

Published in final edited form as:

Ear Hear. 2011 ; 32(4): 511–523. doi:10.1097/AUD.0b013e3182064a87.

Perception of environmental sounds by experienced cochlear implant patients

Valeriy Shafiro¹, Brian Gygi², Min-Yu Cheng¹, Jay Vachhani¹, and Megan Mulvey¹

¹ Rush University Medical Center, Communication Disorders and Sciences, 600 S. Paulina Street Chicago, IL 60612-3833

² Veterans Affairs of Northern California & East Bay Institute for Research and Education 150 Muir Road, Martinez, CA 94553

Abstract

Objectives—Environmental sound perception serves an important ecological function by providing listeners with information about objects and events in their immediate environment. Environmental sounds such as car horns, baby cries or chirping birds can alert listeners to imminent dangers as well as contribute to one's sense of awareness and well being. Perception of environmental sounds as acoustically and semantically complex stimuli, may also involve some factors common to the processing of speech. However, very limited research has investigated the abilities of cochlear implant (CI) patients to identify common environmental sounds, despite patients' general enthusiasm about them. This project (1) investigated the ability of patients with modern-day CIs to perceive environmental sounds, (2) explored associations among speech, environmental sounds and basic auditory abilities, and (3) examined acoustic factors that might be involved in environmental sound perception.

Design—Seventeen experienced postlingually-deafened CI patients participated in the study. Environmental sound perception was assessed with a large-item test composed of 40 sound sources, each represented by four different tokens. The relationship between speech and environmental sound perception, and the role of working memory and some basic auditory abilities were examined based on patient performance on a battery of speech tests (HINT, CNC, and individual consonant and vowel tests), tests of basic auditory abilities (audiometric thresholds, gap detection, temporal pattern and temporal order for tones tests) and a backward digit recall test.

Results—The results indicated substantially reduced ability to identify common environmental sounds in CI patients (45.3%). Except for vowels, all speech test scores significantly correlated with the environmental sound test scores: $r = 0.73$ for HINT in quiet, $r = 0.69$ for HINT in noise, $r = 0.70$ for CNC, $r = 0.64$ for consonants and $r = 0.48$ for vowels. HINT and CNC scores in quiet moderately correlated with the temporal order for tones. However, the correlation between speech and environmental sounds changed little after partialing out the variance due to other variables.

Conclusions—Present findings indicate that environmental sound identification is difficult for CI patients. They further suggest that speech and environmental sounds may overlap considerably in their perceptual processing. Certain spectro-temporal processing abilities are separately associated with speech and environmental sound performance. However, they do not appear to mediate the relationship between speech and environmental sounds in CI patients. Environmental sound rehabilitation may be beneficial to some patients. Environmental sound testing may have potential diagnostic applications, especially with difficult-to-test populations, and could be predictive of speech performance for prelingually deafened patients with cochlear implants.

Introduction

Environmental sound perception is an important component of daily living and a significant concern for cochlear implant (CI) patients (Zhao et al., 1997; Tyler, 1994). Accurate identification of the sources of environmental sounds (defined as all naturally occurring sounds other than speech and music, based on Gygi et al., 2007) provides a number of important benefits. These range from more obvious safety concerns (e.g., avoiding an approaching car, reacting to an alarm or a gunshot) to more subtle effects that include a general sense of environmental awareness and aesthetic satisfaction (enjoying a bird song, waves falling on the beach, or the sound of rain and rustling leaves). Furthermore, accurate recognition of familiar environmental sounds, used as maskers, may also lead to improved speech perception (Oh & Lutfi, 1999). Previous studies have shown that normal hearing (NH) listeners demonstrate accurate identification of a large number of environmental sounds with little difficulty (Ballas, 1993; Marcell et al., 2000; Gygi et al., 2004; Shafiro, 2008a) and can report a wealth of detailed information about the properties of sound sources (Freed, 1990; Li et al., 1991; Kunkler-Peck & Turvey, 2000; Warren & Verbrugge, 1984). On the other hand, environmental sound identification may present considerable difficulty for CI patients who receive distorted sensory input, and after a period of deafness, may often need to relearn many common environmental sounds (Tye-Murray et al., 1992; Reed & Delhorne, 2005; Inverso & Limb, 2010).

In general, CI patients show a great deal of enthusiasm when they “discover” the sources of many environmental sounds following implantation. Although anecdotally poor environmental sound perception is not a common complaint compared to speech perception, available environmental sound tests indicate a generally poor-to-mediocre performance among CI patients (Inverso & Limb, 2010; Proops et al., 1999; Tyler et al., 1989; CHABA, 1991 for summary of earlier research). This contradiction between the apparent lack of overt complaints and unsatisfactory test performance could result from patients not being aware of the richness and multitude of environmental sounds in their environment. Instead of perceiving appropriate meaning of sounds around them, patients may perceive these sounds as generic uninterpretable noise with little informational value. In addition, the main emphasis in implant adjustments and rehabilitation strategies has traditionally been on speech perception. Thus, the absence of frequent complaints may not be a reliable indicator of adequate ability or lack of interest among CI patients.

Among a relatively small number of existing studies that included environmental sound perception testing, many were conducted in the early days of implant development to provide a comprehensive assessment of implant efficacy (Eisenberg et al., 1983; Tyler et al., 1985; Edgerton et al., 1983; Maddox & Porter, 1983; Schindler & Dorcas, 1987). Patients in these studies used single or few-channel devices (e.g., House Ear Institute, Vienna, UCSF/Storz) and had a profound hearing loss for an extensive time period prior to implantation. Early environmental sound test materials typically included a small number (20 or fewer) of arbitrarily selected items, with each sound source typically represented by a single token. Normative data and rationale for stimulus selection and scoring procedures were not typically reported and were not always used consistently across studies. These potential limitations inherent in environmental sound testing in early studies may perhaps explain a large range in CI performance from close to chance to nearly perfect. For example, in the absence of clear and consistent response scoring guidelines, such responses as ‘pot breaking’ and ‘glass’ given to a ‘glass breaking’ stimulus may not be accurately scored. Therefore results from the early studies may give a skewed sense of environmental sound perception abilities, and due to advances in signal processing and implant design, may not be applicable to modern-day implant patients.

Two recent studies have revisited environmental sound perception in CI patients using more comprehensive and rigorously selected sets of materials. First, Reed & Delhorne (2005) examined environmental sound perception in 11 experienced post-lingual CI patients (9 Ineraid and 2 Clarion with 6 and 8 electrodes, respectively). Environmental sound test stimuli consisted of 40 sound sources, grouped by 10 typical settings in which these sounds could be encountered (i.e. Home, Kitchen, Office, Outside). Each sound source was represented by three separate tokens. For each stimulus, patients selected one of 10 response options associated with a specific environmental setting. Patients' average performance across settings was 79.2% with relatively little variation across the settings (range 87.2% 'Office' – 77.6% 'Kitchen'). On the other hand, individual patient performance was more variable, ranging from 45.3% - 93.8% correct. Thus, a number of CI patients tested by Reed and Delhorne seemed to have generally satisfactory environmental sound identification abilities in quiet, while others had greater difficulty.

The findings from a second recent study by Inverso & Limb (2010) showed a greater decrement in environmental sound perception in CI patients. Twenty two experienced post-lingual users of recent Cochlear Corporation (12) and Advance Bionics (10) devices, with 22 and 16 electrodes respectively, were presented with a beta version of the NonLinguistic Sound test (NLST) developed by the authors for clinical testing. The NLST was administered in an open set identification format and consisted of 50 sound sources drawn from five different categories (Animal, Human, Mechanical/Alerting, Nature, Musical instruments). The test sounds were arranged in three 50-item lists (each patient was tested with one list only), with each source represented by a single token within a list, and some sources appearing only in some but not all lists. The results indicated a poor mean sound identification performance of 49% correct across tokens, although 71% of the patient responses belonged to the correct sound source category (among the five categories used to classify sounds during selection). Perhaps, the differences in the stimuli and testing procedures between the study of Reed & Delhorne and that of Inverso & Limb may account for the discrepancies in the reported environmental sound scores. Nevertheless, given the paucity of formal assessments of environmental sound perception in contemporary CI users and variations in the findings of the two studies above, it is difficult to derive definitive conclusions about modern-day CI patients' environmental sound perception abilities.

Furthermore, there are some indications in both recent studies described above, as well as some earlier assessments (Tyler et al., 1989) of an association between patients' environmental sound performance and their speech perception scores. Inverso & Limb (2010) found significant correlations in the 0.5 – 0.58 range between environmental sound identification and HINT and CNC scores in quiet. Reed and Delhorne (2005), while not reporting a specific correlation coefficient, observed that patients with NU-6 scores “greater than roughly 34% correct were able to recognize environmental sounds with a higher degree of accuracy,” and that “the types of confusions made in identifying environmental sounds appear to share a similar basis with those observed in studies of speech recognition through cochlear implants.” Similarly, correlational analyses performed on the speech and environmental sound accuracy data reported by Tyler and colleagues (1989) revealed correlations in the 0.51 – 0.78 range between speech and environmental sound measures.

These findings are not surprising because of a number of notable similarities in the acoustic structure and the perceptual processing of speech and environmental sounds. Both are considerably more complex than physically simple stimuli commonly used in psychoacoustic research. Previously a relationship between perceptual processing of speech and environmental sounds has been shown in studies of patients with cerebral damage (Schindler et al., 1994; Baddeley & Wilson, 1993; Piccirilli et al., 2000). Neuroimaging studies also demonstrated the involvement of similar cortical regions during the processing

of environmental sounds and speech (Lewis et al., 2004), lending some support to earlier claims about similar perceptual and cognitive factors involved in perceptual processing of speech and environmental sounds (Reed & Delhorne, 2005; Gygi et al., 2004; Inverso & Limb, 2010; Watson & Kidd, 2002). For instance, scores on the auditory digit span test commonly used to assess working memory have been found to correlate with speech perception scores of CI patients (Pisoni & Cleary, 2003; Pisoni & Geers, 2001; Pisoni, 2005). On the other hand, given rather severe “peripheral” limitations imposed on sensory input to the auditory system by even the most advanced implant devices to date, it is also possible that an association between speech and environmental sound perception in CI patients is largely driven by more general “peripheral” limiting factors and associated constraints on basic abilities to process auditory sensory input.

Indeed, scores on several psychoacoustic tests that assessed “lower-level” basic spectral and temporal processing were moderately associated with speech perception. In CI patients, measures of temporal envelope processing included gap detection (Tyler et al., 1989; Muchnik et al., 1994), temporal modulation thresholds (Cazals et al., 1994; Fu, 2002) and as a measure of temporal fine structure, detection of opposite Schroeder-phase chirps (Drennan et al., 2008). Other relevant measures included judging the direction of pitch change (Gfeller et al., 1997), place-pitch sensitivity (Donaldson & Nelson, 2000) and detection of spectral ripple density (Won et al., 2007). Overall, these findings support the notion that a relationship between speech and environmental sounds in CI patients might be mediated by basic auditory abilities in processing “low-level” spectral and temporal information as a general limiting factor. Alternatively, if current devices can support sufficiently robust sensory information, the relationship between speech and environmental sounds in CI patients, similar to normal-hearing (NH) listeners, could be also mediated to a considerable degree by more central “higher order” perceptual and cognitive processes.

Another approach to determining the factors that affect CI patients' ability to identify environmental sounds is to examine acoustic parameters that could potentially account for identification performance. The majority of CIs to date use envelope-based signal processing strategies that degrade the acoustic signal in systematic ways, generally preserving dynamic envelope information, but limiting access to acoustic cues to sound identification contained in the narrowband spectra and temporal fine structure. It is possible that some acoustic parameters might predict CI patients' identification performance. Some evidence for this approach was shown in Gygi et al. (2004) in which the identifiability of environmental sounds processed with vocoder-like methods that simulated implant processing was predicted fairly well by a linear combination of four to seven acoustic variables, described below in the methods section. The best model used seven variables and accounted for 62% of the variance in the sounds' identifiability. Shafiro (2008a) found, using discriminant analysis on similarly processed environmental sounds, that only two acoustic parameters (number of bursts in the envelope and standard deviation of centroid velocity) could predict with 83% accuracy whether sounds were identified with fewer spectral channels (<8) or required a greater number of channels to be recognized. Similar factors may be expected to play a role in the perception of environmental sounds by CI patients. Other studies with NH listeners have also used acoustic parameters to predict the similarity ratings of environmental sounds (Gygi, Kidd, and Watson, 2007), semantic attributes of environmental sound (Bonebright, 2001) and musical timbre (Lakatos, 2000).

The present study was conducted to further extend our understanding of environmental sound perception abilities in CI patients. Specifically, the goals of this study were 1) to supplement existing research on the ability of CI patients with modern-day devices to identify common environmental sounds; 2) to examine an association between speech and environmental sound perception and evaluate the role of some low-level basic auditory

abilities in mediating the relationship between environmental sounds and speech; 3) to examine acoustic parameters that may account for environmental sound perception in CI patients.

Methods

Participants

Seventeen CI patients (5 males, 12 females) took part in the study (Table 1). Five participants had Advanced Bionics (Clarion and Auria) and 12 had Cochlear Corporation implants (Freedom and Nucleus 24), all with full electrode insertion. Participants' warble tone thresholds at octave frequencies between 250 Hz – 4000Hz (with CI in sound field) were 36 dB HL or better; average PTA (i.e. for 500 Hz, 1000 Hz, 2000 Hz) of 29 dB HL, SD = 7.87, range 16 - 43). No participant had any re-programming performed on their implant for at least six months prior to participation in the present study. Participants' average age was 58 years, SD = 10, ranging from 40 – 80 years. All participants passed a 30-point Mini Mental State examination (Folstein et al., 1975) with scores of 29 and above to confirm adequate cognitive function. Participants' average implant experience was 3.2 years, with a minimum of one year of daily implant use (range 1 – 7 years). The age of hearing loss onset varied considerably among participants between birth to 50 years of age, with an average of 16 years, SD = 16. However, participants identified themselves as being fluent and proficient oral communicators prior to developing severe-profound hearing loss (i.e. post-lingually implanted). All participants were native English speakers.

Stimuli and procedures

The environmental sound perception abilities were assessed using a previously developed environmental sound test (Shafiro, 2008b). Participants were also administered tests of speech perception and basic auditory abilities as described below. All stimuli were presented in a double-walled sound booth at the average intensity level of 70 dB SPL, using a single speaker positioned at a distance of one meter to the implanted ear at a 45° angle to the participant's midline on the same side as the implant (to accommodate interactive responses on a computer monitor and avoid obstructing the space between the ear and the loud speaker). HINT and CNC tests were administered at a 0° angle. During testing, CI processor settings were those participants found most comfortable during everyday listening. When applicable, participants were instructed to remove hearing aids from the non-implanted ear. During testing the non-implanted ear was also occluded with an E-A-R classic disposable foam ear plug (NRR 29 dB). The total testing time was approximately 1.5 – 2 hours per subject. Participants were encouraged to take breaks between tests to avoid fatigue.

The **environmental sound test** addressed the potential limitations of earlier tests as discussed above by including a large collection of common environmental sounds selected based on their ecological frequency of occurrence (Ballas, 1993), familiarity, and broad representation of different taxonomic categories of environmental sounds (Gaver, 1993). The present test consisted of a collection of 160 environmental sounds which comprised 40 sound sources, each represented by four separate sound tokens. The sounds were drawn from environmental sound sets used in previous research (Ballas, 1993; Gygi et al., 2004; Marcell et al., 2000; Shafiro & Gygi, 2004) and represented five large environmental sound categories identified by Gaver (1993) and included 1) human and animal vocalization, 2) bodily sounds, 3) water sounds, 4) interacting solid objects and 5) aerodynamic sounds. All sounds in the test had been found to be easily identifiable (mean 98%, SD = 2) and highly familiar to NH listeners (mean familiarity was 6.39 on a 1-7 scale where 7 is very familiar). The test was administered in a closed - set format with 60 alternatives, closer approximating the demands of an open set, while reducing potential ambiguity in scoring open set

responses. To accommodate the discrepancy between the larger response set than the stimulus set, and multiple tokens of the same sound, participants were told that some sounds may be heard more than once, while some others may not be present.

During the test, each participant was seated in front of a computer monitor and used a computer mouse to respond and initiate consecutive trials. After hearing a new sound, each participant had to select one of the 60 response options displayed on the screen that best described the sound. All 160 environmental sound stimuli were presented in quiet. Prior to auditory testing, all participants completed a response label familiarization session, during which they provided familiarity ratings for the sounds denoted by each of the 60 sound labels. Each sound label was displayed at the top of the screen, one at a time in a random order and participants had to first provide familiarity ratings for that sound label and then find and mark its location among the 60 response options displayed. The response familiarization session used the same test interface as the environmental sound test and also served to ensure that all participants had adequate computer skills to perform the experimental task.

The **speech tests** consisted of CNC, HINT and two separate closed-set tests of 12 vowels and 20 consonants. Because both CNC and HINT tests are widely used in clinical settings, potential associations between these tests and the environmental sound test would be most clinically relevant. The CNC test assesses the perception of isolated monosyllabic words without the syntactic and semantic cues provided in sentence-based tests. In contrast, HINT involves a greater use of higher-level linguistic and contextual knowledge. HINT sentences were presented both in quiet and in low-level noise white noise. In quiet, CI patients could be expected to have sensory access to all audible acoustic cues. On the other hand, the additional challenge presented by low-level noise may extend beyond the audibility of acoustic cues and require CI patients to segregate speech cues from noise. Similar perceptual processes might be involved in the perception of acoustically complex environmental sounds such as ‘car starting’ or ‘gargling’ when a search for perceptually salient acoustic cues in an acoustically heterogeneous mixture maybe needed. Three lists of CNC words were presented in quiet. The first was a practice list that was not scored, while the last two lists were used to assess performance on this test. Two lists of HINT sentences were then presented first in quiet and then another two lists were presented at 10 dB signal-to-noise ratio (SNR) mixed with white noise (bandlimited at 10 KHz), delivered from the same azimuth position as speech.

The vowel and consonant tests were included to examine whether environmental sound perception that involves processing of a large number of rapidly changing acoustically heterogeneous components is more akin to consonant perception than vowels which are more steady state and acoustically homogenous. The stimuli in the vowel perception test were selected from Hillenbrand et al. (1995) recordings. Each of the 12 vowels was presented in a fixed /h/-vowel-/d/ phonetic context, and was spoken by five male and five female talkers, for a total of 120 stimuli. The consonant stimuli were selected from Shannon et al. (1999) recordings. Each of the 20 consonant sounds were presented in a fixed /a/-consonant-/a/ phonetic context, and was spoken by one male and two female talkers for the total of 60 stimuli, which were presented twice. During both vowel and consonant tests, stimulus presentation was subject-paced, and after hearing each new stimulus the participant had to pick one of the 12 words for vowels, or one of the 20 consonant symbols displayed on the screen.

Basic auditory ability and working memory tests included three psychoacoustic tests – a gap detection test, a test of temporal order for tones, a temporal pattern test (Figure 1) and a backward digit span test. This battery included those tests that have been previously shown

to be associated with speech perception performance in populations with normal and impaired hearing, including CI patients, and could tap into different peripheral and central aspects of auditory perception (Tyler et al., 1989; Gfeller et al., 1997; Surprenant & Watson, 2001; Schindler & Kessler, 1987; Humes & Christopherson, 1991; Christopherson & Humes, 1992; Watson, 1991; Pisoni & Cleary, 2003; Pisoni & Geers, 2001; Pisoni, 2005).

Stimuli in all three psychoacoustic tests were presented using a transformed 3 down – 1 up procedure (Levitt, 1971), with the threshold estimated as the mean of eight reversals, following three initial reversal steps. In the gap detection and temporal order for tones tests, the stimuli were presented in XAB format and participants were instructed to determine which of the last two stimulus intervals was different from the standard, which was presented first on every trial*. For gap detection stimuli, variable duration temporal gaps were centered in a broadband noise for a constant total stimulus duration of 500 ms. Stimuli in the temporal order for tones test, adapted from the Test of Basic Auditory Capabilities (TBAC) II battery (Kidd et al., 2007), consisted of a series of four-tone sequences, one per each interval. The frequency of the first and the fourth trailing tones was always 625 Hz, while the two middle tones had the frequencies of 710 Hz and 550 Hz, with the order of frequency change varying randomly from trial to trial (i.e., either the second tone was 710 Hz and the third tone was 550 Hz, or vice versa). As participants progressed through the test, the duration of all four tones in the sequence was adjusted, requiring participants to judge the order of two center tones with different frequencies in progressively shorter tone sequences.

The stimuli in the temporal pattern test (adopted from Gfeller et al., 1997), consisted of a sequence of six square wave pulses, 1000 Hz each, with four inter-pulse intervals of equal duration and one interval being 10% of that duration. The current version of the test used the same structure and manipulation of a temporal sequence, but a different square wave frequency sampling a spectral region of 1000 Hz and its odd harmonics which decreased in amplitude in 12 dB steps. (The earlier version of the test used by Gfeller et al., 1997 used 440 Hz square wave, approximating middle A note, because the focus of that study was on music perception.) When the short interval occurred during the first or the second inter-pulse interval, the participant was required to press the button labeled “Beginning”, and when the short interval occurred in either of the last two intervals, to press the button labeled “End”. The initial duration of the tones and the regular intervals was 300 ms. The duration of the complete six pulse sequence was varied adaptively based on participants' responses (both the tones and the intervals were shortened by 20% after each successful identification, but one interval was always 10% shorter than the other four).

Finally, participants were also administered a test of working memory – the backward Digit Span test (Koppitz, 1977), during which they were presented strings of digits of increasing length and were asked to repeat the digits they heard in a backward order (e.g. from last to first). The number of digits in the string increased until the listener made an error in two out of four strings with the same number of digits. To prevent participants from using lip reading or other visual cues, single digits were pre-recorded and then concatenated into variable-length strings, presented at the rate of one digit per second. As with the other tests, each participant was seated in a sound booth one meter away facing a loudspeaker, from which the stimuli were presented. Prior to the beginning of the test, participants heard each single digit, and had to repeat each single digit correctly to continue. While the accuracy of participants' single digits responses benefited from the 10-item close-set response format

*There are a number of alternative test formats that could be used to test the underlying auditory abilities (e.g. AXB, AX). These formats differ in the load they place on working memory which may affect the resulting threshold values. The XAB format was selected because it corresponded to one in the earlier research using these tests (e.g. Kidd et al., 2007).

(Miller et al., 1951), this ensured that all participants could identify each single digit when presented separately before beginning the test – which started with 2-digit long strings.

Acoustic analysis

The acoustic analysis of these environmental sounds followed the lines of Gygi et al., 2004 and Shafiro, 2008a. Using Matlab, the corpus of environmental sound stimuli used in this study was measured for the same acoustic variables which were found to be predictive of subject performance in the studies above. The variables obtained reflected different spectral-temporal aspects of the sounds including statistics of the envelope, autocorrelation statistics and moments of the long-term spectrum, and were intended to approximate as fully as possible the broad range of acoustic cues available to listeners. The analyses chosen did not specifically try to approximate the cues that may be available to CI listeners, but rather those cues that may be important for environmental sound identification in general, since the absence of significant effects for some of these cues, like pitch, may also indicate the information that CI patients, in contrast to NH listeners, are not using. The measures and their brief descriptions are listed below. The measures which were found to be most predictive of identification performance in Gygi et al. (2004) and in predicting the most-identifiable and least-identifiable sounds in Shafiro (2008a) are marked with an asterisk (*).

Envelope Measures—(1) Long term RMS/Pause-Corrected RMS (an index of the amount of silence)*, (2) Number of Peaks (transients, defined as a point in an amplitude values' vector that is greater in amplitude than the preceding point by at least 80% of the range of amplitudes in the waveform), (3) Number of Bursts (amplitude increases at least 4 dB sustained for at least 20 ms, based on an algorithm developed by Ballas, 1993)*, (4) Total Duration, (5) Burst Duration/Total Duration (a rough measure of the dynamic variation of the envelope).*

Autocorrelation Statistics—Number of Peaks*, Maximum, Mean Peak, *SD* of the Peaks. Peaks (as defined above) in the autocorrelation function reveal periodicities in the waveform and the statistics of these peaks measure different features of these periodicities, such as the strength of a periodicity and the distribution of periodicities across different frequencies.

Correlogram-Based Pitch Measures (from Slaney, 1995)—Mean pitch, Median pitch, *SD* pitch, Max pitch, Mean pitch salience, Max pitch salience. The correlogram measures the pitch and pitch salience by autocorrelating in sliding 16 ms time windows. This technique captures spectral information and provides measures of the distribution of that information over time.

Moments of the Spectrum—Mean (Centroid), *SD**, skew, kurtosis

RMS energy in octave-wide frequency bands from 63 to 16000 Hz

Spectral Shift in Time Measures—Centroid Mean, Centroid *SD*, Mean Centroid Velocity, *SD* Centroid Velocity*, Max Centroid Velocity. The centroid mean and *SD* are based on consecutive 50-ms time windows throughout the waveform. The spectral centroid velocity was calculated by measuring the change in spectral centroid across sliding 50-ms rectangular time windows.

Cross-Channel Correlation*: This is calculated by correlating the envelopes in octave wide frequency bands (or channels) ranging from 150 to 9600 Hz. It measures the consistency of the envelope across channels.

Modulation Spectrum Statistics: The modulation spectrum, first suggested by Houtgast and Steeneken (1985), reveals periodic temporal fluctuations in the envelope of a sound. The algorithm used here, divides the signal into logarithmically spaced frequency bands, extracts the envelope in each band, filters the envelope with low-frequency bandpass filters (upper f_c ranging from 1 to 32 Hz), and determines the power at that frequency. The result is the depth of modulation by modulation frequency. The statistics measured were the height and frequency of the maximum point in the modulation spectrum, as well as the number, mean and variance of bursts in modulation spectrum (using the burst algorithm described above).

Spectral Flux Statistics: Spectral flux is another measure of the change in the spectrum over time. As described by Lakatos (2000), it is the running correlation of spectra in short (50 ms) time windows. The mean, *SD* and maximum value of the spectral flux were used in this analysis.

Results

The results of environmental sound testing revealed average performance of 45.3% correct, *SD* = 16.2, across all 17 CI participants, suggesting a generally poor environmental sound identification ability (Figure 2). However, the performance of individual patients varied considerably from 16% to 69% correct. Participants' identification performance also varied among the 40 specific sound sources between 1.5% and 91% correct, (see Appendix), and among four specific tokens associated with each sound source. The performance on specific sound tokens associated with the source similarly ranged from 0% to 100% correct overall. Among the sound sources that presented the greatest difficulty to the participants (i.e., accuracy $\leq 25\%$) were 'Brushing teeth', 'Wind', 'Camera', 'Blowing nose', 'Zipper', 'Pouring soda', 'Sighing' and 'Airplane'. The best identified sounds (i.e., $\geq 75\%$) were 'Doorbell', 'Laughing', 'Dog barking' and 'Phone busy signal' (Figure 3). Although there were occasional confusions within specific source categories, e.g., 'Gargling' and 'Pouring soda' responses were given to 'Water bubbling', 'Water draining' and 'Water running' sounds and vice versa, no prominent and consistent confusion patterns could be identified in the data (possibly due to the sparsity of the response matrices). Nevertheless, some systematic variation in sounds' identifiability can be seen in Figure 3, where sounds are arranged from the easiest (bottom) to the most difficult (top) to identify.

Participants' ratings of their familiarity with the sounds corresponding to response labels in the test indicated that they considered these sounds to be generally familiar, with an average familiarity rating of 5.3 (on a 7 point scale, where 7 is extremely familiar and 1 is not familiar at all). The least familiar sound was that of 'Pouring soda' (3.92), while the most familiar was that of 'Phone ringing' (6.45). Participants' ratings of familiarity with the sounds produced by each source correlated moderately and significantly with their source identification accuracy, $r = 0.56$ ($p < 0.01$). A similar correlation between sound familiarity and identification accuracy was previously reported in NH listeners tested with 4-channel vocoded versions of the same environmental sounds (Shafiro, 2008b), indicating that listeners' own memory-based appraisal of sound familiarity often agrees with their ability to identify that sound when presented aurally.

Speech performance scores also varied across other study tests and individual participants (Figure 2). The average HINT accuracy in quiet was 68% correct, (*SD* = 33, range: 3 – 99%), and 55% correct in broadband noise with a 10 dB SNR, (*SD* = 33, range: 0 – 88%). Average monosyllabic words (CNC) performance was 52% correct (*SD* = 30, range: 1 – 89%). A similar average score was obtained for consonant sounds in fixed vocalic contexts, 51% (*SD* = 24, range: 3 – 82%), followed by vowels that were perceived 46% correct on average across patients (*SD* = 18, range: 13 – 67%). Tests of basic auditory abilities

demonstrated an average temporal gap detection threshold of 18 ms (SD = 16, range: 3 – 45 ms). The average component tone duration threshold for accurate discrimination of pitch direction change was 186 ms (SD = 41, range: 95 – 231 ms). The average threshold duration of six-pulse temporal patterns at which participants could determine the location of a shorter inter-pulse interval was 1842 ms (SD = 889, range: 502 – 3296 ms). Finally, the average number of digits that participants could successfully repeat in a reverse order was 3 (SD = 1.2, range: 0 – 6).

To examine the possibility that patients' performance on the study tests was strongly affected by their relatively large age variation as well as experience with their implants, the individual test scores for each test were correlated with participant age and implant experience. However, all correlations were weak and nonsignificant ($0 < r < 0.3$), suggesting that neither age nor implant associated with patient performance within this group of middle-age-to-elderly subjects.

Performance on the environmental sound test significantly correlated with that on speech tests in the moderate-to-strong range (0.64 to 0.73), with an exception of vowels (Table 2)[∞]. CNC word and HINT sentence scores in quiet as well as environmental sound test scores moderately correlated with performance on the temporal order for tones, r ranging from -0.53 to -0.59, $p < 0.05$. The correlation between HINT sentences in noise and the temporal order for tones had a similar magnitude, but had a higher p value (i.e. $r = 0.52$, $p < 0.06$). These correlations suggest a trend linking CI patients' abilities to discriminate a change in frequency order in shorter tones with the accuracy of their speech and environmental sound perception. On the other hand, other tests of basic auditory abilities or auditory working memory were not significantly correlated with speech or environmental sound abilities.

Nevertheless, the finding that both speech and environmental sound scores correlated with at least some aspects of basic spectro-temporal processing abilities, required for the discrimination of the temporal order for tones, may suggest that these abilities are largely responsible for the observed strong correlations between speech and environmental sound performance. In order to test this possibility, the variance contributed by the temporal order for tones test was partialled out from the correlation between speech and environmental sounds. Although this led to a slight decrease in correlation magnitudes, the correlations between most speech and environmental sound tests remained quite pronounced and significant (i.e., $r = 0.61$ for HINT in quiet, 0.57 for HINT in noise, 0.57 for CNC). The marginal decrease in correlation values suggests that underlying auditory abilities measured by the temporal order for tones as well as gap detection and temporal pattern tests, that did not show significant correlations with speech or environmental sounds, had limited involvement in mediating their relationship. On the other hand, because performing these tests also involves a combination of peripheral and central abilities, as working memory must be involved to some degree to make appropriate perceptual comparisons, these results may not be interpreted as reflecting a comprehensive assessment of patients' peripheral auditory abilities. Rather, they indicate that the underlying auditory abilities assessed by these tests did not play a major role in mediating the relationship between speech and environmental sounds.

[∞]Test data from one participant (S9) were excluded from the correlational analysis. This participant, despite reasonable audiometric thresholds, consistently had the lowest scores on most study tests and her inclusion might have biased the correlation analysis. On the other hand, the data from two other participants who had low scores on speech and environmental sound tests (S4, S15) were used in the analysis because these participants had average or better scores on the psychoacoustic tests in the study indicating that they could perform complex auditory tasks. Excluding these two participants from the analysis would imply an *a priori* assumption that speech and environmental sounds are related regardless of the more basic auditory abilities (e.g., it was not *a priori* evident that a CI patient with poor speech scores and good psychoacoustic test scores will not have good environmental sound perception. Setting an exclusion criteria based on speech alone would have eliminated this possibility.) When performing correlational analysis all percent correct scores were converted to rationalized arcsine units (Studebaker, 1985)

Acoustic Factors in Environmental Sound Identification

Two types of analysis were performed on the values obtained for each measure in order to determine the acoustic factors in environmental sound identification by CI listeners. For the first analysis, the identification performance across subjects for all sound tokens were regressed onto all of the acoustic measures listed earlier using both forward and backward stepwise multiple regression. In the second analysis, the tokens were grouped as to whether they were easily identifiable ($p(c) > 0.70$, for 34 tokens), moderately identifiable ($0.35 \leq p(c) < 0.70$, for 74 tokens), or difficult to identify ($p(c) < 0.35$, for 52 tokens). A discriminant analysis was then performed on these groupings to see if some of the acoustic factors could readily predict which category a sound token fell into.

The results of the best-fitting multiple regression solution are shown in Table 3. Five variables were retained in both the forward and backward solutions, all of which had significant beta values: the spectral skew, the burst duration/total duration, the amplitude of the largest burst in the modulation spectrum, the rms energy in the frequency region centered around 8000Hz, and the mean autocorrelation peak. This model accounted for 35% of the variance in the identification results.

The results of the discriminant analysis are listed in Table 4. Note that the three variables retained in discriminant analysis are also present in the multiple regression solution. However, the predictive power of the discriminant analysis was also modest: overall only 57% of the sounds were correctly classified by the model (chance = 33%), with the best performance in the middle third of the distribution ($p(c) = 0.86$) but much worse in the upper and lower thirds ($p(c) = 0.21$ and 0.47 , respectively). Partly the uneven distribution of classification performance may be due to the *a priori* larger number of sounds in the middle group. The difference in classification accuracy between the remaining two groups, although less pronounced, could result from easily identifiable sounds having a multiplicity of cues for the listener, so any one acoustic feature will only be mildly predictive. For the difficult to identify sounds, it is the opposite: listeners are functioning close to chance and so there are few reliable acoustic cues for them to focus on. Regardless of the specific reason, however, it appears that the acoustic factors examined play a limited role in predicting environmental sound identification performance in CI patients.

Discussion

The present findings demonstrate considerable difficulty among modern-day CI patients in the perception of many common environmental sounds in quiet. In fact, participants' performance on the environmental sound test was lower than on any of the speech tests in the study. The finding of 45.3% correct environmental sound identification accuracy closely agrees with a recent study by Inverso & Limb (2010) who found 49% correct identification accuracy. Although Inverso & Limb study used a free-response format, it appears that a large 60-item response set used in the present study provides enough variability in responses to approximate open set performance. On the other hand, the present identification accuracy is inferior to that in Reed & Delhorne's (2005) study who reported an average of 79.2% accuracy for implant patients with somewhat older devices. One potential factor that could have led to the difference in performance may be the use of a smaller 10-response set for each of the four different environmental contexts by Reed & Delhorne. Set size effects have been well-documented in speech perception research (e.g. Miller et al., 1951). It is thus possible that participants' identification performance in the present study that used a large 60-item response set, closely approximated that of an open set test, while smaller response set sizes used in Reed & Delhorne (2005) led to a higher performance. Nevertheless, differences in the stimuli, implant and patients' characteristics might have also contributed to the discrepancies in the results.

Participants' environmental sound performance in the current study correlated significantly with all speech tests that were administered, except for vowel identification (Table 2). The magnitude of the correlation suggests a considerable overlap in perceptual processing of environmental sounds and speech which does not considerably vary with the speech material type. Except for vowels which had the lowest and nonsignificant correlation (i.e. 0.48), isolated words, sentences and consonants had significant moderate-to-strong correlations with participants' environmental sound scores. The results also corroborate previous data indicating an association between speech and environmental sound perception abilities (Inverso & Limb, 2010; Reed & Delhorne, 2005; Tyler et al., 1989) despite considerable differences in the procedures, stimuli, materials and patient populations. Thus, an association between perceptual processing of speech and environmental sounds is a highly robust effect. On the other hand, there are a number of potential peripheral and central factors that may have contributed to this result. For one, there are obvious sensory limitations imposed by CI processing and deficiencies in the auditory periphery. It appears that the present and earlier results of basic auditory ability tests demonstrate some of these deficits.

Scores on all three tests of basic auditory abilities used in the present study revealed substantial deficits in the performance of CI participants relative to that of NH listeners. In line with previous research, average gap detection thresholds were higher than those typically found in NH listeners for broadband sounds (in the range of 1.5 - 3 ms, e.g., Kidd et al., 2007), but comparable to those found in other CI studies when testing was done in a sound field using the patients' speech processors (Wei et al., 2007; Tyler et al., 1989). Although limited normative data exists on the temporal order for tones test (Kidd et al., 2007), the current performance of CI participants is considerably lower than the first decile performance among young NH listeners, i.e., 119 for NH, vs. 186 ms for participants with CIs, falling near the first percentile of the NH performance distribution. For the temporal pattern test, the results (i.e. the threshold duration of the tone patterns) were substantially larger than those obtained by Gfeller et al. (1997) for both NH and CI listeners with older implants (i.e., 1842 ms in this study vs. 606.6 ms for NH, and 1023.8 ms and 1117 ms for users of FOF1F2 and MPEAK strategies, respectively, in Gfeller's et al.). These differences are likely to be due to the differences in test frequencies between Gfeller's et al. and the present version of the test (i.e., 440 Hz vs. 1000 Hz square wave, respectively). Odd harmonics of the lower frequency square wave in Gfeller's test could enable better frequency encoding compared to more sparse odd harmonics of 1000 Hz wave in the present study. If so, this finding indicates that perception of temporal patterns was strongly affected by the sparsity of spectral representation of the stimuli.

Other factors might have also contributed to poor performance of CI patients on gap detection and temporal order for tones tests. Inasmuch as CI listeners might have relied on loudness as a cue to making gap detection judgments (e.g., Shannon, 1989), the results might have been influenced by loudness variation inherent in the present method of testing. Specifically, because participants' processors during testing had their everyday listening settings which were not controlled for level and compression or possible adaptive directional microphone effects, patient-specific sensation levels were not taken into account during the detection procedure. In addition, assuming loudness integration times of 100-200 ms, the longest gap detection threshold of 45 ms would have produced a decrease in loudness (i.e. a drop of about 1 - 3 dB in level). However, for the majority of participants, who had smaller gap detection values, the level change would be smaller, making loudness a weaker cue. Inter-subject loudness variation might have also contributed to the large temporal pattern duration thresholds. Performance on the temporal pattern test, which involves a more global percept of rhythm and is associated with greater uncertainty about the position of a shorter silent interval, may also be affected by some of the factors involved in gap detection, as, for

instance, relying on loudness judgments. These limitations, along with a relatively small number of subjects, might have contributed to the absence of significant correlations between the scores on the gap detection tests and temporal pattern tests, on the one hand, and speech test scores on the other.

Nevertheless, correlations between speech and environmental sound tests with the temporal order for tones test further indicate that the ability to process rapid frequency changes plays a role in the perception of both speech and environmental sounds. During the test the listeners had to judge the direction of pitch change of progressively shorter tones. Because the overall spectrum of tone sequences remains the same, performing this task requires combined spectral and temporal processing. As a measure of excitation pattern over time, it may assess the dynamic current spread in CI listeners, closer approximating the perception of speech where frequency and duration often undergo rapid concurrent changes (e.g. formant transitions), but without relying on overlearned acoustic-phonetic patterns. However, the results of partial correlation analysis suggest that auditory abilities assessed by the temporal order for tones test, while clearly related to the perception of speech and environmental sounds on their own, do not mediate the relationship between these two large classes of sounds.

Lack of auditory working memory correlations in the present results, further indicates that the relationship between speech and environmental sound perception is likely complex and mediated by a number of peripheral and central factors. However, this conclusion is only tentative until a more comprehensive and rigorous assessment of peripheral and central factors is performed, as there seem to be other underlying factors that were not directly tested in the present battery, such as spectral resolution or visual working memory. Furthermore, the digit span scores of the present CI subjects and the lack of significant associations with speech might have been influenced by the very low speech perception abilities of several participants. While the closed-set nature of the test with only ten response options enabled all participants to accurately identify each individual digit at least once, it is likely that some of these correct responses were due to chance and affected the final digit span outcomes.

Environmental sound identification performance of CI patients in this study can be also compared with the performance of NH listeners tested with the same environmental sound stimuli processed with a 4-channel noise-based vocoder to simulate implant processing (Shafiro, 2008b). Current CI patients substantially outperformed the initial performance by NH listeners in that study, i.e., 45.3% vs. 33% correct, respectively. Possible reasons for this difference may be a greater number of spectral channels available to the current CI participants, and also, CI participants' experience with their implants (all had at least one year of experience). In fact, the NH hearing listeners' in Shafiro (2008b) study were able to improve their performance accuracy to 63% following five short training sessions, suggesting a possibility that present CI patients may also be able to benefit from structured training with environmental sounds. A similar conclusion was reached in a recent study by Loebach and Pisoni (2008) who implemented a short course of training with vocoded environmental sounds.

There were also some differences between the CI and NH listeners tested with simulations of sounds that patients found easy or difficult to identify. For example, sounds like 'Doorbell', 'Bird chirping' and 'Laughing' that were identified better than 70% correct by CI participants (i.e., 83%, 71.25% and 88%, respectively), but were initially identified poorly by NH listeners (i.e., 3.5%, 10.5%, and 28.5%, respectively). However, following training, the accuracy in NH listeners improved to 50%, 50% and 71% correct, respectively. On the other hand, sounds of 'Camera' and 'Typing' that presented little difficulty for NH

listeners, were not accurately identified by many CI listeners, i.e., 61% vs. 12% for 'Camera' and 75% vs. 40% for typing. Perhaps this may be due to inherently softer sound levels for such sounds. Other differences, especially for sounds involving the listener's own head (e.g. 'tooth brushing' or 'blowing nose') may be due to the lack of bone conduction that would normally influence the perception of sound. (Perhaps this is one reason why CI performance for 'brushing teeth' is markedly lower than that for NH listeners tested with vocoder in Shafiro, 2008b, i.e. 1.5% correct for CI vs. 68% for NH after training. CI patients may also not have their implant on while brushing teeth, further decreasing the likelihood for recognizing this sound). The substantial differences between CI and NH identification patterns are further demonstrated by only weak ($r < 0.2$) and nonsignificant correlations across test sounds between the performance of the current and NH participants in Shafiro (2008b) regardless of training. One reason for the lack of close correspondence in performance of CI patients and NH listeners tested maybe the use of only 4-channel acoustic simulation to compare with performance of CI patients who maybe expected to have access to a higher number of channels. Nevertheless, generalizing environmental sound findings from simulations to CI performance should be done with caution due to considerable variation in signal- and listener-level factors that could affect environmental sound perception in NH listeners and CI patients.

As can be seen in Figure 3 some sounds such as 'phone busy signal', 'laughing' and 'dog barking' posed little difficulties to most participants and were among the most consistently identified. On the other hand, sounds such as 'brushing teeth', 'wind', 'camera', 'blowing nose' and zipper' were among the most perceptually difficult for the participants. A large number of sounds were identified well by only some of the participants. Compared with previous research by Inverso & Limb (2010) which reported identification accuracy for individual sound sources, there was a close agreement with the present results. For the ten best identified sounds in Inverso's & Limb study, seven were also among the ten most accurately identified in the present study (i.e., 'dog barking', 'bird chirping', 'phone busy signal', 'laughing', 'car horn', 'phone ringing', 'footsteps') with the remaining three sounds ('siren', 'coughing' and 'baby crying') closely following. This consistency in results, despite variation in study stimuli, procedures, and subjects suggests that these sounds are indeed the easiest for experienced post-lingual CI patients to identify[†].

Being able to consistently identify particular environmental sounds may indicate that sensory information necessary for their identification is represented with sufficient accuracy by patients' implant processor. The general envelope-vocoder type signal processing adopted by the participants' implants severely limits sounds' spectral resolution and degrades temporal fine structure, while it preserves the envelope-related temporal characteristics of the input (see Loizou, 2006 for review). However, the sounds that were most accurately identified in this and Inverso's & Limb studies are quite heterogeneous acoustically and differ considerably in their physical sound production mechanisms, thus ruling out a single overriding factor for most accurately identified sounds. On the other hand, it is likely that signal processing limitations inherent in CI processing, along with other stimulus- and subject- specific factors, play an important role in the perception of other poorly identified environmental sounds.

The results of acoustic analysis further highlight the overall limited predictive power of acoustic factors in accounting for environmental sound identification performance in CI patients. Unlike identification performance in NH listeners with full spectrum stimuli, the

[†]Unfortunately, no adequate comparisons with ten lowest identified sounds in Inverso & Limb (2010) could be made since the present test did not include musical instruments among its stimuli, while musical instrument sounds dominated Inverso's lowest accuracy sound list.

identification performance of CI patients might be potentially more dependent on other factors specific to the patient and the device, such as duration of deafness, frequency-to-electrode assignment, the number of surviving ganglion cells and etiology among others. Perhaps substantial variations in these factors among the present patients, in contrast to the homogeneous groups of young NH listeners in the previous studies (Gygi et al., 2004; Shafiro, 2008a), makes the role of acoustic factors less apparent. Somewhat surprisingly, two often implicated factors – patients' age and implant experience did not seem to be associated with scores in the present study. Nevertheless, it is possible that within a more homogeneous group of CI patients, the predictive power of acoustic factors might have been greater. Among factors that had the greatest predictive value for the regression and discriminant analysis, it appears that both spectral and temporal factors can influence environmental sound identification in CI patients.

The predictive power of the best-fitting multiple regression model that accounted for 35% of the variance is similar to that of the multiple regression solution obtained in Gygi et al. (2004) for untrained NH listeners on one-channel vocoder processed environmental sounds, which accounted for about 39% of the variance in the identification results. On the other hand, this solution is much less predictive than the solution in Gygi et al. for the untrained NH listeners on six-channel vocoder processed environmental sounds, which accounted for about half the variance in the identification results in that condition. Moreover, the only variable found in that study to predict NH identification of vocoded sounds which was also found to predict present CI results was the Burst Duration/Total Duration, which is a rough measure of the dynamic variation of the envelope. This measure has consistently been a good predictor of both identification and similarity judgments across different studies with different conditions (Gygi et al., 2004; Shafiro, 2008a; Gygi et al., 2007) and likely captures some fundamental aspect of the sound-producing events. Bursts convey a great deal of information about both the type of object involved in an event (e.g., its configuration) and the nature of the sound-producing interaction (such as rolling, scraping or colliding). Temporal in nature, burst information is also available after envelope-based implant processing and can be used by CI patients.

Apart from that, it does seem that actual CI users attend to quite different acoustic features than NH listeners, even when their implants provide access to the relevant aspects of the acoustic signal (e.g. bursts in the envelope). This in and of itself is not surprising: listeners can change their strategy depending on many factors, such as experience with the sounds or instructions on what to listen to (Gygi et al., 2004; Lutfi & Liu, 2007). Lutfi and Liu also demonstrated that listeners using different strategies can often achieve the same level of performance, even when only one of the strategies is clearly optimal. The acoustic variables that best predicted the NH listeners' performance in Gygi et al. included several that were largely spectral measures, such as the Spectral SD, Spectral centroid velocity SD and Cross-Channel correlation. The only spectrally-related variables CI users seemed to focus on were the Spectral Skew, which is a rough measure of the amount of low-frequency content of a sound; and the RMS energy in the frequency band centered around 8000Hz, which often indicates a quick transient (as in 'ice cubes dropping in a glass', 'footsteps' or 'clock ticking'). The negative association between identification performance and the RMS energy in the 8000 Hz band could also result from masking of lower frequency components of environmental sounds, while the positive correlation between identification accuracy and spectral skew may indicate the importance of transmitting low frequency information. Depending of the settings of individual processors, these features can be expected to be available to some CI users.

Other acoustic predictors of CI performance are all related to the envelope: the two modulation spectrum variables reflect the depth and frequency of the modulation of the

sounds, and the mean autocorrelation peak speaks to the periodicity present in the envelope. This may reflect the role of experience; NH listeners are well-acquainted with the spectral content of these environmental sounds and so perhaps can infer spectral changes even when none are actually present in the signal. This may be due to the fact that in real world sound sources, spectral and temporal changes are almost never independent because of the physics of sound producing objects and events (Attias & Schreiner, 1997). Normal hearing listeners typically have access to both the temporal and spectral properties of real-world sounds. When identifying sounds in envelope-based simulation, they may be able to infer the spectral properties of the sounds from long-term memory. Not surprisingly, correlations with pitch variables were low as implant processing precludes the use of clear pitch cues.

Overall, the present results confirm the relatively poor perceptual accuracy for environmental sounds in CI patients reported in earlier studies. Importantly, the results demonstrate that environmental sound perception abilities of CI patients strongly correlates with their speech perception abilities, seemingly with limited mediation by some of the more general spectro-temporal processing auditory abilities tested in this study. However, further investigation is needed to explore further involvement of additional central and peripheral factors.

One potential clinical implication of strong correlations between environmental sound and speech perception could be found in the development of future diagnostic applications. For example, environmental sound tests can potentially help in determining implant candidacy or in assessing implant performance when speech tests are not available in a patient's language or when they cannot be administered due to the patients' other disabilities. Furthermore, recent CI simulation research has shown that environmental sound training can produce an improvement in speech perception as well as in environmental sound perception (Loebach & Pisoni, 2008). The possibility of gaining an improvement in both environmental sound and speech through non-language-specific environmental sound training may be useful to CI speakers of languages in which no rehabilitative programs have yet been developed. Given an expected exponential increase in the number of CI patients world-wide (Zeng, 2004), this aspect of environmental sound training may be particularly important. In addition to speech, improved environmental sound perception may potentially help CI patients with identification of musical instruments, often included in environmental sound tests and music appreciation in general (Inverso & Limb, 2010). Other hearing impaired patient populations that may benefit from environmental sound testing include children, prelingually deafened CI patients, or patients with brainstem implants. Future research may be able to further clarify the practical advantages and theoretical underpinning of environmental sound perception in cochlear implant patients.

Acknowledgments

We would like to thank Dr. Blythe Holmes for her assistance in this project and Dr. Stanley Sheft for useful comments on an early draft of the manuscript. This project was supported in part by the Deafness Research Foundation, the American Speech Language and Hearing Foundation and NIH-NIDCD (DC008676).

Research support: Support for this study was provided in part by the Deafness Research Foundation, American Speech Language and Hearing Foundation and NIH-NIDCD (DC008676).

Appendix

Table A
Familiarity and identification accuracy of 40 sound sources in the test

Sound Name	Sound Familiarity	Identification Accuracy (%)	Maximum Token Score (%)	Minimum Token Score (%)
Airplane Flying	5.59	24	29	18
Baby crying	5.86	53	59	47
Bird chirping	5.35	71	94	38
Blowing nose	5.39	15	24	0
Bowling	5.16	31	41	12
Brushing teeth	5.00	1.5	6	0
Burping	5.09	53	59	47
Camera taking picture	3.97	12	18	6
Car horn	5.89	62	76	41
Car starting	5.58	35	47	24
Clapping	5.91	43	53	35
Clearing throat	4.36	52	59	41
Clock ticking	5.82	31	41	24
Closing door	5.82	72	88	59
Coughing	5.84	57	76	35
Cow mooing	5.12	56	59	53
Dog barking	6.30	88	100	71
Doorbell	5.40	87	94	76
Footsteps	5.36	69	76	65
Gargling	4.56	41	53	29
Glass breaking	5.38	53	75	35
Helicopter flying	5.14	53	76	24
Horse neighing	4.92	47	59	35
Ice cubes into glass	4.99	35	59	12
Laughing	6.31	88	94	82
Machine gun	4.56	52	82	24
Pour soda into cup	3.92	16	29	6
Rooster	5.56	58	65	53
Sighing	4.05	16	29	0
Siren	6.32	59	76	35
Sneezing	5.54	62	71	41
Snoring	5.52	37	41	29
Telephone	6.45	71	82	59
Telephone busy signal	5.97	91	94	88
Toilet flushing	5.96	33	47	18
Thunder	5.74	35	47	24
Typing	5.61	40	59	24

Sound Name	Sound Familiarity	Identification Accuracy (%)	Maximum Token Score (%)	Minimum Token Score (%)
Wind blowing	5.18	11	18	6
Yawning	4.50	29	35	24
Zipper	4.67	15	24	6

References

- Attias, H.; Schreiner, CE. Temporal low order statistics of natural sounds. In: Mozer, M., editor. *Advances in neural info processing systems*. Vol. 9. Cambridge, MA: MIT Press; 1997. p. 27-33.
- Baddeley AD, Wilson BA. A case of word deafness with preserved span: Implications for the structure and function of short-term memory. *Cortex*. 1993; 29:741–748. [PubMed: 8124947]
- Ballas JA. Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance*. 1993; 19(2):250–267. [PubMed: 8473838]
- Ballas JA, Mullins T. Effects of context on the identification of everyday sounds. *Human Performance*. 1991; 4(3):199–219.
- Bonebright, TL. Perceptual structure of everyday sounds: A multidimensional scaling approach. Paper presented at the 2001 International Conference on Auditory Display; Espoo, Finland. 2001.
- Cazals Y, Pelizzone M, Saudan O, Boex C. Low-pass filtering in amplitude modulation detection associated with vowel and consonants identification in subjects with cochlear implants. *Journal of Acoustical Society of America*. 1994; 96:2048–2054.
- CHABA Working Group on Communication Aids for the Hearing-Impaired. Speech-perception aids for hearing-impaired people: current status and needed research. *The Journal of the Acoustical Society of America*. 1991; 90(2):637–685. [PubMed: 1939883]
- Christopherson LA, Humes LE. Some psychometric properties of the Test of Basic Auditory Capabilities (TBAC). *J Speech Hear Res*. 1992; 35:929–935. [PubMed: 1405548]
- Cummings A, Ceponiene R, Koyama A, Saygin AP, Townsend J, Dick F. Auditory semantic networks for words and natural sounds. *Brain Research*. 2006; 1115(1):92–107. [PubMed: 16962567]
- Donaldson GS, Nelson DA. Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies. *Journal of the Acoustical Society of America*. 2000; 107:1645. [PubMed: 10738818]
- Drennan WR, Longnion JK, Ruffin C, Rubinstein J. Discrimination of Schroeder-Phase Harmonic Complexes by Normal-Hearing and Cochlear-Implant Listeners. *Journal of the Association for Research in Otolaryngology*. 2008; 9:138–149. [PubMed: 18066624]
- Edgerton BJ, Prietto A, Danhauer JL. Cochlear implant patient performance on the MAC battery. *Otolaryngologic Clinics of North America*. 1983; 16(1):267–280. [PubMed: 6687933]
- Eisenberg LS, Berliner KI, House WF, Edgerton BJ. Status of the adults' and children's cochlear implant programs at the house ear institute. *Annals of the New York Academy of Sciences*. 1983:323–331. [PubMed: 6346983]
- Fu Q. Temporal processing and speech recognition in cochlear implant users. *Neuroreport*. 2002; 13(13):1635–1639. [PubMed: 12352617]
- Freed DJ. Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events. *Journal of the Acoustical Society of America*. 1990; 87(1):311–322. [PubMed: 2299041]
- Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*. 1975; 12(3):189–98. [PubMed: 1202204]
- Gaver WW. What in the world do we hear?: An ecological approach to auditory event perception. *Ecological Psychology*. 1993; 5(1):1–29.

- Gfeller K, Woodworth G, Robin DA, Witt S, Knutson JF. Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users. *Ear Hear.* 1997; 18:252–260. [PubMed: 9201460]
- Gygi B. From acoustics to perception: How to listen to meaningful sounds in a meaningful way. *The Journal of the Acoustical Society of America.* 2003; 113(4):2326.
- Gygi B, Kidd GR, Watson CS. Spectral-temporal factors in the identification of environmental sounds. *Journal of Acoustical Society of America.* 2004; 115(3):1252–1265.
- Gygi B, Kidd GR, Watson CS. Similarity and categorization of environmental sounds. *Perception and Psychophysics.* 2007; 69(6):839–855. [PubMed: 18018965]
- Gygi B, Shafiro V. General functions and specific applications of environmental sound research. *Frontiers in Bioscience.* 2007; 12:3167–3176. [PubMed: 17485291]
- Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America.* 1995; 97:3099–3111. [PubMed: 7759650]
- Houtgast T, Steeneken HJM. A review of the mtf concept in room acoustics and its use for estimating speech intelligibility in auditoria. *Journal of the Acoustical Society of America.* 1985; 77(3):1069–1077.
- Humes, LE.; Lee, JH., et al. *The Journal of the Acoustical Society of America.* Vol. 120. 2006. Auditory measures of selective and divided attention in young and older adults using single-talker competition; p. 2926
- Humes LE, Christopherson LA. Speech identification difficulties of hearing-impaired elderly persons: The contributions of auditory processing deficits. *J Speech Hear Res.* 1991; 34:686–693. [PubMed: 2072694]
- Inverso Y, Limb CJ. Cochlear implant-mediated perception of nonlinguistic sounds. *Ear & Hearing.* 2010; 31(4):505–514. [PubMed: 20588119]
- Karlin JE. A factorial study of auditory function. *Psychometrika.* 1942; 7:251–279.
- Kidd GR, Watson CS, Gygi B. Individual differences in auditory abilities. *Journal of the Acoustical Society of America.* 2007; 122(1):418–435. [PubMed: 17614500]
- Koppitz, EM. *The Visual Aural Digit Span Test.* New York: Grune & Stratton; 1977.
- Kunkler-Peck AJ, Turvey MT. Hearing shape. *Journal of Experimental Psychology: Human Perception and Performance.* 2000; 26(1):279–294. [PubMed: 10696618]
- Lakatos S. A common perceptual space for harmonic and percussive timbres. *Perception & Psychophysics.* 2000; 62(7):1426–1439. [PubMed: 11143454]
- Leech, R.; Dick, F.; Aydelott, J.; Gygi, B. *The Journal of the Acoustical Society of America,* 121 pt. Vol. 2. 2007. Naturalistic auditory scene analysis in children and adults; p. 3183
- Levitt H. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America.* 1971; 49:467–477. [PubMed: 5541744]
- Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR, DeYoe EA. Human brain regions involved in recognizing environmental stimuli. *Cerebral Cortex.* 2004; 14:1008–1021. [PubMed: 15166097]
- Li X, Logan RJ, Pastore RE. Perception of acoustic source characteristics: Walking sounds. *Journal of the Acoustical Society of America.* 1991; 90(6):3036–3049. [PubMed: 1787243]
- Loebach JL, Pisoni DB. Perceptual learning of spectrally degraded speech and environmental sounds. *Journal of the Acoustical Society of America.* 2008; 123(2):1126–1139. [PubMed: 18247913]
- Loizou, P. Speech processing in vocoder-centric cochlear implants. In: Møller, A., editor. *Cochlear and Brainstem Implants.* Adv Otorhinolaryngol. Basel: Karger; 2006. p. 64p. 109-143.
- Lutfi RA, Liu CJ. Individual differences in source identification from synthesized impact sounds. *Journal of the Acoustical Society of America.* 2007; 122(2):1017–1028. [PubMed: 17672650]
- Maddox EH, Porter TH. Rehabilitation and results with the single electrode cochlear implant. *Otolaryngologic Clinics of North America.* 1983; 16(1):257–265. [PubMed: 6687932]
- Marcell MM, Borella D, Greene M, Kerr E, Rogers S. Confrontation naming of environmental sounds. *Journal of Clinical and Experimental Neuropsychology.* 2000; 22(6):830–864. [PubMed: 11320440]

- Miller GA, Heise A, Lichten W. The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*. 1951; 41:329–335. [PubMed: 14861384]
- Moore, BCJ. *Introduction to the psychology of hearing*. Academic Press; 2003.
- Muchnik C, Taitelbaum R, Tene S, Hildesheimer M. Auditory temporal resolution and open speech recognition in cochlear implant recipients. *Scandinavian Audiology*. 1994; 23(2):105–9. [PubMed: 8085109]
- Oh EL, Lutfi RA. Information masking by everyday sounds. *Journal of Acoustical Society of America*. 1999; 106(6):3521–3528.
- Piccirilli M, Scarab T, Luzzi S. Modularity of music: Evidence from a case of pure amusia. *Journal of Neurology*. 2000; 69:541–545.
- Pisoni, DB. Speech perception in deaf children with cochlear implants. In: Pisoni, DB.; Remez, RE., editors. *The Handbook of Speech Perception*. Blackwell Publishing, Ltd.; 2005. p. 494-523.
- Pisoni DB, Cleary M. Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear & Hearing*. 2003; 24:106S–120S. [PubMed: 12612485]
- Pisoni DB, Geers A. Working memory in deaf children with cochlear implants: Correlations between digit span and measures of spoken language processing. *Annals of Otolaryngology, Rhinology, and Laryngology*. 2001; 109:92–93.
- Plante E, Van Petten C, Senkfor AJ. Electrophysiological dissociation between verbal and nonverbal semantic processing in learning disabled adults. *Neuropsychologia*. 2000; 38(13):1669–1684. [PubMed: 11099725]
- Proops DW, Donaldson I, Cooper HR, Thomas J, Burrell SP, Stoddart RL, Moore A, Cheshire IM. Outcomes from adult implantation, the first 100 patients. *Journal of Laryngology and Otology*. 1999; 113:5–13. [PubMed: 10664722]
- Reed CM, Delhorne LA. Reception of environmental sounds through cochlear implants. *Ear and Hearing*. 2005; 26(1):48–61. [PubMed: 15692304]
- Schnider A, Benson F, Alexander DN, Schnider-Klaus A. Non-verbal environmental sound recognition after unilateral hemispheric stroke. *Brain*. 1994; 117:281–287. [PubMed: 8186955]
- Schindler RA, Dorcas KK. The UCSF/STORZ cochlear implant: Patient performance. *American Journal of Otolaryngology*. 1987; 8(3):247–255.
- Schindler RA, Kessler DK. The UCSF/Storz cochlear implant: Patient performance. *American Journal of Otolaryngology*. 1987; 8(3):247–255.
- Shannon RV. Detection of gaps in sinusoids and pulse trains by patients with cochlear implants. *J Acoust Soc Am*. 1989; 85:2587–2592. [PubMed: 2745882]
- Shafiro V, Gygi B. How to select stimuli for environmental sound research and where to find them? *Behavior Research Methods, Instruments and Computers*. 2004; 36(4):590–598.
- Shafiro V. Identification of environmental sounds with varying spectral resolution. *Ear and Hearing*. 2008a; 29(3):401–420. [PubMed: 18344871]
- Shafiro V. Development of a large-item environmental sound test and the effects of short-term training with spectrally-degraded stimuli. *Ear and Hearing*. 2008b; 29(5):775–790. [PubMed: 18596641]
- Shannon RV, Jensvold A, Padilla M, Robert ME, Wang X. Consonant recordings for speech testing. *Journal of the Acoustical Society of America*. 1999; 106:L71–L74. [PubMed: 10615713]
- Slaney, M. *Apple Computer Technical Report #45*. 1995. *Auditory Toolbox: A MATLAB toolbox for auditory modeling work*.
- Studebaker GA. A “rationalized” arcsine transform. *Journal of Speech and Hearing Research*. 1985; 28:455–462. [PubMed: 4046587]
- Surprenant AM, Watson CS. Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *Journal of the Acoustical Society of America*. 2001; 110:2085–2095. [PubMed: 11681386]
- Tye – Murray N, Tyler RS, Woodworth GG, Gantz BJ. Performance over time with a Nucleus or Ineraid Cochlear Implant. *Ear and Hearing*. 1992; 13(3):200–209. [PubMed: 1397761]
- Tyler RS. Advantages and disadvantages expected and reported by Cochlear Implant Patients. *American Journal Otolaryngology*. 1994; 15:523–531.

- Tyler RS, Lowder MA, Gantz BJ, Otto SR, McCabe BF, Preece JP. Audiological results with two single channel cochlear implants. *Annals of Otology, Rhinology and Laryngology*. 1985; 94:133–139.
- Tyler RS, Moore B, Kuk F. Performance of some of the better cochlear-implant patients. *Journal of Speech & Language, Hearing Research*. 1989; 32:887–911.
- Van Petten C, Rieffers H. Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*. 1995; 33(4):485–508. [PubMed: 7617157]
- Warren WH, Verbrugge RR. Auditory perception of breaking and bouncing events. *Journal of Experimental Psychology: Human Perception and Performance*. 1984; 10(5):704–712. [PubMed: 6238128]
- Watson CS, Kidd GR. On the lack of association between basic auditory abilities, speech processing and other cognitive skills. *Seminars in Hearing*. 2002; 23:83–93.
- Watson BU. Some relationships between intelligence and auditory discrimination. *J Speech Hear Res*. 1991; 34(3):621–627. [PubMed: 2072686]
- Wei C, Cao K, Jin X, Chen X, Zeng F. Psychophysical performance and Mandarin tone recognition in noise by cochlear implant users. *Ear and Hearing*. 2007; 28(2 Supplemental):62S–65S. [PubMed: 17496650]
- Won JH, Drennan WR, Rubinstein JT. Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. *Journal of the Association for Research in Otolaryngology*. 2007; 8:384–392. [PubMed: 17587137]
- Zeng FG. Trends in cochlear implants. *Trends in Amplification*, 2004. 2004; 8(1):1–34.
- Zhao F, Stephens SDG, Sim SW, Meredith R. The use of qualitative questionnaires in patients having and being considered for cochlear implants. *Clinical Otolaryngology*. 1997; 22:254–259. [PubMed: 9222632]

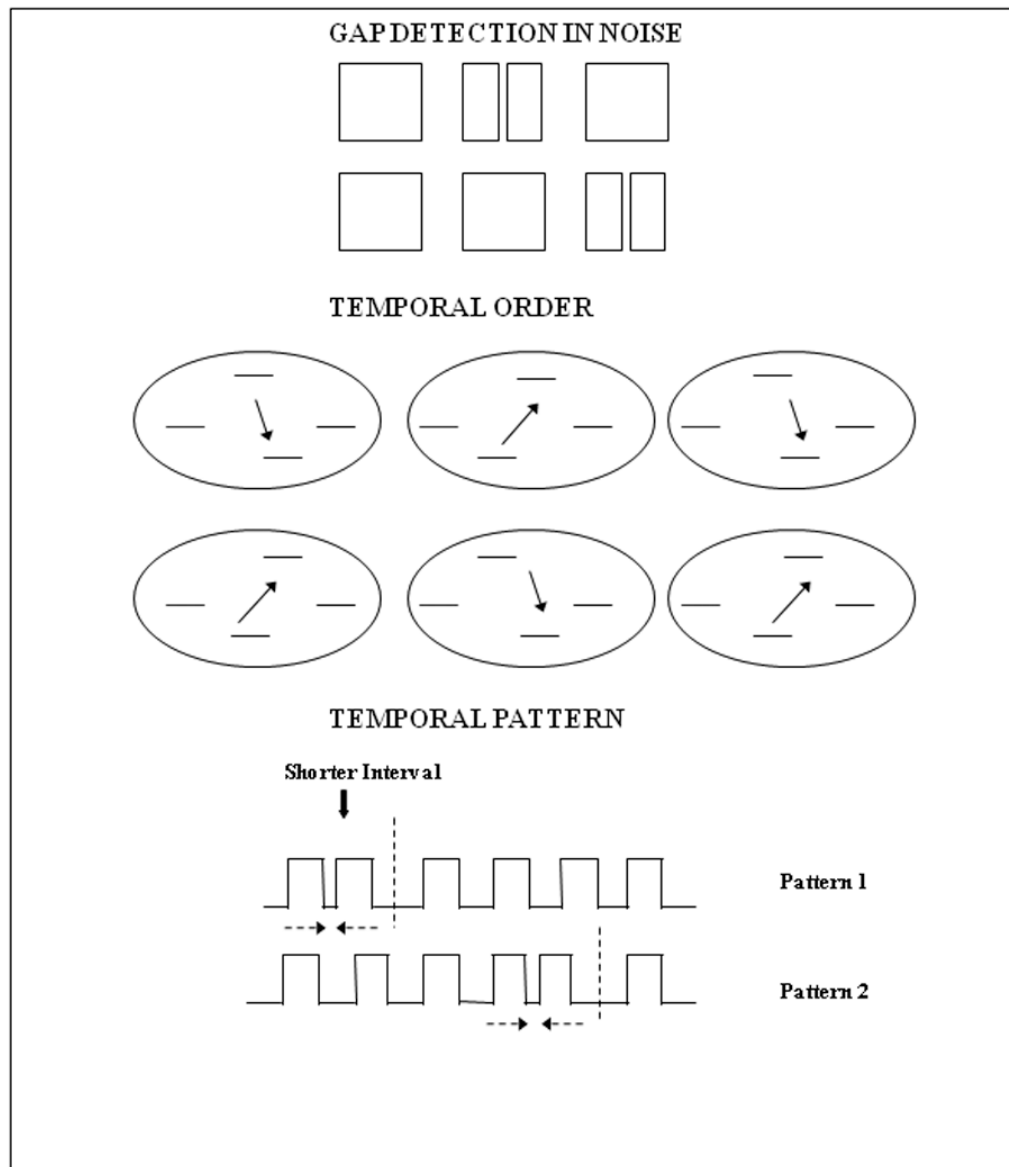


Figure 1.

Schematic depiction of the three tests of basic auditory ability used in the study. In the gap detection and temporal order tests the listeners were asked to pick one of the last two intervals that was different from the first (standard). For the gap detection test (top), a silent gap of variable duration was inserted into one of the last two intervals of broadband noise. For the temporal order test (bottom), the listeners discriminated the stimuli based on the direction of frequency change of the middle two tones. For the temporal pattern test (middle), they were required to determine whether a shorter temporal interval between a series of 6 consecutive tones appeared in the first or the last half of the tone sequence.

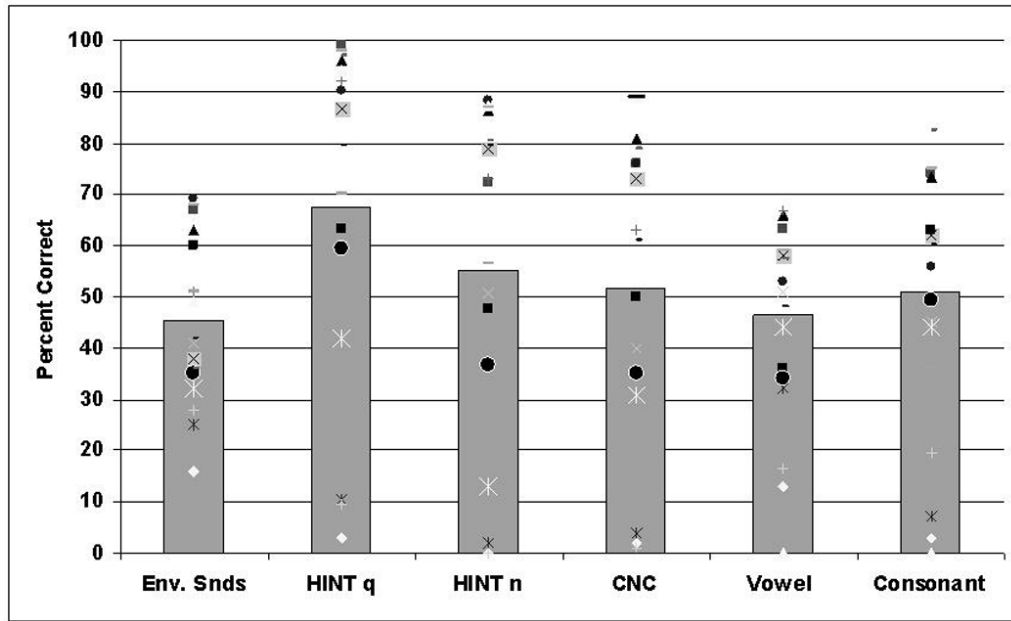


Figure 2. Performance accuracy across environmental sounds and speech test (with individual participant scores as data points). Bar height corresponds to the average performance on each test.

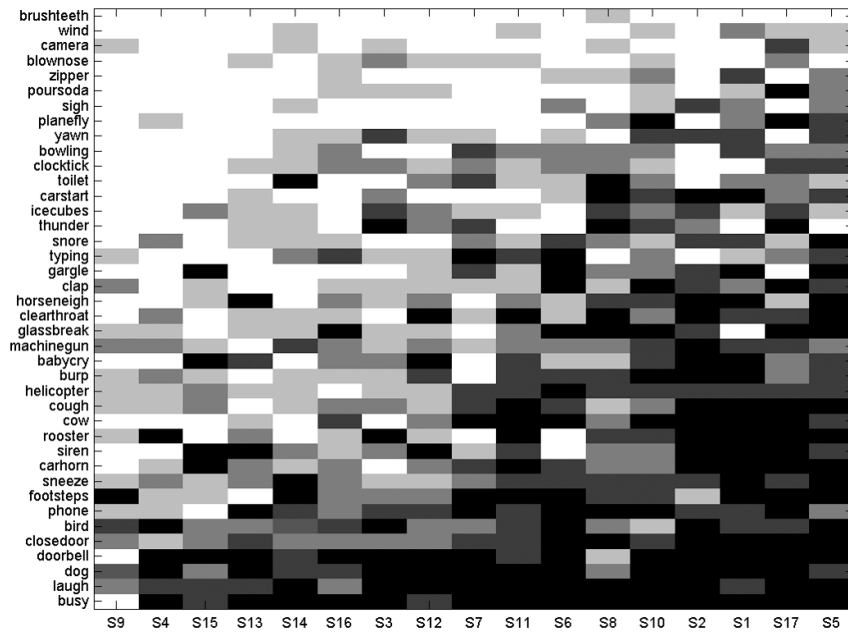


Figure 3. Individual sound source identification accuracy for each study participant. Sound source names are listed on the ordinate and individual participants are on the abscissa. The accuracy is represented by darkness of each cell (the darker the more accurate). The sounds are arranged vertically based on their overall accuracy – those at the bottom were the easiest and those at the top were the most difficult to identify. The participants are arranged based on overall performance accuracy from left to right, starting with the least accurate participant on the left.

Table 1

Participant background characteristics.

S#	Sex	Device	Etiology	Age	Duration hearing loss	Duration CI use	CNC	Env. Sounds
1	F	Nucleus 24	Congenital/Unknown	40	19	7	76	67
2	F	Clarion CII	Measles	56	53	4	81	63
3	F	Clarion CII	Congenital/Unknown	64	39	4	73	38
4	F	Clarion-Auria	Measles	43	37	2	4	25
5	F	Clarion-Auria	Unknown	56	51	2	76	69
6	M	Nucleus-Freedom	Noise-induced	80	35	1	63	51
7	F	Nucleus 24	Unknown	55	35	1	61	42
8	M	Nucleus 24	Unknown	54	51	2	30	51
9	F	Nucleus-Freedom	Measles	42	39	1	2	16
10	F	Nucleus-Freedom	Nerve deafness	56	52	2	50	60
11	F	Nucleus-Freedom	Unknown	70	35	1	89	49
12	M	Nucleus-Freedom	Congenital/Unknown	51	45	2	40	39
13	F	Clarion – Harmony	Unknown	64	64	6	31	32
14	F	Nucleus-Freedom	Autoimmune response	67	61	6	35	35
15	F	Nucleus-Freedom	Unknown	64	14	3	1	28
16	M	Nucleus-Freedom	Unknown	75	40	7	79	36
17	M	Nucleus CI 24R (CS)	Meningitis	53	46	4	89	68

Table 2

Correlations among test scores for 16 CI participants-values with a single asterisk are significant at $p < 0.05$. Values marked with ♣ are significant at $p < 0.06$.

	<u>HINT quiet</u>	<u>HINT noise</u>	<u>CNC</u>	<u>Env Snds</u>	<u>Vowels</u>	<u>Cons</u>	<u>Temp Pat</u>	<u>Gap Det</u>	<u>Temp Ord</u>
<u>HINT quiet</u>	1	0.90*	0.93*	0.73*	0.79*	0.87*	0.47	-0.21	-0.59*
<u>HINT noise</u>		1	0.92*	0.69*	0.73*	0.85*	0.33	-0.31	-0.52♣
<u>CNC</u>			1	0.70*	0.86*	0.91*	0.40	-0.19	-0.56*
<u>Env Snds</u>				1	0.48	0.64*	0.04	-0.11	-0.53*
<u>Vowels</u>					1	0.68*	0.64*	0.05	-0.38
<u>Cons</u>						1	0.31	-0.26	-0.39
<u>Temp Pat</u>							1	0.43	-0.05
<u>Gap Det</u>								1	0.00
<u>Temp Ord</u>									1

Table 3

Best-fitting multiple regression solution for the overall identification performance across sounds. All five acoustic variables had significant beta values.

Regression Summary for Dependent Variable: P(c) R= .589 R²= .347	
	Beta
Mean Autocorrelation Peak	0.312
Spectral Skew	0.245
Burst Duration/Total Duration	0.172
Modulation Spectrum Maximum Burst	0.162
RMS in band Fc = 8000 Hz	-0.160

Table 4

Acoustic variables that enabled most accurate classification of sounds into three identification accuracy group in discriminant analysis.

Acoustic Feature	Wilks' Lamda	Partial r	p-level
Spectral Skew	0.774	0.866	0.000
Burst Duration/Total Duration	0.707	0.948	0.016
Mean Autocorrelation Peak	0.752	0.892	0.000