]. In contrast, previous studies with children showed considerably larger between-subject variability of SRTn compared to adults [32]. Therefore, SRTn was evaluated on an individual basis for each child (mean SRTn = -3.33 dB $\pm$ 3.62 dB). In addition, for the children, we added +5 dB to the SRTn, resulting in a mean SNR of +1.67 dB $\pm$ 3.62 dB for the VD test, to ensure ease of listening based on a pilot study we conducted.

2.6. Study Design

The tests that were carried out during the experimental session are illustrated in Figure 1. Specifically, all participants performed a hearing test at the beginning of the testing session, including air conduction thresholds at 500–4000 Hz in both ears, to assure hearing thresholds less than 20 dB HL. Following the hearing test, only the children performed an SRTn assessment. All participants then continued to perform 12 VD measures: 6 in quiet conditions, 2 with each cue (F0, formants, combined), and 6 in background noise conditions, 2 with each of the cues. The order of presentation of voice cues and sentences for the VD test was controlled across participants, with half of the participants in each group first tested in noisy conditions and the other half in quiet conditions. Prior to the formal testing phase, each participant engaged in a brief familiarization exercise involving sentences that exclusively varied in terms of F0 within a quiet context. This entailed 5–10 trial runs, featuring the most substantial distinction between the reference and comparison stimuli. This preliminary step was implemented to ascertain the participants' comprehension of the task before commencing testing. Overall, the VD assessment, conducted both in quiet and

noisy environments, lasted approximately 75 to 90 min, including brief intervals of rest lasting 5 to 8 min each. It is noteworthy that the adult participants did not receive any form of compensation for their participation in the study, while the children were rewarded with stickers, which served as positive reinforcement.

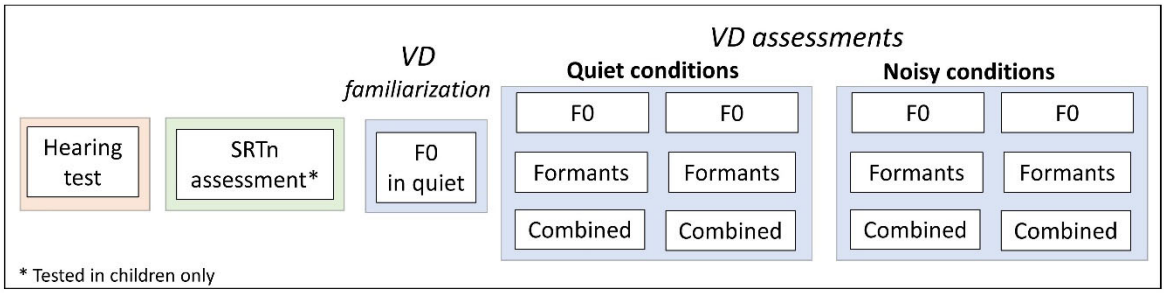

Figure 1. The experimental design. All participants underwent a hearing test and completed 12 VD assessments: 6 in quiet conditions, 2 with each cue (F0, formants, combined), and 6 in background noise, 2 with each of the cues. The order of presentation of voice cues and sentences for the VD test was counterbalanced between participants, with half of the participants first tested in noisy conditions and the other half in quiet conditions. Prior to the testing phase, a brief VD familiarization exercise was carried out with F0 cues in quiet conditions. SRTn = speech reception thresholds in noisy conditions, VD = voice discrimination. Combined = F0 + formants.

2.7. Apparatus

The testing session took place within an acoustically isolated, single-walled chamber designed to mitigate sound interference. The stimuli and background noise (when presented) were delivered using an IBM-compatible computer through a GSI-61 audiometer binaurally via THD-50 headphones.

2.8. Data Analysis

VD thresholds were calculated as the mean of the two measurements conducted for each acoustic cue, separately for the noisy and quiet conditions. Subsequently, all VD thresholds underwent a logarithmic transformation to render their residual distribution suitable for analysis within the framework of a general linear model (as confirmed by the Kolmogorov–Smirnov test with $p > 0.05$). Statistical analysis was performed using SPSS-20 software. To address the issue of multiple comparisons, all post hoc analyses were conducted with Bonferroni corrections applied.

3. Results

The mean ($\pm$1 SE) VD thresholds based on the F0, formants, and combined (F0 + formants) cues in quiet and in noisy conditions for the children and the adults are shown in Table 1. Because there was no significant effect of measurement and no significant interactions with measurement ($p > 0.05$), the two measurements for each acoustic cue, under each condition (quiet; noisy), were averaged. Figure 2 presents box whisker plots for mean measurements 1 and 2 within each group and condition, separately for each acoustic cue.

A repeated measures (RM) ANOVA was conducted with age group (adults, children) as the between-subject variable and condition (quiet, noise) and acoustic cue (F0, formants, combined) as the within-subject variables. The results showed significant main effects of the condition ($F(1,38) = 32.450$ $p < 0.001$, $\eta^2 = 0.461$) and acoustic cue ($[2,38] = 61.840$ $p < 0.001$, $\eta^2 = 0.619$), with no significant main effect of the age group ($F(1,38) = 3.318$ $p = 0.077$). There were significant acoustic cue X age group ($F(2,38) = 4.608$ $p = 0.013$, $\eta^2 = 0.108$) and acoustic cue X condition ($F(2,38) = 19.188$ $p < 0.001$, $\eta^2 = 0.336$) interactions. Post hoc analyses of the interactions are shown in Table 2. In general, the findings indicated that (1) the addition of background noise deteriorated formant VD thresholds for both age groups, and (2) the children exhibited inferior F0 VD thresholds compared to the adults,

irrespective of the testing conditions, with comparable formant and F0+formant VD for both age groups.

Table 1. Mean (±1 SE) voice-discrimination thresholds (in semitones) based on the F0, formants, and combined (F0 + formants) cues in quiet and noisy conditions for children (n = 16) and adults (n = 24).

	Quiet		Noise	
	Adults	**Children**	**Adults**	**Children**
F0 1	0.88 (0.12)	1.66 (0.38)	0.99 (0.12)	1.78 (0.45)
F0 2	0.72 (0.08)	1.86 (0.38)	0.93 (0.11)	1.28 (0.32)
Formant 1	0.62 (0.07)	0.85 (0.14)	1.32 (0.13)	1.39 (0.35)
Formant 2	0.60 (0.07)	0.76 (0.12)	1.20 (0.13)	1.25 (0.31)
Combined 1	0.42 (0.03)	0.59 (0.07)	0.57 (0.07)	0.76 (0.19)
Combined 2	0.43 (0.06)	0.46 (0.06)	0.58 (0.06)	0.80 (0.20)

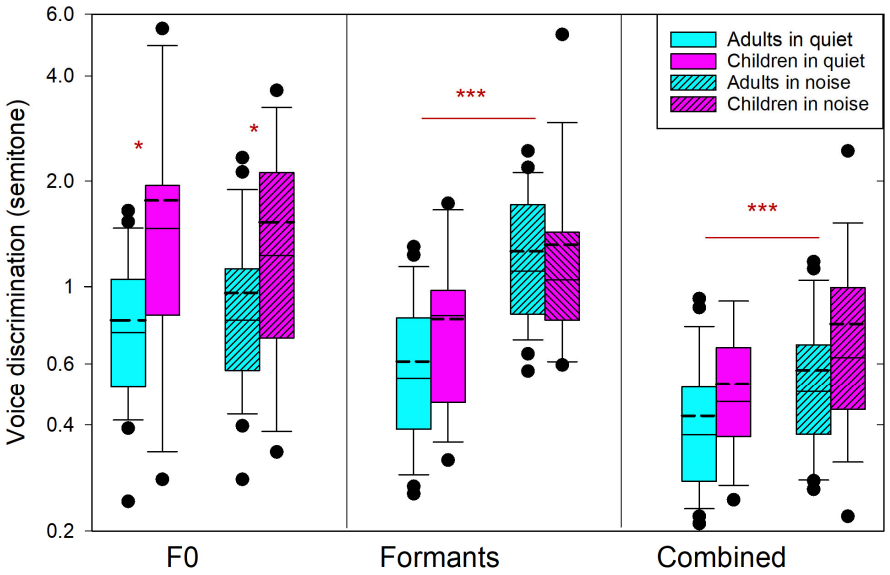

Figure 2. Box plots for mean voice-discrimination thresholds in quiet and noisy conditions for children (n = 16) and adults (n = 24), separately for the F0, formants, and combined (F0 + formants) cues. Box limits encompass the data between the 25th and 75th percentiles, with the median marked by a continuous line within the box. Bars extend to the 10th and 90th percentiles, while outliers are indicated by black dots. The mean is depicted as a dashed line within the box. * = $p < 0.05$ between the children and adults, *** = $p < 0.001$ between the quiet and noisy conditions.

Table 2. Results of the post hoc analyses of the interactions of acoustic cue X age group (first row) and acoustic cue X condition (second row). * = $p < 0.05$, *** = $p < 0.001$. It is important to note that '=' indicates no significant difference between the two cues, and '>' indicates higher (worse) thresholds for the cue indicated on the left.

	F0	**Formants**	**Combined**
Adults and children	Noise = quiet	Noise > quiet ***	Noise > quiet ***
Quiet and noise	Children > adults *	Children = adults	Children = adults

Additional RM ANOVAs separately conducted for each age group revealed significant acoustic cue X condition interactions (for the adults: $F(2,23) = 27.702$ $p < 0.001$, $\eta^2 = 0.546$; for the children: $F(2,15) = 4.420$ $p = 0.027$, $\eta^2 = 0.228$). Post hoc analyses of the interactions showed that in quiet conditions, better VD thresholds were achieved with the formant cues compared with the F0 cues. However, in noisy conditions, for adults, better thresholds were achieved with F0 compared to formant cues, whereas for children, similar thresholds were achieved with both cues. Yet, the best VD thresholds were achieved with the combined cues in both quiet and noisy environments.

Pearson coefficient correlations were conducted for the VDs with each voice cue between the quiet and noisy conditions, separately for the children and adults. The results revealed significant medium-to-strong positive correlations between VD thresholds in quiet and noisy conditions for both groups across voice cues (Figure 3). It is important to note that the correlation based on F0 cues improved ($R = 0.801$ $p < 0.001$) when one outlier was excluded (a child who achieved good VD in quiet conditions but extremely poor VD in noisy conditions). * = $p < 0.05$, *** = $p < 0.001$.

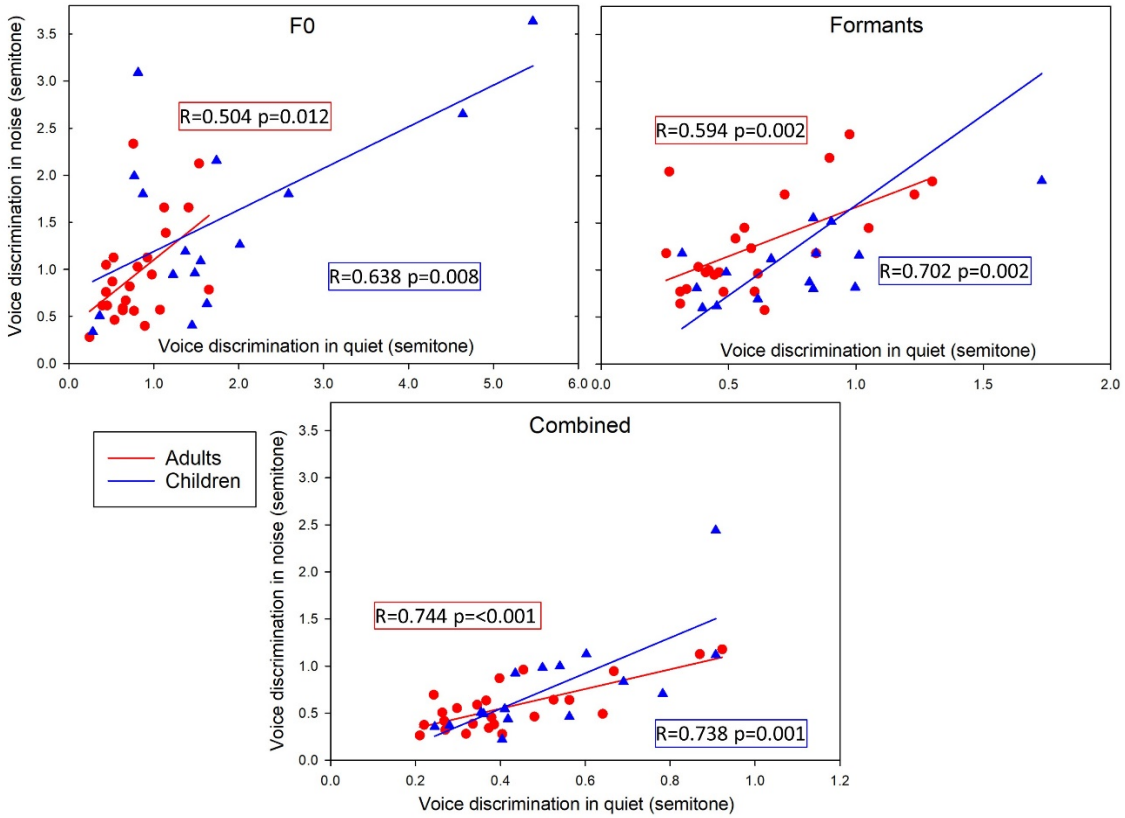

Figure 3. Individual discrimination thresholds for children ($n = 16$) and adults ($n = 24$) in noisy conditions as a function of their discrimination thresholds in quiet conditions, for each acoustic cue separately. The results of Pearson coefficient correlations for each pair are shown in the colored boxes.

4. Discussion

The novel finding of the present study is that background noise had a negative effect on the use of formants for VD, while F0 was less obscured by noise in both children and adults. Specifically, in the quiet condition, children and adults found the formant cues more beneficial for VD compared with F0 cues, whereas in noisy conditions, formant cues did not show this advantage. As a result, a reversed outcome was observed for adults, with F0 becoming the more dominant cue for VD in the presence of noise. For children, noise also obscured formant cues. However, they had difficulty relying on F0 for VD in noisy conditions, probably due to their immature F0 perception, which was already observed in quiet conditions

in this study and others [47]. The current study also found that: (a) the combination of F0 and formants cues produced the best VD over the provision of a single cue (either F0 or formants) in both listening conditions, reflecting the additive value of the formant cues, even in the noisy conditions; (b) there were significant positive correlations found between VD thresholds in quiet and in noisy conditions, suggesting that much of the VD performance in noisy conditions was related to basic spectral and temporal processing mechanisms.

The major finding of this study, that background noise mainly affected formant perception in both children and adults, probably reflects the fact that although the sentences were clearly audible, the noise obscured some of the spectral information contained in the speech stimuli. Because the noise had the same long-term average spectrum as the target sentences, which resulted in a spectral overlap between the target sentences and the noise, it is possible that the listeners had difficulty in identifying the formant peaks in the spectral envelope of the speech signal [43,61,62]. Consequently, larger formant differences were required to accurately discriminate between the voices. This explanation is in line with previous studies that showed degraded vowel formant discrimination when there was background noise [27]. In contrast, VD based on F0 appeared to be less affected by the noise, as shown by the fact that no significant differences were found in VD between quiet and noisy conditions when using F0 for both age groups. This finding is in accordance with studies reporting the robustness of F0 extraction in noisy conditions [63], supporting the notion that F0 perception primarily depends on the temporal envelope of the stimuli [25,26]. Other studies, however, reported the presence of noise to have a detrimental effect on pitch perception [16,17], in line with the theory that noise reduces subcortical neural synchrony, leading to potential disruption of phase locking mechanisms [64,65]. This effect could have also compromised F0 perception due to its partial reliance on fine temporal processing of the speech signal [25,26]. Nonetheless, this potential impact was not observed in the current study, possibly because of the relatively good audibility of the speech signal at the given SNR.

A second important finding is that the children showed significantly poorer VD thresholds based on the F0 cues compared to adults in quiet and noisy conditions. This finding aligns with recent reports on poor VD abilities in children compared to adults [9,21]. Specifically, the finding supports previous evidence indicating that children exhibit non-adult-like F0 discrimination skills until the age of 12 years, while their formant-discrimination abilities reach adult-like levels at approximately 8 years of age [21]. Our findings are also in agreement with observations indicating an age-related enhancement in sensitivity to frequency differences, whether in the context of pure tones or F0 [38–40]. This observed developmental pattern can potentially be attributed to the progressive maturation of temporal processing mechanisms [37]. Other studies suggest, however, that although, as a group, children may show immature temporal processing abilities, adult-like frequency discrimination thresholds [45] and VD thresholds based on F0 cues can be achieved in childhood [8]. This discrepancy can be explained by the notion that an ongoing process of maturation for temporal and central processing occurs in childhood.

The maturational differences in F0 perception between the children and adults seemed to have led to different coping strategies when the formant cues were less accessible, i.e., in the presence of noise. For adults, F0 became the more salient cue, resulting in only a decline of about 35% (from 0.43 to 0.58 semitone) in VD when the combined F0 + formant cues were provided in the noisy conditions (comparing second VD measurements between quiet and noisy conditions). In contrast, children encountered a more significant challenge in noisy conditions, inefficiently utilizing both F0 and formant cues, leading to deterioration of about 74% (from 0.46 to 0.80 semitones) in VD with combined cues. Although the difference between the children and adults' VD in noisy conditions based on the combined cues was not found to be significant, potentially due to the large within-group variance for the children, this outcome may help clarify the difficulties children face when listening in noisy environments.

A substantial portion, ranging from 25% to 55% of the observed variance in VD performance in noisy conditions, could be accounted for by the performance exhibited in quiet

environments by both children and adults. This indicates a strong influence of the basic spectral and temporal processing mechanisms involved in VD under quiet conditions on the performance in noisy environments. These mechanisms encompass the cochlea's ability to resolve frequencies, which provides crucial information regarding the relationship between formant peaks in the spectral envelope of the speech signal [22,23], as well as the temporal processing of intensity fluctuations over time in the speech envelope [25,26]. However, an additional 75% to 45% of the variance in performance in noisy conditions was not explained by performance in quiet conditions. This variance may be attributed to higher top-down cognitive processes such as attention and short-term memory, which have a significant role in the formation of auditory objects and enhance stream segregation [66,67]. Specifically, auditory attention could serve to extract the relevant spectral and temporal signal elements from the competing background noise, temporarily storing them in working memory to facilitate the discrimination task [68]. Subsequent investigations should consider exploring the correlation between VD in noisy conditions and cognitive capabilities of the participants.

The combination of both F0 and formant cues yielded the best discrimination scores in quiet and in noisy conditions. That is, participants from both age groups benefitted from the integration of these cues, even though the formant cues were partly interrupted by noise. A possible explanation for this finding may be related to the fact that children acquire substantial exposure to the integration of these two vocal cues under impoverished conditions from a very early age. In everyday routine, children and adults often hear many voices embedded in background noise. Those are mixed into a unified acoustic stream, necessitating listeners to parse the desired speech from the obstructing background noise to facilitate successful communication. Accurate identification of the F0 and formant frequencies of the relevant talker aids in the segregation process and, thus, enhances speech recognition [69]. This explanation corroborates studies that reported improvements in the segregation of concurrent vowels [70] and sentences [3] when the F0 and formant frequencies of the relevant talker largely differed from those of the distracting talker, further emphasizing the importance of the efficient utilization and integration of both F0 and VTL in speech perception.

5. Limitations of the Current Study and Future Directions

One limitation of the present study stems from the fact that a different method was used to select the SNR for the VD test between the children and adults. Specifically, VD was conducted using the same SNR across the adult participants but was individually adapted for the children based on their SRTn. We believe that this has not influenced the main outcomes of this study because, despite receiving the voice cues at a better SNR than the adults, the children nonetheless showed degraded formant perception compared to quiet conditions. This, however, should be confirmed in future studies where a similar testing method is employed for all participants. In addition, in the present study, the effect of noise on VD was examined using a steady-state speech-shaped noise and not temporally modulated or babble noises in order to avoid intervening top-down processes such as glimpsing and/or linguistic skills that differ between children and adults. Future studies may want to examine the effect of variable noise types to broaden the ecological implications of the current study. To extend the applicability of the current findings, it is important to assess the effect of noise on F0 and formant VD in populations who struggle to understand speech in noisy conditions, such as older adults, very young children, individuals with hearing loss, and those with language impairment.

6. Conclusions

The present study is the first to test the effect of background noise on VD of school-age children and young adults. The findings indicate that the presence of noise, at a level that does not impede the audibility of the speech stimuli, adversely affects the utilization of formants for VD in both school-age children and adults. These findings contribute to our understanding of the acoustic cues involved in the process of perceiving speech in noisy

conditions, emphasizing the importance of accessibility to F0 cues, particularly in challenging auditory environments characterized by ambient noise. These insights may explain part of the difficulties individuals with hearing impairment experience in noisy environments and warrant careful consideration in the development of future signal-processing strategies within the realm of auditory devices and auditory prosthetics (e.g., cochlear implants). Furthermore, the negative effect of noise on VD may pose a particular disadvantage for young school-age children, given their less-mature utilization of F0 cues for VD. Consequently, their ability to identify and follow the voice characteristics of a specific talker of interest as a means to enhance speech understanding in noisy environments may be hindered. This constraint may elucidate the challenges children face, even those without any hearing and/or language impairment, when trying to understand speech in such demanding environments.

Author Contributions: Y.Z. supervised the data collection and analyzed the data. Both Y.Z. and L.K.-R. designed the research, discussed the results, and wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and the research protocol was approved by the Institutional Review Board of Tel Aviv University (approval code 39.19).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors extend their gratitude to Feigy Grinvald, as well as Michal Karman and Haya Skornik from Tel Aviv University's Department of Communication Disorders, for their assistance in the data collection process.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Brungart, D.S.; Chang, P.S.; Simpson, B.D.; Wang, D. Multitalker speech perception with ideal time-frequency segregation: Effects of voice characteristics and number of talkers. *J. Acoust. Soc. Am.* **2009**, *125*, 4006–4022. [CrossRef] [PubMed]
2. Bronkhorst, A.W. The cocktail-party problem revisited: Early processing and selection of multi-talker speech. *Atten. Percept. Psychophys.* **2015**, *77*, 1465–1487. [CrossRef] [PubMed]
3. Darwin, C.J.; Brungart, D.S.; Simpson, B.D. Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *J. Acoust. Soc. Am.* **2003**, *114*, 2913–2922. [CrossRef]
4. Drullman, R.; Bronkhorst, A.W. Speech perception and talker segregation: Effects of level, pitch, and tactile support with multiple simultaneous talkers. *J. Acoust. Soc. Am.* **2004**, *116*, 3090–3098. [CrossRef] [PubMed]
5. Vestergaard, M.D.; Fyson, N.R.C.; Patterson, R.D. The mutual roles of temporal glimpsing and vocal characteristics in cocktail-party listening. *J. Acoust. Soc. Am.* **2011**, *130*, 429–439. [CrossRef]
6. Raphael, L.J.; Borden, G.J.; Harris, K.S. *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*; LWW: Philadelphia, PA, USA, 2007.
7. Zaltz, Y.; Goldsworthy, R.L.; Kishon-Rabin, L.; Eisenberg, L.S. Voice discrimination by adults with cochlear implants: The benefits of early implantation for vocal-tract length perception. *J. Assoc. Res. Otolaryngol.* **2018**, *19*, 193–209. [CrossRef]
8. Zaltz, Y.; Goldsworthy, R.L.; Eisenberg, L.S.; Kishon-Rabin, L. Children with normal hearing are efficient users of fundamental frequency and vocal tract length cues for voice discrimination. *Ear Hear.* **2020**, *41*, 182–193. [CrossRef]
9. Zaltz, Y. The effect of stimulus type and testing method on talker discrimination of school-age children. *J. Acoust. Soc. Am.* **2023**, *153*, 2611. [CrossRef]
10. Fuller, C.D.; Gaudrain, E.; Clarke, J.N.; Galvin, J.J.; Fu, Q.-J.; Free, R.H.; Başkent, D. Gender categorization is abnormal in cochlear implant users. *J. Assoc. Res. Otolaryngol.* **2014**, *15*, 1037–1048. [CrossRef]
11. Hillenbrand, J.M.; Clark, M.J. The role of f(0) and formant frequencies in distinguishing the voices of men and women. *Atten. Percept. Psychophys.* **2009**, *71*, 1150–1166. [CrossRef]
12. Skuk, V.G.; Schweinberger, S.R. Gender differences in familiar voice identification. *Hear. Res.* **2013**, *296*, 131–140. [CrossRef]
13. Meister, H.; Fürsen, K.; Streicher, B.; Lang-Roth, R.; Walger, M. The use of voice cues for speaker gender recognition in cochlear implant recipients. *J. Speech Lang. Hear. Res.* **2016**, *59*, 546–556. [CrossRef]

14. Brungart, D.S. Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.* **2001**, *109*, 1101–1109. [CrossRef] [PubMed]
15. Ezzatian, P.; Li, L.; Pichora-Fuller, M.K.; Schneider, B.A. The effect of energetic and informational masking on the time-course of stream segregation: Evidence that streaming depends on vocal fine structure cues. *Lang. Cogn. Process.* **2012**, *27*, 1056–1088. [CrossRef]
16. Mao, Y.; Xu, L. Lexical tone recognition in noise in normal-hearing children and prelingually deafened children with cochlear implants. *Int. J. Audiol.* **2017**, *56*, S23–S30. [CrossRef] [PubMed]
17. Luo, X. Talker variability effects on vocal emotion recognition in acoustic and simulated electric hearing. *J. Acoust. Soc. Am.* **2016**, *140*, EL497. [CrossRef]
18. Corbin, N.E.; Bonino, A.Y.; Buss, E.; Leibold, L.J. Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers. *Ear Hear.* **2016**, *37*, 55–63. [CrossRef]
19. Wightman, F.L.; Kistler, D.J. Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *J. Acoust. Soc. Am.* **2005**, *118*, 3164–3176. [CrossRef]
20. Neuman, A.C.; Wroblewski, M.; Hajicek, J.; Rubinstein, A. Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. *Ear Hear.* **2010**, *31*, 336–344. [CrossRef]
21. Nagels, L.; Gaudrain, E.; Vickers, D.; Hendriks, P.; Başkent, D. Development of voice perception is dissociated across gender cues in school-age children. *Sci. Rep.* **2020**, *10*, 5074. [CrossRef]
22. Fant, G. *Acoustic Theory of Speech Production*; Mouton: The Hague, The Netherlands, 1960.
23. Lieberman, p.; Blumstein, S.E. Source-filter theory of speech production. In *Speech Physiology, Speech Perception, and Acoustic Phonetics Cambridge Studies in Speech Science and Communication*; Cambridge University Press: Cambridge, UK, 1988; pp. 34–50.
24. Kishon-Rabin, L.; Taitelbaum, R.; Muchnik, C.; Gehtler, I.; Kronenberg, J.; Hildesheimer, M. Development of speech perception and production in children with cochlear implants. *Ann. Otol. Rhinol. Laryngol. Suppl.* **2002**, *189*, 85–90. [CrossRef] [PubMed]
25. Carlyon, R.P.; Shackleton, T.M. Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms? *J. Acoust. Soc. Am.* **1994**, *95*, 3541–3554. [CrossRef]
26. Oxenham, A.J. Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends Amplif.* **2008**, *12*, 316–331. [CrossRef] [PubMed]
27. Liu, C.; Kewley-Port, D. Formant discrimination in noise for isolated vowels. *J. Acoust. Soc. Am.* **2004**, *116*, 3119–3129. [CrossRef] [PubMed]
28. Swanepoel, R.; Oosthuizen, D.J.J.; Hanekom, J.J. The relative importance of spectral cues for vowel recognition in severe noise. *J. Acoust. Soc. Am.* **2012**, *132*, 2652–2662. [CrossRef]
29. Anderson, S.; Kraus, N. Sensory-cognitive interaction in the neural encoding of speech in noise: A review. *J. Am. Acad. Audiol.* **2010**, *21*, 575–585. [CrossRef]
30. Erickson, L.C.; Newman, R.S. Influences of background noise on infants and children. *Curr. Dir. Psychol. Sci.* **2017**, *26*, 451–457. [CrossRef]
31. Elliott, L.L.; Connors, S.; Kille, E.; Levin, S.; Ball, K.; Katz, D. Children's understanding of monosyllabic nouns in quiet and in noise. *J. Acoust. Soc. Am.* **1979**, *66*, 12–21. [CrossRef]
32. Zaltz, Y.; Bugannim, Y.; Zechoval, D.; Kishon-Rabin, L.; Perez, R. Listening in noise remains a significant challenge for cochlear implant users: Evidence from early deafened and those with progressive hearing loss compared to peers with normal hearing. *J. Clin. Med.* **2020**, *9*, 1381. [CrossRef]
33. Doyle, A.B. Listening to distraction: A developmental study of selective attention. *J. Exp. Child Psychol.* **1973**, *15*, 100–115. [CrossRef]
34. Coch, D.; Sanders, L.D.; Neville, H.J. An event-related potential study of selective auditory attention in children and adults. *J. Cogn. Neurosci.* **2005**, *17*, 605–622. [CrossRef] [PubMed]
35. Bonino, A.Y.; Leibold, L.J.; Buss, E. Release from perceptual masking for children and adults: Benefit of a carrier phrase. *Ear Hear.* **2013**, *34*, 3–14. [CrossRef] [PubMed]
36. Leibold, L.J.; Buss, E. Children's Identification of Consonants in a Speech-Shaped Noise or a Two-Talker Masker. *J. Speech Lang. Hear. Res.* **2013**, *56*, 1144–1155. [CrossRef] [PubMed]
37. Moore, J.K.; Linthicum, F.H. The human auditory system: A timeline of development. *Int. J. Audiol.* **2007**, *46*, 460–478. [CrossRef] [PubMed]
38. Moore, D.R.; Cowan, J.A.; Riley, A.; Edmondson-Jones, A.M.; Ferguson, M.A. Development of auditory processing in 6- to 11-yr-old children. *Ear Hear.* **2011**, *32*, 269–285. [CrossRef]
39. Buss, E.; Flaherty, M.M.; Leibold, L.J. Development of frequency discrimination at 250 Hz is similar for tone and /ba/ stimuli. *J. Acoust. Soc. Am.* **2017**, *142*, EL150. [CrossRef]
40. Flaherty, M.M.; Buss, E.; Leibold, L.J. Developmental effects in children's ability to benefit from F0 differences between target and masker speech. *Ear Hear.* **2019**, *40*, 927–937. [CrossRef]
41. Eisenberg, L.S.; Shannon, R.V.; Martinez, A.S.; Wygonski, J.; Boothroyd, A. Speech recognition with reduced spectral cues as a function of age. *J. Acoust. Soc. Am.* **2000**, *107*, 2704–2710. [CrossRef]
42. Mlot, S.; Buss, E.; Hall, J.W. Spectral integration and bandwidth effects on speech recognition in school-aged children and adults. *Ear Hear.* **2010**, *31*, 56–62. [CrossRef]

43. Nishi, K.; Lewis, D.E.; Hoover, B.M.; Choi, S.; Stelmachowicz, P.G. Children's recognition of American English consonants in noise. *J. Acoust. Soc. Am.* **2010**, *127*, 3177–3188. [CrossRef]
44. Hall, J.W.; Grose, J.H. Development of temporal resolution in children as measured by the temporal modulation transfer function. *J. Acoust. Soc. Am.* **1994**, *96*, 150–154. [CrossRef] [PubMed]
45. Zaltz, Y.; Ari-Even Roth, D.; Karni, A.; Kishon-Rabin, L. Long-term training-induced gains of an auditory skill in school-age children as compared with adults. *Trends Hear.* **2018**, *22*, 2331216518790902. [CrossRef] [PubMed]
46. Halliday, L.F.; Taylor, J.L.; Edmondson-Jones, A.M.; Moore, D.R. Frequency discrimination learning in children. *J. Acoust. Soc. Am.* **2008**, *123*, 4393–4402. [CrossRef] [PubMed]
47. Nagels, L.; Gaudrain, E.; Vickers, D.; Matos Lopes, M.; Hendriks, P.; Başkent, D. Development of vocal emotion recognition in school-age children: The EmoHI test for hearing-impaired populations. *PeerJ* **2020**, *8*, e8773. [CrossRef]
48. ANSI/ASA S3.6-2018—Specification for Audiometers. Available online: https://webstore.ansi.org/standards/asa/ansiasas32018 (accessed on 19 August 2023).
49. Bugannim, Y.; Roth, D.A.-E.; Zechoval, D.; Kishon-Rabin, L. Training of speech perception in noise in pre-lingual hearing-impaired adults with cochlear implants compared with normal hearing adults. *Otol. Neurotol.* **2019**, *40*, e316–e325. [CrossRef]
50. Zaltz, Y.; Kishon-Rabin, L. Difficulties experienced by older listeners in utilizing voice cues for speaker discrimination. *Front. Psychol.* **2022**, *13*, 797422. [CrossRef]
51. Moulines, E.; Charpentier, F. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Commun.* **1990**, *9*, 453–467. [CrossRef]
52. Fitch, W.T.; Giedd, J. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *J. Acoust. Soc. Am.* **1999**, *106*, 1511–1522. [CrossRef]
53. Lammert, A.C.; Narayanan, S.S. On short-time estimation of vocal tract length from formant frequencies. *PLoS ONE* **2015**, *10*, e0132193. [CrossRef]
54. Kollmeier, B.; Warzybok, A.; Hochmuth, S.; Zokoll, M.A.; Uslar, V.; Brand, T.; Wagener, K.C. The multilingual matrix test: Principles, applications, and comparison across languages: A review. *Int. J. Audiol.* **2015**, *54* (Suppl. S2), 3–16. [CrossRef]
55. Elliott, L.L. Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *J. Acoust. Soc. Am.* **1979**, *66*, 651–653. [CrossRef] [PubMed]
56. Hall, J.W.; Grose, J.H.; Buss, E.; Dev, M.B. Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear Hear.* **2002**, *23*, 159–165. [CrossRef] [PubMed]
57. Niesen, M.; Bourguignon, M.; Bertels, J.; Vander Ghinst, M.; Wens, V.; Goldman, S.; De Tiège, X. Cortical tracking of lexical speech units in a multi-talker background is immature in school-aged children. *Neuroimage* **2023**, *265*, 119770. [CrossRef] [PubMed]
58. Levitt, H. Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* **1971**, *49* (Suppl. S2), 467. [CrossRef]
59. Sobon, K.A.; Taleb, N.M.; Buss, E.; Grose, J.H.; Calandruccio, L. Psychometric function slope for speech-in-noise and speech-in-speech: Effects of development and aging. *J. Acoust. Soc. Am.* **2019**, *145*, EL284. [CrossRef]
60. Abdel-Latif, K.H.A.; Meister, H. Speech Recognition and Listening Effort in Cochlear Implant Recipients and Normal-Hearing Listeners. *Front. Neurosci.* **2021**, *15*, 725412. [CrossRef]
61. Dubno, J.R.; Dirks, D.D. Evaluation of hearing-impaired listeners using a Nonsense-Syllable Test. I. Test reliability. *J. Speech Hear. Res.* **1982**, *25*, 135–141. [CrossRef]
62. Stelmachowicz, P.G.; Lewis, D.E.; Kelly, W.J.; Jesteadt, W. Speech perception in low-pass filtered noise for normal and hearing-impaired listeners. *J. Speech Hear. Res.* **1990**, *33*, 290–297. [CrossRef]
63. Gockel, H.; Moore, B.C.J.; Plack, C.J.; Carlyon, R.P. Effect of noise on the detectability and fundamental frequency discrimination of complex tones. *J. Acoust. Soc. Am.* **2006**, *120*, 957–965. [CrossRef]
64. Dimitrijevic, A.; Pratt, H.; Starr, A. Auditory cortical activity in normal hearing subjects to consonant vowels presented in quiet and in noise. *Clin. Neurophysiol.* **2013**, *124*, 1204–1215. [CrossRef]
65. Han, J.-H.; Lee, J.; Lee, H.-J. Noise-induced change of cortical temporal processing in cochlear implant users. *Clin. Exp. Otorhinolaryngol.* **2020**, *13*, 241–248. [CrossRef] [PubMed]
66. Best, V.; Gallun, F.J.; Carlile, S.; Shinn-Cunningham, B.G. Binaural interference and auditory grouping. *J. Acoust. Soc. Am.* **2007**, *121*, 1070–1076. [CrossRef] [PubMed]
67. Heinrich, A.; Schneider, B.A.; Craik, F.I.M. Investigating the influence of continuous babble on auditory short-term memory performance. *Q. J. Exp. Psychol.* **2008**, *61*, 735–751. [CrossRef] [PubMed]
68. Johnson, J.A.; Zatorre, R.J. Attention to simultaneous unrelated auditory and visual events: Behavioral and neural correlates. *Cereb. Cortex* **2005**, *15*, 1609–1620. [CrossRef]
69. Başkent, D.; Gaudrain, E. Musician advantage for speech-on-speech perception. *J. Acoust. Soc. Am.* **2016**, *139*, EL51–EL56. [CrossRef]
70. Vestergaard, M.D.; Fyson, N.R.C.; Patterson, R.D. The interaction of vocal characteristics and audibility in the recognition of concurrent syllables. *J. Acoust. Soc. Am.* **2009**, *125*, 1114–1124. [CrossRef]