Relationship between multipulse integration and speech recognition with cochlear implants

Ning Zhou^{a)}

Department of Communication Sciences and Disorders, East Carolina University, Greenville,
North Carolina 27834

Bryan E. Pfingst

Kresge Hearing Research Institute, Department of Otolaryngology, University of Michigan,
Ann Arbor, Michigan 48109-5616

(Received 19 November 2013; revised 25 June 2014; accepted 3 July 2014)

Comparisons of performance with cochlear implants and postmortem conditions in the cochlea in humans have shown mixed results. The limitations in those studies favor the use of within-subject designs and non-invasive measures to estimate cochlear conditions. One non-invasive correlate of cochlear health is multipulse integration, established in an animal model. The present study used this measure to relate neural health in human cochlear implant users to their speech recognition performance. The multipulse-integration slopes were derived based on psychophysical detection thresholds measured for two pulse rates (80 and 640 pulses per second). A within-subject design was used in eight subjects with bilateral implants where the direction and magnitude of ear differences in the multipulse-integration slopes were compared with those of the speech-recognition results. The speech measures included speech reception threshold for sentences and phoneme recognition in noise. The magnitude of ear difference in the integration slopes was significantly correlated with the magnitude of ear difference in speech reception thresholds, consonant recognition in noise, and transmission of place of articulation of consonants. These results suggest that multipulse integration predicts speech recognition in noise and perception of features that use dynamic spectral cues.

© 2014 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4890640>]

PACS number(s): 43.66.Ts, 43.66.Cb, 43.71.Ky [DB]

Pages: 1257–1268

I. INTRODUCTION

Cochlear implants are the treatment of choice to provide hearing in patients with severe to profound sensorineural hearing loss. Many factors contribute to the outcome with a cochlear implant including the insertion depth of the electrode array, the electrode-neuron interface, the health of the neural population, cognitive skills required to understand degraded speech signals, and experience with the device.

Because the devices bypass the damaged organ of Corti and directly stimulate auditory nerves with patterned electrical pulses, one of the conditions assumed to be important for cochlear implant function is the health of the target auditory nerve population. The anatomical measure that is commonly used to estimate health of the auditory nerve is the number or density of surviving, healthy-looking spiral ganglion cell bodies. In non-implanted deaf human ears, the spiral ganglion survival range is approximately 4% to 92% of normal depending on etiology, duration of deafness, and other variables (Hinojosa and Marion, 1983; Incesulu and Nadol, 1998). In deaf implanted ears, the survival range of the spiral ganglion cells is only moderately lower, i.e., 2% to 72% (Nadol *et al.*, 2001; Fayad and Linthicum, 2006). While it is intuitive to assume that the function of a cochlear implant depends on the status of neural survival in the inner ear, comparisons of postmortem

conditions in the human cochlea with the implant functional data collected in these humans prior to death do not necessarily support this notion. Khan *et al.* (2005) reported that the counts of spiral ganglion cells did not predict speech recognition performance measured before death in 15 cochlear implant users. A negative correlation between spiral ganglion cell counts and NU-6 words was demonstrated by Nadol *et al.* (2001). Interpretation of the lack of significant correlation or even negative correlation between implant function and neural survival is difficult. The results might have been confounded by differences in cognitive function across subjects, such that two subjects with the same neural input from the cochlear implant might differ greatly in the cognitive ability to use that information. An additional complication in interpreting the relationship between human spiral ganglion cell count and speech recognition is that the condition of the nerve at the time the speech measures are taken might be markedly different from the condition at the time that the temporal bones are harvested, depending on the duration of that interval and intervening variables, such as the cause of death.

Animal models, where these confounding factors can be better controlled, have provided a clearer picture of the relationship between neural health and implant function than is possible with humans. Simpler and more objective measures of implant function used in the animal models can also help reduce the confounding effects of cognitive ability. Several psychophysical and electrophysiological measures have been developed to probe sensory and neural health in

^{a)}Author to whom correspondence should be addressed. Electronic mail: zhoun@ecu.edu

implanted animals. These measures include psychophysical threshold functions such as temporal integration of electrical pulse trains (threshold-versus-stimulus-duration functions) as well as electrophysiological measures such as electrically evoked compound action potentials and auditory brainstem responses (Hall, 1990; Prado-Guitierrez *et al.*, 2006; Smith and Simmons, 1983; Kang *et al.*, 2010; Pfungst *et al.*, 2011a; Pfungst *et al.*, 2011b; Ramekers *et al.*, 2014). Advantages of an animal model include that histological analysis can be conducted a short time after these functional measurements are made and that the conditions of the inner ear of the animals can be systematically manipulated. Most importantly, if these psychophysical and electrophysiological measures are non-invasive, they can be used for probing the inner ear conditions in humans with cochlear implants and relating these conditions to the speech recognition performance of the implantees.

One of the psychophysical measures that showed significant correlation with spiral ganglion cell counts is a multipulse-integration measure (Kang *et al.*, 2010; Pfungst *et al.*, 2011b). The multipulse-integration measure is based on psychophysical detection thresholds as a function of pulse rate for pulse trains fixed in duration. With the duration of the stimulus fixed, the increase in pulse rate increases the number of pulses in the pulse train. Guinea pigs with greater than 70% nerve survival tended to be able to integrate the multiple pulses, demonstrating a decrease of psychophysical detection thresholds with increasing pulse rate on the order of -1 to -3 dB per doubling of pulse rate (i.e., steeper slopes for multipulse-integration functions). It is important to note that in these studies, animals with steep multipulse-integration slopes had high spiral ganglion cell counts as well as high counts of residual inner hair cells (IHCs). However, this does not mean that steep multipulse integration depends on electrical or electrophonic stimulation of IHCs. Data from human subjects in the current study as well as a previous study (Zhou *et al.*, 2012) showed that the range of multipulse-integration functions was very similar to that found in guinea pigs with a known large range of nerve survival despite the fact that none of the human subjects had any measurable hearing and thus, they probably had no functional IHCs. An alternative interpretation for the role of IHCs is that they contribute to the survival and possibly the health of auditory nerve fibers. These findings in animals and humans suggest that multipulse-integration functions

could be a psychophysical candidate for estimating neural survival across stimulation sites in human ears.

The purpose of the current study was to examine whether multipulse integration predicts speech recognition in humans with cochlear implants. We hypothesized that the most direct effect of sparse auditory nerve survival would be a decrease of spatial selectivity leading to less independence between neural populations stimulated by individual spectral channels from the speech processor. The decreased independence of neural channels would in turn affect speech recognition in noise and the transmission of speech features that rely on spectral cues such as formant frequencies, formant transitions, and low frequency periodicity.

Attempts to relate psychophysical data to speech recognition performance across individuals often fail, probably because speech recognition can depend strongly on other subject variables such as cognitive ability. In the current study, a within-subject design was used to eliminate such across-subject confounding factors by comparing performance of the two ears of the same bilaterally implanted subject. The general approach was to compare the direction as well as magnitude of the ear difference in the multipulse-integration measure of neural survival to the direction and magnitude of ear difference in various speech-recognition measures. It is important to note that in addition to neural health, ear-specific variables such as insertion depth and electrode-neuron interface can also impact the performance of the two ears of the same subject independently.

II. METHODS

A. Subjects

Eight subjects implanted bilaterally with the Nucleus (Cochlear Corporation, Englewood, CO) cochlear implants participated in this study. Their demographic information is provided in Table I. The subjects were all sequentially implanted. All ears were tested for residual acoustic hearing using narrow-band noise stimuli with center frequencies at octave intervals between 250 and 8000 Hz via insertion earphones. The extended stimulus levels [120 dB hearing level (HL) for all frequencies] were used and none of the ears had measurable acoustic hearing. The use of human subjects in this study was reviewed and approved by the University of Michigan Medical School Institutional Review Board.

TABLE I. Subject demographics.

Subject	Gender	Age (yr)	CI use L (yr)	CI use R (yr)	Duration of deafness L (yr)	Duration of deafness R (yr)	Implant type (L/R)	Etiology (L/R)
S52	F	58.5	1.9	12.0	54.0	21.5	CI24R(CA)/CI24M	Trauma/Hereditary
S60	M	71.4	8.2	2.2	0.2	6.2	CI24R(CS)/CI24RE(CA)	Hereditary/Progressive
S69	M	70.8	6.4	1.6	4.4	9.2	CI24R(CA)/CI512	Noise exposure/Noise exposure
S81	F	59.8	5.2	3.7	2.5	4.0	CI24RE(CA)/CI24RE(CA)	Progressive/Progressive
S88	M	60.0	2.2	8.5	4.5	33.5	CI24RE(CA)/CI24RE(CA)	Progressive/Progressive
S89	M	66.4	2.3	6.5	64.1	59.9	CI24RE(CA)/CI24R(CA)	Hereditary/Hereditary
S95	F	19.9	0.9	1.4	19.0	18.4	CI24RE(CA)/CI512	Hereditary/Hereditary
S96	F	65.7	0.2	10.3	29.5	25.4	CI24RE(CS)/CI24R(CS)	Progressive/Progressive

B. Measuring multipulse integration

A laboratory owned Freedom speech processor (Cochlear Corporation, Englewood, CO) was used for the psychophysical testing. Stimuli for measuring multipulse integration were trains of symmetric biphasic pulses of two stimulation rates (80 and 640 pps). The pulse duration was 25 μ s/phase with an interphase interval of 8 μ s, and the pulse-train duration was 500 ms. The pulse rates used were chosen to be below 1000 pps because threshold change at pulse rates above 1000 pps is believed to involve a mechanism of residual cell charge and the animal studies showed that spiral ganglion density was more strongly correlated with the slopes of the multipulse-integration function below 1000 pps than above (Pfungst *et al.*, 2011b). For each ear, detection thresholds were measured three times for each of the two pulse rates for all functioning electrodes in a monopolar configuration (MP 1 + 2). Functioning electrodes were those that were turned on in the subject's clinical MAP, had normal impedances, and did not cause perceptual discomfort. Trials were fully randomized for electrode and pulse-rate conditions. The method of adjustment was used for measuring the thresholds, where the subjects adjusted the stimulation levels on their own to find thresholds. Subjects had access to two sets of buttons on a computer screen; one set allowed adjustment in a step size of 5 current level units (CLUs) and the other was for a finer adjustment in 1 CLU steps. The subjects were instructed to first use the larger step size to approximate the threshold region and then fine tune the stimulus level using the smaller step size (1 CLU) until the level at which the signal was just audible was determined. The stimulus was presented repeatedly with a 600 ms inter-stimulus interval as the subject adjusted the level to determine the threshold. Thresholds obtained in CLUs were first converted to microamperes and then expressed in decibels relative to 1 mA. Mean and range of the slopes of threshold-versus-pulse-rate functions were determined for each stimulation site.

C. Speech recognition tests

For each bilaterally implanted subject, speech recognition was measured for the left and right ear alone in randomized order. The speech measures consisted of a speech reception threshold (SRT) test for CUNY sentences (Boothroyd *et al.*, 1985) and phoneme recognition at various signal-to-noise ratios (SNRs). Speech performance was measured using the subjects' everyday clinical processors and settings, including Smart-sound features. The Smart-sound features were the same between ears for each subject. Tests were conducted in a double-walled sound-attenuating booth through a loudspeaker positioned 1 m away from the subject at 0° azimuth. The levels of the speech stimuli were calibrated and presented at 60 dB sound pressure level (SPL).

The 72 CUNY lists (each containing 12 sentences) were divided into blocks of 4 lists, one block for each SRT test. The CUNY sentences are meaningful utterances spoken by a male speaker. Contextual cues and lengths of sentences were not equated, but the total number of words was the same

across lists. The target CUNY sentences were presented in white noise amplitude modulated at 100% modulation depth with a 4 Hz sinusoid, a frequency similar to the slow modulation frequency in speech envelopes. A fluctuating noise was used because listening in fluctuating noises requires good spectral resolution to segregate signal from noise to take advantage of the spectral or temporal minima where SNR is favorable. The amplitude-modulated noise was presented alone for 1.5 s before the target sentence, during the target, and for 0.5 s alone after the target. The onset and the offset, each measuring 5% of the stimulus duration, were applied with raised cosine ramps. SNR was calculated for the time where the target and noise overlapped. The mixed signal (target + noise) was normalized to its peak amplitude resulting in similar stimulation levels from trial to trial. SNR started at 20 dB at the beginning of the test and adapted in a one-down one-up procedure with a step size of 2 dB. The stimulus was presented one time to the subject who was instructed to repeat the sentence to the researcher. SNR was lowered by 2 dB if the subject repeated all words in the sentence correctly or increased by 2 dB for an incorrect response. The one-down one-up procedure estimated the 50% correct point on the psychometric function. The SRT was taken as the mean of the SNRs at the last 6 reversals of a total of 12 reversals. SRTs were measured twice and the mean of the two SRTs was calculated.

Consonant and vowel recognition were measured in quiet and in a speech-shaped noise at 10 and 0 dB SNRs. Consonant and vowel materials were chosen from Shannon *et al.* (1999) presented in a consonant-/a/ context and Hillenbrand *et al.* (1995) presented in a /h/-vowel-/d/ context, respectively. Subjects were instructed to choose the token they heard from a closed set of 12 vowels or 20 consonants. The speech tokens were shown on a computer screen and selected using a mouse. Each SNR condition was tested twice and the average percent correct was calculated.

D. Information transmission analysis

In addition to deriving a percentage score for correctly recognized consonant and vowel sounds, error patterns were analyzed to examine transmission of various articulatory features for consonants and vowels (Miller and Nicely, 1955). The consonant articulatory features analyzed included voicing, place of articulation, and manner of articulation. The labials, alveolars, velars, and palatals were then binary coded and treated as subcategory features of place of articulation (see Zhou *et al.*, 2010 for a review). The three articulatory features analyzed for vowels were the first formant (F1), second formant (F2), and duration.

E. Linear regression analysis

For each ear, the slopes of the multipulse-integration functions were averaged for the whole array to derive an across-site mean (ASM) slope. The slopes were also averaged over specific frequency regions to derive regional means. The ear differences (left ear minus right ear) were then calculated for the ASM slopes, regional slopes, and various speech-recognition measures. The relationship between

the ear differences in the ASM slopes or regional slopes and the ear differences in various speech-recognition measures was examined using two approaches. First, a binary correlation was examined where the direction of ear differences in the slopes and ear differences in the speech-recognition measures was compared. Second, the relationship between the magnitude of ear differences in the slopes and the magnitude of ear differences in speech recognition was examined. For correlating magnitudes of ear differences in estimated neural survival with those in the speech-recognition measures, linear regressions were performed to force through the origin such that intercept was excluded in these linear models (Casella, 1983). The independent and dependent variables in these linear regressions were ear differences where

there was no inherent directional bias. Therefore interpretation of the linear regressions would be only meaningful if the intercept (constant) were excluded. Statistical significance for a linear regression without intercept would indicate that the two variables were proportionally related.

III. RESULTS

Figure 1 shows the ranges and means of the multipulse-integration slopes measured at each available stimulation site in each ear for the eight subjects. The slopes for multipulse integration varied across stimulation sites, to a greater degree in some ears than others. This across-site variation was statistically significant for all ears [one way repeated-

FIG. 1. Means and ranges of the multipulse-integration slopes for all available stimulation sites measured in two ears. Each panel shows data measured from the left ear as stars and right ear as squares for one bilaterally implanted subject. Subject numbers are given in the lower left corner of each panel. Across-site mean slopes (in decibels per doubling of pulse rate) for each ear and statistics for comparing the ear differences for each subject are reported in the lower right corner.

TABLE II. Correlations between the absolute detection threshold at two rates and the multipulse-integration slopes across sites.

		S52		S60		S69		S81		S88		S89		S95		S96	
		L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R
80 pps	<i>r</i>	-0.56	-0.42	-0.67	-0.64	-0.45	-0.73	-0.26	-0.16	-0.67	-0.34	-0.73	-0.45	0.25	-0.61	0.01	0.35
	<i>p</i>	0.01*	0.08	0.00*	0.00*	0.05	0.00*	0.25	0.49	0.00*	0.12	0.00*	0.04	0.27	0.00*	0.97	0.17
640 pps	<i>r</i>	-0.23	0.01	-0.16	0.09	0.89	-0.15	0.32	0.44	0.09	0.38	0.34	0.16	0.83	0.19	0.85	0.44
	<i>p</i>	0.29	0.96	0.47	0.70	0.00*	0.52	0.14	0.05	0.70	0.08	0.14	0.48	0.00*	0.40	0.00*	0.08

p values < 0.025 (Bonferroni adjusted criterion *p* value) are marked with an asterisk.

measures analysis of variance (ANOVA), $p < 0.05$]. For some subjects, the steeper slopes were in the same ear across most of the array (e.g., S96), whereas for other subjects, the better ear for multipulse integration differed across stimulation sites. Ear differences in the across-site mean slopes were only statistically different from zero for subjects 52, 89, 69, and 81 (one sample *T* test). This is not problematic for later analysis because, in theory, if the hypothesized relationship between the multipulse-integration slopes and speech recognition is true, minimal ear differences in the slopes should predict corresponding minimal ear differences in speech recognition. For each ear, analysis was conducted to examine the correlation between the integration slopes and the absolute detection threshold across the stimulation sites. The relationships between the integration slopes and the absolute detection thresholds at the two pulse rates are

shown in Table II (Pearson's *R*). Except for subject 81, one or both ears of the subjects showed a significant correlation between the absolute detection thresholds at one or both pulse rates and the multipulse-integration slopes across the stimulation sites. In cases of statistical significance, stimulation sites that required higher current for activation also tended to have shallow integration slopes.

Individual data for the ASM multipulse-integration slopes as well as the SRTs and consonant and vowel recognition at various SNRs are shown in Fig. 2. Three demographic variables including age of the subject, duration of deafness, and the duration of CI use were examined. None of these demographic variables alone, or the single principle component derived from the three demographic variables via a factor analysis, accounted for variance in the multipulse-integration slopes or variance in speech performance across

FIG. 2. The across-site mean (ASM) multipulse-integration slopes and various speech recognition measures measured for individual ears. Data from the ears with steeper ASM slopes are shown in the black bars, and the data from the ears with shallower ASM slopes are shown in gray bars.

the 16 ears (Pearson's R , $p > 0.05$). In addition, neither duration of CI use nor duration of deafness accounted for the direction of ear asymmetry in speech performance in the eight subjects.

Figure 3 shows the eight subjects' ear differences in SRTs plotted against their ear differences in the ASM slopes. Based on the direction that the ear differences were obtained (left ear minus right ear), the upper left quadrant of the figure contains data points from subjects who had both better SRT and steeper ASM slope in the left ear, whereas data points in the lower right quadrant were from subjects who had better SRT and steeper ASM slope in the right ear. Except for one subject who had no ear difference in sentence recognition in noise but demonstrated a better ASM slope in the right ear, all subjects' better ear for SRT was also the ear that showed a steeper (more negative) ASM multipulse-integration slope. A linear regression fit to the data through the origin revealed statistical significance, suggesting that not only was the direction of the ear differences consistent between the two measures but also that the magnitude of ear difference in subjects' SRTs was proportionally related to the magnitude of ear difference in the ASM slopes.

Figure 4 shows the scatter plots of the ear differences in the ASM slopes against ear differences in consonant and vowel recognition percent-correct scores at various SNRs. The majority of the data points was confined within the upper left and lower right quadrants, especially for consonants, indicating that steeper multipulse-integration slope and better consonant recognition were found in the same ear for those subjects. Interestingly, in the most challenging SNR condition (0 dB SNR), the amount of slope differences between ears was found to predict almost perfectly (Pearson's R , $r = 0.95$) the amount of ear differences in

consonant recognition. However, at other SNR conditions for consonants and at all tested SNR conditions for vowels, the magnitude of ear differences in phoneme recognition was not significantly correlated with the magnitude of ear differences in the ASM slopes.

The relationships between the ASM slopes and transmission of the voicing, manner of articulation and place of articulation features for consonants were also examined with results shown in Fig. 5. Percentages of information transmitted for the three features were averaged across the SNR conditions. Direction of ear differences was found to be the most consistent for the place-of-articulation feature. A few data points fell outside of the upper left and lower right quadrants for voicing, although the majority of the data points did suggest consistent ear differences. The least consistent ear differences were found for manner of articulation. The magnitude of ear differences in the ASM slopes significantly predicted the magnitude of ear differences in perceiving place of articulation (Pearson's R , $p < 0.016$, Bonferroni adjusted criterion), but not for voicing or manner of articulation (Pearson's R , $p > 0.016$, Bonferroni adjusted criterion).

For most subjects, multipulse integration was not steeper in the same ear across all stimulation sites; i.e., the better ear for multipulse integration differed as a function of place along the tonotopic axis of the cochlea. This led to the question of whether the ear differences in the regional slopes would predict the ear differences in articulatory features of consonant sounds that were spectrally specific. The acoustic correlates for place of articulation of consonants are spectrally specific. The location of the constriction made during consonant production (i.e., place of articulation) is perceived by detecting spectral peaks, or in cases of nasals, absence of spectral peaks (e.g., anti-formants) in the short-term spectrum at consonantal release (Fujimura, 1962; Stevens and Blumstein, 1978). The spectral characteristics cuing place of articulation lie in the highest frequency region for alveolars followed by velars, palatals, and labials. Labials are produced with virtually no front vocal cavity, demonstrating a diffused spectrum with most energy distributed in the low frequencies. For the current experiment, the multipulse-integration slopes were averaged for electrodes 1–3, 4–6, 7–16, and 15–22, which were assigned with frequencies that roughly corresponded to acoustic correlates for alveolars, velars, palatals, and labials, respectively (Table III). The ear differences in the slopes were significant for all subjects in all four frequency regions ($p < 0.05$) except for the frequency region corresponding to the velar feature for subject 81.

Figure 6 shows the scatter plots of ear differences in transmission of these frequency-specific place-of-articulation features at three SNRs in relation to the ear differences in the regional slopes. Results showed that the transmission of features with acoustic correlates in specific frequency regions were significantly correlated with multipulse-integration slopes averaged across the corresponding frequency regions under all SNR conditions (Pearson's R , $p < 0.016$, Bonferroni adjusted criterion). The trend was that the correlation became stronger as the SNR became more challenging.

FIG. 3. Scatter plot of ear differences (left minus right) in the speech reception thresholds (SRTs) against the ear differences in the ASM multipulse-integration slopes. Ear-difference data in the left upper quadrant are from subjects whose better ear is the left ear for both variables (L), and data in the right lower quadrant are from subjects whose better ear is the right ear for both variables (R). Each data point represents one bilaterally implanted subject. Regression through the origin is indicated with a dashed line. The r and p values for the regression are shown in the lower left corner of the figure.

FIG. 4. Scatter plot of ear differences (left minus right) in consonant recognition (top row) and vowel recognition (bottom row) against the ear differences in the ASM multipulse-integration slopes. The letters “L” and “R” indicate the direction of ear differences. Each data point represents one bilaterally implanted subject (refer to legend in Fig. 3). Regression through the origin is indicated with a dashed line. The r and p values for the regressions are shown in the lower left corner of each panel.

The relationship between the transmissions of articulatory features for vowels and the multipulse-integration slopes is shown in Fig. 7. The frequency ranges for the first and second formants of the vowels were determined based

on an acoustic analysis of the stimuli used in the study, which was close to the averaged formant frequencies for American English (Lieberman and Blumstein, 1988). Transmission of the second formant was compared with the

FIG. 5. Scatter plot of ear differences (left minus right) in transmitting the consonant features against the ear differences in the ASM multipulse-integration slopes. Symbols indicate various signal-to-noise ratios (SNRs) and the mean collapsed across the three SNRs. L and R indicate the direction of ear differences. Each data point represents one condition for one bilaterally implanted subject. In each panel, a regression through the origin for the mean data is indicated with a dashed line with statistics shown in the lower left corner.

TABLE III. Frequency allocations associated with electrodes selected for the regional slopes.

	Electrodes	Frequency allocation (Hz)	Features
Consonants	1–3	5313–7938	Alveolar
	4–6	3563–5413	Velar
	7–16	938–3563	Palatal
	15–22	188–1063	Labial
Vowels	16–22	125–1036	F1
	9–16	938–2688	F2

regional multipulse-integration slopes averaged across electrodes 9–16, and the first formant was compared with the regional slopes averaged across electrodes 16–22 (Table III). The ear differences in the transmission of the duration feature were compared with the ear differences in the slopes averaged across the whole array. The ear differences in the slopes were significant for all subjects for the two frequency regions analyzed ($p < 0.05$). The direction of ear differences in the ASM or regional slopes was not always consistent with that in the various articulatory features of the vowels nor did the magnitudes of ear differences correlate.

IV. DISCUSSION

The relationship between a psychophysical measure, which was correlated with auditory neuron survival in an animal model, and speech recognition in humans, was examined. The results of the present study indicate that neural survival estimated by slopes of the multipulse-integration functions predicts ear asymmetry in human listeners' performance for sentence recognition in noise, consonant recognition at challenging SNRs, and reception of various consonant features that are dependent on the transmission of spectral information.

A. Across-site variation

Results of the present study showed that the slopes of multipulse-integration functions varied across the

stimulation sites in patterns that were subject and ear specific. This suggests an uneven distribution of pathology along the tonotopic axis. Across-site variation has been found for various psychophysical measures (Zwolan *et al.*, 1997; Pfingst *et al.*, 2004; Pfingst and Xu, 2005; Bierer, 2007; Bierer and Faulkner, 2010; Pfingst *et al.*, 2008; Zhou *et al.*, 2012; Garadat *et al.*, 2012), and the patterns in which these psychophysical measures vary across the stimulation sites are also unique to each measure (Pfingst *et al.*, 2011a). The unique patterns suggest that the various psychophysical measures are not mediated by a single mechanism.

For multipulse integration, the animal studies have demonstrated that a significant amount of the variance in the slopes of the functions could be accounted for by the density of the spiral ganglion cell bodies of the auditory neurons in Rosenthal's canal near the stimulation sites (Kang *et al.*, 2010; Pfingst *et al.*, 2011b). The mechanism for the effects of auditory fiber density on the ear's ability to integrate multiple electrical pulses is not fully understood. In McKay and colleagues' model of temporal processing (McKay and McDermott, 1998; McKay *et al.*, 2013), detection of pulsatile stimulation in electric hearing is influenced by a combined effect of reduced neural response due to refractory effects and temporal integration of neural activity in the integration window. Previous reports have shown that guinea pigs with sparse neural survival also tend to have low spontaneous activity in the remaining fibers, probably due to the loss of inner hair cells (Kang *et al.*, 2010; Pfingst *et al.*, 2011b). The lack of spontaneous activity can lead to abnormal across-fiber synchrony. That is, if the auditory nerve is silent when the first pulse in a pulse train arrives, all of the fibers will tend to fire at once and then enter a refractory state so that there is no response to the closely following subsequent pulses. After recovery from refraction, the sequence would repeat (Wilson *et al.*, 1997). Above about 200 pps, stimuli would be undersampled due to the abnormal across-fiber synchrony, leading to shallow threshold-versus-pulse-rate functions. Neural health in a broader sense could include

FIG. 6. Scatter plot of ear differences (left minus right) in transmitting the frequency-specific place of articulation features (in different symbols) against the ear differences in the multipulse-integration slopes averaged across the corresponding frequency regions. Results for three SNR conditions are shown. L and R indicate the direction of ear differences. Regression for all articulation features is indicated with a dashed line through the origin with statistics shown in the lower left corner.

FIG. 7. Scatter plot of ear differences (left minus right) in transmitting the vowel features against the ear differences in the ASM slopes for the duration feature and regional slopes for the F2 and F1 features. Symbols indicate various SNRs and the mean collapsed across the three SNRs. L and R indicate the direction of ear differences. Each data point represents one condition for one bilaterally implanted subject. Regression through the origin for the mean data is indicated with a dashed line with statistics shown in the lower left corner.

not only the number of surviving auditory nerve fibers but also the condition of those fibers (myelination, membrane properties, metabolic function, etc.), all of which could affect the responses of the auditory nerve population to electrical stimulation. One possibility is that the surviving neurons in ears that have considerable damage might have prolonged refractoriness due to loss of myelin (Zhou *et al.*, 1995a; Zhou *et al.*, 1995b), so that increases in stimulus pulse rate would not result in comparable increases in evoked discharge rate. This would also lead to shallower pulse rate functions in ears that have greater pathology.

B. Effects of neural survival on speech recognition in noise

Spiral ganglion cells and their peripheral processes degenerate upon the damage of cochlea and loss of hair cells. The amount of nerve loss is generally believed to increase with the duration of deafness without the support from hair cells. Many studies have demonstrated that speech recognition with cochlear implants is related to a subject's duration of deafness (e.g., Rubinstein *et al.*, 1999). However, there is little evidence in the literature that has linked the dependence of speech recognition in electrical hearing with neural survival in the inner ear.

In damaged ears, thresholds for activation can be elevated due to loss of the most sensitive fibers or increased distance between the implant electrodes and the most sensitive nerve fibers (Bierer and Faulkner, 2010; Long *et al.*, 2014). Our data showed that in some ears, stimulation sites that had shallow slopes did also tend to have higher absolute detection thresholds. In such cases, the elevated stimulation levels are likely to cause greater spread in neural excitation, which in turn could result in a reduction in spatial tuning. Our first hypothesis was that nerve loss would have a detrimental effect on speech recognition in noise. This hypothesis was developed under the assumption that speech recognition in noise requires much better spectral resolution than does speech recognition in quiet, which is discussed in detail in

the following section. Evidence from the present study supported this hypothesis by showing that subjects who had steeper slopes of multipulse integration in one ear also performed better for sentence recognition in amplitude-modulated noise in the same ear. The consistent ear differences were found for consonant and vowel recognition in speech-shaped noise for most cases as indicated by data points confined in the two quadrants (Fig. 4). It should be noted that the fact that one ear is better at both measures does not seem to be due to ear differences in the amount of practice listening with the implants because the ear with more device experience was not always the ear that performed better on the two measures. More interestingly, the magnitude of the differences in estimated nerve loss between the two ears predicted proportionally the differences in the amount of noise that the two ears were able to tolerate for understanding sentences at the 50% correct level (Fig. 3). Such a proportional correlation was also found for consonant recognition at 0 dB SNR (Fig. 4, upper left panel) but not for the two other easier SNR conditions.

Many studies in the literature have suggested that speech recognition in noise is dependent on spectral resolution to a greater degree than speech recognition in quiet. Vocoder studies have suggested that a substantially increased number of channels is needed for listening in noise backgrounds (Dorman *et al.*, 1998; Xu and Zheng, 2007), particularly for vowel and voice pitch perception (approximately > 20 channels) (Kong and Zeng, 2006). The effective number of channels in a cochlear implant has found to be much lower (i.e., ~8) than the number of implanted electrodes (Friesen *et al.*, 2001). This explains for example why implant users do not show masking release when a steady-state noise is replaced with an amplitude-modulated noise or a competing talker (Nelson *et al.*, 2003; Fu and Nogaki, 2005) as normal hearing listeners typically do (Festen and Plomp, 1990). The difficulties with listening in fluctuating noise via cochlear implants are attributed to the listeners' inability to use the temporal or spectral minima since frequency selectivity falls short of supporting segregation of

target from the noise. [Fu and Nogaki \(2005\)](#) provided direct evidence that spectral smearing as a result of the loss of spatial tuning in electrical hearing is the reason that implant users cannot segregate speech information from fluctuating noises. They demonstrated that as the carrier filter slope was decreased to -6dB/octave , sentence recognition in normal-hearing subjects listening to the acoustic simulation of a cochlear implant became equivalent to the performance on the same task measured in implanted listeners. Nerve loss is often hypothesized to be the underlying reason for the spread of neural excitation that leads to poor spatial tuning, and the results of the present study, which suggested a link between neural survival estimated psychophysically and speech recognition performance in noise, provided evidence consistent with that hypothesis.

C. Effects of neural survival on transmission of articulatory features

Our second hypothesis was that nerve loss would have a detrimental effect on the transmission of speech features that rely on spectral information. Evidence to support this hypothesis came from our results that showed strong correlations between ear differences in the multipulse-integration slopes and transmission of place of articulation, a feature that is primarily spectral (Fig. 5, right panel). For stops, place of articulation is perceived by detecting the locations of peaks or minima in the short-term spectrum at consonantal release or the direction of F2 transition ([Stevens and Blumstein, 1978](#)). It is important to note that the acoustic correlates for place of articulation are either based on short-term spectra or dynamic changes of spectral information that are different from the more steady-state formant frequencies of vowels. Our results indicate that nerve loss would result in difficulties in perceiving these more dynamic spectral cues. More importantly, the multipulse-integration slopes almost perfectly predicted the amount of information lost in the transmission of the place of articulation feature. This is consistent with the assumption that nerve loss would result in reduced salience of the spectral representation in the neural population response or increased spread of excitation that would blur the independence of neural response channels ([Baer and Moore, 1994](#)).

Given this strong effect observed for place of articulation, we considered a second research question that involved the dependence of frequency-specific subcategory features of place of articulation on nerve survival in the corresponding frequency regions. The question was whether the transmission of alveolars, velars, palatals, and labials, each treated as a binary feature with acoustic correlates in specific frequency regions, could be predicted by neural survival near the corresponding stimulation sites.

Note that for most subjects, the direction and magnitude of ear differences in the multipulse-integration slopes differed across stimulation sites (Fig. 1). Correlations between the ear differences in the transmission of frequency-specific features and those in the regionally averaged slopes were significant for all three SNRs. The strength of the correlations, not surprisingly, was found to

be the strongest for the most challenging SNR and weaker for the 10 dB SNR and the quiet conditions (Fig. 6). These results suggested that nerve survival is important for perceiving place of articulation of consonants, and the regions where neural survival are poor would affect perceiving place of articulation of consonants that have acoustic correlates at those frequencies.

The dependence of transmission of the voicing feature on nerve survival was weaker. The correlation was not significant after a Bonferroni correction (Fig. 5, first panel). The acoustic correlates for the voicing distinction can be both temporal and spectral. The temporal cues include the duration of the oral constriction or closure, durations of preceding vowels, and the absence or presence of transient release ([Pickett, 1999](#)). The spectral cues for voicing are the presence of low frequency periodicity during constriction and correlated periodicity in the middle and high frequencies that is sometimes superimposed on the random amplitude fluctuations of the turbulence ([Pickett, 1999](#)). Therefore it seems reasonable to expect a somewhat weaker relationship between the perception of voicing and the multipulse-integration slopes. Manner of articulation is coded mainly with temporal information relative to the voicing and place of articulation features. These cues could be murmur intensity or rise time of formant transitions. Abrupt or gradual changes in spectral envelopes could also be related to perception of manner of articulation ([Pickett, 1999](#)). These results suggested that the dependence on nerve survival varied systematically with the degree to which the consonant features were coded spectrally.

A puzzling finding of the present study was the weak relationship between vowel recognition and the estimated neural survival. No significant correlations for magnitude were found between multipulse-integration slopes and percent correct scores for vowel recognition at all three SNRs, although direction of ear differences was consistent between the two measures in most cases (Fig. 4). There were no compelling theoretical reasons to predict a relationship between the perception of the duration feature of vowels and neural survival. However, transmission of the formant features did not seem to be dependent on multipulse-integration slopes averaged from the corresponding frequency regions either. Similar to place of articulation of consonants, formant frequencies of vowels reveal information regarding the position of tongue in the production of vowels ([Pickett, 1999](#)). One of the possible reasons that could account for the weaker relationship is that the spectral cues for perceiving vowels are more steady state than the spectral cues for consonant features, which are typically more transient in time, such as the short-term spectrum of consonantal release or formant transitions.

The variance in speech recognition in noise and perception of speech cues that are primarily spectral apparently are not completely explained by the integration slopes. Although comparing ear differences eliminated some of the subject variables such as cognition, other subject variables specific to ears including insertion depth and electrode-neuron interface could also greatly affect the speech perception outcome.

D. The relationship between slopes of multipulse integration functions and other measures

One interpretation of the results of this study is that the loss of auditory neurons results in elevation of electrical thresholds, which leads to greater spread of neural excitation. To test this interpretation one could compare the slopes of multipulse integration, a potential measure for survival, to psychophysical measures that explicitly assess spatial specificity of electric stimulation such as forward masking (Nelson *et al.*, 2011; Chatterjee *et al.*, 2006). As was discussed earlier, nerve survival density describes one aspect of neural health and the surviving neurons can vary in many aspects such as myelination, membrane properties, or metabolic function that all potentially affect the functional response to cochlear implant stimulation. It would be interesting to relate the psychophysical measure for nerve survival to measures that assess the temporal processing fidelity of the auditory neurons such as amplitude modulation detection (Garadat *et al.*, 2012; Zhou and Pfungst, 2012). It remains to be tested whether temporal processing acuity would be dependent on the density of nerve fibers or is a function of other properties of the surviving neurons. Future studies are warranted to examine other candidate measures for estimating neural status in humans.

E. Clinical implications

In the current study, the relationship between slopes of multipulse-integration functions and speech recognition was examined using a within-subject design that compared ear differences in individual subjects. An alternative approach would be to compare speech recognition using experimental processor maps with stimulation sites for one map selected to have steep slopes and stimulation sites for the other map having shallow slopes. An example of this approach using another psychophysical measure has been published previously (Garadat *et al.*, 2012). In the clinic, stimulation sites with steep integration slopes could be selectively favored in the processor map in an effort to improve speech recognition, similar to the experiment by Garadat *et al.* (2013) that selected sites based on modulation detection acuity. A potential advantage of multipulse-integration slopes over some other psychophysical measures is that the multipulse-integration measure is simpler and relatively quick and it requires little or no training of the patients since threshold measures are frequently used during clinical visits.

ACKNOWLEDGMENT

We thank our dedicated subjects with cochlear implants. This work was supported by NIH-NIDCD Grant Nos. R01 DC010786, T32 DC00011, and P30 DC05188.

- Baer, T., and Moore, B. C. (1994). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," *J. Acoust. Soc. Am.* **95**, 2277–2280.
- Bierer, J. A. (2007). "Threshold and channel interaction in cochlear implant users: Evaluation of the tripolar electrode configuration," *J. Acoust. Soc. Am.* **121**, 1642–1653.
- Bierer, J. A., and Faulkner, K. F. (2010). "Identifying cochlear implant channels with poor electrode-neuron interface: Partial tripolar, single-

- channel thresholds and psychophysical tuning curves," *Ear Hear.* **31**, 247–258.
- Boothroyd, A., Hanin, L., and Hnath, T. (1985). *A Sentence Test of Speech Perception: Reliability, Set Equivalence, and Short Term Learning* (City University of New York, New York).
- Casella, G. (1983). "Leverage and regression through the origin," *Am. Stat.* **37**, 147–152.
- Chatterjee, M., Galvin, J. J. III, Fu, Q. J., and Shannon, R. V. (2006). "Effects of stimulation mode, level and location on forward-masked excitation patterns in cochlear implant patients," *J. Assoc. Res. Otolaryngol.* **7**, 15–25.
- Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (1998). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels," *J. Acoust. Soc. Am.* **104**, 3583–3585.
- Fayad, J. N., and Linthicum, F. H., Jr. (2006). "Multichannel cochlear implants: Relation of histopathology to performance," *Laryngoscope* **116**, 1310–1320.
- Festen, J. M., and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *J. Acoust. Soc. Am.* **88**, 1725–1736.
- Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.* **110**, 1150–1163.
- Fu, Q. J., and Nogaki, G. (2005). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," *J. Assoc. Res. Otolaryngol.* **6**, 19–27.
- Fujimura, O. (1962). "Analysis of nasal consonants," *J. Acoust. Soc. Am.* **34**, 1865–1875.
- Garadat, S. N., Zwolan, T. A., and Pfungst, B. E. (2012). "Across-site patterns of modulation detection: Relation to speech recognition," *J. Acoust. Soc. Am.* **131**, 4030–4041; **134**, 715(E) (2013).
- Garadat, S. N., Zwolan, T. A., and Pfungst, B. E. (2013). "Using temporal modulation sensitivity to select stimulation sites for processor MAPs in cochlear implant listeners," *Audiol. Neurotol.* **18**, 247–260.
- Hall, R. D. (1990). "Estimation of surviving spiral ganglion cells in the deaf rat using the electrically evoked auditory brainstem response," *Hear. Res.* **49**, 155–168.
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.
- Hinojosa, R., and Marion, M. (1983). "Histopathology of profound sensorineural deafness," *Ann. N.Y. Acad. Sci.* **405**, 459–484.
- Incesulu, A., and Nadol, J. B., Jr. (1998). "Correlation of acoustic threshold measures and spiral ganglion cell survival in severe to profound sensorineural hearing loss: Implications for cochlear implantation," *Ann. Otol. Rhinol. Laryngol.* **107**, 906–911.
- Kang, S. Y., Colesa, D. J., Swiderski, D. L., Su, G. L., Raphael, Y., and Pfungst, B. E. (2010). "Effects of hearing preservation on psychophysical responses to cochlear implant stimulation," *J. Assoc. Res. Otolaryngol.* **11**, 245–265.
- Khan, A. M., Handzel, O., Burgess, B. J., Damian, D., Eddington, D. K., and Nadol, J. B., Jr. (2005). "Is word recognition correlated with the number of surviving spiral ganglion cells and electrode insertion depth in human subjects with cochlear implants?," *Laryngoscope* **115**, 672–677.
- Kong, Y. Y., and Zeng, F. G. (2006). "Temporal and spectral cues in Mandarin tone recognition," *J. Acoust. Soc. Am.* **120**, 2830–2840.
- Lieberman, P., and Blumstein, S. (1988). *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge University Press, New York), Chap. 10, pp. 221–241.
- Long, C. J., Holden, T. A., McClelland, G. H., Parkinson, W. S., Shelton, C., Kelsall, D. C., and Smith, Z. M. (2014). "Examining the electro-neural interface of cochlear implant users using psychophysics, CT scans, and speech understanding," *J. Assoc. Res. Otolaryngol.* **15**, 293–304.
- McKay, C. M., Lim, H. H., and Lenarz, T. (2013). "Temporal processing in the auditory system: Insights from cochlear and auditory midbrain implantees," *J. Assoc. Res. Otolaryngol.* **14**, 103–124.
- McKay, C. M., and McDermott, H. J. (1998). "Loudness perception with pulsatile electrical stimulation: The effect of interpulse intervals," *J. Acoust. Soc. Am.* **104**, 1061–1074.
- Miller, G. A., and Nicely, P. E. (1955). "An analysis of perceptual confusions among some English consonants," *J. Acoust. Soc. Am.* **27**, 338–352.

- Nadol, J. B., Jr., Shiao, J. Y., Burgess, B. J., Ketten, D. R., Eddington, D. K., Gantz, B. J., Montandon, P., Coker, N. J., Roland, J. T., Jr., and Shallop, J. K. (2001). "Histopathology of cochlear implants in humans," *Ann. Otol. Rhinol. Laryngol.* **110**, 883–891.
- Nelson, D. A., Kreft, H. A., Anderson, E. S., and Donaldson, G. S. (2011). "Spatial tuning curves from apical, middle, and basal electrodes in cochlear implant users," *J. Acoust. Soc. Am.* **129**, 3916–3933.
- Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (2003). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **113**, 961–968.
- Pfingst, B. E., Bowling, S. A., Colesa, D. J., Garadat, S. N., Raphael, Y., Shibata, S. B., Strahl, S. B., Su, G. L., and Zhou, N. (2011a). "Cochlear infrastructure for electrical hearing," *Hear. Res.* **281**, 65–73.
- Pfingst, B. E., Burkholder-Juhász, R. A., Xu, L., and Thompson, C. S. (2008). "Across-site patterns of modulation detection in listeners with cochlear implants," *J. Acoust. Soc. Am.* **123**, 1054–1062.
- Pfingst, B. E., Colesa, D. J., Hembrador, S., Kang, S. Y., Middlebrooks, J. C., Raphael, Y., and Su, G. L. (2011b). "Detection of pulse trains in the electrically stimulated cochlea: Effects of cochlear health," *J. Acoust. Soc. Am.* **130**, 3954–3968.
- Pfingst, B. E., and Xu, L. (2005). "Psychophysical metrics and speech recognition in cochlear implant users," *Audiol. Neurotol.* **10**, 331–341.
- Pfingst, B. E., Xu, L., and Thompson, C. S. (2004). "Across-site threshold variation in cochlear implants: Relation to speech recognition," *Audiol. Neurotol.* **9**, 341–352.
- Pickett, J. M. (1999). *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology* (Allyn and Bacon, Boston), pp. 1–395.
- Prado-Guitierrez, P., Fewster, L. M., Heasman, J. M., McKay, C. M., and Shepherd, R. K. (2006). "Effect of interphase gap and pulse duration on electrically evoked potentials is correlated with auditory nerve survival," *Hear. Res.* **215**, 47–55.
- Ramekers, D., Versnel, H., Strahl, S. B., Smeets, E. M., Klis, S. F. L., and Grolman, W. (2014). "Auditory-nerve responses to varied inter-phase gap and phase duration of the electric pulse stimulus as predictors for neuronal degeneration," *J. Assoc. Res. Otolaryngol.* **15**, 187–202.
- Rubinstein, J. T., Parkinson, W. S., Tyler, R. S., and Gantz, B. J. (1999). "Residual speech recognition and cochlear implant performance: Effects of implantation criteria," *Am. J. Otol.* **20**, 445–452.
- Shannon, R. V., Jansvold, A., Padilla, M., Robert, M. E., and Wang, X. (1999). "Consonant recordings for speech testing," *J. Acoust. Soc. Am.* **106**, L71–74.
- Smith, L., and Simmons, F. B. (1983). "Estimating eighth nerve survival by electrical stimulation," *Ann. Otol. Rhinol. Laryngol.* **92**, 19–23.
- Stevens, K. N., and Blumstein, S. E. (1978). "Invariant cues for place of articulation in stop consonants," *J. Acoust. Soc. Am.* **64**, 1358–1368.
- Wilson, B. S., Finley, C. C., Lawson, D. T., and Zerbi, M. (1997). "Temporal representations with cochlear implants," *Am. J. Oto.* **18**, 30–34.
- Xu, L., and Zheng, Y. (2007). "Spectral and temporal cues for phoneme recognition in noise," *J. Acoust. Soc. Am.* **122**, 1758–1764.
- Zhou, N., and Pfingst, B. E. (2012). "Psychophysically based site selection coupled with dichotic stimulation improves speech recognition in noise with bilateral cochlear implants," *J. Acoust. Soc. Am.* **132**, 994–1008.
- Zhou, N., Xu, L., and Lee, C. Y. (2010). "The effects of frequency-place shift on consonant confusion in cochlear implant simulations," *J. Acoust. Soc. Am.* **128**, 401–409.
- Zhou, N., Xu, L., and Pfingst, B. E. (2012). "Characteristics of detection thresholds and maximum comfortable loudness levels as a function of pulse rate in human cochlear implant users," *Hear. Res.* **284**, 25–32.
- Zhou, R., Abbas, P. J., and Assoulin, J. G. (1995a). "Electrically evoked auditory brainstem response in peripherally myelin-deficient mice," *Hear. Res.* **88**, 98–106.
- Zhou, R., Assouline, J. G., Abbas, P. J., Messing, A., and Gantz, B. J. (1995b). "Anatomical and physiological measures of auditory system in mice with peripheral myelin deficiency," *Hear. Res.* **88**, 87–97.
- Zwolan, T. A., Collins, L. M., and Wakefield, G. H. (1997). "Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects," *J. Acoust. Soc. Am.* **102**, 3673–3685.