Research Article

Correlations Between Pitch and Phoneme Perception in Cochlear Implant Users and Their Normal Hearing Peers

RAYMOND L. GOLDSWORTHY¹

¹*University of Southern California, Los Angeles, CA, USA*

Received: 26 August 2014; Accepted: 30 August 2015; Online publication: 15 September 2015

ABSTRACT

This study examined correlations between pitch and phoneme perception for nine cochlear implant users and nine normal hearing listeners. Pure tone frequency discrimination thresholds were measured for frequencies of 500, 1000, and 2000 Hz. Complex tone fundamental frequency (F0) discrimination thresholds were measured for F0s of 110, 220, and 440 Hz. The effects of amplitude and frequency roving were measured under the rationale that individuals who are robust to such perturbations would perform better on phoneme perception measures. Phoneme identification was measured using consonant and vowel materials in quiet, in stationary speech-shaped noise (SSN), in spectrally notched SSN, and in temporally gated SSN. Cochlear implant pure tone frequency discrimination thresholds ranged between 1.5 and 9.9 %, while cochlear implant complex tone F0 discrimination thresholds ranged between 2.6 and 28.5 %. On average, cochlear implant users had 5.3 dB of masking release for consonants and 8.4 dB of masking release for vowels when measured in temporally gated SSN compared to stationary SSN. Correlations with phoneme identification measures were generally higher for complex tone discrimination measures than for pure tone discrimination measures. Correlations with phoneme identification measures were also generally higher for pitch perception measures that included amplitude and frequency roving. The strongest correlations were observed for measures of complex tone F0 discrimination with phoneme identification in temporally gated SSN. The results of this study suggest that musical training or

signal processing strategies that improve F0 discrimination should improve consonant identification in fluctuating noise.

Keywords: cochlear implants, pitch perception, speech reception, psychophysics

INTRODUCTION

Cochlear implants (CIs) have improved to the point that it is common for recipients to comprehend speech in quiet without relying on speech reading cues. While this level of speech reception is remarkable, performance deteriorates rapidly when listening in noisy environments. Music perception also tends to be poor for CI users suggesting that the technology may not provide sufficient resolution to convey the nuances of musical expression. The hypothesis under investigation in the present article is that both music perception and speech reception in noise depend on the available cues for pitch perception.

Two types of pitch perception were considered in the present study—pure and complex tone frequency discrimination—which depend upon different physiological responses in the auditory system. Pure tones excite a relatively narrow region of the auditory nerve with response synchrony decreasing as stimulus frequency increases. In comparison, complex tones excite a broader region of the auditory nerve with a degree of synchrony to the fundamental frequency (F0) observed across neural populations (Cariani and Delgutte 1996). Cochlear implants stimulate the auditory nerve in a manner that attempts to elicit similar responses to those observed in normal hearing, with pure tones encoded as constant-amplitude

Correspondence to: Raymond L. Goldsworthy · University of Southern California · Los Angeles, CA, USA. email: Raymond.Goldsworthy@med.usc.edu

electrical stimulation over a narrow electrode region and complex tones encoded as amplitude-modulated stimulation delivered over a wider electrode region.

Pure tone frequency and complex tone F0 discrimination have been characterized for normal hearing listeners (Sek and Moore 1995; Micheyl et al. 2012). Pure tone frequency discrimination thresholds for normal hearing listeners have been shown to be less than 1 % for frequencies between 250 and 8000 Hz (Sek and Moore 1995), with musically trained listeners achieving thresholds near 0.2 % (Micheyl et al. 2006). Similarly, complex tone F0 discrimination thresholds for normal hearing listeners have been shown to be less than 1 % for F0s between 100 and 250 Hz (Kaernbach and Bering 2001), with musically trained listeners achieving thresholds near 0.1 % (Micheyl et al. 2006); however, discrimination depends on the resolvability of the harmonics with thresholds increasing to between 1 and 3 % for unresolved harmonics (Kaernbach and Bering 2001).

Compared to their normal hearing peers, pure tone frequency and complex tone F0 discrimination have not been as well characterized for CI users. Studies (Pretorius and Hanekom 2008; Goldsworthy et al. 2013) have found CI discrimination thresholds between 5 and 30 %; furthermore, Vandali et al. (2014) demonstrated that training can improve CI discrimination thresholds by more than a factor of 2. While the use of synthetic tones presented through the CI processor is relatively unexplored, there is a rich literature investigating CI pitch perception based on electrode psychophysics (Tong et al. 1982; McDermott and McKay 1997; Hanekom and Shannon 1998; Zeng 2002; Henry and Turner 2003; Carlyon et al. 2010; Hughes and Goulson 2011; Goldsworthy and Shannon 2014) as well as using more natural stimuli (Fujita and Ito 1999; Gfeller et al. 2002; McDermott 2004; Looi et al. 2008). In the present article, the suggestion put forth by Pretorius and Hanekom (2008) that pure and complex tones can be used to probe different aspects of perception is extended toward understanding correlations with phoneme identification in noise.

The hypothesis that the availability of pitch cues affects speech reception in noise can be subdivided into separate suppositions concerning pure and complex tones. Since pure tones are relatively narrow probes of spectral resolution, it is supposed that pure tone pitch perception is relevant to phonemic identification that requires spectral resolution (e.g., phonemic contrasts associated with formant trajectories or listening within spectral dips). Since complex tones are relatively broadband probes of perception with a consistent acoustic cue available across the spectrum, it is supposed that complex tone pitch perception is relevant for phoneme identification that requires

voicing detection or for voicing-mediated sound source segregation. The present study examined phoneme identification in different types of background noise designed to evaluate different spectral and temporal aspects of noise on phoneme identification. Specifically, three noise types were examined: stationary speech-shaped noise (SSN), spectrally notched SSN, and temporally gated SSN.

Previous studies have considered the effect of hearing loss and of cochlear implant use on speech reception in spectrally and/or temporally fluctuating noise (Bacon et al. 1998; Peters et al. 1998; Jin and Nelson 2006; Nelson et al. 2003). For CI users, studies have demonstrated that spectral resolution as measured by spectral-ripple discrimination (Henry and Turner 2003) and forward-masked tone detection (Goldsworthy et al. 2013) is correlated to speech reception in quiet as well as in stationary and temporally fluctuating noise. In the present study, pure tone frequency discrimination was examined as a measure of spectral resolution that probes perception in a way that differs from spectral-ripple discrimination and from forward-masked detection. Specifically, pure tone frequency discrimination requires interpreting spectral differences as perceived pitch, whereas spectral-ripple discrimination and forward-masked detection probe energetic masking of adjacent spectral regions. In the present study, the effects of spectrally notched SSN on phoneme identification was measured under the rationale that pure tone pitch, since it measures pitch perception associated with small spectral changes, would be better correlated to performance in such noise conditions that provide relatively narrow spectral glimpses of the target speech.

Previous studies have also considered relationships between complex tone perception with speech reception in noise. One line of research uses CI simulations toward understanding this relationship. For example, Qin and Oxenham (2003) found that normal hearing individuals listening through CI simulations performed worse on speech reception in noise tasks when the background noise was a competing talker compared to stationary SSN. They hypothesized that the reduction in available F0 voicing cues makes sound source segregation difficult. Additional support for this hypothesis is found in studies of bimodal hearing in which the CI user has residual hearing in the non-implanted ear (Looi and Radford 2011; Rader et al. 2013). The general finding of these studies is that the residual acoustic hearing improves both complex tone pitch perception as well as speech reception in noise even when the acoustic hearing does not provide sufficient cues for speech reception by itself. The hypothesis emerging from these findings is that F0 perception derived from residual hearing

enhances sound source segregation and, hence, speech reception in noisy environments.

A motivating hypothesis for the present study is that there is a strong relationship between complex tone discrimination and speech reception in fluctuating noise. This expectation is based on reasoning that, to perform well in temporally gated noise, the listener must have a strong perceptual cue for when the gaps in the noise occur. While that statement might seem obvious, it is non-trivial for a CI user to disambiguate envelope fluctuations that should be attributed to the target speech from those that should be attributed to the background noise. Hypothetically, the ability to perceive voice pitch might allow experienced CI users to more readily parse phoneme information during the regular periods of silence within fluctuating noise.

A second motivating hypothesis for the present study is that there is a strong relationship between pure tone frequency discrimination and speech reception in spectrally notched noise. This expectation is based on the reasoning that, to perform well in spectrally notched noise, the listener must depend on underlying spectral resolution to separate speech and noise into independent perceptual channels. It is expected that such perceptual abilities can be measured using both speech reception in spectrally notched noise as well as using pure tone frequency discrimination.

The present study examined correlation strength between pure tone frequency and complex tone F0 discrimination with phoneme identification in quiet and noise for CI users and their normal hearing peers. The effect of amplitude and frequency roving on pitch perception was considered under the hypothesis that individuals who are more robust to such perturbations will perform better on speech reception measures. This hypothesis linking the robustness of pitch perception to a potential robustness in speech reception has been suggested by others (Sucher and McDermott 2007; Vandali et al. 2014); however, clear evidence of such a relationship has not been previously demonstrated. The results provide a direct comparison of pitch discrimination and phoneme identification measures toward understanding relationships between these two areas of perception.

METHODS

Subjects

The subjects for this study included nine CI users and nine age- and gender-matched normal hearing listeners. The St. Vincent Medical Center's Institute Review Board approved the study protocol. All subjects provided informed consent on their first visit to the laboratory and were paid for their participation.

Demographic and audiologic information for participating CI users is listed in Table 1. All subjects were native English speakers without any formal musical training. For the normal hearing subjects, a hearing test was administered before they began the study to document that they had normal hearing defined as having 20 dB HL or better at octave frequencies between 125 and 8000 Hz. CI users were required to have had at least 1 year of experience with their currently used sound processing strategy and that they could obtain at least 60 % phoneme identification in quiet. The latter criterion was necessary to exclude individuals who would not be able to perform the measures of phoneme identification in noise, which converge to 50 % identification accuracy. Of the nine CI users, seven used implants manufactured by Cochlear Corporation and two used implants manufactured by Advanced Bionics Corporation.

Measures of Frequency Discrimination

Frequency discrimination was measured using pure and complex tones. All tones were 400 ms in duration with 20-ms raised-cosine on/off ramps generated in the time domain and played out at a sampling rate of 44,100 Hz. Complex tones were generated by summing all harmonics of the F0 lower than 22,050 Hz and then filtering through a second-order band-pass filter centered at 1000 Hz with 3-dB attenuation points at 707 and 1414 Hz. Band-pass filter specifications were selected such that the filter was centered on a fairly important region for speech reception (Pavlovic 1987).

Discrimination was measured using a three-interval, three-alternative, forced-choice procedure. Two of the stimulus intervals were defined equivalently (except for amplitude roving of stimuli as described below) and are referred to as standard intervals; the other interval, referred to as the target interval, had an adaptively higher frequency. A 200-ms silent gap was inserted between intervals. The interval ordering of standards and target was randomly assigned with equal probability. Subjects were instructed: "Which interval was higher in pitch?"

Pure tone frequency discrimination thresholds (FDTs) were measured for standard frequencies of 500, 1000, and 2000 Hz. Complex tone F0 discrimination thresholds (F0DTs) were measured for standard F0s of 110, 220, and 440 Hz. The initial frequency or F0 difference between the standard and target was 10 %; this difference was decreased by a factor of $\sqrt[3]{2}$ after each correct response and was increased by a factor of 2 after each incorrect response, which theoretically converges to 75 % discrimination accuracy (Kaernbach 1991). The tracking procedure stopped after 11 reversals, and the discrimination threshold for the run was calculated as

TABLE 1

Demographic and audiologic information for participating CI users

Subject	Sex	Age at testing (years)	Age at onset of hearing loss (years)	Age at implantation (years)	Implant use (years)	Implant model	Processing strategy
CI1	M	82	50s-progressive	65	17	N22	SPEAK
CI2	M	75	50s-progressive	71	4	Freedom	ACE
CI3	F	56	30s-progressive	49	7	HiRes90K	HiRes
CI4	F	66	40s-progressive	59	7	Freedom	ACE
CI5	M	63	30s-progressive	50	13	HiRes90K	HiRes
CI6	M	21	Birth-progressive	20	1	Freedom	ACE
CI7	F	26	Birth-progressive	20	6	Freedom	ACE
CI8	F	29	Birth-progressive	24	5	Freedom	ACE
CI9	F	68	40s-progressive	55	13	Freedom	ACE

Normal hearing subjects were age-matched to within ± 2 years of a corresponding CI user

the geometric mean of the frequency differences of the last 8 reversals.

Discrimination thresholds were measured under four conditions: no roving, amplitude roving, frequency roving, and amplitude and frequency roving. For the no roving conditions, the presentation level for each interval was 65 dB of sound pressure level (dB SPL) and the standard frequency was precisely the condition frequency (i.e., if the condition frequency was 1000 Hz, then the standard frequency was precisely 1000 Hz for every standard interval in the run). For the amplitude roving conditions, the presentation level for each interval was selected randomly from a uniform distribution between 59 and 71 dB SPL. For the frequency roving conditions, the standard frequency was selected randomly from a uniform distribution ranging between $\frac{1}{2}$ octave below and above the condition frequency (i.e., if the condition frequency was 1000 Hz, then the standard frequency was selected from a uniform distribution between 707 and 1414 Hz). The two standard intervals had the same roved frequency, and the target interval was always higher as controlled by the adaptive frequency difference. For the amplitude and frequency roving condition, both of these aforementioned variations were implemented independently. Visual feedback indicating whether the response was correct was given by flashing the selected button green (correct) or red (incorrect) for 400 ms following the response.

Measures of Phoneme Identification

Eight measures of phoneme identification were tested including consonant and vowel identification in quiet, in stationary speech-shaped noise (SSN), in spectrally notched SSN, and in temporally gated SSN. Consonants were drawn from speech samples collected by Shannon et al. (1999) for five male and five female talkers and consisted of 20 phonemes /b t f d g d₃ k l m n p r s j t d_h v w y z/, presented in /a/-C-/a/ context (aba, acha, ada, afa, aga, aja, aka, ala, ama, ana, apa, ara, asa, asha, ata,

atha, ava, awa, aya, aza). Vowels were drawn from speech samples collected by Hillenbrand et al. (1995) for five male and five female talkers and consisted of ten monophthongs (/i I e æ u u a ɔ ʌ ɜ:/) and two diphthongs (/əʊ eɪ/), presented in /h/-V-/d/ context (heed, hid, head, had, who'd, hood, hod, hud, hawed, heard, hoed, hayed). Listeners responded using a graphical user interface with 20 (consonants) or 12 (vowels) alternatives with the appropriately labeled phonemes. Feedback was provided in the form of the pressed button flashing green for correct answers and flashing red for incorrect answers.

For all of the consonant identification measures, a run consisted of 200 trials corresponding to one presentation of each of the consonant tokens in the database. Likewise, all of the vowel identification measures used runs consisting of 120 trials. For the measures in noise, the initial signal-to-noise ratio was 0 dB and thereafter was decreased/increased by 2 dB after every correct/incorrect answer. The overall presentation level of the combined speech plus noise was 65 dB SPL. The speech reception threshold (SRT) for each run was calculated as the average signal-to-noise ratio, excluding the first 20 trials. While it is less common to use monosyllabic words, or closed-set phoneme materials, within adaptive SRT procedures, previous studies have demonstrated that monosyllabic words can be used to measure SRTs so long as a sufficient number of trials are used (Liu and Eddins, 2012).

For the noise conditions, SSN was generated by filtering random noise drawn from a uniform distribution through a speech-shaping filter. This filter was generated by estimating the power spectral density of the corresponding speech corpus (i.e., consonants or vowels) using Welch's periodogram method and converting this density to an eighth-order IIR filter using Prony's method (Parks and Burrus 1987). For the spectrally notched SSN conditions, SSN was filtered through an eighth-order bandstop filter having a stopband centered at 1000 Hz and having a 1-octave 3-dB bandwidth (i.e., 3-dB attenuation points at 707 and 1414 Hz). The center frequency of the notch was

selected such that it centered on a fairly important region for speech reception (Pavlovic 1987). For the temporally gated SSN conditions, SSN was gated on and off at 10 Hz with 4-ms rise/fall times.

Experimental Procedures

The measures of frequency discrimination and phoneme identification were combined into four test protocols referred to as the pure tone protocol, the complex tone protocol, the consonant identification protocol, and the vowel identification protocol. The pure and complex tone protocols each consisted of four subtests (i.e., no roving, amplitude roving, frequency roving, and amplitude and frequency roving) with each of these subtests including two repetitions of the three standard frequencies tested. Both the consonant and vowel identification protocols included four subtests with identification measured in quiet, stationary SSN, spectrally notched SSN, and temporally gated SSN. The order of subtests within each protocol instance was randomized. These protocols were each administered three times generally over the course of three test sessions each lasting approximately 4 h with breaks as required.

All testing was conducted in a double-walled sound-proof booth. Sounds were presented through an Alesis M1 Active 320 USB speaker connected to the PC using an ESI U24 XL USB digital audio interface. Subjects were seated 1 m in front of the speaker and all sounds were presented at 65 dB SPL (calibrated in free-field where the center of the listener's head would be), except for the amplitude roving described above for the frequency discrimination measures. Cochlear implant users used their clinical sound processor with their self-selected most-commonly-used sound processing strategy. Cochlear implant users were allowed to adjust their processor sensitivity before commencing the study but used the same settings throughout the study.

RESULTS

Pure Tone Frequency Discrimination

The completed study included nine CI users and nine normal hearing listeners each tested on six repetitions of four roving conditions and three standard frequencies. Figure 1 plots geometric means of the measured FDTs averaged across repetitions.

Measured FDTs were analyzed in logarithmic space¹ using an analysis of variance (ANOVA) with subject as a random factor, with subject group (CI

user or normal hearing listener) as a between-subjects factor, and with roving (no roving, amplitude roving, frequency roving, or both amplitude and frequency roving) and standard frequency (500, 1000, or 2000 Hz) as within-subjects factors. As expected, CI users had higher FDTs with geometric means across all conditions of 4.2 % compared to 1.3 % for normal hearing listeners, which was a significant difference ($F_{9, 1182}=109.2, p<0.001$).

Roving was significant ($F_{3, 1182}=86.3, p<0.001$) as was the interaction between roving and subject group ($F_{27, 1182}=5.8, p<0.001$); however, the interaction between roving and standard frequency was not significant ($F_{6, 1182}=1.5, p=0.30$). Further analyzing this finding, the geometric means averaged across frequencies for CI users were 2.8, 3.5, 5.6, and 6.0 % (i.e., for no roving, amplitude roving, frequency roving, and both amplitude and frequency roving), respectively, and for the normal hearing listeners 0.80, 0.98, 1.9, and 2.1 %.

Standard frequency was significant ($F_{2, 1182}=10.6, p=0.0012$), as was the interaction between standard frequency and subject group ($F_{18, 1182}=10.8, p<0.001$). Further analyzing this finding, the geometric means averaged across roving conditions for CI users were 6.0, 3.8, and 3.4 % (i.e., for 500, 1000, and 2000 Hz), respectively, and for the normal hearing listeners 1.5, 1.3, and 1.2 %.

Complex Tone F0 Discrimination

A similar analysis, as for the pure tone frequency discrimination described in the previous section, was calculated for the complex tone F0 discrimination results. Figure 2 plots geometric means of the measured F0DTs averaged across repetitions.

Measured F0DTs were analyzed in logarithmic space using an ANOVA with subject as a random factor, with subject group (CI user or normal hearing listener) as a between-subject factor, and with roving (no roving, amplitude roving, frequency roving, or both amplitude and frequency roving) and standard frequency (110, 220, 440 Hz) as within-subjects factors. As expected, CI users had, on average, much higher F0DTs with geometric means across all conditions of 12.5 % compared to 1.4 % for normal hearing listeners, which was a significant difference ($F_{9, 1182}=131.1, p<0.001$).

Roving was significant ($F_{3, 1182}=28.2, p<0.001$) as was the interaction between roving and subject group ($F_{27, 1182}=3.4, p<0.001$); however, the interaction between roving and standard frequency was not significant ($F_{6, 1182}=1.1, p=0.34$). Further analyzing this finding, the geometric means across frequencies for CI users were 8.3, 12.0, 15.1, and 16.3 % (i.e., for no roving, amplitude roving, frequency roving, and

¹ See Micheyl et al. (2006) for a rationale for using logarithmic space when analyzing discrimination thresholds using similar methods.

FIG. 1. Pure tone FDTs for each condition tested. *Shaded and unshaded boxes indicate normal hearing and CI listener data, respectively. Data points represented by big dots indicate mean performance averaged across runs and subjects. The edges of the surrounding box indicate 25th and 75th*

percentile. The horizontal line within the mark indicates the median. Whiskers extend to the most extreme data points not considered outliers and outliers are plotted individually as plus sign.

both amplitude and frequency roving), respectively, and for the normal hearing listeners 0.95, 1.0, 2.1, and 2.1 %.

Standard frequency was significant ($F_{2, 1182}=22.3$, $p=0.023$) as was the interaction between standard frequency and subject group ($F_{18, 1182}=16.2$, $p<0.001$).

FIG. 2. Complex tone F0DTs for each condition tested. *Shaded and unshaded boxes indicate normal hearing and CI listener data, respectively. Data points represent by big dots indicate mean performance averaged across runs and subjects. The edges of the surrounding box indicate 25th*

and 75th percentile. The horizontal line within the mark indicates the median. Whiskers extend to the most extreme data points not considered outliers and outliers are plotted individually as plus sign.

Further analyzing this finding, the geometric means across roving conditions for CI users were 11.6, 18.9, and 9.0 % (i.e., for 110, 220, and 440 Hz), respectively, and for the normal hearing listeners 1.8, 1.4, and 1.2 %.

Phoneme Identification

Phoneme identification scores were calculated as the percent correct of phonemes correctly identified based on phonetic label (i.e., these were measures of overall phonetic identification as opposed to average identification across a particular dimension such as place of articulation or voicing). Consonant identification in quiet ranged between 67.6 and 93.3 % for CI users and between 91.7 and 98.2 % for normal hearing listeners. Vowel identification scores in quiet ranged between 67.9 and 90.3 % for CI users and between 88.0 and 95.0 % for the normal hearing listeners.

Figure 3 plots SRTs for spectrally notched versus stationary SSN for each subject averaged across runs. The dashed line represents where performance in spectrally notched and stationary SSN are equal; data points falling below this line indicate masking release for spectrally notched compared to stationary SSN. The average consonant SRT for CI users was 0.7 dB in stationary and 1.3 dB in spectrally notched SSN for an average masking release of -0.6 dB. The average vowel SRT for CI users was -2.8 dB in stationary and

FIG. 3. Spectrally notched versus stationary SSN SRTs. *Smaller symbols without error bars indicate SRTs averaged across runs for each subject. The numbers adjacent to the CI data points indicate the subject number of Table 1. Larger symbols indicate subject averages, with error bars indicating the standard errors of the means.*

-3.8 dB in spectrally notched SSN for an average masking release of 1.0 dB. The average consonant SRT for normal hearing listeners was -11.3 dB in stationary and -10.7 dB in spectrally notched SSN for an average masking release of -0.6 dB. The average vowel SRT for normal hearing listeners was -10.4 dB in stationary and -12.3 dB in spectrally notched SSN for an average masking release of 1.9 dB.

Figure 4 plots SRTs for temporally gated versus stationary SSN for each subject averaged across runs. The average consonant SRT for CI users was 0.7 dB in stationary and -4.6 dB in temporally gated SSN for an average masking release of 5.3 dB. The average vowel SRT for CI users was -2.8 dB in stationary and -11.2 dB in temporally gated SSN for an average masking release of 8.4 dB. The average consonant SRT for normal hearing listeners was -11.3 dB in stationary and -24.6 dB in temporally gated SSN for an average masking release of 13.3 dB. The average vowel SRT

FIG. 4. Temporally gated versus stationary SSN SRTs. *Smaller symbols without error bars indicate SRTs averaged across runs for each subject. The numbers adjacent to the CI data points indicate the subject number of Table 1. Larger symbols indicate subject averages, with error bars indicating the standard errors of the means.*

for normal hearing listeners was -10.4 dB in stationary and -28.6 dB in temporally gated SSN for an average masking release of 18.2 dB.

Correlations Between Measures

There were 24 measures of frequency discrimination collected for each subject accounting for two stimulus types (pure and complex tones), four levels of roving, and three frequencies tested for each stimulus type. There were eight measures of phoneme identification collected for each subject accounting for two phoneme types (consonants and vowels) and four noise types (quiet, stationary SSN, spectrally notched SSN, and temporally gated SSN). Four additional phoneme identification measures are of interest, specifically the masking release differences between spectrally notched and also between temporally gated SSN with stationary SSN (for both consonants and vowels). Consequently, there are 288 correlations of interest between frequency discrimination and phoneme identification measures. Analysis is implemented to examine trends in the data concerning the effects of stimulus type, roving, and frequencies tested.

For all of the collected measures—except for phoneme identification in quiet—a lower score corresponds to better performance. To adjust for this in the correlation analysis, and facilitate the overall comparison, phoneme identification in quiet is transformed into rationalized arcsine units and then is negated such that a lower score corresponds to better performance after this transformation. This allows the correlation figures (Figs. 5 and 6) that summarize trends to be more readily interpreted (i.e., a positive correlation can be interpreted as better performance on one measure predicting better performance on the other).

To identify general trends, an ANOVA with first-order interactions was calculated on the 288 correlations of interest using stimulus type, roving, and phoneme identification condition as factors. Stimulus type was significant ($F_{1, 225}=33$, $p<0.01$) with a higher average correlation for complex F0 discrimination (0.31) than for pure tone frequency discrimination (0.14). Roving was significant ($F_{3, 225}=4.4$, $p<0.01$) with correlations increasing with roving: no roving (0.16), amplitude roving (0.19), frequency roving (0.26), and amplitude and frequency roving (0.29). Note that this is the same ordering that roving affected discrimination thresholds (Figs. 1 and 2). This trend held for both pure and complex tones as quantified by a lack of interaction between stimulus type and roving ($F_{3, 225}=0.047$, $p=0.51$). As expected, condition was significant ($F_{11, 225}=8.3$, $p<0.001$); the details of which will be analyzed in subsequent analyses. The interaction between condition and

stimulus type was significant ($F_{11, 225}=2.5$, $p=0.0054$) indicating different correlation trends for pure and complex tones, which will also be considered in subsequent analyses. Finally, the interaction between condition and roving was not significant ($F_{33, 225}=0.67$, $p=0.91$) indicating that while correlations were generally higher for the roved frequency discrimination measures, the general correlation trends were similar.

The preceding analysis indicated general trends: Correlations with phoneme identification measures were higher for complex F0 discrimination than for pure tone frequency discrimination and were higher for roved discrimination measures, and that this effect of roving did not interact with either stimulus type or with phoneme identification condition. Consequently, subsequent analyses focus on the correlations for the measures collected with both amplitude and frequency roving toward examining detailed trends in the correlations analysis.

Correlation distributions and associated confidence intervals were estimated using a bootstrapping method that has been validated for correlation analysis (Sievers 1996; Li et al. 2011; Nilsson and Castro 2012). Estimating confidence intervals, rather than adjusting p values, has been recommended by a number of authors and scientific journals (Rothman 1990; Feise 2002; Haukoos and Lewis 2005; Gelman et al. 2012).

Distributions were estimated for the correlations between pure tone FDTs (with amplitude and frequency roving) with measures of phoneme identification. Correlation distributions were estimated for each frequency to examine how frequency might affect correlation strength. Estimated correlation distributions are shown in Figure 5. For these analyses, a correlation will be described as significant if the 5 % confidence interval is to the right of the 0-correlation line, that is, at least 95 % of the estimated distribution is greater than 0. With that measure of significance, none of the correlations between pure tone frequency discrimination and phoneme identification in quiet were significant. There is a general trend for higher correlations between pure tone frequency discrimination and phoneme identification in stationary SSN, but only the correlation between pure tones measured at 2 kHz with consonant identification in stationary SSN reached statistical significance. Similar trends were observed for both spectrally notched and temporally gated SSN with estimated correlation distributions having modes greater than 0, but not always reaching statistical significance. Correlation distributions for measures of spectrally notched and temporally gated masking release (relative to stationary SSN) generally did not reach statistical significance; the only exception was for the estimated

FIG. 5. Estimated correlation distributions between pure tone FDTs (measured with amplitude and frequency roving) for each phoneme identification measure. Plotted circles indicate the 5 and 95 % edge

for each distribution. Distributions having a 5 % edge greater than 0 are noted as significant and are shaded for emphasis.

correlation distributions between frequency discrimination at 1 kHz with masking release for spectrally notched SSN for the vowel identification measure, which corresponds reasonably to the fact that the spectral notch in the SSN was located at 1 kHz.

A similar analysis was calculated for the estimated correlation distributions between complex tone F0DTs (with amplitude and frequency roving) with

measures of phoneme identification. Estimated correlation distributions are shown in Figure 6. All of the estimated correlation distributions between complex tone F0 discrimination and consonant identification in quiet were significant, but none of the corresponding comparisons reach statistical significance for vowel identification in quiet. Similarly, correlations were stronger for consonant

FIG. 6. Estimated correlation distributions between complex tone F0DTs (measured with amplitude and frequency roving) for each phoneme identification measure. *Plotted circles* indicate the 5 and

95 % edge for each distribution. Distributions having a 5 % edge greater than 0 are noted as significant and are shaded for emphasis.

identification in both stationary and spectrally notched SSN compared to the corresponding vowel identification measures. For temporally gated SSN, all correlations between complex F0 discrimination and phoneme identification reached statistical significance. None of the correlations for masking release in spectrally notched SSN reached statistical

significance, while the majority of the correlations for masking release in temporally gated SSN did reach statistical significance.

The above analyses emphasize certain trends for pure tone frequency and complex tone F0 discrimination as predictors of phoneme identification. Pure tone frequency discrimination was not as strong of a predictor of

phoneme identification compared to complex tone F0 discrimination. The significant correlations for pure tone frequency discrimination tended to occur for the spectrally notched noise, although weaker—generally non-significant—correlations were observed for stationary and temporally gated SSN. In contrast, complex tone F0 discrimination tended to be a stronger predictor of phoneme identification especially for consonant identification and for either consonant or vowel identification when measured in temporally gated SSN. The higher strength of correlation observed between frequency discrimination measures with consonant identification relative to vowels might be associated with the fact that the CI users tended to perform worse on the consonant identification measures (i.e., had higher SRTs). This relative difficulty between consonant and vowel measures might be a result of the consonant materials having a wider distribution of acoustic intensity cues.

DISCUSSION

The hypothesis considered in the present article is that frequency discrimination is correlated with speech reception in noise. This hypothesis was subdivided into two suppositions: that F0 frequency discrimination is particularly correlated with speech reception in fluctuating noise and that pure tone frequency discrimination is particularly correlated with speech reception in spectrally notched noise. The present study provided evidence supporting the first supposition, but the evidence supporting the latter supposition is weak. The measures of speech reception in spectrally notched noise used in the present study did not sufficiently determine performance differences between subjects to allow characterization of masking release. It is possible that other speech reception measures based on spectral resolution such as detection of formant trajectories might yield insightful correlations. For the present study, however, the most interesting conclusions are related to the relatively strong correlations observed between F0 discrimination and phoneme identification in temporally gated SSN.

It has been shown that F0 discrimination can be improved through auditory training for both cochlear implant (Gfeller et al. 2002; Vandali et al. 2012) and normal hearing listeners (Demany and Semal 2002, Micheyl et al. 2006; Amitay et al. 2010). The results of the present study indicate that the relationship between F0 discrimination with speech reception in noise is nuanced with the strongest correlations observed between F0 discrimination and consonant identification in temporally gated SSN. It is reasonable to suppose that auditory training of F0 discrimination may lead to improved speech reception in

challenging listening environments; the results from the present study suggest that successful detection of such transfer of pitch training to speech reception would benefit from focusing on consonant identification in temporally fluctuating noise. The present results also suggest that such pitch training includes relatively broad training of pitch across loudness and pitch dimensions, which would naturally occur in a music therapy program and has been recommended by others (Looi et al. 2012; Vandali et al. 2014).

The present study is also relevant for improving CI signal processing. A number of authors have suggested that psychophysical stimuli can be used to understand individual differences between CI users (Tong et al. 1982, Shannon 1983) as well as to probe the detailed effects of CI signal processing (Henry et al. 2000). The results from the present study are relevant for CI signal processing developments that attempt to improve CI pitch perception. For example, it has been shown that enhancing the temporal representation of the F0 in the envelopes used to modulate electrical pulse trains significantly improves CI pitch perception (Vandali and van Hoesel, 2012). The results from the present study suggest that such algorithm improvements might translate to speech reception benefits if the appropriate dimensions of speech reception are considered.

In addition to these primary directions of improving CI hearing through auditory rehabilitation and signal processing, the results of the present study also clarify the thresholds of pure and complex tone perception in CI users. Typically, CI pitch perception has been assessed using musical stimuli that are designed with constant frequency differences (see, e.g., Looi et al. 2012 and McDermott et al. 2004 for reviews). For example, Sucher and McDermott (2007) considered 1 and 6 semitone differences in complex F0 discrimination, which correspond to F0 differences of 6 and 41 %. They found that, as a group, the CI users performed at chance level for the 6 % difference but that a subset of CI users could correctly rank stimuli with greater than 70 % accuracy for the 41 % difference. The present study provides a more precise description of discrimination thresholds in CI users using adaptive procedures. Specifically, the present study found that 75 % discrimination thresholds for pure tones generally ranged between 2 and 10 %, while the corresponding thresholds for complex tones ranged between 5 and 20 %. However, it is worth noting that Vandali et al. (2012) demonstrated that the average complex tone F0 discrimination threshold for CI users could be improved to at least 5 % through auditory training.

The results from the present study provided further evidence of masking release in temporally gated SSN for CI users on consonant and vowel identification (Goldsworthy et al. 2013). While the magnitude of the observed masking release in temporally gated SSN was

smaller for CI users (5.3 and 8.4 dB on consonant and vowel materials, respectively) compared to their normal hearing peers (13.3 and 18.2 dB on consonant and vowel materials, respectively), the fact that CI users do show some masking release in temporally gated SSN is promising. Other studies have generally found little or no masking release in temporally gated, or fluctuating, noise for CI users (Nelson et al. 2003; Stickney et al. 2004; Rader et al. 2013). In fact, some evidence indicates an opposite phenomenon where CI users perform worse in the fluctuating noise condition (Kwon and Turner 2001; Nelson et al. 2003; Stickney et al. 2004). However, Kwon et al. (2012) demonstrated that the better performing CI users—as measured by speech reception in stationary noise—tend to take advantage of temporal gaps in temporally gated SSN.

CONCLUSIONS

Pure tone frequency discrimination thresholds for CI users were generally between 1 and 10 % for frequencies of 500, 1000, and 2000 Hz. Complex tone F0 discrimination thresholds for CI users were generally between 5 and 20 % for F0s of 110, 220, and 440 Hz. Cochlear implant users exhibited masking release in temporally gated noise, although to a lesser degree than their normal hearing peers. Correlations with phoneme identification measures were generally higher for complex tone F0 discrimination than for pure tone frequency discrimination. Furthermore, such correlations were generally higher for frequency discrimination measures that included amplitude and frequency roving. The results indicate that complex pitch perception in CI users is particularly relevant to phoneme identification in temporally fluctuating noise.

ACKNOWLEDGMENTS

This research was supported by NIH grant DC010524-02. The author thanks Louis D. Braida and Andrew E. Vandali for helpful comments on an early draft of this article. The author also thanks Amy Martinez for assistance in collecting subject data.

Conflict of interest The author declares that he has no conflict of interest.

REFERENCES

AMITAY S, HALLIDAY L, TAYLOR J, SOHOGLU E, MOORE DR (2010) Motivation and intelligence drive auditory perceptual learning. *PLoS One* 5:e9816

- BACON SP, OPIE JM, MONTONA DY (1998) The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *J Speech Lang Hear Res* 41:549–563.
- CARIANI P, DELGUTTE B (1996) Neural correlates of the pitch of complex tones. Pitch and pitch salience. *I J Neurophys* 76:1698–1716
- CARIYON RP, DEEKS JM, MCKAY CM (2010) The upper limit of temporal pitch for cochlear-implant listeners: stimulus duration, conditioner pulses, and the number of electrodes stimulated. *J Acoust Soc Am* 127:1469–1478
- DEMANY L, SEMAL C (2002) Learning to perceive pitch differences. *J Acoust Soc Am* 111:1377–1388
- FEISE RJ (2002) Do multiple outcome measures require p-value adjustment? *BMC Med Res Methodol* 2:8
- FUJITA S, ITO J (1999) Ability of nucleus cochlear implantees to recognize music. *Ann Oto Rhinol Laryn* 108:634–640
- GELMAN A, HILL J, YAJIMA M (2012) Why we (usually) don't have to worry about multiple comparisons. *J Res Educ Effect* 5:189–211
- GFELLER KE, TURNER C, MEHR M, WOODWORTH G, FEARN R, KNUTSON JF, STORDAHL J (2002) Recognition of familiar melodies by adult cochlear implant recipients and normal hearing adults. *Cochlear Implants Int* 3:29–53
- GOLDSWORTHY RL, DELHORNE LA, BRAIDA LD, REED CM (2013) Psychoacoustic and phoneme identification measures in cochlear-implant and normal hearing listeners. *Trends Amplif* 17:27–44
- GOLDSWORTHY RL, SHANNON RV (2014) Training improves cochlear implant rate discrimination on a psychophysical task. *J Acoust Soc Am* 135:334–341
- HANEKOM JJ, SHANNON RV (1998) Gap detection as a measure of electrode interaction in cochlear implants. *J Acoust Soc Am* 104:2372–2384
- HAUKOOS JS, LEWIS RJ (2005) Advanced statistics: bootstrapping confidence intervals for statistics with “difficult” distributions. *Acad Emerg Med* 12:360–365
- HENRY BA, MCKAY CM, McDERMOTT HJ, CLARK GM (2000) The relationship between speech perception and electrode discrimination in cochlear implantees. *J Acoust Soc Am* 108:1269–1280
- HENRY BA, TURNER CW (2003) The resolution of complex spectral patterns by cochlear implant and normal hearing listeners. *J Acoust Soc Am* 113:2861–2873
- HILLENBRAND J, GETTY LA, CLARK MJ, WHEELER K (1995) Acoustic characteristics of american english vowels. *J Acoust Soc Am* 97:3099–3111
- HUGHES ML, GOULSON AM (2011) Electrically evoked compound action potential measures for virtual channels versus physical electrodes. *Ear Hear* 32:323–330
- JIN SH, NELSON PB (2006) Speech perception in gated noise: the effects of temporal resolution. *J Acoust Soc Am* 119:3097–3108
- KAERNBACH C (1991) Simple adaptive testing with the weighted up-down method. *Percept Psychophys* 75:227–230
- KAERNBACH C, BERING C (2001) Exploring the temporal mechanism involved in the pitch of unresolved harmonics. *J Acoust Soc Am* 110:1039–1048
- KWON BJ, TURNER CW (2001) Consonant identification under maskers with sinusoidal modulation: masking release or modulation interference? *J Acoust Soc Am* 110:1130–1140
- KWON BJ, PERRY TT, WILHELM CL, HEALY EW (2012) Sentence recognition in noise promoting or suppressing masking release by normal hearing and cochlear-implant listeners. *J Acoust Soc Am* 131:3111–3119
- LI JC, CHAN W, CUI Y (2011) Bootstrap standard error and confidence intervals for the correlations corrected for indirect range restriction. *Brit J Math Stat Psy* 5:367–387

- LIU C, EDDINS DA (2012) Measurement of stop consonant identification using adaptive tracking procedures. *J Acoust Soc Am* 132:EL250–256
- LOOI V, McDERMOTT H, McKAY C, HICKSON L (2008) Music perception of cochlear implant users compared with that of hearing aid users. *Ear Hear* 29:421–434
- LOOI V, RADFORD CJ (2011) A comparison of the speech recognition and pitch ranking abilities of children using a unilateral cochlear implant, bimodal stimulation or bilateral hearing aids. *Int J Pediatr Otorhinolaryngol* 75:472–482
- LOOI V, GFELLER K, DRISCOLL VD (2012) Music appreciation and training for cochlear implant recipients: a review. *Semin Hear* 33:307–334
- McDERMOTT HJ (2004) Music perception with cochlear implants: a review. *Trends Amplif* 8:49–82
- McDERMOTT HJ, McKAY CM (1997) Musical pitch perception with electrical stimulation of the cochlea. *J Acoust Soc Am* 101:1622–1631
- MICHEYL C, DELHOMMEAU K, PERROT X, OXENHAM AJ (2006) Influence of musical and psychoacoustical training on pitch discrimination. *Hear Res* 219:36–47
- MICHEYL C, XIAO L, OXENHAM AJ (2012) Characterizing the dependence of pure-tone frequency difference limens on frequency, duration, and level. *Hear Res* 292:1–13
- NELSON PB, JIN SH, CARNEY AE, NELSON DA (2003) Understanding speech in modulated interference: cochlear implant users and normal hearing listeners. *J Acoust Soc Am* 113:961–968
- NILSSON W, CASTRO B (2012) Bootstrap confidence interval for a correlation curve. *Stat Prob Lett* 82:1–6
- PARKS TW, BURRUS CS (1987) *Digital filter design*. Wiley-Interscience, New York, NY
- PAVLOVIC CV (1987) Derivation of primary parameters and procedures for use in speech intelligibility predictions. *J Acoust Soc Am* 82:413–422
- PETERS RW, MOORE BCJ, BAER T (1998) Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *J Acoust Soc Am* 103:577–587
- PRETORIUS LL, HANEKOM JJ (2008) Free field frequency discrimination abilities of cochlear implant users. *Hear Res* 244:77–84
- QIN MK, OXENHAM AJ (2003) Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *J Acoust Soc Am* 114:446–454
- RADER T, FASTL H, BAUMANN U (2013) Speech perception with combined electric-acoustic stimulation and bilateral cochlear implants in a multisource noise field. *Ear Hear* 34:324–332
- ROTHMAN KJ (1990) No adjustments are needed for multiple comparisons no adjustments are needed. *Epidemiology* 1:43–46
- SEK A, MOORE BCJ (1995) Frequency discrimination as a function of frequency, measured in several ways. *J Acoust Soc Am* 97:2479–2486
- SHANNON RV (1983) Multichannel electrical stimulation of the auditory nerve in man. I basic psychophysics. *Hear Res* 11:157–189
- SHANNON RV, JENSVOED A, PADILLA M, ROBERT ME, WANG X (1999) Consonant recordings for speech testing. *J Acoust Soc Am* 106:L71–74
- SIEVERS W (1996) Standard and bootstrap confidence intervals for the correlation coefficient. *Brit J Math Stat Psy* 49:381–396
- STICKNEY GS, ZENG FG, LITOVSKY R, ASSMANN P (2004) Cochlear implant speech recognition with speech maskers. *J Acoust Soc Am* 116:1081–1091
- SUCHER CM, McDERMOTT HJ (2007) Pitch ranking of complex tones by normally hearing subjects and cochlear implant users. *Hear Res* 230:80–87
- TONG YC, CLARK GM, BLAMEY PJ, BUSBY PA, DOWELL RC (1982) Psychophysical studies for two multiple-channel cochlear implant patients. *J Acoust Soc Am* 71:153–160
- VANDALI AE, SIY D, COWAN R, VAN HOESEL RJM (2014) Training of cochlear implant users to improve pitch perception in the presence of competing place cues. *Ear Hear* 36:e1–e13
- VANDALI AE, VAN HOESEL RJM (2012) Enhancement of temporal cues to pitch in cochlear implants: effects on pitch ranking. *J Acoust Soc Am* 132:392–402
- ZENG FG (2002) Temporal pitch in electric hearing. *Hear Res* 174:101–106