International Journal of Audiology 2014; 53: 270–284

Original Article

Low-frequency signals support perceptual organization of implant-simulated speech for adults and children

Susan Nittrouer, Eric Tarr, Virginia Bolster, Amanda Caldwell-Tarr, Aaron C. Moberly & Joanna H. Lowenstein

Department of Otolaryngology, The Ohio State University, Columbus, USA

Abstract

Objective: Using signals processed to simulate speech received through cochlear implants and low-frequency extended hearing aids, this study examined the proposal that low-frequency signals facilitate the perceptual organization of broader, spectrally degraded signals. *Design:* In two experiments, words and sentences were presented in diotic and dichotic configurations as four-channel noise-vocoded signals (VOC-only), and as those signals combined with the acoustic signal below 0.25 kHz (LOW-plus). Dependent measures were percent correct recognition, and the difference between scores for the two processing conditions given as proportions of recognition scores for VOC-only. The influence of linguistic context was also examined. *Study sample:* Participants had normal hearing. In all, 40 adults, 40 seven-year-olds, and 20 five-year-olds participated. *Results:* Participants of all ages showed benefits of adding the low-frequency signal. The effect was greater for sentences than words, but no effect of diotic versus dichotic presentation was found. The influence of linguistic context was similar across age groups, and did not contribute to the low-frequency effect. Listeners who had poorer VOC-only scores showed greater low-frequency effects. *Conclusion:* The benefit of adding a low-frequency signal to a broader, spectrally degraded signal derives in some part from its facilitative influence on perceptual organization of the sensory input.

Key Words: Electric-acoustic stimulation; perceptual organization; hearing loss; bimodal stimulation; hybrid implants

William House is generally recognized as the inventor of the modern-day cochlear implant (Niparko & Wilson, 2000). Even though that first device, implanted in two patients in 1961, did not support especially good speech recognition, his attempts are credited with igniting scientific inquiry into how an auditory prosthesis might be designed to provide interpretable signals to users. Eventually those efforts proved fruitful, and in 1995 a panel of experts convened by the National Institutes of Health declared the device and the signal processing it employed to be successes. Their evidence was that most postlingually deafened adults using the devices available at that time could correctly recognize about 80% of words in highly predictable sentences, when presented in quiet. That marked a tremendous advance in speech recognition over what had previously been possible for individuals with severe-to-profound hearing loss through hearing aids. Nonetheless, those outcomes did not represent unmitigated success. In particular, even when listeners shared endogenous traits, used the same devices, and were tested in identical conditions, speech recognition scores were highly variable, with some listeners doing quite poorly (e.g. Helms et al, 1997; Müller et al, 2002; Skinner et al, 1994; Wilson, 2006). As a consequence, efforts have continued to try to find ways of improving further the speech recognition performance of listeners with severe-to-profound hearing loss.

One of those efforts has involved looking at potential benefits of combining electric and acoustic stimulation. Until roughly the turn of the century, no attention had been paid to trying to preserve or use residual hearing when patients received cochlear implants. Candidacy for an implant required that patients have no residual hearing whatsoever, meaning that no single auditory threshold could be better than 100 dB hearing level (NIH Consensus Development Panel, 1995). Some of the reason for that requirement was that surgical techniques were not sufficiently refined to prevent the loss of whatever hearing might remain, due to cochlear trauma, high noise during surgery, and perilymphatic fluid loss (Cohen, 2004; Gantz & Turner, 2003; Skarzynski et al, 2002). In any event, common wisdom at that time was that listeners would have difficulty integrating auditory input provided by combined electric and acoustic stimulation. Specifically there was concern that combining the two sorts of signals would create informational masking for users because of a lack of auditory integration across signal types and frequencies (e.g. Dooley et al, 1993).

However, as advances were made in the design of cochlear implants themselves and in the signal processing strategies used with those devices, outcomes of patients with severe-to-profound hearing loss who received cochlear implants started outpacing those of

Correspondence: Susan Nittrouer, Department of Otolaryngology, The Ohio State University, 915 Olentangy River Road, Suite 4000, Columbus, OH 43212, USA. E-mail: Nittrouer.1@osu.edu

(Received 2 May 2013; accepted 30 November 2013)

ISSN 1499-2027 print/ISSN 1708-8186 online © 2014 British Society of Audiology, International Society of Audiology, and Nordic Audiological Society
DOI: 10.3109/14992027.2013.871649

patients with less severe hearing loss who wore hearing aids (Boothroyd, 1997; NIH Consensus Development Panel, 1995; Osberger, 1994). That trend, in turn, led to changes in requirements for implant candidacy, such that patients with some residual hearing began to be considered. The change in the population of individuals receiving implants led to concomitant changes in surgery, to preserve residual hearing, and in perspectives about combining electric and acoustic stimulation. Two new options for configuring auditory prostheses emerged: bimodal, in which a traditional cochlear implant is worn on one ear and a hearing aid on the other ear; and hybrid, in which a specially designed implant and an attached hearing aid are worn on the same ear. These hybrid devices can be implanted bilaterally. Thus, the combination of electric and acoustic stimulation can be dichotic or diotic in current clinical practice. Both of these treatment approaches have demonstrated improvements over what is available with implants alone in terms of speech recognition for adults (e.g. Ching et al, 2004; Dorman & Gifford, 2010; Gantz et al, 2006; Gifford et al, 2007). For example, a recent review of outcome studies on hybrid implants by Incerti and colleagues (2013) revealed that average improvement in word recognition is about 10% for combined electric-acoustic stimulation over electric alone. That outcome was demonstrated for a variety of electrode arrays, and across patients with variety in the amount of remaining low-frequency hearing. Outcomes of a similar magnitude have been reported for patients who use bimodal stimulation: that is, recognition improves by roughly 10% with the use of a hearing aid on the ear contralateral to the cochlear implant, compared to the implant alone (e.g. Hamzavi et al, 2004). For children, language learning has been shown to be facilitated by the combination of electric and acoustic stimulation (Ching et al, 2001; Nittrouer & Chapman, 2009).

Most studies reviewed above involved patients with some residual hearing, defined as auditory thresholds of 60 dB hearing level or better in frequencies up to at least 0.5 kHz, but possibly as high as 1 kHz (von Ilberg et al, 2011). As a consequence, these patients generally could hear the first formant of speech signals through their hearing aid. They also could hear five to eight harmonics of the fundamental frequency. Access to these two properties of the speech signal can help with speaker identification, signal segregation in noise, and cues to both vowel identity and consonant manner. In contrast to that work, the current study was concerned with questions related to electric-acoustic stimulation when the acoustic signal would only be very low frequency. The current study simulated listening conditions in which a traditional cochlear implant would be used, with a low-frequency cut-off of 0.25 kHz, and the only acoustic signal that could be heard was lower than 0.25 kHz. This situation meant that typically only the first harmonic (i.e. fundamental frequency) was available to listeners in this study. As would be expected, this single harmonic is not interpretable as any kind of linguistic unit by itself.

An essential proposition of the work conducted here was that an explanation for the broad variability in speech recognition outcomes observed for users of traditional implants might rest with how well patients organize the degraded signals they receive through those implants. It has been known for quite some time that signals that are inherently non-speech can be organized perceptually so that phonetic qualities can be recovered (e.g. Risberg & Agelfors, 1982). Referred to here are signals that lack the periodic structure imposed on speech by glottal pulsing or the broad bandwidths associated with vocal-tract resonances. The most commonly recognized demonstration of the principle at stake involves sine-wave speech. In the first use of these speech analogs, Remez and colleagues (1981) processed speech signals to preserve only the center frequencies of each of the lowest three formants, and presented those frequencies as time-varying sine waves. When no instructions were given to study participants, most reported hearing whistles or bird chirps or some other nonspeech signal. When participants were instructed that they would be hearing sentences that they were to transcribe, however, most were able to recognize the original sentences, indicating that they were able to integrate the separate sine waves into unitary phonetic objects. Thus, it is fair to conclude that listeners were imposing organization on the sensory information they were receiving.

The phenomenon described above is well studied in visual psychology. The definition in that work of perceptual organization is that it refers to the processes structuring visual information into coherent units (e.g. Kimchi, 2009), and that definition can be readily transferred to speech perception where listeners structure acoustic information into coherent phonetic units. In visual psychology, a common example of the phenomenon is to be found in Ruben's vase. This pattern of light and dark is seen as either two faces (in profile) on opposing sides of the image or as a single vase in the middle. The same sensory information is available in each case, but the form that is recovered is determined by how that information is organized in the perceptual system. As this example illustrates, the critical role of organizing sensory input in perception is easy to demonstrate for situations where the input is sparse, or degraded. Nonetheless, that kind of organization surely takes place on a regular basis in our everyday interactions with the world. Where speech is concerned, for example, there is no ready rationale for why the disparate formants, fricative noises, and release bursts produced over the course of production get harnessed together into phonetic forms. Yet they do quite efficiently, at least for most of us. But even with the best signal processing algorithms, the signals provided by cochlear implants are highly degraded by comparison to natural speech. Therefore it is reasonable to propose that some patients with implants may have difficulty organizing these signals such that they can recover phonetic forms.

This proposal—that a primary challenge for users of cochlear implants rests with developing perceptual strategies that permit the recovery of phonetic form—is not new. In 1983, in a presentation to the New York Academy of Sciences, Studdert-Kennedy made a similar proposal, only more specifically cautioning that the structure preserved by implant processing needs to specify the time-varying articulatory structure underlying speech. The proposal tested in the current work was that the very low-frequency components of speech can facilitate the perceptual organization of spectrally degraded speech signals in order to improve recognition. But again, this idea is not entirely new. Chang et al (2006) tested a similar claim using sentences that were noise-vocoded into four channels. Those signals were presented in background noise, either alone or with the low-pass filtered signal (below 0.3 kHz) from the original sentences. Using an adaptive procedure, thresholds of intelligibility (i.e. where word recognition was 50% correct) were obtained. Results showed that those thresholds could be obtained at signal-to-noise ratios that were 10 to 15 dB poorer when the low-frequency signal was present than when it was not. Because that low-pass signal was completely unrecognizable on its own, the authors concluded that the effect could not have been due to a linear addition of low- and high-frequency signal components that separately inform phonetic or lexical decisions. Rather, they decided, the low-pass signal facilitated appropriate grouping of the noise-vocoded signals. A similar outcome was subsequently obtained by Başkent and Chatterjee (2010), who additionally observed that the benefit accrued by the low-frequency signal component was greatest when recognition with the vocoded-alone signals was poorest.

The first issue addressed by the current study was whether the beneficial effect of adding a low-frequency signal to a vocoded signal reported in these earlier studies could be replicated in the current study with slightly different procedures.

A second issue addressed by the current study was whether the benefit of the low-frequency signal on recognition would be stronger when the low-pass and vocoded signals were presented to the same ears, in a diotic manner, rather than to different ears, in a dichotic manner. The practice of presenting some spectral components of the speech signal to one ear and other components to the other ear in recognition tasks has been widely used since its inception (Kimura, 1961). Experiments employing this paradigm typically show that recognition of syllables, words, or sentences is not impaired under dichotic conditions of presentation for listeners with normal hearing (e.g. Fox & Jacewicz, 2010; Liberman et al, 1981; Mann & Liberman, 1983; Studdert-Kennedy et al, 1972; Whalen & Liberman, 1987). However, in those experiments, the spectral components divided between ears for presentation were all linguistically relevant: typically, different formants were presented to different ears. In the current experiment, the signal was divided such that most of the spectrum, so most of the phonetically informative region, was presented to one ear, and the generally uninformative, low-frequency component was presented to the other ear. That division might not matter to outcomes, because spectral integration is believed to occur centrally in the auditory system, rather than at the auditory periphery (Chistovich, 1985; Fox et al, 2008). But not all studies have found spectral integration for speech signals presented dichotically. For example, Loizou et al (2003) created sine-vocoded analogs of speech, and presented them in both diotic and dichotic presentation modes to listeners with normal hearing. Results showed that recognition of sentences was significantly hampered when the lower-frequency channels were presented to one ear and the higher-frequency channels were presented to the other ear. Furthermore, tests of consonant and vowel recognition with deaf patients using bimodal stimulation have revealed evidence of poorer integration across frequencies and ears than what is observed for listeners with normal hearing (Kong & Braida, 2011; Yang & Zeng, 2013). In the present study, both the vocoded-only as well as the low-frequency enhanced signals were presented in both diotic and dichotic conditions to see if a finding of better recognition with the diotic configuration, as reported by Loizou et al (2003), would be replicated. The outcome of this test should have strong clinical implications. If an advantage for the diotic condition were found, it could mean that bilateral hybrid implants should be viewed as preferable to bimodal stimulation, even for patients with severe-to-profound hearing loss. Currently, hybrid implants are used only with patients who have enough residual hearing to be able to get more than just the first harmonic through acoustic amplification.

A third issue addressed in the current study concerned the materials used. Although Chang et al (2006) described the low-frequency signal component of speech as completely uninformative (see also Cullington & Zeng, 2010), that claim is not entirely accurate. Even just the first harmonic can provide information about how to segment sentences into constituent words and about sentence prosody. That means recognition might be better facilitated for words presented in sentences, rather than in isolation. To examine that possibility, both sentences and words were used in the current study, and word recognition was compared across materials. Potential contributions of linguistic factors (lexical and syntactic) were measured. If recognition was more accurate for words presented in sentences than in isolation, and if that finding could be explained primarily by linguistic context effects, it would suggest that the low-frequency signal has its influence through the additional information about linguistic structure that it provides. However, if an advantage for recognition in sentences was observed that could not be explained by linguistic context effects, it would suggest that the low-frequency signal facilitates perceptual organization and that organization is more readily achieved with longer stretches of signal. Calling on visual analogies again, it is easy to appreciate that a critical amount of sensory information is required to achieve perceptual organization. In the example of Ruben's vase, it would be difficult to recognize either form (the vase or the faces) without a sufficiently broad perspective. Outcomes of this manipulation involving materials should have clinical implications. Regardless of the basis, if a greater advantage of adding the low-frequency signal to the implant-simulated signal was found for sentences than for words, it would suggest that clinical tests should make use of sentence materials to evaluate potential benefits to individual patients of electric-acoustic stimulation.

Finally, the fourth issue addressed in the current study was whether children would benefit more from the addition of the low-frequency acoustic signal than adults. The prediction going into data collection was that children likely would benefit more, and that followed from the fact that children are not as sensitive to linguistic structure in general as are adults, either at the level of morphosyntactic structure (Chomsky, 1969) or word-internal phonemic structure (Liberman et al, 1974; Walley et al, 1986). In particular, children are still learning about structure at supra-lexical levels through the first decade of life. That suggests that they might be especially attentive to signal structure that can help with dividing the ongoing speech signal into constituent words.

In summary, the study reported here was designed to evaluate whether the addition of a very low-frequency (i.e. < 0.25 kHz) signal to an implant simulation (using noise vocoding) would improve speech recognition. In all, four hypotheses were tested:

1. Adding the low-frequency component of the speech signal to an implant-simulated signal would improve speech recognition.
2. The advantage would be greater in magnitude when both signal components were presented diotically, rather than dichotically.
3. The advantage would be greater in magnitude for words presented in sentences rather than in isolation.
4. Children would demonstrate a greater advantage than adults.

The outcomes of this work should have significant clinical implications by facilitating our collective ability to answer the *who*, *what*, *when*, and *why* questions surrounding the electric-acoustic stimulation advantage observed for patients with hearing loss: Who benefits? What is it that they gain? When—or under what conditions—do they obtain the advantages? And why are those benefits observed?

Preliminary experiment

Before the main experiment was conducted, a preliminary experiment was done to see if there was any evidence of a benefit when the low-frequency signal component was combined with the vocoded signal. In this experiment, only the conditions predicted to best facilitate perceptual organization with the addition of a low-frequency signal were included. This meant that stimulus materials were sentences, and presentation mode was diotic only. Outcomes of this preliminary experiment were subsequently used to establish reliability for the dependent measures used in the main experiment.

Method

Participants

Previous investigations using noise-vocoded sentences have shown large differences in recognition for adults and children. In particular, Nittrouer et al (2009) presented similar sentences to those used in this study. For four-channel vocoded sentences, the Cohen's *d* indexing the magnitude of the difference between adults' and children's scores was 1.47. With that difference as a precedent, it was decided that 20 participants per group in the current study would provide adequate power for detecting group differences, with an alpha of .05.

Listeners were recruited through the distribution of flyers to pupils in local public schools and to university students. In all, sixty listeners participated in this experiment: 20 adults between the ages of 18 and 39, 20 seven-year-olds (ranging from 7 years; 0 months to 7 years; 11 months) and 20 five-year-olds (ranging from 5 years; 0 months to 5 years; 11 months). All listeners were native speakers of American English, and all passed hearing screenings at 25 dB hearing level for the frequencies 0.5, 1, 2, 4, and 6 kHz. All listeners had histories of normal speech and language skills.

Equipment

All sentence materials were recorded in a sound booth, directly onto the computer hard drive, via an AKG C535 EB microphone, a Shure M268 amplifier, and a Creative Laboratories Soundblaster soundcard. Perceptual testing took place in a sound booth, with the computer that controlled the experiment in an adjacent room. Stimuli were stored on a computer and presented through a Samson headphone amplifier and AKG-K141 headphones. The hearing screening was done with a Welch Allyn TM262 audiometer and TDH-39 headphones.

Stimuli

Fifty-six four-word sentences (six for practice, 50 for testing) were created, according to principles used previously (e.g. Boothroyd & Nittrouer, 1988; Nittrouer et al, 2009; Stelmachowicz et al, 2000). These sentences are comprised entirely of monosyllabic content words and are syntactically appropriate. However, they are semantically anomalous. These sentences, which are listed in the Supplementary material to be found online at http://informahealthcare.com/doi/abs/10.3109/14992027.2013.871649, provide lexical and syntactic constraints, but no semantic constraints. Using such sentences restricts the extent to which top-down linguistic factors can influence speech recognition, thus allowing a more sensitive examination of the role of perceptual organization. They were recorded by an adult male speaker of American English at a 44.1-kHz sampling rate with 16-bit digitization. The top panel of Figure 1 shows a waveform of the original sentence *Hot slugs pick boats*, and the next panel shows a spectrogram of that sentence made with a 0.05-kHz analysing filter.

To create the vocoded stimuli, the same MATLAB routine was used as in previous experiments (e.g. Nittrouer & Lowenstein, 2010, Nittrouer et al, 2009). All signals were first band-pass filtered with a low-frequency cut-off of 0.25 kHz and a high-frequency cut-off of 8 kHz. Cutoff frequencies between channels for vocoding were 0.8, 1.6, and 3.2 kHz. All filtering was done with digital filters that had greater than 50-dB attenuation in stop bands, and had 0.01-kHz transition bands between pass- and stop-bands. Each channel was half-wave rectified and filtered using a 0.16-kHz high-frequency cut-off. The temporal envelopes derived for separate channels were subsequently used to modulate white noise, limited to the same channels as those used to divide the speech signal. The resulting bands of amplitude-modulated noise were combined to create vocoded sentences. Root-mean-square amplitude was equalized across sentences. These stimuli consisting of only the noise-vocoded signals are described here as the VOC-only stimuli. The middle spectrogram of Figure 1 is from the VOC-only version of the sentence shown in the top spectrogram.

Low-frequency signals were also created for each sentence using MATLAB. To do that, the original sentences were low-pass filtered

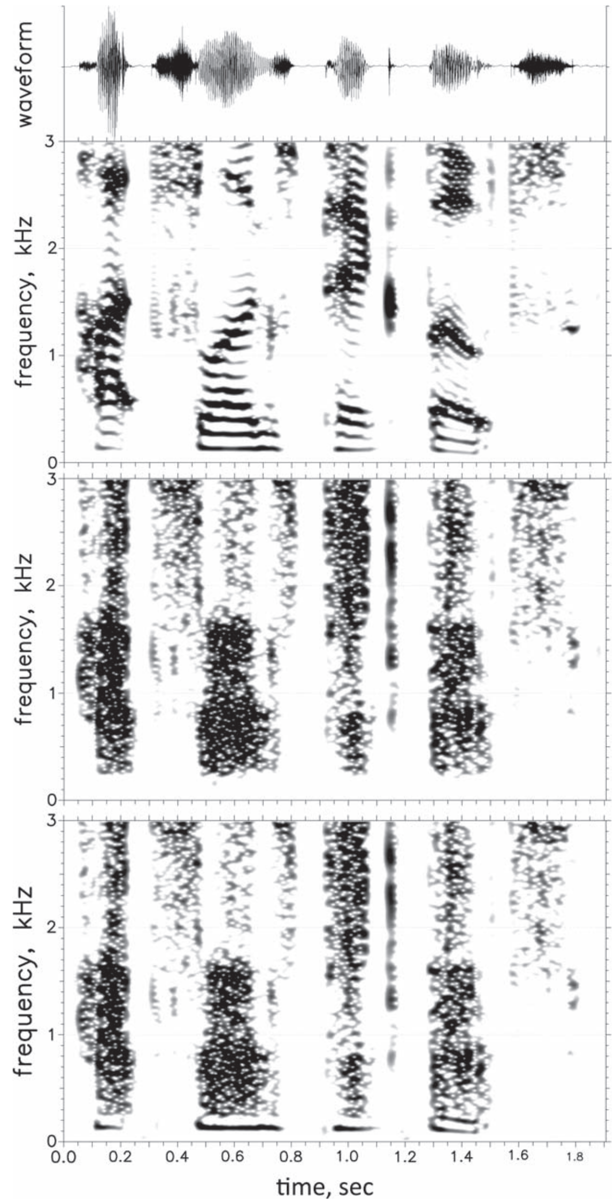

Figure 1. Waveform and three spectrograms (created with a 0.05-kHz analysing filter) of the sentence *Hot slugs pick boats:* Unprocessed sample (top); VOC-only sample (middle); and LOW-plus sample (bottom).

with a high-frequency cut-off of 0.25 kHz, using the same type of digital filter described above. The root-mean-square amplitude of each low-pass filtered signal was adjusted to equal that of the matching vocoded signal, and those signals were combined to create the stimuli termed LOW-plus in this experiment. The bottom spectrogram of Figure 1 is of the LOW-plus version of the sentence shown in the spectrograms above. Unlike the VOC-only version, the first harmonic can be seen across most of this stimulus. It is clear how this property marks lexical units. For the last word, the lowest two harmonics are seen, reflecting the lowered fundamental frequency that accompanies sentence-final declination.

Procedures

All procedures were approved by the Institutional Review Board of the Ohio State University. After participants (along with their parents, in the case of children) signed the consent form, the hearing screening was administered.

Stimuli for testing were presented under headphones at 68 dB sound pressure level, and all presentation was done diotically. The task consisted of repeating the sentences heard. Prior to testing, practice was conducted with the six practice sentences, with VOC-only processing. For each practice sentence, the natural version was played first, and the listener was asked to repeat it. Then the listener was told that a robot was also recorded saying each sentence. The VOC-only version was presented, and the listener was asked to repeat that version, as well.

During testing, the order of presentation of the sentences was randomized independently for each listener. Listeners heard half the stimuli as VOC-only and half as LOW-plus. The selection of sentences to be presented with each type of processing was randomly made for each listener by the software, and presentation of these two kinds of stimuli was intermingled during testing with the stipulation that no more than two VOC-only or two LOW-plus stimuli could be presented in a row. Each sentence was played once, and the listener repeated it as best as possible. The number of incorrect words for each sentence was entered into the program interface by the examiner.

After hearing all 50 test sentences in their processed forms (either VOC-only or LOW-plus), listeners heard all sentences in their unprocessed forms. Listeners could get no more than 10% of the words wrong on this task to have their data entered in the final analysis. All responses were scored by the last author, who is a trained linguist.

Analyses

The percentage of words recognized correctly in each processing condition was used as the primary dependent measure. Data were screened for normal distributions and homogeneity of variance prior to conducting any statistical tests. For inferential tests, arcsine transformations were applied. A significance level of .05 was applied. Nonetheless, in reporting outcomes, precise significance levels are given when $p < .10$; for $p > .10$, outcomes are reported simply as not significant.

In order to examine the contribution to recognition made by linguistic context effects (which for these sentences primarily meant syntactic structure) j factors were computed using the formula described by Boothroyd and Nittrouer (1988). This metric derives from the fact that the probability of recognizing words in a sentence is dependent on the probabilities of recognizing the separate words that comprise the sentence. If linguistic context played no role in recognition, then the probability of recognizing a whole sentence correctly would be directly dependent upon the probability of recognizing each word separately such that:

$$p_s = p^n_w \tag{1}$$

where p_s is the probability of recognizing the complete sentence, p_w is the probability of recognizing each word, and n is the number of words in the sentence. However, linguistic context does contribute to recognition, so the relationship above does not strictly apply. A more useful way of viewing the relationship is that some number of independent channels of information is required for recognition, and the greater the contribution of linguistic context, the fewer channels that are required. Thus, equation (1) goes to

$$p_s = p^j_w \tag{2}$$

where j is the number of independent channels of information, and is between 1 and n. We now have a way of solving for the effective number of information channels in the sentence:

$$j = \log(p_s)/\log(p_w) \tag{3}$$

In this formulation, the independent channels indexed by j are not appropriately viewed as actual words. Rather, j is a dimensionless factor that serves as an index of how strongly sentence context influences recognition. The smaller j is, the greater the effect of sentence context on recognition.

Results

All listeners were able to recognize all the words in the unprocessed sentences with better than 90% accuracy, so data from all listeners were included in analyses.

Figure 2 shows mean correct word recognition for each group in the top panel. A two-way, repeated-measures analysis of variance (ANOVA) was performed, with age as the between-subjects factor, and processing as the within-subjects factor. The main effect of age was significant, $F\ (2,57) = 111.77$, $p < .001$, $\eta^2 = .797$, as were all *post hoc* contrasts among age groups ($p < .001$ for all contrasts using a Bonferroni adjustment for multiple contrasts). The main effect of processing (VOC-only or LOW-plus) was also significant, $F\ (1,57) = 58.55$, $p < .001$, $\eta^2 = .507$. The Age × Processing interaction was not significant. These results indicate that participants were better at recognizing words for the LOW-plus stimuli than for the VOC-only stimuli, and that recognition generally improved with increasing age.

Context effects

To examine the contributions of linguistic context effects, j factors were computed for individual listeners using word and sentence recognition scores. However, these factors can only be computed when both word and sentence recognition is between 5% and 95% correct. Recognition scores for whole sentences are shown on the bottom of Figure 2, and indicate that scores of better than 5% correct were not attained in either processing condition for many listeners. In fact, it was only for the LOW-plus stimuli that a majority of adults and seven-year-olds had better than 5% correct sentence recognition scores. Most five-year-olds did not reach this criterion for either processing condition. Thus, j factors were calculated only for the 17 adults and 11 seven-year-olds who met this criterion, and only for the LOW-plus stimuli. Mean j factors and standard deviations (SDs) were 3.00 (.60) for adults

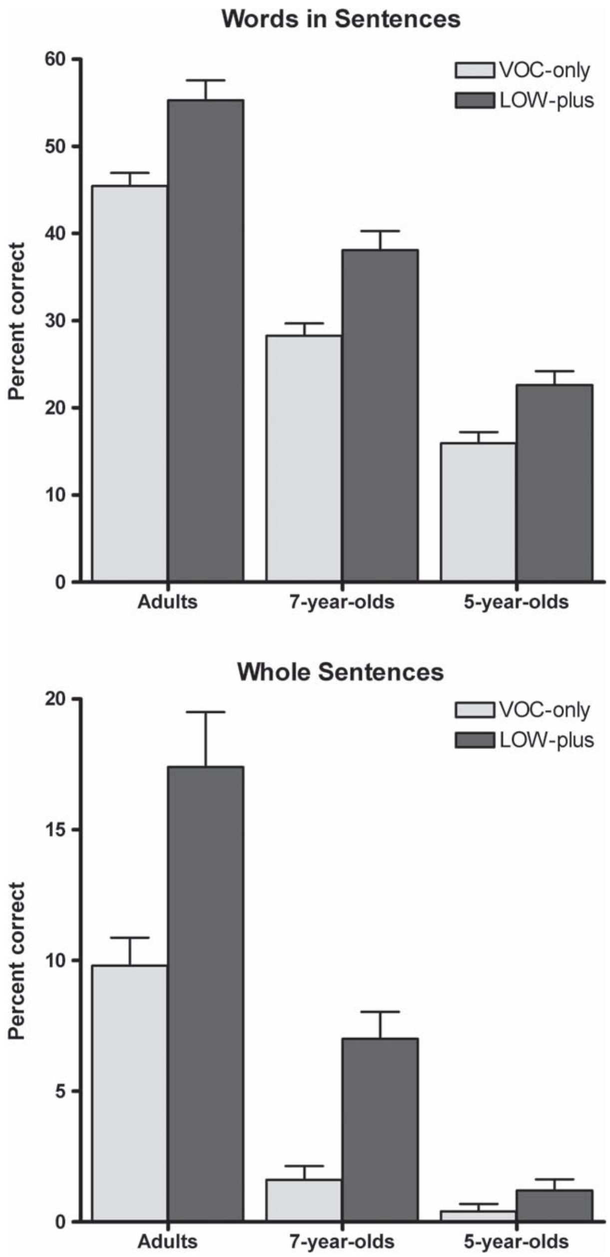

Figure 2. Mean recognition scores for words in sentences (top) and whole sentences (bottom) for adults, seven-year-olds, and five-year-olds in the preliminary experiment. Error bars are standard errors of the means.

and 2.66 (.34) for seven-year-olds. A *t* test performed on these individual *j* factors did not reveal a significant age effect, so it may be concluded that listeners in both age groups used sentence (syntactic) context to a comparable extent. Although *j* factors could not be computed for five-year-olds in the current experiment, Nittrouer and Boothroyd (1990) were able to compute these factors for children between 4 and 6 years of age in a study on speech recognition in noise, and found similar values for them as for adults. Returning to this experiment, the finding that listeners across age groups apparently made use of syntactic context to the same extent means that the significant age effect found for word scores must be explained by some other aspect of perception.

Magnitude of the low-frequency effect across recognition probabilities

A final problem addressed in this preliminary study was that mean recognition probabilities differed across listener groups. That situation gives rise to the question of whether a difference between the two processing conditions that is consistent in absolute size across groups represents an equivalent effect of adding the low-frequency signal component. In order to normalize for recognition probabilities, a metric of effect size was computed using the formula:

$$\text{EFFECT} = (p_{\text{LOW-plus}} - p_{\text{VOC-only}})/p_{\text{VOC-only}} \tag{4}$$

where $p_{\text{LOW-plus}}$ is the recognition probability for the LOW-plus condition and $p_{\text{VOC-only}}$ is the recognition probability for the VOC-only condition.

Mean EFFECT scores (and SDs) were .53 (.52) for five-year-olds, .39 (.38) for seven-year-olds, and .23 (.23) for adults. However, in spite of the appearance of a developmental trend to smaller EFFECT scores with increasing age, a one-way ANOVA performed on these scores was not significant for age, $F\ (2,57) = 2.84$, $p =\ .067$.

Next, Pearson product-moment correlation coefficients were computed between *j* factors and the EFFECT scores for the 28 participants for whom *j* factors were computed. This analysis permitted examination of whether the magnitude of the low-frequency effect could be explained to any significant extent by participants' abilities to use top-down linguistic context. However, the obtained coefficient of .21 was not significant.

Discussion

This preliminary experiment was undertaken to see if a benefit to speech recognition would be observed when a low-frequency signal is combined with a spectrally degraded signal for adults or children, in conditions predicted to strongly facilitate such an effect. Results showed that under these conditions, differences of roughly 10% in absolute recognition scores were found for all age groups. That outcome matches what others have found for patients with either hybrid implants or bimodal electric-acoustic stimulation (Hamzavi et al, 2004; Incerti et al, 2013). When given as proportions, effects on the order of 20% to 50% improvement were observed for combination signals over spectrally degraded signals alone.

The contributions of top-down linguistic context were also examined in this preliminary study. These effects were found to be similar in magnitude for adults and seven-year-olds. That outcome matches what has been observed in the past (Nittrouer & Boothroyd, 1990; Nittrouer et al, 2009): At least when the linguistic structure at stake is simple, as the syntax was for these four-word sentences, children are perfectly capable of applying that structure to aid recognition. In spite of showing similar linguistic context effects, however, the magnitude of the low-frequency effect was greater for children than for adults, when the difference score was given as a proportion of the recognition probability for the VOC-only stimuli.

All outcomes of this preliminary study suggested that the main experiment that was designed would be productive. Consequently, that experiment was conducted, with an expanded design.

Main experiment

This experiment was undertaken to examine the four issues described in the Introduction. In order to do this, the experimental conditions used in the preliminary experiment were expanded. In this main experiment, a dichotic configuration was used, as well as the diotic configuration of the preliminary experiment. Lists of isolated words were presented, as well as the simple four-word sentences used in the preliminary experiment. Adults and seven-year-olds participated in this experiment; five-year-olds were not included because the expanded protocol made it too demanding for children so young. In particular, listening to noise-vocoded versions of isolated words was found to be too hard for these children, creating a situation in which any data that were obtained from them were deemed to be unreliable.

Method

Participants

Forty new listeners were tested in this experiment: 20 adults between the ages of 18 and 38 years, and 20 children between 7 years; 0 months and 7 years; 8 months. All children were free from significant histories of otitis media, defined as more than five episodes before the age of three years. All participants (or in the case of children, their parents on their behalf) reported having normal hearing, speech and language. All participants passed hearing screenings of the frequencies of 0.5, 1, 2, 4, and 6 kHz presented at 25 dB hearing level to each ear separately.

Even though participants all reported normal speech and language, age-appropriate screenings were administered because it was considered especially important to ensure that all participants had language abilities within the normal range, in order to minimize concern that variability in either recognition probabilities or in the magnitude of the low-frequency effect might be due to differences in language skills. Adults were given the reading subtest of the wide range achievement test 4 (WRAT; Wilkinson & Robertson, 2006) because the ability to read draws on multiple language skills for adults, so serves as an index of general language ability. All demonstrated at least a 12th-grade reading level. Children were given the Goldman Fristoe 2 test of articulation (Goldman & Fristoe, 2000) and were required to score at or better than the 30th percentile for their age in order to participate. Their performance ranged from 0 to 3 errors on this measure, which resulted in a mean ranking of the 54th percentile (SD = 8). The lowest-scoring child was at the 32nd percentile.

All listeners were given the expressive one-word picture vocabulary test - 4th edition (EOWPVT; Martin & Brownell, 2011), and were required to achieve a standard score of at least 92 (30th percentile) for their age. The authors of this test have demonstrated that scores on this instrument correlate well ($r = .77$) with broader measures of language ability, so this test served as a criterion to ensure that language abilities for all participants were within the normal range. In addition, this vocabulary measure was used as a predictor variable to see if the magnitude of the low-frequency effect could be explained by language abilities. The mean EOWPVT standard score for adults was 102 (SD = 9), which corresponds to the 55th percentile. The mean EOWPVT standard score for children was 111 (SD = 10), which corresponds to the 77th percentile. These scores indicate that the adult listeners had expressive vocabularies that were just above the mean of the normative sample used by the authors of the EOWPVT, and children had expressive vocabularies closer to one SD above the normative mean.

Equipment

The same equipment was used as in the preliminary experiment. For this experiment, all test sessions were video-recorded using a Sony HDR-XR550V video recorder so that scoring could be done later. Participants wore Sony FM microphones that transmitted speech signals directly to the line input of the camera in order to ensure good sound quality for all recordings.

Stimuli

Forty-eight of the 50 sentences from the preliminary experiment were used, as well as 16 word lists from Mackersie et al. (2001). Each word list consisted of 10 phonetically balanced CVC words. These words were recorded by an adult male speaker of American English at a 44.1-kHz sampling rate with 16-bit digitization, just as the sentences had been. The same signal processing methods as those employed in the preliminary experiment were used to create noise-vocoded and low-frequency signals for all stimuli. These signals were then used to create VOC-only and LOW-plus stimuli (both words and sentences) that could be presented in both a diotic and dichotic manner. In the diotic configuration, the same signal was presented to both ears, regardless of whether it was the VOC-only or the LOW-plus stimuli. This configuration simulated bilateral cochlear implants in the diotic VOC-only condition (VOC-only signals to both ears) and bilateral hybrid implants in the diotic LOW-plus condition (combined vocoded and low-frequency signals to both ears). In the dichotic configuration, each ear was presented with a different signal. This configuration simulated a unilateral cochlear implant in the dichotic VOC-only condition (a VOC-only signal to just one ear) and a unilateral cochlear implant with a contralateral hearing aid in the dichotic LOW-plus condition (a VOC-only signal to one ear and the low-frequency signal to the other ear). Half of the listeners in each group heard the VOC-only signals in their right ears in the dichotic condition; the other half heard those signals in their left ears. Root-mean-square amplitude of the vocoded and low-frequency signals was equalized in all LOW-plus conditions.

Four word lists (40 words total) were randomly selected for presentation in each of the four stimulus conditions (VOC-only diotic, LOW-plus diotic, VOC-only dichotic, LOW-plus dichotic) for each participant. Similarly, twelve sentences were randomly selected for each of the four stimulus conditions, for each participant.

Procedures

Participants completed the experiment in a single listening session. The hearing screening was completed at the beginning of the listening session.

There were four blocks in the experiment based on configuration and materials: diotic words, dichotic words, diotic sentences, and dichotic sentences. Equal numbers of stimuli of each processing type (VOC-only and LOW-plus) were presented in each block. For the word materials, VOC-only and LOW-plus stimuli alternated between each ten-word list. For the sentences, VOC-only and LOW-plus stimuli were mixed, with the only rule being that no more than two VOC-only or LOW-plus stimuli could be presented in a row. Thus each word block consisted of eight-word lists (4 VOC-only and 4 LOW-plus), and each sentence block consisted of 24 sentences (12 VOC-only and 12 LOW-plus).

The kind of stimuli presented in the first block was randomly selected from the four possible types (diotic words, dichotic words, diotic sentences and dichotic sentences), with the stipulation that each starting condition was evenly distributed across subjects in each age group. Block presentation alternated between sentences and words. The fourth block was the same configuration (diotic or dichotic) as the first block. The second and third blocks were the alternative configuration. Thus, an example of condition order was: diotic words, dichotic sentences, dichotic words, diotic sentences.

Each block of testing was preceded by a set of training stimuli. For the blocks consisting of words, the listener first heard and repeated five unprocessed words, and then 10 processed words (5 VOC-only and 5 LOW-plus), with the software randomly selecting whether VOC-only or LOW-plus stimuli were presented first. For blocks consisting of sentences, the listener first heard and repeated two unprocessed sentences, and then two processed sentences, one VOC-only and one LOW-plus.

During testing, each word or sentence was played once, and the listener repeated what was heard. Children moved a game piece along a four-space game board after each block. This procedure provided some reinforcement, and served as a visible indicator of progress.

After testing was completed with the four blocks of test stimuli, the two screening tasks were administered: WRAT and EOWPVT for adults; Goldman-Fristoe test and EOWPVT for seven-year-olds. For this main experiment, a post test consisting of unprocessed materials was not used, both because in the preliminary experiment adults and seven-year-olds performed so close to 100% correct (mean scores of 99% and 96% correct, respectively) and because it would have required adding a second session to the protocol. There seemed no need to do so.

Scoring and analyses

Four separate measures were obtained. Dependent measures for the word materials were the percent of phonemes and whole words recognized correctly. Dependent measures for the sentences were the percent of words and whole sentences recognized correctly.

For this experiment, all responses were scored by the last author. In addition, the second and third authors each scored a different half of the children's responses and 25% of the adults' responses. Pearson product-moment correlation coefficients were computed between scores of the last author and each of the other scorers as measures of inter-rater reliability. This procedure was done for data from seven-year-olds and adults separately. As a metric of inter-subject reliability, both word and whole-sentence scores for the VOC-only and LOW-plus sentences in the diotic condition were compared to those from the preliminary experiment. These are the two conditions that the two experiments had in common.

All statistical analyses were performed similarly to those of the preliminary experiment. However, because word lists were used as stimuli in this experiment, *j* factors were computed for these materials, as well as for sentences. For words, context effects mainly represented the influence of lexicality on recognition, and the appropriate formula is:

$$j = \log(p_w)/\log(p_p), \qquad (5)$$

where p_w is the probability of correct word recognition, p_p is the probability of correct phoneme recognition, and j is the number of independent channels of information required for word recognition. Again, precise significance levels are provided when $p < .10$; for $p > .10$, outcomes are reported simply as not significant.

Results

Inter-rater reliability was between .94 and .99 across conditions, for data from adults and children. This was considered sufficient, and scores of the last author were used in analyses.

Figure 3 shows recognition scores obtained for words, with phoneme scores on the top and whole-word scores on the bottom. Figure 4 shows recognition scores obtained for sentences, with word scores on the top and whole-sentence scores on the bottom. When *t* tests were performed on scores from this experiment and those from comparable conditions in the preliminary experiment, for adults and children separately, none of the tests resulted in statistical significance. Thus it was concluded there was adequate inter-subject reliability.

The next question addressed was whether differences were found in recognition probabilities dependent upon which ear heard the VOC-only signal in the dichotic configuration. To answer this question, *t* tests were done for adults and children separately, with groups defined by which ear was presented with the VOC-only stimuli. Tests were done on all four measures obtained, for both the VOC-only and LOW-plus processing conditions, making a total of eight *t* tests per age group. For seven-year-olds, none of these *t* tests resulted in statistically significant outcomes, so it was concluded there was no effect of which ear heard the VOC-only signal in the dichotic configuration for children. For adults, none of the tests for sentence materials resulted in a statistically significant ear effect. Neither did the tests involving LOW-plus stimuli for word materials. However, for the VOC-only stimuli involving isolated words, statistically significant differences in recognition were found based on which ear received the VOC-only signal: for phonemes in words, $t\,(18) = 3.28$, $p = .004$; for whole words, $t\,(18) = 2.28$, $p = .035$. Mean recognition scores are shown in Table 1, and reveal a clear right-ear advantage. Nonetheless, scores in the dichotic configuration were collapsed across listeners, regardless of which ear heard the VOC-only signal because none of the other scores showed an ear advantage for these listeners, and none was found for any scores for seven-year-olds.

Separate analyses on dependent measures

Separate analyses were performed on each of the four dependent measures: phonemes in words, whole words, words in sentences, and whole sentences. These are the scores shown in Figures 3 and 4. For each score, a three-way, repeated-measures ANOVA was performed, with age as the between-subjects factor and processing and configuration as within-subjects factors. Results of the analyses for word materials are shown in Table 2, along with effect sizes in the form of η^2. The primary findings for these word materials were that adults performed better than seven-year-olds, scores were better in the LOW-plus processing condition than in the VOC-only processing condition, but the main effect of configuration was not significant. None of the two-way interactions reached statistical significance, but there was a significant three-way interaction for both scores (phoneme and whole word). Examination of Figure 3 reveals that it was due to adults showing a larger benefit of the low-frequency signal in the dichotic than the diotic configuration, and children showing the opposite pattern: a stronger benefit of the low-frequency signal in the diotic than the dichotic configuration. However, effect sizes were small for this three-way interaction for both scores, so it is difficult to make much of this finding.

Table 3 shows the outcomes of statistical analyses for sentence materials. The same pattern is seen here as for word materials: primarily the only effects that were significant were those of age and processing. The one exception is a significant Age × Processing interaction for whole sentences. Again, however, the effect size is small.

Words versus sentences: The materials' effect

In the analyses above, separate statistics were performed on each of the four measures, prohibiting investigation of potential materials'

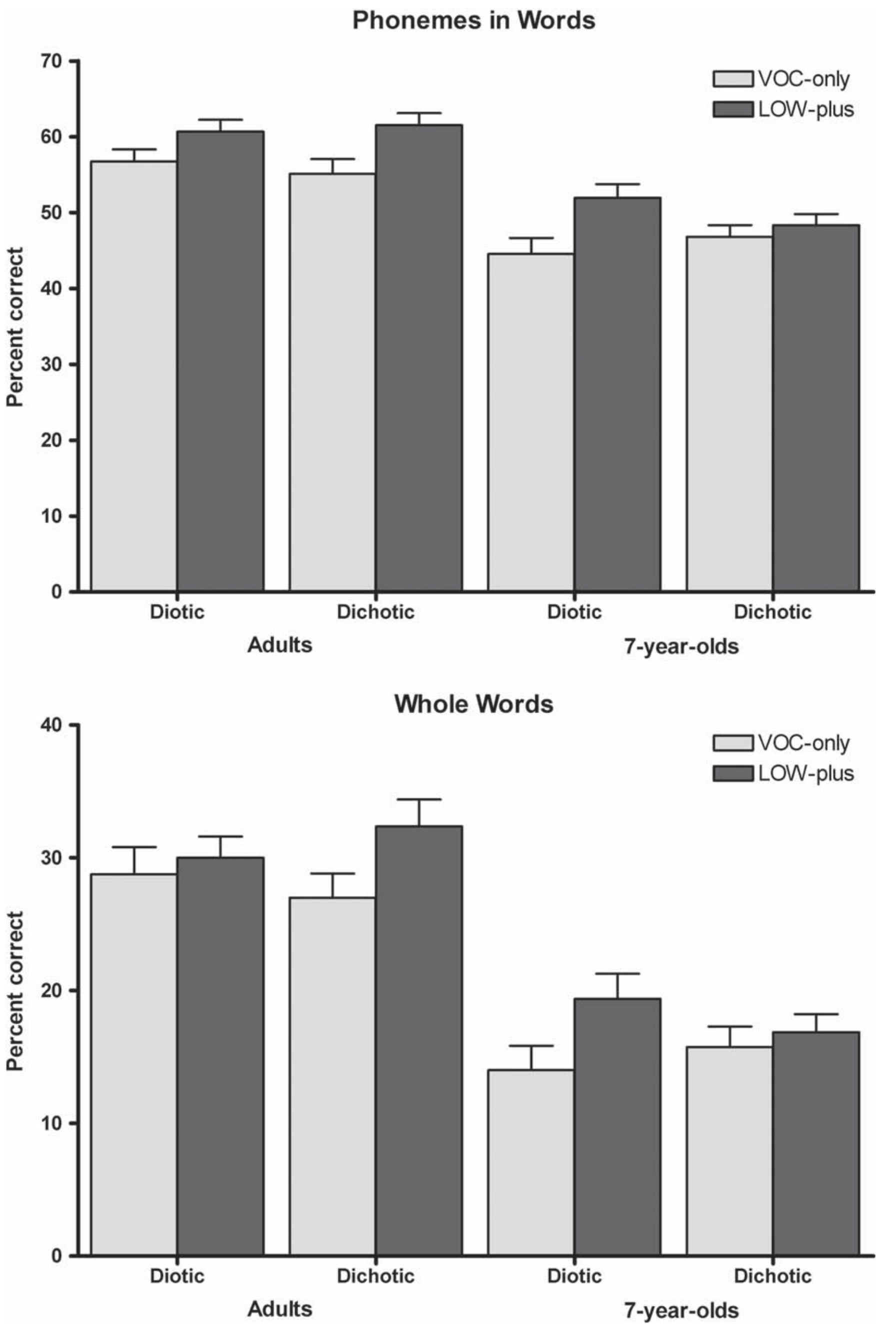

Figure 3. Mean recognition scores for phonemes in words (top) and whole words (bottom) for adults and seven-year-olds in the main experiment. Error bars are standard errors of the means.

effects. To obtain a test of the materials' effect, a four-way ANOVA was performed on word scores: percent correct recognition of whole words (bottom of Figure 3) and of words in sentences (top of Figure 4). Although the same words were not used in both kinds of speech materials, all words were monosyllabic. Results of this analysis revealed significant main effects of age, F (1,38) = 99.42, $p < .001$, $\eta^2 = .723$, and processing, F (1,38) = 67.95, $p < .001$, $\eta^2 = .641$, matching what had been found for the separate analyses. And again, the main effect of configuration was not statistically significant. Of primary interest, the main effect of materials was significant, F (1,38) = 159.49, $p < .001$, $\eta^2 = .808$, indicating that words were more readily recognized in a sentence context than in isolation. The only two-way interaction that was significant was the Materials × Processing interaction, F (1,38) = 11.77, $p =$ .001, $\eta^2 = .236$. Examination of Figures 3 and 4 reveals that this outcome was due to the low-frequency effect being greater when words were presented in sentences, rather than in isolation. Finally, the three-way interaction of Age × Processing × Configuration was significant, F (1,38) = 6.27, $p =$ .017, $\eta^2 = .142$. That is the same three-way interaction found to be significant when words in isolation were analysed separately, and that earlier outcome appears to account entirely for the significant interaction found here.

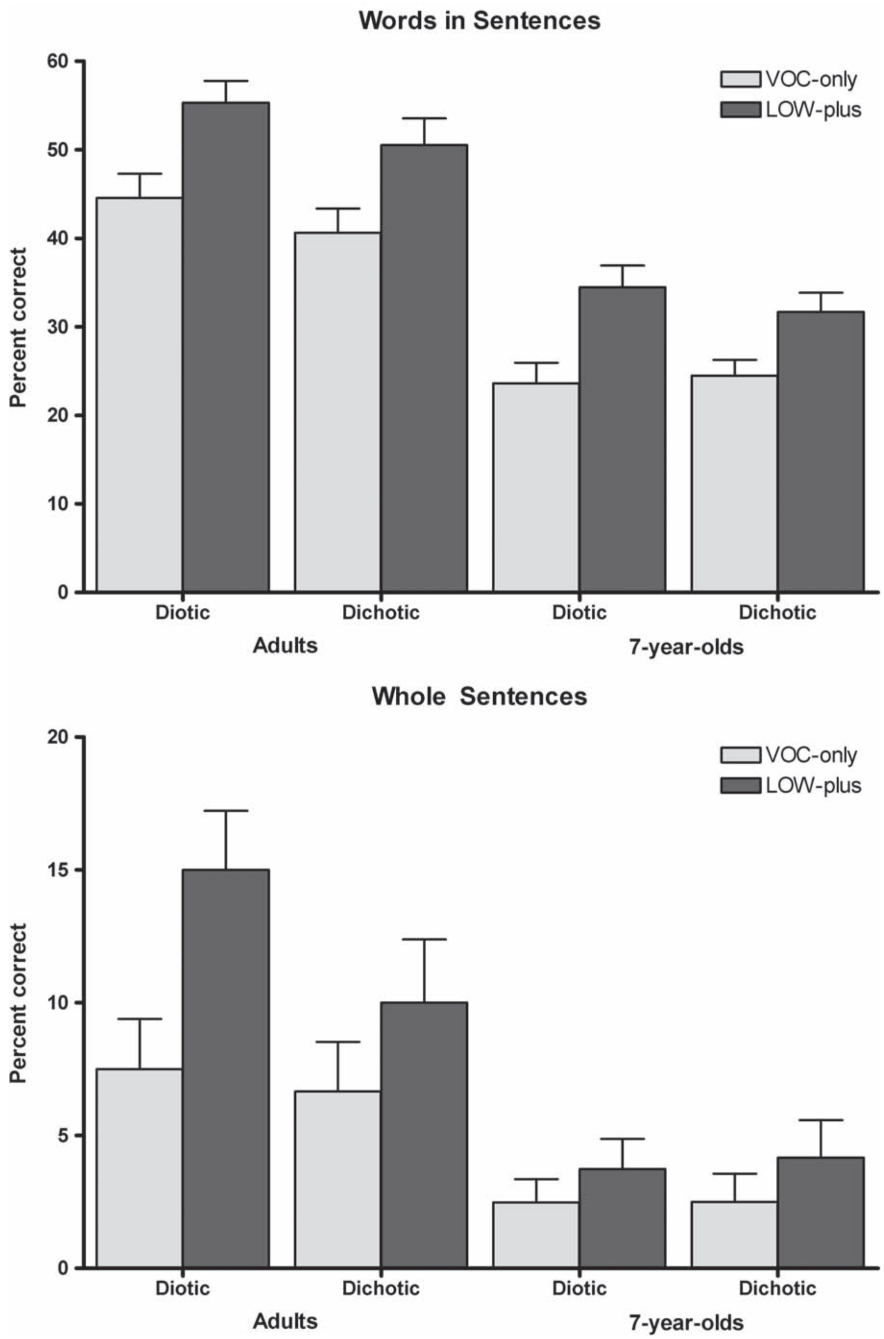

Figure 4. Mean recognition scores for words in sentences (top) and whole sentences (bottom) for adults and seven-year-olds in the main experiment. Error bars are standard errors of the means.

Table 1. Mean recognition probabilities (and SDs) for adults in the main experiment, for VOC-only, isolated word stimuli presented in the dichotic configuration. *Ear* indicates which ear heard the VOC-only signal.

	Right ear	*Left ear*
Phonemes in words	60.33 (6.28)	49.93 (7.82)
Whole words	30.75 (6.88)	23.25 (7.82)

Context effects

For words, j factors could be computed for adults and seven-year-olds for all four stimulus conditions: VOC-only diotic, LOW-plus diotic, VOC-only dichotic, LOW-plus dichotic. These values are shown in Table 4, with the numbers of participants in each group who had both phoneme and whole-word scores above 5% given. When t tests were performed on each j factor, only the test for the VOC-only, diotic condition (first column) was found to be significant, $t\ (36) = 2.22$, $p = .033$. That result indicated that adults

Table 2. Statistical outcomes of three-way ANOVAs for word materials. Degrees of freedom = 1,38.

	F	*p*	η^2
Phonemes in words			
Age	31.05	<.001	.450
Processing	44.26	.001	.538
Configuration	0.15	NS	
Age × Processing	0.24	NS	
Age × Configuration	0.01	NS	
Processing × Configuration	1.67	NS	
Age × Processing × Configuration	9.68	.004	.203
Whole words			
Age	54.56	<.001	.589
Processing	13.28	.001	.259
Configuration	0.02	NS	
Age × Processing	0.31	NS	
Age × Configuration	0.001	NS	
Processing × Configuration	0.17	NS	
Age × Processing × Configuration	5.38	.026	.124

used lexical context to a greater extent than seven-year-olds in the recognition of these words. Nonetheless, because none of the other tests were significant it must be concluded that adults and seven-year-olds showed similar context effects for word recognition across conditions.

For sentence materials, whole-sentence recognition was somewhat poorer for seven-year-olds in this main experiment than in the preliminary experiment, even though those differences across experiments did not reach statistical significance. In particular, most seven-year-olds scored less than 5% correct on whole-sentence recognition, so *j* factors could not be computed. For adults, however, *j* factors could be computed for more than half of the group for sentence scores, and mean values were similar to those obtained in the preliminary experiment: For adults in this main experiment, the four *j* factors obtained for sentence materials ranged from 2.97 to 3.41, which matches the *j* factor of 3.00 obtained in the preliminary experiment. Based on the finding in the preliminary experiment of no age effect, and the fact that no significant age effects have been observed for similar materials in previous studies

Table 3. Statistical outcomes of three-way ANOVAs for sentence materials. Degrees of freedom = 1,38.

	F	*p*	η^2
Words in sentences			
Age	74.18	<.001	.664
Processing	66.02	<.001	.635
Configuration	1.28	NS	
Age × Processing	0.03	NS	
Age × Configuration	0.67	NS	
Processing × Configuration	1.12	NS	
Age × Processing × Configuration	0.56	NS	
Whole sentences			
Age	24.91	<.001	.396
Processing	20.60	<.001	.352
Configuration	0.85	NS	
Age × Processing	4.10	.050	.097
Age × Configuration	0.79	NS	
Processing × Configuration	1.10	NS	
Age × Processing × Configuration	1.79	NS	

Table 4. Mean *j* factors (and SDs) computed on recognition probabilities for isolated words in the main experiment. Numbers in italics above each *j* factor indicate the numbers of participants contributing to the mean values.

	Diotic		*Dichotic*	
	VOC-only	*LOW-plus*	*VOC-only*	*LOW-plus*
Adults	*20*	*20*	*20*	*20*
	2.28 (0.42)	2.47 (0.29)	2.26 (0.26)	2.39 (0.34)
7-year-olds	*18*	*20*	*17*	*19*
	2.58 (0.40)	2.65 (0.50)	2.40 (0.34)	2.46 (0.25)

(e.g. Nittrouer & Boothroyd, 1990), it was considered unlikely that linguistic context effects differed in magnitude for adults and seven-year-olds in the current experiment.

Magnitude of the low-frequency effect across recognition probabilities

The EFFECT scores were computed as they had been in the preliminary experiment to see if the magnitude of the low-frequency effect differed for adults and children, once the overall difference in recognition probabilities was taken into account. In this main experiment, two kinds of materials were used, and stimuli were presented in two configurations. Thus, four EFFECT scores were computed, using recognition scores for whole words (for words in isolation) and words in sentences, and scores for the diotic and dichotic configurations. From these four scores, two summary EFFECT scores were computed for whole words and words in sentences by taking the means across the diotic and dichotic configurations. This process was considered appropriate because no effects of configuration had been observed.

Looking first at the EFFECT scores for whole words, means (and SDs) were .50 (.73) for seven-year-olds and .19 (.31) for adults. A *t* test performed on these scores failed to reveal a significant age effect, $t(38) = 1.79$, $p = .081$. For words in sentences, mean EFFECT scores were .57 (.79) for seven-year-olds and .29 (.21) for adults. This difference was not statistically significant. In both cases, the lack of significant finding could likely be attributed to the large standard deviations. Thus it was concluded that there was at least a trend of smaller low-frequency effects for older listeners, compared to younger, but it was not statistically significant.

Başkent and Chatterjee (2010) had reported the observation that for adults there was a trend towards greater low-frequency effects with poorer recognition for the VOC-only stimuli. In this study, seven-year-olds had lower recognition scores generally than adults. Consequently, it was possible that the observation of higher EFFECT scores for seven-year-olds than for adults, even though the difference was never significant, might reflect the trend

Table 5. Pearson product-moment correlation coefficients (and *p* values) for EFFECT scores (i.e. effect of combining the low-frequency signal with the VOC-only signal) and recognition probabilities for the VOC-only stimuli, for whole words and words in sentences. Numbers in parentheses indicate the numbers of participants.

	All listeners (40)	*Adults (20)*	*7-year-olds (20)*
Whole words	–.603 (<.001)	–.564 (.010)	–.727 (<.001)
Words in sentences	–.387 (.014)	–.443 (.050)	–.364 (NS)

reported by Başkent and Chatterjee. To examine that possibility, Pearson product-moment correlation coefficients were computed between the EFFECT scores and recognition probabilities for the VOC-only whole words and words in sentences. In order to make the comparison equivalent, mean VOC-only scores had to be computed across the diotic and dichotic configurations. These correlation coefficients are shown in Table 5, computed for all 40 listeners, and for each age group separately. For the whole words, there is a clear trend to larger effects of combining the low-frequency signal with the VOC-only stimuli, across groups and within group, as well. For sentence materials, there is an effect, but it is not as strong, nor as consistent. In particular, it does not reach statistical significance for the seven-year-olds. Nonetheless, it is possible that the enhanced low-frequency effect found for children, as compared to adults, may simply reflect the trend of stronger effects for listeners who are poorer at recognizing vocoded stimuli.

Explaining variance in the low-frequency effect

The overarching proposition tested by the work reported here was that the low-frequency component of speech could facilitate effective perceptual organization of the higher frequency speech components, when they are spectrally degraded. This situation generally matches that of patients with traditional cochlear implants, supplemented by hearing aids with extended low-frequency amplification. However, the possibility existed that the low-frequency signal component provided phonetically significant information in its own right. In that case, rather than this signal portion aiding perceptual organization, the correct perspective would be that each signal component is providing separate linguistic information that gets added together. Although difficult to design direct tests to decide between these two possibilities, language abilities were correlated with outcome measures in the current work as one way to try to address this issue. The rationale for this approach was the following: If linguistic factors explain either recognition per se with these spectrally degraded signals, or the advantage of combining the two signals, then participants with better language abilities should show greater benefit from adding the low-frequency signal because presumably they would be more sensitive to linguistically relevant structure. Accordingly, both raw and standard scores for the vocabulary test (EOWPVT) were correlated with each of ten measures: the eight recognition probabilities associated with word scores (i.e. whole words and words in sentences, for both VOC-only and LOW-plus stimuli, in both the diotic and dichotic configuration), and the two EFFECT scores (whole words and words in sentences). These Pearson product-moment correlation coefficients were computed for all listeners together, and for each group separately. Not one of these 60 correlation coefficients was statistically significant, which means that language abilities (as indexed by vocabulary skill) did not explain any variability in word recognition per se, or in the magnitude of the low-frequency effect. Thus no support was found for the proposition that the low-frequency signal has its effect by adding more linguistic information.

Discussion

Cochlear implants have significantly improved the comprehension of spoken language for patients with severe-to-profound hearing loss. However, these devices have not completely solved the communication problems faced by these individuals. Although the best-performing patients do quite well with implants, many continue to struggle with speech recognition. As a consequence, research efforts have continued to investigate options for more effective implants and signal processing strategies for those implants. Another approach that has been taken to try to improve speech recognition is to combine amplified acoustic signals with the electric signals received through cochlear implants. Typically these efforts have focused on patients with substantial residual hearing (e.g. 60 dB hearing level or better in the frequencies of 0.5 kHz to possibly as high as 1 kHz). Amplifying the spectral components of the speech signal in these frequencies means that linguistically significant information is being provided by that acoustic signal. Consequently, it is reasonable that access to this signal structure—with the spectrally clear signals available through acoustic hearing—would improve speech recognition.

The work reported in the current study took a different perspective on the potential advantage of combining a low-frequency acoustic signal with a spectrally degraded electric signal. In this work, the possibility was tested that even a very low-frequency signal (defined as that signal portion below 0.25 kHz) would have beneficial effects on the recognition of spectrally degraded speech. The overarching hypothesis tested in this work was that the low-frequency signal portion would facilitate better perceptual organization of the spectrally degraded speech signal. This phenomenon was proposed to be independent of linguistic context effects on speech recognition. That very low-frequency signal provides some linguistic information, mostly about voicing in the case of single words, and about word segmentation and sentence prosody in the case of sentences. Nonetheless, the major contribution of this signal component was hypothesized to be that it adds a specifically speech-like quality to the sensory information reaching the listener, which in turn promotes the kind of perceptual organization that supports the recovery of speech-like form.

In total, four specific hypotheses were tested:

1. Adding the very low-frequency component of the speech signal to an implant-simulated signal would improve speech recognition.
2. The advantage would be greater in magnitude when both signal components were presented diotically, rather than dichotically.
3. The advantage would be greater in magnitude for words presented in sentences than for words presented in isolation.
4. Children would demonstrate a greater advantage than adults.

These hypotheses were tested using isolated word and four-word sentence materials, with noise-vocoded signals (high-pass filtered above 0.25 kHz), and noise-vocoded plus low-frequency signals. Stimuli were presented in diotic and dichotic configurations to adults and seven-year-olds.

Results of the main experiment reported here provided strong evidence to support the first hypothesis: For both sets of speech materials and both configurations, an advantage was found for the addition of the low-frequency signal component. This effect varied from roughly 20% to 60% improvement in recognition scores over what was obtained for the noise-vocoded signals alone. These outcomes suggest that even if patients with hearing loss can hear only the lowest frequencies in the speech signal, there is likely an advantage to be obtained to amplifying those frequencies with hearing aids.

The second hypothesis was not supported by the data collected here. Diotic and dichotic configurations resulted in effects of equivalent magnitude.

The third hypothesis was supported by the finding of a significant Materials × Processing interaction for word recognition scores. That outcome indicates that the magnitude of the effect of adding the low-frequency signal component to the noise-vocoded signal was greater for word recognition when words were heard in sentences rather than in isolation. It is tempting to attribute that outcome to the idea that the low-frequency signal contributes more linguistic information to sentence than word recognition, but the failure to find any effect of linguistic influences on recognition scores themselves or on the magnitude of the low-frequency effect contradicts that conclusion. Instead, it is suggested that longer sequences of materials make it easier to achieve the perceptual organization required to recover speech-like form.

Regarding the fourth hypothesis, it is tricky to reach a firm conclusion as to whether the current data supported it or not. Absolute differences in recognition scores between VOC-only and LOW-plus conditions were consistent across listener age, in both the preliminary and main experiments. However, when those differences are transformed to proportions of recognition scores for the VOC-only condition, it appears as if children gain more by the addition of the low-frequency signal component. That outcome might reflect the finding that Başkent and Chatterjee (2010) reported that the benefit of adding a low-frequency signal to an otherwise spectrally degraded signal was greater for listeners who were poorer at recognizing those spectrally degraded signals to begin with. That interpretation finds support in the fact that significant correlation coefficients were obtained between the EFFECT scores and recognition scores for the VOC-only stimuli, both across and within listener groups. Thus, it seems that the poorer listeners' abilities were to recognize the spectrally degraded speech signals, the more they benefited from the addition of the low-frequency signal component, regardless of age, but children tended to have poorer recognition of the VOC-only signals than adults. Viewing this outcome from the opposite perspective, it may be that listeners who were already successful at perceptually organizing the spectrally degraded (vocoded) signals in order to recover speech-like form did not require the aid of the very low-frequency signal.

Clinical implications and limitations

The study reported here was conducted under ideal conditions, especially with regard to listener characteristics. All participants in the two experiments had normal hearing. Simulations of signals received through cochlear implants were created by using noise-vocoded signals. Thus, the signals were spectrally degraded in the sense that they lacked the kind of detailed frequency structure that typically contributes to speech recognition. Nonetheless, these signals covered the usual range of speech frequencies, and frequencies in the signal aligned with locations along the basilar membrane that are matched to those frequencies for processing by listeners with normal hearing. This frequency-place match can not be assumed for listeners with cochlear implants (Rosen et al, 1999). That fact could negatively impact how signal components are integrated across frequencies for these listeners (e.g. Goupell et al, 2013). Furthermore, patterns of survival of spiral ganglion cells in listeners who have suffered significant hair cell loss could influence integration across the spectrum (Blamey, 1997; Carlson et al, 2011; Fayad & Linthicum, 2006).

The finding of a strong effect of the materials used has implications for clinical practice. Word lists, or even syllables differing by a single segment, are more commonly used in audiological evaluations than sentence materials. That practice could underestimate the benefit that individual patients might demonstrate in the clinic when low-frequency acoustic signals are combined with the electric signal they get through their cochlear implants. Thus, sentence materials would be preferable for evaluating whether a patient might benefit from electric-acoustic stimulation.

This study provides insight concerning which patients could potentially benefit from the addition of amplified low-frequency acoustic signals to the electric signals they receive through their cochlear implants. In particular, the results indicate that even patients with profound hearing loss might be candidates for electric-acoustic stimulation. And those patients who have the most difficulty recognizing speech with their cochlear implants might be the best candidates for the addition of low-frequency acoustic amplification. Finally, both adults and children are able to benefit from the combination of a very low-frequency signal with their cochlear implant signal.

General Conclusions

The two experiments reported here were undertaken to evaluate issues related to the combination of a low-frequency signal with a spectrally degraded signal, as might be the configuration for a patient with profound hearing loss who uses a cochlear implant and a hearing aid with extended low-frequency amplification, either in the same or separate ears. The study was designed to answer the *who*, *what*, *when*, and *why* questions surrounding the electric-acoustic advantage observed for many patients with hearing loss in clinical studies. Results were able to adequately address each of these questions: It turns out that both pediatric and adult patients can benefit from this kind of configuration, but patients with the poorest speech recognition through a cochlear implant will likely benefit the most from the addition of a very low-frequency signal. In this study, it was found that speech recognition in quiet can be improved with this combination of signal types by anywhere from 20% to 60%, depending on the individual patient and type of material used for testing. These benefits can be obtained in both diotic and dichotic configurations. That means that both hybrid and bimodal stimulation should support the kind of integration required to achieve the benefit of low-frequency signals. Finally, evidence from this study suggests that the benefit is not achieved because of any linguistic structure provided by the very low-frequency signal. Rather it appears that very low-frequency speech signals facilitate the perceptual organization of spectrally degraded signals that are communicatively significant so that phonetic form can be recovered.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

This work was supported by Grant No. R01 DC000633 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health.

References

Başkent D. & Chatterjee M. 2010. Recognition of temporally interrupted and spectrally degraded sentences with additional unprocessed low-frequency speech. *Hear Res*, 270, 127–133.

Blamey P.J. 1997. Are spiral ganglion cell numbers important for speech perception with a cochlear implant? *Am J Otol*, 18, S11–S12.

Boothroyd A. 1997. Auditory capacity of hearing-impaired children using hearing aids and cochlear implants: Issues of efficacy and assessment. *Scand Audiol Suppl*, 46, 17–25.

Boothroyd A. & Nittrouer S. 1988. Mathematical treatment of context effects in phoneme and word recognition. *J Acoust Soc Am*, 84, 101–114.

Carlson M.L., Driscoll C.L., Gifford R.H., Service G.J., Tombers N.M. et al. 2011. Implications of minimizing trauma during conventional cochlear implantation. *Otol Neurotol*, 32, 962–968.

Chang J.E., Bai J.Y. & Zeng F.G. 2006. Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise. *IEEE Trans Biomed Eng*, 53, 2598–2601.

Ching T.Y., Incerti P. & Hill M. 2004. Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. *Ear Hear*, 25, 9–21.

Ching T.Y., Psarros C., Hill M., Dillon H. & Incerti P. 2001. Should children who use cochlear implants wear hearing aids in the opposite ear? *Ear Hear*, 22, 365–380.

Chistovich L.A. 1985. Central auditory processing of peripheral vowel spectra. *J Acoust Soc Am*, 77, 789–805.

Chomsky C. 1969. *The Acquisition of Syntax in Children from 5 to 10*. Cambridge, USA: M.I.T. Press.

Cohen N.L. 2004. Cochlear implant candidacy and surgical considerations. *Audiol Neurootol*, 9, 197–202.

Cullington H.E. & Zeng F.G. 2010. Bimodal hearing benefit for speech recognition with competing voice in cochlear implant subject with normal hearing in contralateral ear. *Ear Hear*, 31, 70–73.

Dooley G.J., Blamey P.J., Seligman P.M., Alcantara J.I., Clark G.M. et al. 1993. Combined electrical and acoustical stimulation using a bimodal prosthesis. *Arch Otolaryngol Head Neck Surg*, 119, 55–60.

Dorman M.F. & Gifford R.H. 2010. Combining acoustic and electric stimulation in the service of speech recognition. *Int J Audiol*, 49, 912–919.

Fayad J.N. & Linthicum F.H. Jr. 2006. Multichannel cochlear implants: Relation of histopathology to performance. *Laryngoscope*, 116, 1310–1320.

Fox R.A. & Jacewicz E. 2010. Auditory spectral integration in nontraditional speech cues in diotic and dichotic listening. *Percept Motor Skills*, 111, 543–558.

Fox R.A., Jacewicz E. & Feth L.L. 2008. Spectral integration of dynamic cues in the perception of syllable-initial stops. *Phonetica*, 65, 19–44.

Gantz B.J. & Turner C. W. 2003. Combining acoustic and electrical hearing. *Laryngoscope*, 113, 1726–1730.

Gantz B.J., Turner C. & Gfeller K.E. 2006. Acoustic plus electric speech processing: Preliminary results of a multicenter clinical trial of the Iowa/Nucleus Hybrid implant. *Audiol Neurootol*, 11 Suppl. 1, 63–68.

Gifford R.H., Dorman M.F., McKarns S.A. & Spahr A.J. 2007. Combined electric and contralateral acoustic hearing: Word and sentence recognition with bimodal hearing. *J Speech Lang Hear Res*, 50, 835–843.

Goldman R. & Fristoe M. 2000. *Goldman Fristoe 2: Test of Articulation*. Circle Pines, USA: American Guidance Service, Inc.

Goupell M.J., Stoelb C., Kan A. & Litovsky R.Y. 2013. Effect of mismatched place-of-stimulation on the salience of binaural cues in conditions that simulate bilateral cochlear-implant listening. *J Acoust Soc Am*, 133, 2272–2287.

Hamzavi J., Pok S.M., Gstoettner W. & Baumgartner W.D. 2004. Speech perception with a cochlear implant used in conjunction with a hearing aid in the opposite ear. *Int J Audiol*, 43, 61–65.

Helms J., Müller J., Schön F., Moser L., Arnold W. et al. 1997. Evaluation of performance with the COMBI40 cochlear implant in adults: A multicentric clinical study. *ORL J Otorhinolaryngol Relat Spec*, 59, 23–35.

Incerti P.V., Ching T.Y. & Cowan R. 2013. A systematic review of electric-acoustic stimulation: Device fitting ranges, outcomes, and clinical fitting practices. *Trends Amplif*, 17, 3–26.

Kimchi R. 2009. Perceptual organization and visual attention. *Prog Brain Res*, 176, 15–33.

Kimura D. 1961. Cerebral dominance and the perception of verbal stimuli. *Can J Psychol*, 15, 166–171.

Kong Y.Y. & Braida L.D. 2011. Cross-frequency integration for consonant and vowel identification in bimodal hearing. *J Speech Lang Hear Res*, 54, 959–980.

Liberman A.M., Isenberg D. & Rakerd B. 1981. Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Percept Psychophys*, 30, 133–143.

Liberman I.Y., Shankweiler D., Fischer F.W. & Carter B. 1974. Explicit syllable and phoneme segmentation in the young child. *J Exp Child Psychol*, 18, 201–212.

Loizou P.C., Mani A. & Dorman M.F. 2003. Dichotic speech recognition in noise using reduced spectral cues. *J Acoust Soc Am*, 114, 475–483.

Mackersie C.L., Boothroyd A. & Minniear D. 2001. Evaluation of the computer-assisted speech perception assessment test (CASPA). *J Am Acad Audiol*, 12, 390–396.

Mann V.A. & Liberman A.M. 1983. Some differences between phonetic and auditory modes of perception. *Cognition*, 14, 211–235.

Martin N. & Brownell R. 2011. *Expressive One-Word Picture Vocabulary Test (EOWPVT)*. (4th ed.) Novato, USA: Academic Therapy Publications, Inc.

Müller J., Schön F. & Helms J. 2002. Speech understanding in quiet and noise in bilateral users of the MED-EL COMBI 40/40 + cochlear implant system. *Ear Hear*, 23, 198–206.

NIH Consensus Development Panel. 1995. Cochlear implants in adults and children. *JAMA*, 274, 1955–1961.

Niparko J.K. & Wilson B.S. 2000. History of cochlear implants. In: J.K. Niparko K.I. Kirk, N.K. Mellon, A.M. Robbins, D.L. Tucci & B.S. Wilson (eds.), *Cochlear Implants: Principles and Practices*. Philadelphia: Lippincott Williams and Wilkins, pp. 103–107.

Nittrouer S. & 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. & Chapman C. 2009. The effects of bilateral electric and bimodal electric-acoustic stimulation on language development. *Trends Amplif*, 13, 190–205.

Nittrouer S. & Lowenstein J.H. 2010. Learning to perceptually organize speech signals in native fashion. *J Acoust Soc Am*, 127, 1624–1635.

Nittrouer S., Lowenstein J.H. & Packer R. 2009. Children discover the spectral skeletons in their native language before the amplitude envelopes. *J Exp Psychol Human*, 35, 1245–1253.

Osberger M.J. 1994. Speech perception and production results in children with multichannel cochlear implants. *J Acoust Soc Am*, 95, 2905.

Remez R.E., Rubin P.E., Pisoni D.B. & Carrell T.D. 1981. Speech perception without traditional speech cues. *Science*, 212, 947–949.

Risberg A. & Agelfors E. 1982. Speech perception based on non-speech signals. In: R. Carlson and B. Granstrom (eds.), *The Representation of Speech in the Peripheral Auditory System*. Elsevier Biomedical Press, pp. 209–215.

Rosen S., Faulkner A. & Wilkinson L. 1999. Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *J Acoust Soc Am*, 106, 3629–3636.

Skarzynski H., Lorens A., D Haese P., Walkowiak A., Poitrowska A. et al. 2002. Preservation of residual hearing in children and postlingually deafened adults after cochlear implantation: An initial study. *ORL J Otorhinolaryngol Relat Spec*, 64, 247–253.

Skinner M.W., Clark G.M., Whitford L.A., Seligman P.M., Staller S.J. et al. 1994. Evaluation of a new spectral peak coding strategy for the Nucleus 22 Channel Cochlear Implant System. *Am J Otol*, 15 Suppl. 2, 15–27.

Stelmachowicz P.G., Hoover B.M., Lewis D.E., Kortekaas R.W. & Pittman A.L. 2000. The relation between stimulus context, speech audibility, and perception for normal-hearing and hearing-impaired children. *J Speech Lang Hear Res*, 43, 902–914.

Studdert-Kennedy M. 1983. Limits on alternative auditory representations of speech. *Ann NY Acad Sci*, 405, 33–38.

Studdert-Kennedy M., Shankweiler D. & Pisoni D. 1972. Auditory and phonetic processes in speech perception: Evidence from a dichotic study. *Cogn Psychol*, 3, 455–466.
von Ilberg C.A., Baumann U., Kiefer J., Tillein J. & Adunka O.F. 2011. Electric-acoustic stimulation of the auditory system: A review of the first decade. *Audiol Neurootol*, 16 Suppl. 2, 1–30.
Walley A.C., Smith L.B. & Jusczyk P.W. 1986. The role of phonemes and syllables in the perceived similarity of speech sounds for children. *Mem Cognit*, 14, 220–229.
Whalen D.H. & Liberman A.M. 1987. Speech perception takes precedence over nonspeech perception. *Science*, 237, 169–171.
Wilkinson G.S. & Robertson G.J. 2006. *The Wide Range Achievement Test (WRAT)*, (4th ed.). Lutz, USA: Psychological Assessment Resources.
Wilson B.S. 2006. Speech processing strategies. In: H.R. Cooper & L.C. Craddock (eds.), *Cochlear Implants: A Practical Guide* (2nd ed). Hoboken, USA: John Wiley and Sons, pp. 21–69.
Yang H.I. & Zeng F.G. 2013. Reduced acoustic and electric integration in concurrent-vowel recognition. *Sci Rep*, 3, 1419.

Supplementary material available online

Supplementary materials. Sentences used in the preliminary and main experiment.