To cite this chapter:
Başkent, D., Gaudrain, E., Tamati, T. N., and Wagner, A. (2016). "Perception and psychoacoustics of speech in cochlear implant users,"
In A. T. Cacace, E. de Kleine, A. G. Holt, and P. van Dijk (Eds.), Scientific foundations of audiology: perspectives from physics, biology, modeling, and medicine*, Plural Publishing, Inc, San Diego, CA, pp. 285–319. ISBN13:978-1-59756-652-0.*

Chapter 12

Perception and Psychoacoustics of Speech in Cochlear Implant Users

Deniz Başkent, Etienne Gaudrain, Terrin N. Tamati, and Anita Wagner

Introduction

Cochlear implants (CIs) are prosthetic devices that restore hearing in deaf individuals via electric stimulation of the auditory nerve through an electrode array inserted in the cochlea. The device consists of a microphone and an externally worn speech processor that converts acoustic signals into electric signals. The speech signal coded with electric pulses is then delivered via a wireless transmission system through the scalp to an electrode array, implanted in the cochlea, traditionally in the scala tympani (Grayden & Clark, 2006).

While research with CIs dates back to 1950s (Djourno & Eyriès, 1957), the Food and Drug Administration (FDA) has approved CI use in adults in 1984, in children 2 years and older in 1989, and in children 12 months and older in 2000. The FDA reports that approximately 324,000 people worldwide had received CIs by 2012.

Despite the relatively long history of the implant, the speech signal transmitted via the modern CIs is still inherently degraded in fine spectrotemporal details (e.g., Loizou, 1998; Rubinstein, 2004). The device mainly delivers slow-varying amplitude envelopes of speech modulating (usually) fixed-rate digital pulses, delivered at a small number of contact points (electrodes). This degraded signal is recognized and reinterpreted by the brain as speech (Fu, 2002; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). One of the main forms of degradation is the reduced spectral resolution (Friesen, Shannon, Başkent, & Wang, 2001; Fu & Nogaki, 2005; Henry, Turner, & Behrens, 2005). This reduction does not come from the small number of electrodes per se but instead from the channel interactions caused by the spatial overlap of the broad stimulation from individual electrodes (e.g., Shannon, 1983; Stickney et al., 2006). Coding of temporal fine structure is also limited, mainly caused by the characteristics of the electric stimulation of the auditory nerve (Rubinstein & Hong, 2003). Other factors that can affect the quality of the speech signal delivered via the CI include the position of the electrode array, such as the insertion depth (Başkent & Shannon, 2005; Dorman, Loizou, & Rainey, 1997; Hochmair, Hochmair, Nopp, Waller, & Jolly, 2015; Skinner et al., 2002) or the proximity to the spiral ganglia (Holden et al., 2013), potential mismatch in the frequency-place mapping (Başkent & Shannon, 2004; Siciliano, Faulkner, Rosen, & Mair, 2010; Venail et al., 2015), limited dynamic range of electric hearing (Zeng et al., 2002), presence of acoustic low-frequency hearing in the implanted or contralateral ear (Cullington & Zeng, 2010; Gantz, Turner, Gfeller, & Lowder, 2005; Gifford, Dorman, McKarns, & Spahr, 2007), robustness of the electrode-nerve interface (Bierer & Faulkner, 2010), neural survival patterns (Khan et al., 2005) and potential dead regions (Kasturi, Loizou, Dorman, & Spahr, 2002; Shannon, Galvin, & Başkent, 2002), cochlear abnormalities and surgical factors (Finley & Skinner, 2008; Sennaroğlu, 2010), and device-related factors such as sound-processing strategy (Wilson et al., 1991), electrode design, configuration, and stimulation mode (Stickney et al., 2006; Zwolan, Kileny, Ashbaugh, & Telian, 1996) and stimulation rate (Friesen, Shannon, & Cruz, 2005; Vandali, Whitford, Plant, & Clark, 2000).

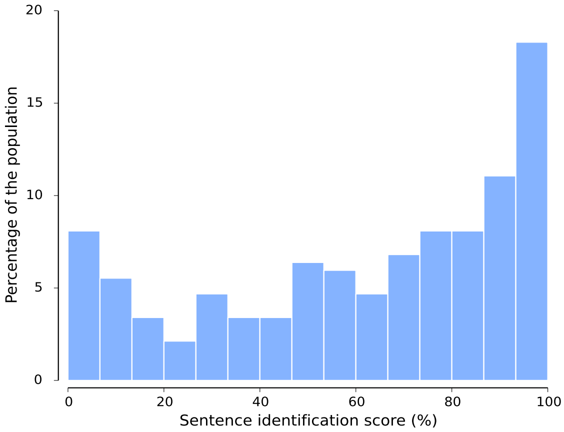

Fig. 12-1: Distribution of percent correct scores for post-implantation sentence identification in the CI user population. Adapted from Blamey et al. (2013).

After a postimplantation adaptation period (Lazard, Innes-Brown, & Barone, 2014), many CI users achieve a good level of speech understanding in quiet, ideal listening conditions (Rouger et al., 2007). However, performance across individual CI users is still highly variable (Blamey et al., 2013, also see Figure 12–1). Additionally, comprehension of speech further degraded by other, external factors, for example, due to interfering sounds or poor room acoustics, remains to be a challenge for these individuals (Friesen et al., 2001; Fu & Nogaki, 2005; P. Nelson, Jin, Carney, & Nelson, 2003; Stickney, Zeng, Litovsky, & Assmann, 2004).

Historical Perspective

Speech perception research with CIs dates back to the single-channel devices in the 1970s and 1980s (Danley & Fretz, 1982; Douek, Fourcin, Moore, & Clark, 1977; Hochmair & Hochmair-Desoyer, 1983; Merzenich, Michelson, Schindler, Pettit, & Reid, 1973). These devices were only able to deliver the slow-varying amplitude envelope of the broadband speech signal, with a severely limited dynamic range and spectral information (Millar, Tong, & Clark, 1984). Due to using a single point of stimulation in the cochlea, the implant could not evoke any place pitch percept (i.e., the pitch percept evoked by stimulating different locations of the tonotopically organized cochlea partition). Some temporal pitch percept could be achieved via the rate of stimulation. Yet, this was thought to be limited by the nerve refractory period, allowing only partial transmission of voice pitch (defined by fundamental frequency, F0), and first formant (F1) information (Dorman & Spahr, 2006).

Despite these limitations, these devices succeeded in providing some segmental and supra-segmental speech cues. Overall intensity fluctuations can already provide segmental information and stress patterns (Rosen, Walliker, Brimacombe, & Edgerton, 1989). Voice pitch, even if partial, can help with perceiving information on voicing (i.e., discrimination of voiced consonants from unvoiced ones, such as /z/ vs. /s/) and manner (i.e., discrimination of stop consonants, such as /p/, from fricatives, such as /s/). Pitch fluctuations can provide sentence prosody (i.e., discrimination of a question from a statement). Through F1, some vowel identification can be achieved (Dorman & Spahr, 2006). However, lack of spectral resolution means place of articulation cues are lost (i.e., consonants such as /p, t, k/ that mostly differ in their place cue could not be discriminated). Lack of higher formant cues leads to confusions in discriminating different vowels from each other (e.g., /i/ vs. /u/) and also in discriminating fricatives from each other (e.g., /s/ vs. /ʃ/). With such limited transmission of speech cues, as a result, while some CI users showed open-set speech recognition abilities with auditory input only (Berliner, Tonokawa, Dye, & House, 1989), most CI users could only derive useful speech perception benefit in closed-set phoneme or word discrimination, or in combination with visual speech cues (Rosen et al., 1989).

In modern multichannel CIs, the cochlear tonotopic organization is taken into account (Greenwood, 1990); low-frequency components of speech are delivered to apical electrodes while high-frequency components are delivered to basal electrodes. Multichannel CIs thus present a significant improvement over single-channel devices in transmitting speech cues. As a result, drastic improvements have been observed in speech perception performance of CI users in general (Clark, 2015; Zeng, 2004; also see Figure 12–2). Even in the same users, who, after using a single-channel device, were reimplanted with a multichannel device (for example due to device failure), an immediate improvement in speech perception was observed. In such a CI user, only 3 months after implantation, Spillman and Dillier (1989) observed an improvement in recognition of vowels and sentences, and further, the patient could also make use of second formant information, F2, in addition to F1.

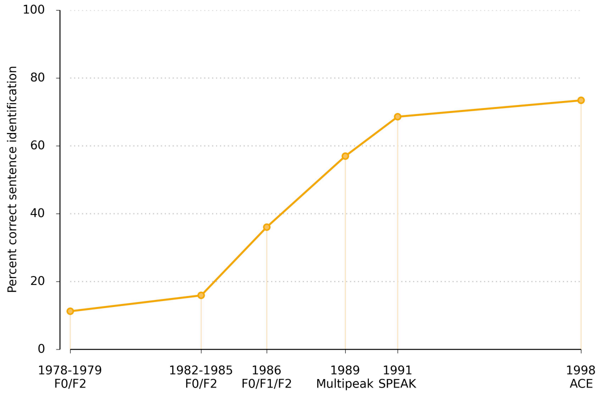

Fig. 12-2: Evolution of sentence identification scores through the history of cochlear implants. Adapted from Clark (2015).

In modern CI devices, despite the improvements over the years in device design, surgical techniques, and speech coding strategies, speech information transmitted via electric stimulation still only partially mimics that of acoustic hearing (Loizou, 1998; also see an example of a sentence processed with an acoustic simulation of a CI, and represented in electrodogram in Figure 12–3). The modern CI processor bandpass filters the acoustic input of broadband speech signal into a number of frequency bands, to be delivered to distinct electrodes for tonotopical stimulation of the auditory nerve. However, the stimulating carrier current is usually a fixed-rate digital pulse sequence. As a result, what is eventually delivered to the nerve is fixed-rate current pulses at each electrode, modulated by the slow-varying amplitude envelopes of the corresponding spectral band. In the signal used for electric stimulation, the slow-varying temporal envelopes and gross spectral details are preserved; however, all spectrotemporal fine structure is lost. While Shannon et al. (1995) have shown early on that even a small number of bands with slow varying envelopes are sufficient for basic level of speech perception, this level of speech detail seems to be insufficient for more advanced levels of speech perception, such as in background noise or talkers (Friesen et al., 2001; Fu & Nogaki, 2005; Stickney et al., 2004), or to achieve other speech-related tasks, such as vocal emotion perception (Chatterjee et al., 2015; Luo, Fu, & Galvin, 2007). Perception (and enjoyment) of more complex, and potentially pleasurable, sounds, such as music, seems to be minimally available to CI users (Crew, Galvin, & Fu, 2012; Drennan et al., 2015; Fuller, Free, Maat, & Başkent, 2012; Limb & Rubinstein, 2012).

Perception of Vocal Characteristics

One of the most important speech cues that helps with higher level perception of speech is voice characteristics, namely, voice pitch, directly related to the glottal pulse rate of the speaker, that is, F0, and vocal tract length (VTL), directly related to the size of the speaker (Fitch & Giedd, 1999). While the perception of the former relies both on temporal and place coding of speech (i.e., harmonic structure), the perception of the latter relies mostly on the perception of spectral characteristics, namely, the formant structure (Smith, Patterson, Turner, Kawahara, & Irino, 2005). The way these cues are coded in the acoustic speech signal is shown in Figure 12–4.

A robust perception of voice characteristics is not only important for identifying the speaker characteristics but also plays an important role in everyday life speech perception. In such scenarios, one hears many sounds mixed into one signal where the target speech has to be segregated from the interfering background sounds and the individual audible speech segments be grouped into a meaningful speech stream. Voice characteristics make it possible for the listener to lock onto as acoustic cues for such segregation. This was shown by the

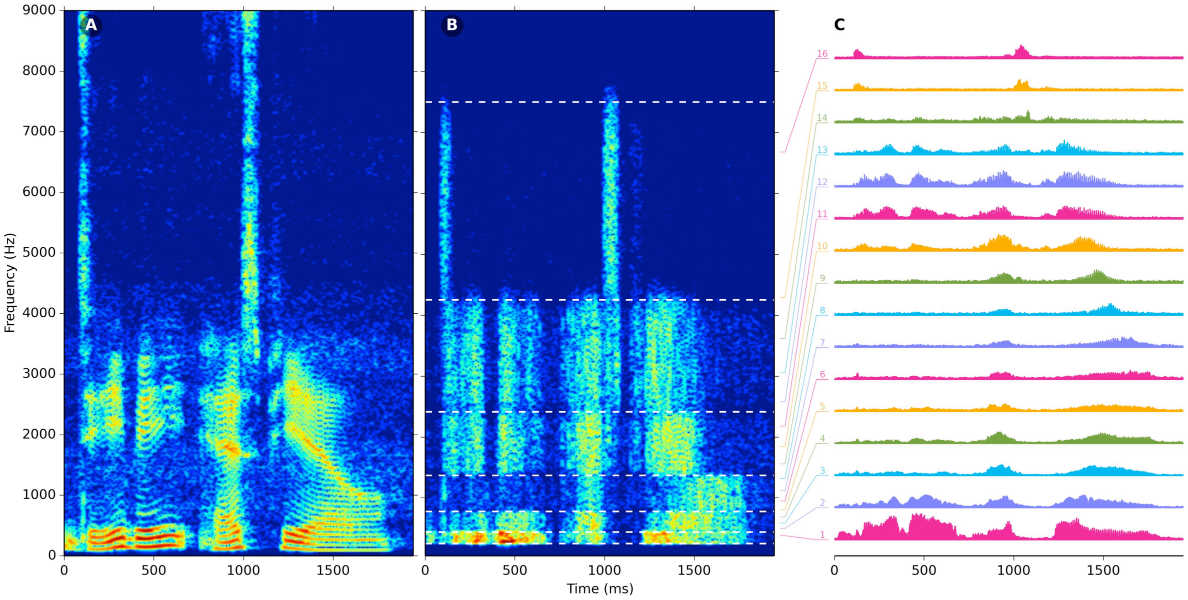

Fig. 12-3: (**A**) Spectrogram of a sentence. (**B**) Spectrogram of the sentence processed with an acoustic simulation of a CI. Here a six-channel noise-band vocoder is used (based on Shannon et al., 1995). The white dashed lines indicate the band cutoff frequencies of the vocoder bandpass filters. (**C**) Electrodogram of the same signal, showing electrical activity in a 16-channel implant using a CIS strategy. Each row corresponds to an electrode, and is connected to the frequency axis of panel B to indicate its center frequency.

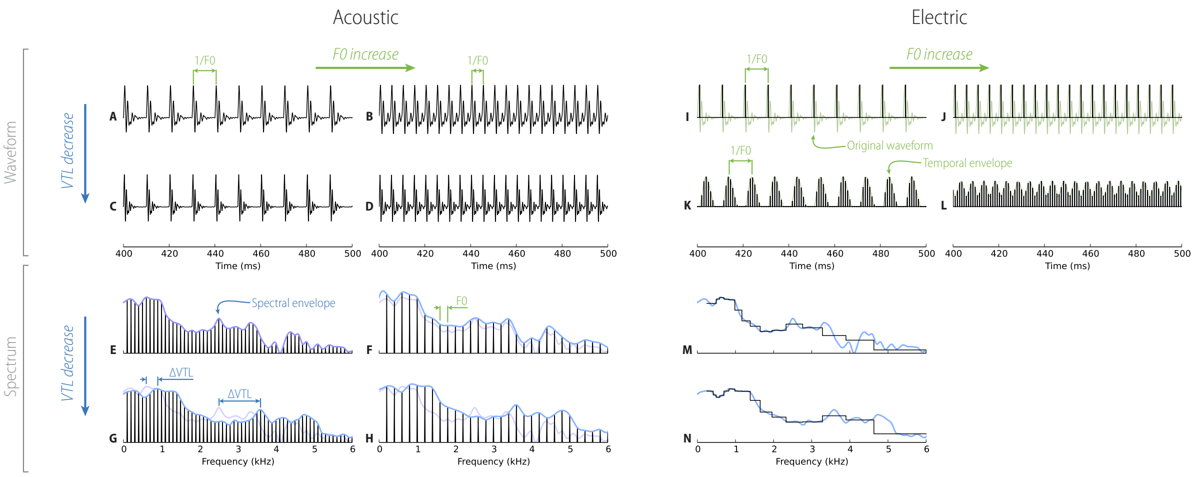

Fig. 12-4: Left column—schematic representations of the waveform (**A–D**) and spectrum (**E–H**) of the vowel /a/ for different combinations of F0 and VTL. A decrease in VTL results in shrinking the temporal pattern produced by a single glottal pulse (A vs. C), which corresponds to an expansion of the spectral envelope (E vs. G). An increase in F0 results in glottal pulses being more frequent (A vs. B), or to harmonic components to be more spaced while they follow the same spectral envelope (E vs. F). Right column—Schematic representation of the coding of F0 (**I–L**) and VTL (**M–N**) in the implant. F0 can be coded by modifying the stimulation pulse rate (I, J) or by using a fixed-rate pulse train with the signal's temporal envelope (K, L). The panels M and N illustrate how spectral quantization affects VTL representation, but do not show the additional effect of spectral smearing that result from current spread.

increase in F0 and VTL difference leading to better segregation of concurrent vowels (Vestergaard, Fyson, & Patterson, 2009) and sentences (Darwin, Brungart, & Simpson, 2003). Hence, perception of target speech in a cocktail party seems to heavily rely on effective use of vocal cues.

Even in multichannel CIs the vocal characteristics delivered by the device are considerably weak. In principle, voice pitch (i.e., F0) can be coded through stimulation rate, temporal pattern of stimulation, or place of stimulation, as explained before. However, all have limitations. The coding of voice pitch via stimulation rate is not used in most current processors as they use a relatively high, but fixed, pulse rate. Research has shown that strategies capturing the F0 or the temporal fine structure by timing individual pulses accordingly have some potential to improve speech and music perception (Arnoldner et al., 2007; Laneau, Wouters, & Moonen, 2006). To date, only one clinically used processor is exploiting this form of coding. Another form of pitch coding is through the amplitude modulation pattern of high rate pulse trains. This form of pitch coding, like the previous one, is limited by the rate at which auditory nerve fibers can fire in response to electrical stimulation as well as by the restricted dynamic range offered by electrical stimulation. Consequently, although CI users seem to be able to experience a pitch percept through this mechanism, this percept is consistently reported as being weak (Moore & Carlyon, 2005). To enhance that pitch percept, researchers have developed strategies that code the F0 by explicitly modulating the stimulation pattern rather than relying on natural amplitude modulations of the original stimulus (Milczynski, Wouters, & Van Wieringen, 2009; Vandali & van Hoesel, 2011). Finally, the coding of voice pitch via place of stimulation in CIs is primarily limited by the low spectral resolution available through the implant.

In contrast to large number of studies on F0 perception in CIs, perception of the other dominant voice characteristic, namely VTL, has been only minimally studied. Only recently, Gaudrain and Başkent (2015a) showed that VTL perception is severely impaired in CI users. Using acoustic simulations of CIs, Gaudrain and Başkent (2015b) showed that this limitation was likely due to channel interactions and smeared spectral resolution, similar to the limitation of the place percept of F0. Because VTL is a cue used in many situations by normal-hearing (NH) listeners, this specific impairment of CI users has repercussions on many speech-related tasks. One instance of such consequence that has been demonstrated to date is how speaker gender categorization is impaired in CI users because they can only rely on F0 and cannot access the VTL cue to conduct the task (Fuller et al., 2014).

While impairment in gender categorization per se might not have dramatic consequences in real situations, as other cues are often available to perform this task, its consequences on the ability to hear a specific speaker among other talkers are very real. Brungart (2001) observed that when presenting two competing sentences to NH listeners, intelligibility was much higher if the two sentences were uttered by two speakers of opposite sex, than when the two sentences were uttered by the same speaker, or by speakers of the same gender. When the target and masker had the same intensity level, this talker-gender difference provided an advantage of 52 percentage points. Darwin et al. (2003) later showed that most of this advantage can be explained by F0 and VTL differences. Thus, considering the importance of vocal acoustic cues for the perception of speech on speech, the weakness of their representation in CIs is likely a major limitation for perception of speech in interfering background sounds by CI users.

Speech Perception in Background Interference

The perception of speech in background noise, or in general in presence of any competing sound source, is undoubtedly considered the strongest limitation to CIs, and thus poses the greatest challenge for the research community. Many reports unambiguously show how little robustness speech perception has to competing sound sources in CI users, compared to reference NH listeners (Friesen et al., 2001; Fu & Nogaki, 2004; Stickney et al., 2004).

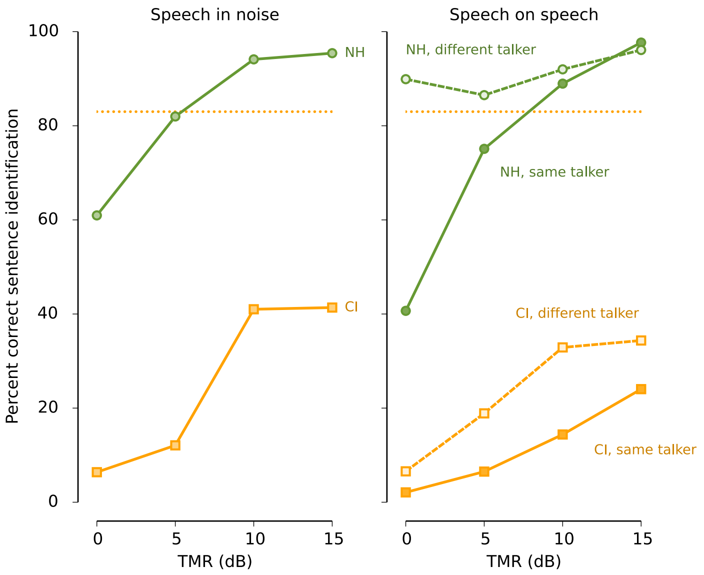

Fig. 12-5: Left panel: Identification scores for sentences presented in speech-shaped noise for NH and CI listeners, as a function of target-to-masker ratio (TMR). The dotted line shows baseline performance of the CI listeners in quiet. Adapted from Friesen et al. (2001). Right panel: Same for speech-on-speech, where the target and masker are either the same talker (solid lines) or where the target is a male voice but the target is a female voice (dashed lines). Adapted from Stickney et al. (2004).

When the masker is stationary noise, like the one used in most clinically used speech tests, vowel and consonant identification are relatively moderately affected by noise level, both in CI and NH listeners. However, word and sentence identification performance quickly becomes challenging for CI users, even at positive signal to noise ratios. The nature of the masker also plays an important role. Maskers that are more ecological than stationary noise, such as speech from other speakers, can reveal even more dramatic differences between NH and CI listeners. In the speech-on-speech study by Stickney et al. (2004), where a target sentence was presented simultaneously with a masker sentence, the advantage that NH listeners could derive from voice gender differences reached 49 percentage points (similar to the study by Brungart, 2001). However, in the same conditions, the largest advantage that CI listeners could derive from talker gender differences was only 19 percentage points (Figure 12–5).

Speech-on-speech perception and speech-in-noise perception involve different types of masking, which may explain why CI and NH listeners differ in the way they cope with the two situations. Speech-in-noise perception mostly involves *energetic masking* (i.e., the loss of information representation either in the peripheral auditory system or, in the case of CI listeners, also in the auditory device). Speech-on-speech, on the other hand, involves relatively little energetic masking but instead involves a collection of masking phenomena often gathered under the umbrella term *informational masking*. In fact, the whole phenomenon of perceiving speech in an interferer can be decomposed in order to identify which component mechanisms are affected by the implant limitations.

A component that captures most of the energetic masking is simultaneous segregation (Bregman, 1990) and concerns the perceptual separation of two sound events occurring at the same time. Simultaneous segregation can by studied with the “double vowel” paradigm or with the “concurrent syllables” paradigm. While NH listeners can use F0 (De Cheveigne, 1999) and VTL (Vestergaard et al., 2009), CI listeners do not seem to be able to do so. Luo, Fu, Wu, and Hsu (2009) observed that, similar to competing sentences, CI listeners do not benefit from voice gender differences (which include F0 and VTL differences) in identifying concurrent vowels or syllables. These limitations are thought to be directly related to the poor spectral resolution available through the implant (Qin & Oxenham, 2005).

Another component concerns the way successive speech elements are stringed together to be processed at the linguistic level. In presence of multiple sound sources, successive speech elements must undergo "triage" (i.e., be assigned to the foreground and background auditory streams). This process, known as sequential segregation, can be induced by any perceptual cue that allows discrimination of the two streams (Moore & Gockel, 2002). While CI listeners have been shown to be able to use this mechanism when segregation cues are preserved (Chatterjee et al., 2006; Hong & Turner, 2009), the degradation of the segregation cues that are available in natural speech (like F0 and VTL) also yields weaker stream separation (Gaudrain et al., 2007, 2008).

The literature also described a phenomenon, referred to as *glimpsing*, that is very much related to the segregation mechanisms explained above. Glimpsing is the ability to exploit temporal or spectral sparseness of a masker to "glimpse" unmasked portions of the target signal. While spectral glimpsing is severely limited by the poor spectral resolution of the implant, temporal glimpsing should, in principle, remain possible, just like sequential segregation is in principle less affected by electric hearing than simultaneous segregation. However, using interrupted maskers, researchers have consistently reported that CI listeners seem less able to benefit from glimpses than NH listeners, even when only temporal glimpses are to be used (Gnansia et al., 2010; Nelson & Jin, 2004; Nelson et al., 2003; Qin & Oxenham, 2003).

Simultaneous and sequential segregation, or spectral and temporal glimpsing, are thus all affected by electric hearing, but to different degrees. One could thus venture that different types of maskers have different effects on speech perception in CI than in NH listeners because they involve simultaneous and sequential segregation to different degrees. Kwon, Perry, Wilhelm, and Healy (2012) investigated this hypothesis by creating maskers that either maximized or minimized the need for simultaneous segregation versus sequential segregation. They found that while the sequential condition was easier for NH listeners, CI listeners displayed no such benefit, suggesting that sequential processing is either more degraded in CIs than previously estimated in other studies, or that another component of speech-in-noise perception is also impaired by electrical hearing. Gaudrain and Carlyon (2013) used a more direct approach by developing a purely sequential speech mixture, Zebra-speech, which they compared to a normal speech mixture requiring both simultaneous and sequential segregation (see Figure 12–6). Using noise-band vocoders to simulate some aspects of electrical stimulation, they found sequential segregation was less affected than simultaneous segregation at lower spectral resolutions but also concluded that another mechanism involved in concurrent speech perception must be impaired in CIs.

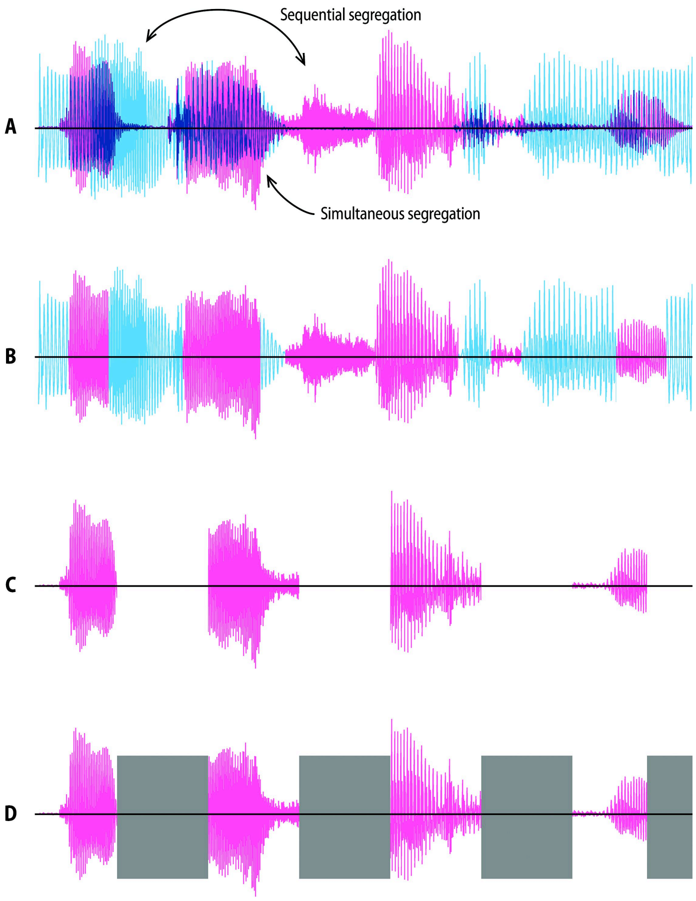

Fig. 12-6: (**A**) Waveforms of two simultaneous sentences overlapping. The arrows indicate speech segments whose processing would involve sequential and simultaneous segregation. (**B**) Waveform of the Zebra-speech version of the same overlapping sentences. (**C**) Waveform of one of the sentences periodically interrupted, similar to the ones used in interrupted speech experiments. (**D**) Same as C but where the silent intervals are filled with noise.

Indeed, simultaneous and sequential segregation interact together with linguistic processes, whose function it is to transform all accumulated auditory evidence into meaningful semantic content. It has been studied either on its own using interrupted speech (Nelson et al., 2004) or using the phonemic restoration paradigm (Bhargava et al., 2014). This component mechanism strongly hinges on the linguistic and cognitive capacities of the listener (Benard et al., 2014).

Cognitive Factors

In order to get insight into speech processing with CIs, it is important to look beyond the initial sensory input from the device (device- and physiology-related factors described above) and to focus on what the CI users are able to do with this degraded sensory information. Cognitive functions, such as executive functions (e.g., the ability to control and regulate attention, speed of processing, sequential integration) and verbal working memory (the ability to retain and manipulate acoustic signals along the way of the mapping to meaning), are fundamental to speech perception (e.g., Cleary et al., 2000; Nittrouer et al., 2013).

Speech perception requires the ability to adapt the processing of acoustic information toward different speakers and different surrounding acoustics, that is, the allocation of attention to relevant acoustic events and selective inhibition of information that is redundant to a conversation (e.g., Cherry, 1953). Understanding speech also requires the ability to retain acoustic information in memory (Baddeley, 1997) and to integrate these acoustic events with other sources of information, such as grammatical structure, and semantic context from preceding sentential context (e.g., Dahan & Tanenhaus, 2004). To obtain these goals, perception relies on higher-level cognitive processes, which allow top-down processing to enhance the acoustic bottom-up information. Our scientific knowledge of these processes is primarily based on NH populations. However, by now, there is enough evidence that these processes are not equivalent for CI populations, including children and postlingually deafened adults, who in addition show also a great deal of interindividual variation.

The development of cognitive functions in normally developing NH children takes its course in parallel to their speech development (Singer & Bashir, 1999); auditory deprivation during critical developmental phases leads to atypical development of executive functions (Kronenberger et al., 2014). In NH infants, this parallel development of speech perception and executive functions leads to an optimization of processing of speech in adult listeners. Many prelingually deaf pediatric CI users, however, score significantly below their age-matched peers in a variety of neurocognitive processing measures, including those assessing short-term and working memory, sequential learning, verbal rehearsal, and executive functioning (AuBuchon et al., 2015; Beer et al., 2011; Conway et al., 2011; Harris et al., 2013; Kronenberger et al., 2014; Kronenberger et al., 2013; Pisoni et al., 2011; Van Wieringen & Wouters, 2015). Further, language outcomes for prelingually deaf, pediatric CI users have been at least partially attributed to individual differences in these domains, even when statistically controlling for several potentially confounding variables such as age, duration of deafness, duration of device use, age at onset of deafness, number of active electrodes, and communication mode (e.g., Cleary & Pisoni, 2002; Geers & Sedey, 2011; Marschark et al., 2007; Pisoni & Cleary, 2003; Pisoni & Geers, 2000).

For adult NH populations, understanding native speech is fast, robust, and effortless. This is thanks to a native language specialization of linguistic processes (Cutler, 2012) that are based on and develop in parallel with higher-level cognitive functions. Such perceptual specialization enables listeners to quickly attend to acoustic information that is distinctive and reliable within their native language (Iverson et al., 2003; Wagner et al., 2006). These mechanisms of automatic selection of acoustic cues appear to be limited to speech and do not easily generalize to stimuli that simulate the signals transmitted via CIs (Iverson et al., 2011; Iverson et al., 2016). Furthermore, self-regulatory mechanism of attention shifts between speech signals and masking noise also appears to be guided by the acoustic details that are present in natural speech (Wöstmann et al., 2015). It is hence unclear whether and which of these native perceptual strategies can also be applied by CI listeners when processing spectrotemporally impoverished speech signals. Furthermore, individuals' ability to perceptually adapt to CI signals depends also on listeners' cognitive abilities.

Even among NH young adults, listeners with stronger neurocognitive skills might be better able to understand spoken words, especially in adverse conditions, such as with poor room acoustics, accented or reduced speech, or increased cognitive load (e.g., Francis & Nusbaum, 2009; Tamati et al., 2013), though findings of this link are so far not entirely conclusive (Akeroyd et al., 2014). For postlingually deafened CI users, individual differences in neurocognitive processing mechanisms, for example, after long period of sensory deprivation, have been found to contribute to speech perception and recognition skills (e.g., Collison et al., 2004; Heydebrand et al., 2007; Holden et al., 2013; Lazard et al., 2010; Lazard et al., 2013).

Taken together, what the CI users are able to do with the information received through a CI is as important for speech and language outcomes as the sensory information itself. Neurocognitive processing skills related to information-processing operations used in the encoding, storage, rehearsal, and retrieval of the phonological and lexical representations of spoken words seem to contribute to the vast amounts of individual differences in the spoken language outcomes of adult and pediatric CI users.

Top-Down Compensation

The fact that CI users can understand speech, given the limitations of the device, as well as the demands that the spectrotemporally impoverished signal set on listeners' cognitive functions, demonstrates a great deal of plasticity of the perceptual system. This plasticity suggests that long-term exposure to speech transmitted via CIs by itself can lead to successful adaptation of the perceptual system, an adjustment of the processing toward the demands of the degraded signal (Svirsky et al., 2001). This implies that compensation mechanisms that help NH listeners to cope with degraded signals could, maybe in adapted ways, also be employed by CI listeners. Among such compensation mechanisms is phonemic restoration, the ability to perceptually complete masked parts of the speech signal by top-down interpretations.

One way to test this idea in the lab is to measure intelligibility performance for *interrupted speech* (see Figure 12–6, panel C). Interrupted, or "gated," sentences are produced by applying a square wave modulation to the original sentence's waveform, which thus periodically turns some segments silent. The parameters are the interruption rate, typically varied from about 1 to 32 Hz, and the duty cycle (i.e., the proportion of remaining speech to the silence). Interruption rate has a very clear effect on interrupted speech perception. Nelson and Jin (2004) observed that, at 2-Hz interruptions, NH listeners' performance drops from nearly 100% to 30% correct, while with 32-Hz interruptions, performance only drops to about 80%. In contrast, CI listeners' performance drops from 80% correct in the uninterrupted case to about 5% and 10% correct for 2- and 32-Hz interruptions, respectively. Bhargava et al. (2016) found similar differences in how CI and NH cope with interruptions at different rates but also found that CI listeners were less able to take advantage of increased duty cycle than NH listeners. In a subsequent experiment, Bhargava et al. selected NH participants whose age was matching individually that of the CI participants, and used noise band vocoders (Shannon, 1995) that were adjusted so that the intelligibility of uninterrupted speech of each NH listener was also matching that of their paired CI listener. They then found that interruptions had a similar or more deleterious effect on performance for these NH listeners. In other words, the small loss of intelligibility that signal degradations—either caused by electric hearing or artificially imposed by a noise-band vocoder—induce in uninterrupted speech translates into large loss of intelligibility when the signal is interrupted.

Two phenomena happen when speech is interrupted. First, some speech elements are removed. This phenomenon is similar to the one that takes place when listening to speech in noise, where some speech elements can be masked and thus made inaccessible to the listener. But a second phenomenon also takes place; by introducing silences, spurious speech cues are introduced. Indeed, sharp onsets and offsets and pauses all carry phonetic and linguistic information. By artificially inserting silences in sentences, speech elements are not only lost but also replaced with potentially misleading speech cues. Dealing with these spurious cues requires more active cognitive processes than dealing with the loss of information alone.

This active top-down restoration can be investigated in the lab using the *phonemic restoration* (PR) paradigm (Warren, 1970; Warren & Sherman, 1974). PR is measured as the difference in intelligibility between interrupted sentences (Figure 12–6, panel C) and sentences interrupted in the same way except that the silent interruptions are filled with noise bursts (Figure 12–6, panel D). In NH listeners, adding the noise in the silent interruptions protects against the apparition of spurious cues, allowing a more faithful interpretation of the remaining speech segments and thus resulting in an increase in intelligibility (see Figure 12–7). As such, PR is considered a measure of top-down restoration.

Studies using vocoders to simulate CI processing in NH listeners have led to the conclusion that spectral and temporal degradations such as those occurring in electrical stimulation hinder phonemic restoration (Benard & Başkent, 2014; Başkent, 2012). However, Bhargava et al. (2014) showed that in actual CI users, phonemic restoration can be observed when the duty cycle is made more favorable (Figure 12–7, left panel).

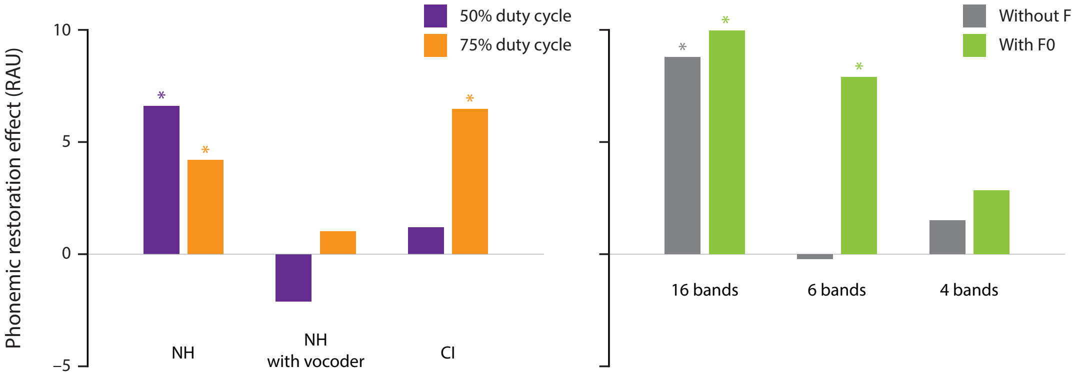

Fig. 12-7: Left panel: Phonemic restoration effect in rationalized arcsine units (RAU) for NH listeners, NH listeners presented with noise-band vocoded stimuli, and CI listeners. The two colors correspond to two duty cycles (see legend). Adapted from Bhargava et al. (2014). Right panel: Phonemic restoration effect (RAU) for NH listeners as a function of number of bands in a vocoder. The colors correspond to whether the F0 was preserved or discarded in the vocoder (see legend). Adapted from Clarke, Başkent, and Gaudrain (2016).

The discrepancy between the acoustic CI simulations and actual CI listeners seems to come from the type of vocoder—a noise-band vocoder—that was used both by Başkent (2012) and by Bhargava et al. (2014). In order for phonemic restoration to take place, the noise segments must be clearly identified as a potential masker of the speech. When spectral resolution is reduced, and when a noise carrier is used to excite the vocoder, the noisy interruptions become difficult to distinguish from the vocoded speech, which is then itself noisy in nature. Clarke et al. (2016) showed that when the F0 information is restored in the vocoder, phonemic restoration is also restored, for intermediate spectral resolutions (Figure 12–7, right panel). Although F0 information is severely degraded in CI listeners, they generally remain able to distinguish noisy signals from periodic ones, which give them the ability to use top-down restoration for phonemic restoration.

Speech Perception Mechanisms with Degraded Speech

Adverse conditions, such as hearing impairment or noisy surroundings, may change the functioning of cognitive mechanisms that underlie automatic speech perception in ideal conditions (Mattys et al., 2012). The current audiological assessment methods of speech perception only provide a measure of intelligibility, expressed in a single number, such as percent correct score or speech reception threshold. Hence, they do not fully reveal the underlying mechanisms of speech perception and the potential changes in them as a result of the speech degradation.

One of such changes is an increase in cognitive processing load (listening effort) to decode the degraded speech (Winn et al., 2015). While an increase in cognitive processing can be a good compensation mechanism to enhance perception of degraded speech, if sustained, it can lead to mental fatigue (Bess & Hornsby, 2014). Further, taking up more of the cognitive resources for speech comprehension may lead to fewer cognitive resources left for other mental tasks, such as remembering previously heard message and applying it for predictive processing of successive speech contents (Wagner, Pals, et al., 2016).

While CI users are able to understand the speech signal transmitted via CIs, speech perception is likely a more effortful task for them than NH individuals. Studies with CI simulations show that reducing the spectral resolution of the speech signal increases the effort involved in processing it. This was shown both with pupil dilation, a physiological measure of listening effort (Winn et al., 2015), as well as a dual-task paradigm (Pals et al., 2013), in which response times to a secondary task (either linguistic or nonlinguistic mental manipulation) paired with primary speech perception task reflect listening effort. The rationale behind the latter paradigm is that simultaneous cognitive tasks compete for cognitive resources, which are limited (Kahneman, 1973). Hence, increasing the number of channels in the signal reduces the response time needed by listeners to perform the secondary task. This confirms the presence of competition between parallel cognitive processes and shows that processing of degraded speech takes up cognitive resources, making speech perception a more effortful task.

Recently, new studies started investigating changes in speech perception mechanisms, more comprehensively and not limited to listening effort only. An eye-tracking study that combined the measure of the time course of lexical access (i.e., the mapping of the signal to meaning) and of the effort involved in speech perception by NH listeners presented with CI simulations showed that degradation of the signal obscured listeners' use of cues transmitted in the signal, slowed down lexical access, and increased the effort involved in listening (Wagner, Toffanin, et al., 2016). A similarly designed study with CI listeners showed that despite great individual differences among CI users, in general, experienced CI users are able to adapt their use of acoustic cues that are also employed by NH listeners. Durational cues are in particular susceptible to reweighting since they are reliably transmitted through the CI (Wagner, Opie, et al., 2016). Similar adaptation toward higher cue weighting of durational cues in experienced CI users relative to NH participants has been reported by Winn et al. (2012). However, Wagner, Opie, et al. (2016) additionally found that despite an adaptation toward a stronger reliance on reliably transmitted cues, the process of lexical access was still delayed and prolonged for experienced CI users. In the same vain, Moberly et al. (2014) found evidence for reweighting of the perceptual use of acoustic cues, such as duration and spectral cues. This study, however, also found individual differences in the reweighting of these cues in experienced CI users and concluded that CI individuals with cue-weighting most similar to NH listeners, who thus relied more on (degraded) spectral cues, showed better performance in word identification.

Potentially, compensation for such phonetic processing could come from a stronger reliance on semantic and contextual information. Effects of such top-down filling in of information based on sources other than the acoustic signal alone were investigated by Bhargava et al. (2014), mentioned above. This study shows that CI users benefit from phonemic restoration, but to a lesser degree than NH listeners, and this benefit is more limited by conditions of testing. Wagner, Pals, et al. (2016) investigated the time course of integration of semantic information from sentential context by means of eye-tracking, using CI simulations. In this study, degradation of the signal reduced listeners' ability to benefit from previously heard information, delayed the integration of contextual information, and these changes came at the cost of a longer and more effortful lexical access. For CI listeners this implies limitations to their ability to enhance the speech signal through top-down information. Whereas NH listeners can use the sentential information preceding the target to anticipate upcoming words, this ability is restricted for CI users. Figure 12–8, right panel, shows the patterns of integration of information in a sentence that differed between the processing of natural and degraded speech. The figure shows the gaze fixation patterns when recognizing a target word, such as *tree,* when it is presented within the sentence "Since when grows a tree so fast." The sentence verb in this sentence disambiguates the following target as something animate that can grow. In the experimental paradigm, the picture of a tree (target) was presented together with the picture of a child (presenting ambiguity as a child can also grow), and the pictures of two inanimate distractor objects (Figure 12–8, left panel).

In the figure, the thick solid lines depict the proportion of the gaze fixations toward the target tree, the dashed lines depict the proportion of the gaze fixations toward the competitor child, and the thin dotted lines show the proportion fixations toward the inanimate distractor objects. During the processing of natural speech (black line), listeners integrate the semantic information from the verb, which restricts their gaze fixations toward the two animate objects and discards the possibility that the inanimate objects could be the target word. When listening to CI simulations (red line), listeners presented with degraded speech take about 300 ms longer to clearly identify the tree as the target, and they tend to fixate the inanimate objects to the same

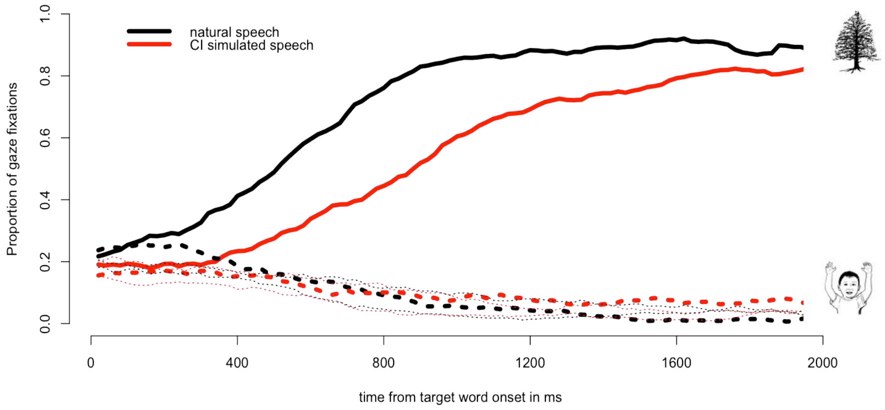

Fig. 12-8: The time curves of gaze fixations towards the target picture (tree), the semantically viable picture (baby), and the two inanimate distractor pictures (saw and headphone).

degree as the animate object child. This shows that the integration of semantic information from the sentence context is delayed and cannot be used timely to limit the search for the target from objects that are excluded by the verb (inanimate objects in this example). And note that the experimental setting in this experiment mimicked optimal conditions. Listeners' need to integrate semantic and grammatical sources of information is likely to be even greater in real-life situations.

Real-Life Speech Perception

Speech communication in real life can be very challenging. The adverse conditions of everyday speech communication not only involve the background noise, competing talkers or signals, or poor room acoustics, referred to above, but also involve natural variability in the speech signal (e.g., Mattys et al., 2012). In normal, everyday environments, a talker makes stylistic changes to their speech depending on the speech environment, speaking rate, or the surrounding phonetic or phonologic context. Additionally, individuals or social groups also have diverse speech patterns, reflecting their diverse language and developmental histories (Abercrombie, 1967). Thus, real-life speech is characterized by a great deal of variability in the realization of sounds and words.

In everyday, real-life listening environments, listeners draw upon linguistic knowledge and cognitive resources in order to achieve the perceptually and cognitively demanding task of adapting to and using these sources of variability in communication. While some speaking styles, talkers, and accents are relatively easy to understand, others may be more difficult, such as unfamiliar regional or foreign accents (e.g., Clopper & Bradlow, 2008; Mason, 1946), fast speech (e.g., Bradlow et al., 1996; Picheny et al., 1989), or reduced speech involving syllable and segment reduction or deletion (e.g., Ernestus et al., 2002; Janse et al., 2007). Further, the presence of multiple talkers and multiple sources of variability can present additional challenges to successful speech recognition (e.g., Mullennix et al., 1989). However, in normal hearing, speech communication in real-life environments is not entirely hindered. NH individuals are able to rapidly adapt to and use variation in the talker's speech characteristics and draw upon prior linguistic knowledge to successfully understand the utterance, while also extracting information about the environment, context, and talker (e.g., Johnson & Mullennix, 1997; Johnsrude et al., 2013; Tamati et al., 2014).

CI users, similar to NH listeners, are also faced with multiple sources of variability in real-life listening environments. However, because CI users must rely on a signal that is less detailed in acoustic-phonetic information than is typically available to NH listeners, their capacity to encode fine context-sensitive episodic information to reliably perceive and use subtle variation may be limited. Additionally, due to periods of auditory depravation, some CI users may have to rely upon disrupted perceptual and linguistic systems, and atypical neurocognitive skills, both of which seem to be associated with poor speech perception performance, as discussed above. Thus, the adverse effects of speech variability may be exacerbated for CI users who are receiving limited information from a degraded signal, and additionally may have poor spoken language skills and/or limited or atypically developed cognitive functions.

Despite these concerns, in contrast to our growing knowledge of NH perception of real-life speech variability, the speech perception skills of CI populations in these conditions are relatively unknown. A simple approach to studying CI perception of real-life speech variability is to assess their speech recognition with materials that more closely reflect real-life speech. Some recent studies indicate that highly variable speech, more reflective of real-life conditions, presents a significant challenge for CI users. A few recent studies have shown that CI listeners are disproportionately less accurate than NH listeners at recognizing speech produced by multiple talkers from different regions of origin (Faulkner, Tamati, & Pisoni, 2016; Faulkner, Tamati, Gilbert, & Pisoni, 2015), fast speech (Ji et al., 2013), foreign-accented speech (Ji et al., 2014), and reduced speech (Tamati et al., 2015).

Another way to study how CI users perceive, encode, and store robust information about real-life speech variability is to examine the discrimination or identification of different sources of speech variability, and the influence of linguistic information on these nonlinguistic judgments. Studies using this approach have suggested that CI listeners may not be able to make use of detailed talker-specific acoustic-phonetic information to the same extent as normal-hearing (NH) listeners to discriminate or identify different sources of real-life variability. In a recent study, Tamati and Pisoni (2015) used a foreign-accent intelligibility rating task to investigate the perception of foreign-accented speech by prelingually deaf long-term CI users. The CI users, and age-matched NH listeners, rated the intelligibility of short sentences produced by native and nonnative speakers of American English. Both the CI and NH listeners perceived the foreign-accented sentences as less intelligible than native sentences. However, compared to the NH listeners, the CI listeners perceived a smaller difference in intelligibility for foreign-accented and native speech. These findings suggest that although the CI users are sensitive to some subtle acoustic-phonetic differences between foreign-accented and native speech, they are much less so than NH listeners.

The discrimination of different sources of speech variability has also been found to be closely linked to the perception of the linguistic information in the utterance. Examining talker discrimination abilities of pediatric CI users, Cleary and colleagues (Cleary & Pisoni, 2002; Cleary et al., 2005) found that CI users were able to make more accurate talker discrimination judgments when the linguistic content was constant across two items, compared to when the sentences differed. This suggests that the pediatric CI users were poor at attending to the relevant dimension (i.e., talker voices) and ignoring the irrelevant dimension (i.e., the linguistic content). Further, they found that children who were better on the talker discrimination tasks were also more accurate at recognizing spoken words. Similarly, Tamati and Pisoni (2015) further analyzed the foreign-accent intelligibility ratings by comparing individual differences in accent discrimination scores (i.e., difference in intelligibility ratings for foreign-accented speech compared to native speech). They also found that the accent discrimination scores were related to several measures of sentence recognition abilities. Figure 12–9 displays the relation between the discrimination scores on the foreign-accent rating task and scores on the PRESTO sentence recognition test for prelingually deaf, long-term CI users. Taken together, CI users' ability to deal with real-life speech variability may reflect the development and use of basic speech and language-processing skills. Thus, CI users may benefit from better basic speech and language skills. In addition, some case studies with postlingually deaf adults suggest that CI users may also be able to use prior linguistic knowledge and experience to improve their perception of some sources of real-life speech (e.g., Tamati et al., 2014).

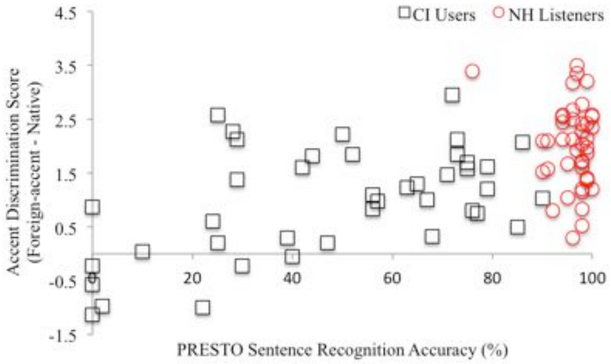

Fig. 12-9: The relation between accent discrimination scores (y-axis) and PRESTO sentence recognition scores (x-axis) for CI users (open black squares) and NH listeners (open red circles). Adapted from Tamati and Pisoni (2015).

Despite the challenges of real-life speech, most previous research studies and clinical assessments with CI users have concerned the effects of sources of environmental degradation, such as background noise or competing talkers, and only a few tests used to assess benefits and outcomes of CI implantation in adults contain speech more characteristic of everyday speech environments. As a consequence, we do not have a full picture of the potential communicative challenges CI users face in their everyday lives, nor do we have the appropriate tools available to treat or improve their communicative abilities.

New Assessment Techniques

The conventional tests of spoken word recognition commonly used in the clinics have been developed with very simple, familiar materials, slowly and clearly produced by a single talker with no discernable accent. The Hearing in Noise Test (HINT; Nilsson, Soli, & Sullivan, 1994) is an example of a widely used conventional, low-variability sentence recognition test. In United States clinics, it is commonly used as a tool to determine CI candidacy and measure outcome and benefit for CI users (e.g., Fabry et al., 2009). The sentences in the HINT are short and syntactically simple, selected or modified to have roughly the same length and the same intelligibility at a fixed noise level, and were produced by a single male talker, with a standard unmarked General American regional dialect. Because listeners benefit from the simple structure of the sentence and the lack of natural talker variability, such sentence tests, like the HINT, likely result in artificially high sentence recognition scores and ceiling effects (e.g., Gifford et al., 2008). Additionally, these simple clinical tests likely do not capture the actual challenges CI users may face in real life and as such fail to identify listeners who may struggle in real-life complex listening environments. Using these materials may also mask potential beneficial outcomes of some new device features or training approaches, leading manufactures and clinicians to discount them.

With advances in device technology, signal-processing strategies, and the expansion of implant candidacy criteria to listeners with greater residual hearing (Gifford et al., 2010), more sensitive tests are thus needed to assess candidacy, as well as real-life outcomes and benefit from CI use. Recently, several new, more challenging sentence recognition tests have been developed with materials that incorporate more sources of natural speech variability, such as the Az-Bio test (Spahr & Dorman, 2004), the Az-TIMIT test (King et al., 2012), the STARR test (Sentence Test with Adaptive Randomized Roving levels; Boyle, Nunn, O'Connor, & Moore, 2013), and the PRESTO test (Perceptually Robust English Sentence Test Open-Set; Gilbert et al., 2013). These new tests require a listener to adapt to a talker's individual idiolectal speech characteristics. The PRESTO test, for example, was designed to contain a great deal of natural variability to reflect foundational components of real-life listening environments. The test includes sentences produced by unique (i.e., no talker was repeated within a list) male and female talkers from several different dialect regions of the United States. The PRESTO test has been shown to be more challenging for CI users, compared to conventional low-variability materials, better reflecting the basic speech perception and neurocognitive abilities (Faulkner, Tamati, & Pisoni, 2016; Tamati & Pisoni, 2015). Figure 12–10 displays individual scores of prelingually deaf long-term CI users on HINT and PRESTO. Although listeners who performed well on the HINT sentences also tended to perform well on the PRESTO sentences, group and individual patterns of performance differed by test. As can be seen in the figure, the CI users were near ceiling on the HINT sentences in the quiet, also demonstrating little listener variability. The CI users were less accurate at recognizing words on the PRESTO sentences, and they also showed a vast amount of individual variability.

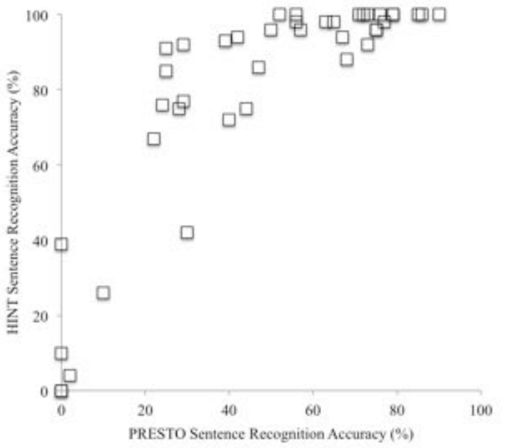

Fig. 12-10: The relation between HINT (y-axis) and PRESTO (x-axis) sentence recognition scores for CI users. Adapted from Tamati and Pisoni (2015) and Faulkner, Tamati, et al. (2016).

Despite development of these new tests that include natural variability more reflective of real-life environments, most research and clinical tests on CI speech perception still focus on a patient's ability to recognize ideal speech. In addition, few high-variability sentence recognition tests have been developed for different languages, and there is a lack of high-variability materials and resources for perceptual training tools in the clinic. As a result, our knowledge of the mechanisms underlying CI speech perception and spoken word recognition remains limited. More widespread use of these tests that incorporate more natural variability, and future materials, within the clinic and research labs, will thus allow us to better understand CI users' real-life speech perception abilities, as well as the vast amounts of individual variability in these abilities. Further, future improvements on clinical and research materials for CI users will allow for the development of more effective rehabilitation programs to meet the real-life needs of this patient population.

New Training Approaches

The basic auditory and cognitive processes underlying speech perception performance, as outlined above, have also been a focus of training programs developed for CI users. The basic premise of these programs is that improving CI users' skills in these areas will also lead to improvements in real-life speech communication. Conventional training programs aiming at improving auditory and perceptual skills have focused on challenging areas for CI users, such as fine-grained discrimination of phonemes or pitch, or speech in noise perception. Auditory training using a more bottom-up approach involves training listeners to attend to fine-grained acoustic details to identify a linguistic category, such as a vowel or consonant, or attend to the relevant dimension of a target item, such as a target word in noise. Studies using this approach have been successful at improving vowel and consonant recognition (Dawson & Clark, 1997) and lexical tone recognition in Mandarin, a tonal language (Wu et al., 2007), in CI users. Other studies have demonstrated that speech-in-noise training yielded improvements on speech recognition in noise for CI users (Fu & Galvin, 2008; Ingvalson et al., 2013).

These studies have shown that there is great potential in providing targeted training programs to CI users in improving speech perception and related auditory skills. However, as is often seen in training, most improvement has usually been observed in the skills that are specifically used in the training, and a transfer of learning to other related skills has been limited. Such a transfer of learning has been often seen with musicians, suggesting an overlap between music and speech recognition networks in the auditory system, leading to better use of acoustic cues for speech perception, and perhaps also an improvement in general cognitive skills related to auditory perception that can both be applied to speech and music perception (Besson et al., 2011; Micheyl et al., 2006). As a result, several studies have shown that musically trained individuals perform better not only in music-related tasks but also in perception of speech in noise, compared to nonmusicians (e.g., Başkent & Gaudrain, 2016; Parbery-Clark et al., 2009). However, this benefit seems to depend on the specific form of interfering noise (Fuller et al., 2014; Swaminathan et al., 2014). In CI users, music training can improve performance in music-related tasks (Galvin et al., 2007, 2012; Gfeller et al., 2002), but a transfer of learning to speech perception has not yet been shown. Only with CI simulations, recently, Fuller et al. (2014) have shown that the musician advantage can persist with degraded stimuli, but again the benefit seemed to become smaller from music- to speech-related perceptual tasks. On the other hand, music, specifically its perception and appreciation, is another challenging area for CI users (e.g., Fuller et al., 2012; Galvin et al., 2009; Gfeller et al., 2002), on which training with music could also potentially have a positive effect (Looi et al., 2012; Yücel et al., 2009). An additional advantage of training with music could be the fun factor, encouraging the CI users to participate more actively. A pilot study from our group has shown that CI users seem to be more willing to participate in an interactive and fun training program, such as music therapy, than in a scientifically proven but less interactive program, such as computer-based training (Free et al., 2014). Perhaps a combination of two would provide a both fun but also beneficial training option to CI users.

These same principles motivate cognitive training programs with CI users as well. Speech and language outcomes of CI users have been linked to individual differences in underlying neurocognitive processes. In particular, studies on language and cognitive development in CI users, as well as individual differences among CI users, have identified working memory as a key component of speech development and processing. As such, a new direction for cognitive training for CI users is to use training to modify working memory capacity in order to improve speech and language outcomes of CI users. That is, skills that are acquired through the working memory training can be transferred to speech perception skills. Kronenberger and colleagues (2011) used working memory training with nine prelingually deaf children with CIs. They found that the children improved on most of the training exercises and also improved, compared to baseline scores, on measures of verbal and nonverbal working memory, parent reports of the child's memory behavior, and sentence repetition immediately following training. Improvements declined after training for the working memory measures, with no improvements remaining after 6 months, but improvements with sentence repetition were more lasting with improvements being maintained after 6 months. In a larger study, Ingvalson, Young, and Wong (2014) trained the phonologic awareness and working memory skills of 10 prelingually deaf children with CIs, with an additional 9 children with CIs serving as controls without training. They found that the children who completed the training showed improvements on oral expressive language and spoken language composite scores, compared to the untrained control listeners. These studies suggest that cognitive training may be a useful new direction for improving speech and language outcomes of CI users.

Taken together, training CI users to make better use of the degraded sensory information they are receiving, in combination with advances in improving the quality of this sensory input, may help CI users achieve greater speech and language outcomes. Further, combining the current and more traditional training methods with new approaches based on new findings on CI speech perception in adverse conditions may offer a path to achieve real-life speech communication benefits for current and future CI users.

Summary

Cochlear implants have provided many deaf individuals the function of hearing, and hence, the ability to communicate. The speech signal transmitted via CIs differs from that of normal hearing and is spectrotemporally degraded. Many CI users seem to adapt to and learn to make use of this degraded speech for communication. However, speech perception in adverse conditions still seems to be a challenge, and there is also a large amount of variability in CI outcomes across individuals. Demographics and device- and physiology-related factors have already been identified as contributing to this interindividual variability in adaptation success and limitations. Recent research shows that, among acoustic factors, the voice-related cues seem to contribute greatly to speech perception, and these are also the cues that are not delivered properly by the device. Further, cognitive processes of speech perception and potential top-down enhancement mechanisms may be altered by the degraded speech input of a CI, and possibly, individual cognitive abilities also contribute to the variability across individuals. Traditionally, the tools used for research on CIs and the rehabilitation for CI users have focused mainly on measures of speech intelligibility, capturing only the end result of speech perception. These, however, do not reveal changes or interactions of speech perception and comprehension mechanisms with degraded speech, especially in more realistic real-life listening environments where the CI users have to deal with different types of background noises, as well as various realizations of speech, such as reduced speech, regional accents, or other types of speaker-induced variability. In this chapter, we present the most contemporary research on these areas. We cover new techniques and methods, which also take into account cognitive factors that can more fully identify the performance and limitations of speech perception and comprehension by CI users, especially in more realistic listening conditions.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh, UK: Edinburgh University Press..

Akeroyd, M. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology, 47*(Suppl. 2), S53–S71.

Arnoldner, C., Riss, D., Brunner, M., Durisin, M., Baumgartner, W. D., & Hamzavi, J. S. (2007). Speech and music perception with the new fine structure speech coding strategy: Preliminary results. *Acta Oto-laryngologica, 127*(12), 1298–1303.

AuBuchon, A. M., Pisoni, D. B., & Kronenberger, W. G. (2015). Verbal processing speed and executive functioning in long-term cochlear implant users. *Journal of Speech, Language, and Hearing Research, 58*(1), 151–162.

Baddeley, A. D. (1997). *Human memory: Theory and practice*. Hove, UK: Psychology Press.

Başkent, D. (2012). Effect of speech degradation on top-down repair: Phonemic restoration with simulations of cochlear implants and combined electric–acoustic stimulation. *Journal of the Association for Research in Otolaryngology, 13*(5), 683–692.

Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America (in press).*

Başkent, D., & Shannon, R. V. (2004). Frequency-place compression and expansion in cochlear implant listeners. *The Journal of the Acoustical Society of America, 116*(5), 3130–3140.

Başkent, D., & Shannon, R. V. (2005). Interactions between cochlear implant electrode insertion depth and frequency-place mapping. *The Journal of the Acoustical Society of America, 117*(3, Pt. 1), 1405–1416.

Beer, J., Kronenberger, W. G., & Pisoni, D. B. (2011). Executive function in everyday life: Implications for young cochlear implant users. *Cochlear Implants International, 12* (Suppl. 1), S89–S91.

Benard, M. R., & Başkent, D. (2014). Perceptual learning of temporally interrupted and spectrally degraded speech. *The Journal of the Acoustical Society of America, 136*(3), 1344–1351.

Benard, M. R., Mensink, J. S., & Başkent, D. (2014). Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities. *The Journal of the Acoustical Society of America, 135*(2), EL88–EL94.

Berliner, K. I., Tonokawa, L. L., Dye, L. M., & House, W. F. (1989). Open-set speech recognition in children with a single-channel cochlear implant. *Ear and Hearing, 10*(4), 237–242.

Bess, F. H., & Hornsby, B. W. (2014). Commentary: Listening can be exhausting—Fatigue in children and adults with hearing loss. *Ear and Hearing, 35*(6), 592–599.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in psychology,* 2, 94..

Bhargava, P., Gaudrain, E., & Başkent, D. (2014). Top–down restoration of speech in cochlear-implant users. *Hearing Research, 309,* 113–123.

Bhargava, P., Gaudrain, E., & Başkent, D. (2016). The intelligibility of interrupted speech: Cochlear implant users and normal hearing listeners. *Journal of the Association for Research in Otolaryngology, under revision*.

Bierer, J. A., & Faulkner, K. F. (2010). Identifying cochlear implant channels with poor electrode-neuron interface: Partial tripolar, single-channel thresholds and psychophysical tuning curves. *Ear and Hearing, 31*(2), 247–258.

Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., . . . Gallégo, S. (2013). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients. *Audiology and Neurotology, 18*(1), 36–47.

Boyle, P. J., Nunn, T. B., O'Connor, A. F., & Moore, B. C. (2013). STARR: A speech test for evaluation of the effectiveness of auditory prostheses under realistic conditions. *Ear and Hearing, 34*(2), 203–212.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*(3), 255–272.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109* (3), 1101–1109.

Chatterjee, M., Sarampalis, A., & Oba, S. I. (2006). Auditory stream segregation with cochlear implants: A preliminary report. *Hearing Research, 222*(1), 100–107.

Chatterjee, M., Zion, D. J., Deroche, M. L., Burianek, B. A., Limb, C. J., Goren, A. P., . . . Christensen, J. A. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hearing Research, 322*, 151–162.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America, 25*(5), 975–979.

Clark, G. M. (2015). The multi-channel cochlear implant: Multi-disciplinary development of electrical stimulation of the cochlea and the resulting clinical benefit. *Hearing Research, 322*, 4–13.

Clarke, J., Başkent, D., & Gaudrain, E. (2016). Pitch and spectral resolution: A systematic comparison of bottom-up cues for top-down repair of degraded speech. *The Journal of the Acoustical Society of America 139, 395-405.*

Cleary, M., & Pisoni, D. B. (2002). Talker discrimination by prelingually deaf children with cochlear implants: Preliminary results. *The Annals of Otology, Rhinology & Laryngology. Supplement, 189*, 113–118.

Cleary, M., Pisoni, D. B., & Kirk, K. I. (2000). Working memory spans as predictors of spoken word recognition and receptive vocabulary in children with cochlear implants. *The Volta Review, 102*(4), 259–280.

Cleary, M., Pisoni, D. B., & Kirk, K. I. (2005). Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. *Journal of Speech, Language, and Hearing Research, 48*(1), 204–223.

Clopper, C. G., & Bradlow, A. R. (2008). Perception of dialect variation in noise: Intelligibility and classification. *Language and Speech, 51*(3), 175–198.

Collison, E. A., Munson, B., & Carney, A. E. (2004). Relations among linguistic and cognitive skills and spoken word recognition in adults with cochlear implants. *Journal of Speech, Language, and Hearing Research, 47*(3), 496–508.

Conway, C. M., Karpicke, J., Anaya, E. M., Henning, S. C., Kronenberger, W. G., & Pisoni, D. B. (2011). Nonverbal cognition in deaf children following cochlear implantation: Motor sequencing disturbances mediate language delays. *Developmental Neuropsychology, 36*(2), 237–254.

Crew, J. D., Galvin, J. J., III, & Fu, Q.-J. (2012). Channel interaction limits melodic pitch perception in simulated cochlear implants. *The Journal of the Acoustical Society of America, 132*(5), EL429–EL435.

Cullington, H., & Zeng, F. (2010). Bimodal hearing benefit for speech recognition with competing voice in cochlear implant subject with normal hearing in contralateral ear. *Ear and Hearing, 31*(1), 70–73.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology Learning Memory and Cognition, 30*(2), 498–513.

Danley, M., & Fretz, R. (1982). Design and functioning of the single-electrode cochlear implant. *The Annals of Otology, Rhinology & Laryngology. Supplement, 91*(2, Pt. 3), 21.

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America, 114*(5), 2913–2922.

Dawson, P., & Clark, G. M. (1997). Changes in synthetic and natural vowel perception after specific training for congenitally deafened patients using a multichannel cochlear implant. *Ear and Hearing, 18*(6), 488–501.

De Cheveigné, A. (1999). Waveform interactions and the segregation of concurrent vowels. *The Journal of the Acoustical Society of America, 106*, 2959–2972.

Djourno, A., & Eyriès, C. (1957). Prothese auditive par excitation electrique a distance du nerf sensoriel a laide dun bobinage inclus a demeure. *Presse médicale, 65*(63), 1417–1417.

Dorman, M., & Spahr, A. (2006). *Speech perception by adults with multichannel cochlear implants*. New York: Thieme Medical Publishers.

Dorman, M. F., Loizou, P. C., & Rainey, D. (1997). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *The Journal of the Acoustical Society of America, 102*(5), 2993–2996.

Douek, E. E., Fourcin, A. J., Moore, B. C. J., & Clark, G. P. (1977). A new approach to the cochlear implant. *Proceedings of the Royal Society of Medicine, 70*, 379–383.

Drennan, W. R., Oleson, J. J., Gfeller, K., Crosson, J., Driscoll, V. D., Won, J. H., . . . Rubinstein, J. T. (2015). Clinical evaluation of music perception, appraisal and experience in cochlear implant users. *International Journal of Audiology, 54*(2), 114–123.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language, 81*(1), 162–173.

Fabry, D., Firszt, J., Gifford, R., Holden, L., & Koch, D. (2009). Evaluating speech perception benefit in adult cochlear implant recipients. *Audiology Today, 21*, 36–43.

Faulkner, K. F., Tamati, T. N., Gilbert, J. L., & Pisoni, D. B. (2015). List equivalency of PRESTO for the evaluation of speech recognition. *Journal of the American Academy of Audiology, 26*(6), 582–594.

Faulkner, K. F., Tamati, T. N., & Pisoni, D. B. (2016). Individual differences in high-variability speech recognition: Some new findings using PRESTO in clinical and non-clinical populations. *In preparation.*

Finley, C., & Skinner, M. (2008). Role of electrode placement as a contributor to variability in cochlear implant outcomes. *Otology & Neurotology, 29*(7), 920–928.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America, 106*(3), 1511–1522.

Francis, A. L., & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, & Psychophysics, 71*(6), 1360–1374.

Free, R., Fuller, C., Maat, B., & Baskent, D. (2014). *The effect of music therapy and training on speech and music perception in cochlear-implant users.* Paper presented at the 37th Annual Mid-winter Meeting of the Association for Research in Otolaryngology, San Diego, CA.

Friesen, L. M., Shannon, R. V., Başkent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America, 110*(2), 1150–1163.

Friesen, L. M., Shannon, R. V., & Cruz, R. J. (2005). Effects of stimulation rate on speech recognition with cochlear implants. *Audiology & Neurotology, 10*(3), 169–184.

Fu, Q.-J. (2002). Temporal processing and speech recognition in cochlear implant users. *Neuroreport, 13*(13), 1635–1639.

Fu, Q.-J., & Galvin, J. J., III. (2008). Maximizing cochlear implant patients' performance with advanced speech training procedures. *Hearing Research, 242*(1–2), 198–208.

Fu, Q.-J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology, 6*(1), 19–27.

Fuller, C., Free, R., Maat, B., & Başkent, D. (2012). Musical background not associated with self-perceived hearing performance or speech perception in postlingual cochlear-implant users. *The Journal of the Acoustical Society of America, 132*(2), 1009–1016.

Fuller, C. D., Galvin, J. J., III, Maat, B., Free, R. H., & Başkent, D. (2014). The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience, 8,* 179.

Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., III, Fu, Q.-J., Free, R. H., & Başkent, D. (2014). Gender categorization is abnormal in cochlear implant users. *Journal of the Association for Research in Otolaryngology, 15*(6), 1037–1048.

Galvin, J. J., III, Fu, Q. J., & Shannon, R. V. (2009). Melodic contour identification and music perception by cochlear implant users. *Annals of the New York Academy of Sciences, 1169*(1), 518–533.

Gantz, B. J., Turner, C., Gfeller, K. E., & Lowder, M. W. (2005). Preservation of hearing in cochlear implant surgery: Advantages of combined electrical and acoustical speech processing. *The Laryngoscope, 115*(5), 796–802.

Gaudrain, E., & Başkent, D. (2015a February). *Discrimination of vocal characteristics in cochlear implants.* Paper presented at the 38th Annual Mid-winter Meeting of the Association for Research in Otolaryngology, Baltimore, MD.

Gaudrain, E., & Başkent, D. (2015b). Factors limiting vocal-tract length discrimination in cochlear implant simulations. *The Journal of the Acoustical Society of America, 137*(3), 1298–1308.

Gaudrain, E., & Carlyon, R. P. (2013). Using zebra-speech to study sequential and simultaneous speech segregation in a cochlear-implant simulation. *The Journal of the Acoustical Society of America, 133*(1), 502–518.

Gaudrain, E., Grimault, N., Healy, E. W., & Béra, J. C. (2007). Effect of spectral smearing on the perceptual segregation of vowel sequences. *Hearing Research, 231*, 32–41.

Gaudrain, E., Grimault, N., Healy, E. W., & Béra, J. C. (2008). Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation. *The Journal of the Acoustical Society of America, 124*(5), 3076–3087.

Geers, A. E., & Sedey, A. L. (2011). Language and verbal reasoning skills in adolescents with 10 or more years of cochlear implant experience. *Ear and Hearing, 32*(1 Suppl.), 39S–48S.

Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J. F., . . . Stordahl, J. (2002). Recognition of familiar melodies by adult cochlear implant recipients and normal hearing adults. *Cochlear Implants International, 3*(1), 29–53.

Gifford, R.H., Dorman, M., McKarns, S., & Spahr, A. (2007). Combined electric and contralateral acoustic hearing: Word and sentence recognition with bimodal hearing. *Journal of Speech, Language and Hearing Research, 50*(4), 835–843.

Gifford, R. H., Shallop, J. K., & Peterson, A. M. (2008). Speech recognition materials and ceiling effects: Considerations for cochlear implant programs. *Audiology and Neurotology, 13*(3), 193–205.

Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, reliability and validity of PRESTO: A new high-variability sentence recognition test. *Journal of the American Academy of Audiology, 24*(1), 26–36.

Gnansia, D., Pressnitzer, D., Péan, V., Meyer, B., & Lorenzi, C. (2010). Intelligibility of interrupted and interleaved speech for normal-hearing listeners and cochlear implantees. *Hearing Research, 265*(1), 46–53.

Grayden, D. B., & Clark, G. M. (2006). Implant design and development. In H. Cooper & L. Craddock (Eds.), *Cochlear implants: A practical guide* (pp. 1–20). Chichester, UK: Whurr.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *The Journal of the Acoustical Society of America, 87*, 2592–2605.

Harris, M. S., Kronenberger, W. G., Gao, S., Hoen, H. M., Miyamoto, R. T., & Pisoni, D. B. (2013). Verbal short-term memory development and spoken language outcomes in deaf children with cochlear implants. *Ear and Hearing, 34*(2), 179–192.

Henry, B. A., Turner, C. W., & Behrens, A. (2005). Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners. *The Journal of the Acoustical Society of America, 118*(2), 1111–1121.

Heydebrand, G., Hale, S., Potts, L., Gotter, B., & Skinner, M. (2007). Cognitive predictors of improvements in adults’ spoken word recognition six months after cochlear implant activation. *Audiology and Neurotology, 12*(4), 254–264.

Hochmair, E. S., & Hochmair-Desoyer, I. J. (1983). Percepts elicited by different speech coding strategies. In C. W. Parkins & S. W. Anderson (Eds.), *Cochlear prostheses—An international symposium* (Vol. 405). New York: Annals of the New York Academy of Sciences.

Hochmair, I., Hochmair, E., Nopp, P., Waller, M., & Jolly, C. (2015). Deep electrode insertion and sound coding in cochlear implants. *Hearing Research, 322*, 14–23.

Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., . . . Skinner, M. W. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear Hear, 34*, 342–360.

Hong, R. S., & Turner, C. W. (2009). Sequential stream segregation using temporal periodicity cues in cochlear implant recipients. *The Journal of the Acoustical Society of America, 126*(1), 291–299.

Ingvalson, E. M., Lee, B., Fiebig, P., & Wong, P. C. (2013). The effects of short-term computerized speech-in-noise training on postlingually deafened adult cochlear implant recipients. *Journal of Speech, Language, and Hearing Research, 56*(1), 81–88.

Ingvalson, E. M., Young, N. M., & Wong, P. C. (2014). Auditory–cognitive training improves language performance in prelingually deafened cochlear implant recipients. *International Journal of Pediatric Otorhinolaryngology, 78*(10), 1624–1631.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. i., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*(1), B47–B57.

Iverson, P., Wagner, A., Pinet, M., & Rosen, S. (2011). Cross-language specialization in phonetic processing: English and Hindi perception of/w/-/v/speech and nonspeech. *The Journal of the Acoustical Society of America, 130*(5), EL297–EL303.

Iverson, P., Wagner, A., & Rosen, S. (2016). Effects of language experience on pre-categorical perception: Distinguishing general from specialized processes in speech perception. *Journal of the Acoustical Society of America* (in revision).

Janse, E., Nooteboom, S. G., & Quené, H. (2007). Coping with gradient forms of/t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes, 22*(2), 161–200.

Ji, C., Galvin, J. J., III, Xu, A., & Fu, Q.-J. (2013). Effect of speaking rate on recognition of synthetic and natural speech by normal-hearing and cochlear implant listeners. *Ear and Hearing, 34*(3), 313–323.

Ji, C., Galvin, J. J., Chang, Y.-P., Xu, A., & Fu, Q.-J. (2014). Perception of speech produced by native and nonnative talkers by listeners with normal hearing and listeners with cochlear implants. *Journal of Speech, Language, and Hearing Research, 57*(2), 532–554.

Johnson, K., & Mullennix, J. W. (1997). *Talker variability in speech processing*. San Francisco, CA: Morgan Kaufmann Publishers Inc.

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science, 24* (10)*,* 1995–2004.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kasturi, K., Loizou, P. C., Dorman, M., & Spahr, T. (2002). The intelligibility of speech with "holes" in the spectrum. *The Journal of the Acoustical Society of America, 112*(3, Pt. 1), 1102–1111.

Khan, A. M., Handzel, O., Burgess, B. J., Damian, D., Eddington, D. K., & Nadol, J. B., Jr. (2005). Is word recognition correlated with the number of surviving spiral ganglion cells and electrode insertion depth in human subjects with cochlear implants? *The Laryngoscope, 115*(4), 672–677.

King, S. E., Firszt, J. B., Reeder, R. M., Holden, L. K., & Strube, M. (2012). Evaluation of TIMIT sentence list equivalency with adult cochlear implant recipients. *Journal of the American Academy of Audiology, 23*(5), 313–331.

Kronenberger, W., Beer, J., Castellanos, I., Pisoni, D., & Miyamoto, R. (2014). Neurocognitive risk in children with cochlear implants. *Journal of the American Medical Society Otolaryngology-Head & Neck Surgery, 140,* 608–615.

Kronenberger, W. G., Pisoni, D. B., Henning, S. C., & Colson, B. G. (2013). Executive functioning skills in long-term users of cochlear implants: A case control study. *Journal of Pediatric Psychology, 38,* 902–914.

Kronenberger, W. G., Pisoni, D. B., Henning, S. C., Colson, B. G., & Hazzard, L. M. (2011). Working memory training for children with cochlear implants: A pilot study. *Journal of Speech, Language, and Hearing Research, 54*(4), 1182–1196.

Kwon, B. J., Perry, T. T., Wilhelm, C. L., & Healy, E. W. (2012). Sentence recognition in noise promoting or suppressing masking release by normal-hearing and cochlear-implant listeners. *The Journal of the Acoustical Society of America, 131*(4), 3111–3119.

Laneau, J., Wouters, J., & Moonen, M. (2006). Improved music perception with explicit pitch coding in cochlear implants. *Audiology & Neurotology, 11*(1), 38–52.

Lazard, D., Lee, H., Gaebler, M., Kell, C., Truy, E., & Giraud, A.-L. (2010). Phonological processing in post-lingual deafness and cochlear implant outcome. *Neuroimage, 49*(4), 3443–3451.

Lazard, D. S., Innes-Brown, H., & Barone, P. (2014). Adaptation of the communicative brain to post-lingual deafness. Evidence from functional imaging. *Hearing Research, 307*, 136–143.

Lazard, D. S., Lee, H. J., Truy, E., & Giraud, A. L. (2013). Bilateral reorganization of posterior temporal cortices in post-lingual deafness and its relation to cochlear implant outcome. *Human Brain Mapping, 34*(5), 1208–1219.

Limb, C. J., & Rubinstein, J. T. (2012). Current research on music perception in cochlear implant users. *Otolaryngologic Clinics of North America, 45*(1), 129–140.

Loizou, P. C. (1998). Mimicking the human ear: An overview of signal-processing strategies for converting sound into electrical signals in cochlear implants. *IEEE Signal Processing Magazine*, September, 101–130.

Looi, V., Gfeller, K., & Driscoll, V. (2012). Music appreciation and training for cochlear implant recipients: A review. *Seminars in Hearing*, *33*(04), 307–334.

Luo, X., Fu, Q.-J., & Galvin, J. J. (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification, 11*(4), 301–315.

Luo, X., Fu, Q.-J., Wu, H.-P., & Hsu, C.-J. (2009). Concurrent-vowel and tone recognition by Mandarin-speaking cochlear implant users. *Hearing Research, 256*(1), 75–84.

Marschark, M., Rhoten, C., & Fabich, M. (2007). Effects of cochlear implants on children's reading and academic achievement. *Journal of Deaf Studies and Deaf Education, 12*(3), 269–282.

Mason, H. M. (1946). Understandability of speech in noise as affected by region of origin of speaker and listener. *Communications Monographs, 13*(2), 54–58.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7–8), 953–978.

Merzenich, M. M., Michelson, R. P., Schindler, R. A., Pettit, C. R., & Reid, M. (1973). Neural encoding of sound sensation evoked by electrical stimulation of the acoustic nerve. *Annals of Otology, 82*, 486–503.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219*(1–2), 36–47.

Milczynski, M., Wouters, J., & Van Wieringen, A. (2009). Improved fundamental frequency coding in cochlear implant signal processing. *The Journal of the Acoustical Society of America, 125*, 2260–2271.

Millar, I., Tong, Y., & Clark, G. M. (1984). Speech processing for cochlear implant prostheses. *Journal of Speech, Language, and Hearing Research, 27*(2), 280–296.

Moberly, A. C., Lowenstein, J. H., Tarr, E., Caldwell-Tarr, A., Welling, D. B., Shahin, A. J., & Nittrouer, S. (2014). Do adults with cochlear implants rely on different acoustic cues for phoneme perception than adults with normal hearing? *Journal of Speech, Language, and Hearing Research, 57*(2), 566–582.

Moore, B. C. J., & Carlyon, R. P. (2005). Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In C. J. Plack, A. J. Oxenham, R. R. Fay, & A. N. Popper (Eds.), *Pitch perception* (pp. 234–277). New York: Springer.

Moore, B. C. J., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica, 88*, 320–333.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America, 85*(1), 365–378.

Nelson, P. B., & Jin, S. H. (2004). Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America, 115*(5, Pt. 1), 2286–2294.

Nelson, P. B., Jin, S. H., Carney, A. E., & Nelson, D. A. (2003). Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America, 113*(2), 961–968.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America, 95*, 1085–1099.

Nittrouer, S., Caldwell-Tarr, A., & Lowenstein, J. H. (2013). Working memory in children with cochlear implants: Problems are in storage, not processing. *International Journal of Pediatric Otorhinolaryngology, 77*(11), 1886–1898.

Pals, C., Sarampalis, A., & Başkent, D. (2013). Listening effort with cochlear implant simulations. *Journal of Speech, Language, and Hearing Research, 56,* 1075–1084.

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing, 30*(6), 653–661.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1989). Speaking clearly for the hard of hearing: III. An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech, Language, and Hearing Research, 32*(3), 600–603.

Pisoni, D., Kronenberger, W., Roman, A., & Geers, A. (2011). Measures of digit span and verbal rehearsal speed in deaf children following more than 10 years of cochlear implantation. *Ear and Hearing, 32,* 60S–74S.

Pisoni, D. B., & Cleary, M. (2003). Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear and Hearing, 24*(1, Suppl.), 106S–120S.

Pisoni, D. B., & Geers, A. E. (2000). Working memory in deaf children with cochlear implants: Correlations between digit span and measures of spoken language processing. *The Annals of Otology, Rhinology & Laryngology Supplement, 185*, 92–93.

Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *The Journal of the Acoustical Society of America, 114*, 446–454.

Qin, M. K., & Oxenham, A. J. (2005). Effects of envelope-vocoder processing on f0 discrimination and concurrent-vowel identification. *Ear and Hearing, 26*(5), 451–460.

Rosen, S., Walliker, J., Brimacombe, J. A., & Edgerton, B. J. (1989). Prosodic and segmental aspects of speech perception with the House/3M single-channel implant. *Journal of Speech, Language, and Hearing Research, 32*(1), 93–111.

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., & Barone, P. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proceedings of the National Academy of Sciences, 104*(17), 7295–7300.

Rubinstein, J. T. (2004). How cochlear implants encode speech. *Current Opinion in Otolaryngology & Head and Neck Surgery, 12*(5), 444–448.

Rubinstein, J. T., & Hong, R. (2003). Signal coding in cochlear implants: Exploiting stochastic effects of electrical stimulation. *The Annals of Otology, Rhinology & Laryngology Supplement, 191*, 14–19.

Schoof, T., & Rosen, S. (2014). The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. *Frontiers in Aging Neuroscience, 6*, 307.

Sennaroğlu, L. (2010). Cochlear implantation in inner ear malformations—a review article. *Cochlear Implants International, 11*(1), 4–41.

Shannon, R. V. (1983). Multichannel electrical stimulation of the auditory nerve in man: II. Channel interaction. *Hearing Research, 12*(1), 1–16.

Shannon, R. V., Galvin, J. J., III, & Başkent, D. (2002). Holes in hearing. *Journal of the Association for Research in Otolaryngology, 3*(2), 185–199.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*, 303–304.

Siciliano, C., Faulkner, A., Rosen, S., & Mair, K. (2010). Resistance to learning binaurally mismatched frequency-to-place maps: Implications for bilateral stimulation with cochlear implants. *The Journal of the Acoustical Society of America, 127*, 1645–1660.

Singer, B. D., & Bashir, A. S. (1999). What are executive functions and self-regulation and what do they have to do with language-learning disorders? *Language, Speech, and Hearing Services in Schools, 30*(3), 265–273.

Skinner, M. W., Ketten, D. R., Holden, L. K., Harding, G. W., Smith, P. G., Gates, G. A., & Blocker, B. (2002). CT-derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus-22 recipients. *Journal of the Association for Research in Otolaryngology, 3*(3), 332–350.

Smith, D. R., Patterson, R. D., Turner, R., Kawahara, H., & Irino, T. (2005). The processing and perception of size information in speech sounds. *The Journal of the Acoustical Society of America, 117*(1), 305–318.

Spahr, A. J., & Dorman, M. F. (2004). Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices. *Archives of Otolaryngology-Head & Neck Surgery, 130*(5), 624–628.

Spillman, T., & Dillier, N. (1989). Comparison of single-channel extracochlear and multichannel intracochlear electrodes in the same patient. *British Journal of Audiology, 23*, 25–31.

Stickney, G. S., Loizou, P. C., Mishra, L. N., Assmann, P. F., Shannon, R. V., & Opie, J. M. (2006). Effects of electrode design and configuration on channel interactions. *Hearing Research, 211*(1–2), 33–45.

Stickney, G. S., Zeng, F. G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *The Journal of the Acoustical Society of America, 116*(2), 1081–1091.

Svirsky, M. A., Silveira, A., Suarez, H., Neuburger, H., Lai, T. T., & Simmons, P. M. (2001). Auditory learning and adaptation after cochlear implantation: A preliminary study of discrimination and labeling of vowel sounds by cochlear implant users. *Acta Oto-Laryngologica, 121*(2), 262–265.

Swaminathan, J., Mason, C., Streeter, T., Best, V., Kidd, G., Jr., & Patel, A. (2014). Musical training, individual differences and the cocktail party problem. *Scientific Reports, 5*, 11628.

Tamati, T. N., Janse, E., & Başkent, D. (2015, April). *The perception of real-life speaking styles under cochlear implant simulation.* Paper presented at the Improving Cochlear Implant Performance Meeting 2015, London, UK.

Tamati, T. N., & Pisoni, D. B. (2015). The perception of foreign-accented speech by cochlear implant users. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences,* Glasgow, UK: the University of Glasgow. Tamati, T. N., Gilbert, J. L., & Pisoni, D. B. (2013). Some factors underlying individual differences in speech recognition on PRESTO: A first report. *Journal of the American Academy of Audiology, 24*(7), 616–634.

Tamati, T. N., Gilbert, J. L., & Pisoni, D. B. (2014). Influence of early linguistic experience on regional dialect categorization by an adult cochlear implant user: A case study. *Ear and Hearing, 35*(3), 383–386.

Van Wieringen, A., & Wouters, J. (2015). What can we expect of normally-developing children implanted at a young age with respect to their auditory, linguistic and cognitive skills? *Hearing Research, 322*, 171–179.

Vandali, A. E., & van Hoesel, R. J. (2011). Development of a temporal fundamental frequency coding strategy for cochlear implants. *The Journal of the Acoustical Society of America, 129*(6), 4023–4036.

Vandali, A. E., Whitford, L. A., Plant, K. L., & Clark, G. M. (2000). Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 cochlear implant system. *Ear and Hearing, 21*(6), 608–624.

Venail, F., Mathiolon, C., Menjot de Champfleur, S., Piron, J. P., Sicard, M., Villemus, F., & Uziel, A. (2015). Effects of electrode array length on frequency-place mismatch and speech perception with cochlear implants. *Audiology and Neurotology, 20*(2), 102–111.

Vestergaard, M. D., Fyson, N. R., & Patterson, R. D. (2009). The interaction of vocal characteristics and audibility in the recognition of concurrent syllables. *The Journal of the Acoustical Society of America, 125*(2), 1114–1124.

Wagner, A., Ernestus, M., & Cutler, A. (2006). Formant transitions in fricative identification: The role of native fricative inventory. *The Journal of the Acoustical Society of America, 120*(4), 2267–2277.

Wagner, A., Pals, C., de Blecourt, C., Sarampalis, A., & Başkent, D. (2016). Does signal degradation affect top-down processing of speech? In van Dijk, P., Başkent, D., Gaudrain, E., de Kleine, E., Wagner, A., Lanting, C. (Eds.), *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*. New York, NY: Springer.

Wagner, A., Toffanin, P., & Başkent, D. (2016). The timing and effort of lexical access in natural and degraded speech. *Frontiers (in revision).*

Wagner, A., Opie, J., & Başkent, D. (2016). The time-course of pre-lexical and lexical processing in experienced CI users. In preparation.

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

Warren, R. M., & Sherman, G. L. (1974). Phonemic restorations based on subsequent context. *Attention, Perception, & Psychophysics, 16*(1), 150–156.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., & Rabinowitz, W. M. (1991). Better speech recognition with cochlear implants. *Nature, 352*, 236–238.

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. *The Journal of the Acoustical Society of America, 131*(2), 1465–1479.

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear & Hearing, 36* (4) , e153–e165.

Wöstmann, M., Schröger, E., & Obleser, J. (2015). Acoustic detail guides attention allocation in a selective listening task. *Journal of Cognitive Neuroscience, 27*(5)*,* 988–1000.

Wu, J.-L., Yang, H.-M., Lin, Y.-H., & Fu, Q.-J. (2007). Effects of computer-assisted speech training on Mandarin-speaking hearing-impaired children. *Audiology and Neurotology, 12*(5), 307–312.

Yücel, E., Sennaroğlu, G., & Belgin, E. (2009). The family oriented musical training for children with cochlear implants: Speech and musical perception results of two year follow-up. *International Journal of Pediatric Otorhinolaryngology, 73*(7), 1043–1052.

Zeng, F.-G. (2004). Trends in cochlear implants. *Trends in Amplification, 8*(1), 1–34.

Zeng, F.-G., Grant, G., Niparko, J., Galvin, J. J., III, Shannon, R., Opie, J., & Segel, P. (2002). Speech dynamic range and its effect on cochlear implant performance. *The Journal of the Acoustical Society of America, 111*(1, Pt. 1), 377–386.

Zwolan, T. A., Kileny, P. R., Ashbaugh, C., & Telian, S. A. (1996). Patient performance with the Cochlear Corporation “20+2” implant: Bipolar versus monopolar activation. *Otology & Neurotology, 17*(5), 717–723.