A mathematical model of vowel identification by users of cochlear implants

Elad Sagi^{a)}

Department of Otolaryngology, New York University School of Medicine, New York, New York 10016

Ted A. Meyer

Department of Otolaryngology–HNS, Medical University of South Carolina, Charleston, South Carolina 29425

Adam R. Kaiser and Su Wooi Teoh

Department of Otolaryngology, Head and Neck Surgery, DeVault Otologic Research Laboratory, Indiana University School of Medicine, Indianapolis, Indiana 46202

Mario A. Svirsky

Department of Otolaryngology, New York University School of Medicine, New York, New York 10016

(Received 1 January 2009; revised 25 November 2009; accepted 30 November 2009)

A simple mathematical model is presented that predicts vowel identification by cochlear implant users based on these listeners' resolving power for the mean locations of first, second, and/or third formant energies along the implanted electrode array. This psychophysically based model provides hypotheses about the mechanism cochlear implant users employ to encode and process the input auditory signal to extract information relevant for identifying steady-state vowels. Using one free parameter, the model predicts most of the patterns of vowel confusions made by users of different cochlear implant devices and stimulation strategies, and who show widely different levels of speech perception (from near chance to near perfect). Furthermore, the model can predict results from the literature, such as Skinner, *et al.* [(1995). *Ann. Otol. Rhinol. Laryngol.* **104**, 307–311] frequency mapping study, and the general trend in the vowel results of Zeng and Galvin's [(1999). *Ear Hear.* **20**, 60–74] studies of output electrical dynamic range reduction. The implementation of the model presented here is specific to vowel identification by cochlear implant users, but the framework of the model is more general. Computational models such as the one presented here can be useful for advancing knowledge about speech perception in hearing impaired populations, and for providing a guide for clinical research and clinical practice.

© 2010 Acoustical Society of America. [DOI: 10.1121/1.3277215]

PACS number(s): 43.71.An, 43.66.Ts, 43.71.Es, 43.71.Ky [MSS]

Pages: 1069–1083

I. INTRODUCTION

Cochlear implants (CIs) represent the most successful example of a neural prosthesis that restores a human sense. The last two decades have been witness to systematic improvements in technology and clinical outcomes, yet substantial individual differences remain. The reference to the individual CI user is important because typical fitting procedures for CIs are guided primarily by the listener's preference, by what "sounds better," independent of their speech perception (which does not always correlate perfectly with subjective preference; Skinner *et al.*, 2002). Several researchers have suggested that one of the factors limiting performance in many CI users is precisely this lack of performance-based fitting. If CI users were fit according to their specific perceptual and physiological strengths and weaknesses clinical outcomes might improve significantly (Shannon, 1993). Yet, assessing the effect of all possible fitting parameters on a given CI user's speech perception is not

feasible. In this regard, quantitative models may prove a useful aid to clinical practice. In the present study we propose a mathematical model that explains a CI user's vowel identification based on their ability to identify average formant center frequency values, and assess this model's ability to predict vowel identification performance under two CI device setting manipulations.

One example that demonstrates how such a model might guide clinical practice relates to the CI user's "frequency map," i.e., the frequency bands assigned to each stimulation channel. More than 20 years after the implantation of the first multichannel CIs the optimal frequency map remains unknown, either on average or for each specific CI user. The lack of evidence in this case is not total, however. Skinner *et al.* (1995) reported that a certain frequency map (frequency allocation table or FAT No. 7) used with the Nucleus-22 device resulted in better speech perception scores for a group of CI users than the frequency map that was the default for the clinical fitting software, and also the most widely used map at the time (FAT No. 9). Skinner *et al.*'s (1995) study resulted in a major shift and FAT No. 7 became much more commonly used by CI audiologists. Yet,

^{a)}Author to whom correspondence should be addressed. Electronic mail: elad.sagi@nyumc.org

with the large number of possible combinations, testing the whole parametric space of frequency map manipulations is both time and cost prohibitive. A possible alternative would be to use a model that provides reasonable predictions of speech perception under each FAT, and test a listener's performance using only the subset of FATs that the model deems most promising.

Several acoustic cues have been shown to influence vowel perception by listeners with normal hearing, including steady-state formant center frequencies (Peterson and Barney, 1952), formant frequency ratios (Chistovich and Lublinskaya, 1979), fundamental frequency, formant trajectories during the vowel, and vowel duration (Hillenbrand *et al.*, 1995; Syrdal and Gopal, 1986; Zahorian and Jagharghi, 1993), as well as formant transitions from and into adjacent phonemes (Jenkins *et al.*, 1983). That is, listeners with normal hearing can utilize the more subtle, dynamic changes in formant content available in the acoustic signal. Supporting this notion is the observation that listeners with normal hearing are highly capable of discriminating small changes in formant frequency. Kewley-Port and Watson (1994) found that listeners with normal hearing could detect differences in formant frequency of about 14 Hz in the range of F1 and about 1.5% in the range of F2. Hence, when two vowels consist of similar steady-state formant values, listeners with normal hearing have sufficient acuity to differentiate between these vowels based on small differences in formant trajectories.

In contrast, due to device and/or sensory limitations, listeners with CIs may only be able to utilize a subset of these acoustic cues (Chatterjee and Peng, 2008; Fitzgerald *et al.*, 2007; Hood *et al.*, 1987; Iverson *et al.*, 2006; Kirk *et al.*, 1992; Teoh *et al.*, 2003). For example, in terms of formant frequency discrimination, Fitzgerald *et al.* (2007) found that users of the Nucleus-24 device could discriminate about 50–100 Hz in the F1 frequency range and about 10% in the F2 frequency range, i.e., roughly five times worse than the normal hearing data reported by Kewley-Port and Watson (1994). Hence, some of the smaller formant changes that help listeners with normal hearing identify vowels may not be perceptible to CI users. Indeed, Kirk *et al.* (1992) demonstrated that when static formant cues were removed from vowels, normal hearing listeners were able to identify these vowels at levels significantly above chance whereas CI users could not. Furthermore, little or no improvement in vowel scores was found for the CI users when dynamic formant cues were added to static formant cues. In more recently implanted CI users, Iverson *et al.* (2006) found that CI users could utilize the larger dynamic formant changes that occur in diphthongs in order to differentiate these vowels from monophthongs, but it was also found that normal hearing listeners could utilize this cue to a far greater extent than CI users.

CI users' limited access to these acoustic cues gives us the opportunity to test a very simple model of vowel identification that relies only on steady-state formant center frequencies. Clearly, such a simple model would be insufficient to explain vowel identification in listeners with normal hearing, but it may be adequate to explain vowel identification in

current CI users. The model employed in the present study is an application of the multidimensional phoneme identification or MPI model (Svirsky, 2000, 2002), which was developed as a general framework to predict phoneme identification based on measures of a listener's resolving power for a given set of speech cues. In the present study, the model is tested on four experiments related to vowel identification by CI users. The first two were conducted by us and consist of vowel and first-formant identification data from CI listeners. The purpose of these two data sets was to test the model's ability to account for vowel identification by CI users, and to assess the model's account of relating vowel identification to listeners' ability to resolve steady-state formant center frequencies. The third and fourth data sets were extracted from Skinner *et al.*, 1995 and Zeng and Galvin, 1999, respectively. These two data sets were used to test the MPI model's ability to make predictions about how changes in two CI device fitting parameters (FAT and electrical dynamic range, respectively) affect vowel identification in these listeners.

II. GENERAL METHODS

A. MPI model

The mathematical framework of the MPI model is a multidimensional extension of Durlach and Braida's single-dimensional model of loudness perception (Durlach and Braida, 1969; Braida and Durlach, 1972), which is in turn based on earlier work by Thurstone (1927a, 1927b) among others. The MPI model is more general than the Durlach–Braida model not only due to the fact that it is multidimensional, but also because loudness need not be one of the model's dimensions. Let us first define some terms and assumptions that underlie the MPI model. We assume that a *phoneme* (vowel or consonant) is identified based on several *acoustic cues*. A given acoustic cue assumes characteristic values for each phoneme along the respective *perceptual dimension*. A subject's resolving power, or just-noticeable-difference (JND), along this perceptual dimension can be measured with appropriate psychophysical tests. The JNDs for all dimensions are subject-specific inputs to the MPI model. Because listeners have different JND values along any given dimension, the model's predictions can be different for each subject.

1. General implementation: Three steps

The implementation of the MPI model in the present study can be summarized in three steps. *First*, we must hypothesize what the relevant perceptual dimensions are. These hypotheses are informed by knowledge about acoustic-phonetic properties of speech, and about the auditory psychophysical capabilities of CI users (Teoh *et al.*, 2003). *Second*, we have to measure the mean location of each phoneme along each postulated perceptual dimension. These locations are uniquely determined by the physical characteristics of the stimuli and the selected perceptual dimensions. *Third*, we must measure the subjects' JNDs along each perceptual dimension using appropriate psychophysical tests, or leave the JNDs as free parameters to determine how well the model could fit the experimental data. Because there are several

FIG. 1. Electrodegram of the vowel in “had” obtained with the Nucleus device. Higher electrode numbers refer to more apical or low-frequency encoding electrodes. Charge magnitude is depicted as a gray-scale from 0% (light) to 100% (dark) of dynamic range. Rectangle centered at 200 ms represents the time window used to compile histogram on the right, which represents a weighted count of the number of times each electrode was stimulated. Locations of mean formant energies (F1, F2, and F3 in millimeters from most basal electrode) extracted from histogram.

ways to measure JNDs, these two approaches could yield JND values that are related, but not necessarily the same.

Step 1. The proposed set of relevant perceptual dimensions for the present study of vowel identification by CI users is the mean locations along the implanted electrode array of stimulation pulses corresponding to the first three formant frequencies, i.e., F1, F2, and F3. These dimensions are measured in units of distance along the electrode array (e.g., mm from most basal electrode) rather than frequency (Hz). In experiment 1, different combinations of these dimensions are explored to determine a set of dimensions that best describe each CI subject’s vowel confusion matrix. In experiments 3 and 4, the F1F2F3 combination is used exclusively.

Step 2. Locations of mean formant energy along the electrode array were obtained from “electrodegrams” of vowel tokens. The details of how electrodegrams were obtained are in Sec. II B. An electrodegram is a graph that includes information about which electrode is stimulated at a given time, and at what current amplitude and pulse duration. Depending on the allocation of frequency bands to electrodes, an electrodegram depicts how formant energy becomes distributed over a subset of electrodes. The left panel of Fig. 1 is an example of an electrodegram of the vowel “had” obtained with the Nucleus device where higher electrode numbers refer to more apical or low-frequency encoding electrodes. For each pulse, the amount of electrical charge (i.e., current times pulse duration) is depicted as a gray-scale from 0% (light) to 100% (dark) of the dynamic range, where 0% represents threshold stimulation level and 100% represents the maximum comfortable level. We are particularly concerned with how formant energies F1, F2, and F3 are distributed along the array over a time window centered at the middle portion of the vowel stimulus (rectangle in Fig. 1). The right panel of Fig. 1 is a histogram of the number of times each electrode was stimulated over this time window, weighted by the amount of electrical charge above threshold for each current pulse (measured with the percentage of the dynamic range described above). The histogram’s vertical axis is in units of millimeters from the most basal electrode as measured along the length of the electrode

array. These units are inferred from the inter-electrode distance of a given CI device (e.g., 0.75 mm for the Nucleus-22 and Nucleus-24 CIs and 2 mm for the Advanced Bionics Clarion 1.2 CI). To obtain the location of mean formant energy along the array for each formant, the histogram was first partitioned into regions of formant energies (one for each formant) and then the mean location for each formant was calculated from the portion of the histogram within each region. The frequency ranges selected to partition histograms into formant regions, based on the average formant measurements of Peterson and Barney (1952) for male speakers, were $F1 \leq 800 \text{ Hz} < F2 \leq 2250 \text{ Hz} < F3 \leq 3000 \text{ Hz}$ for all vowels except for “heard,” for which $F1 \leq 800 \text{ Hz} < F2 \leq 1700 \text{ Hz} < F3 \leq 3000 \text{ Hz}$. In Fig. 1, the locations of mean formant energies are indicated to the right of the histogram. Whereas each electrode is located at discrete points along the array, the mean location of formant energy varies continuously along the array.

Step 3. JND was varied as a free parameter with one degree of freedom until a predicted matrix was obtained that “best-fit” the observed experimental matrix. That is, in a given best-fit model matrix, JND was assumed to be equal for each perceptual dimension.

2. MPI model framework

Qualitative description. The MPI model is comprised of two sub-components, an *internal noise model* and a *decision model*. The internal noise model postulates that a phoneme produces percepts that are represented by a *Gaussian* probability distribution in a *multidimensional* perceptual space. For the sake of simplicity it is assumed that perceptual dimensions are *independent* (orthogonal) and distances are *Euclidean*. These distributions represent the assumption that successive presentations of the same stimulus result in somewhat different percepts, due to imperfections in the listener’s internal representation of the stimulus (i.e., *sensory noise* and *memory noise*). The center of the Gaussian distribution corresponding to a given phoneme is determined by the physical characteristics of the stimulus along each dimension. The standard deviation along each dimension is equal to the listener’s JND for the stimulus’ physical characteristic along that dimension. Smaller JNDs produce narrower Gaussian distributions and can result in fewer confusions among different sounds.

The decision model employed in the present study is similar to the approach employed by Braida (1991) and Ronan *et al.* (2004), and describes how subjects categorize speech sounds based on the perceptual input. According to the decision model, the multidimensional perceptual space is subdivided into non-overlapping *response regions*, one for each phoneme. Within each response region there is a *response center*, which represents the listener’s expectation about how a given phoneme should sound. One interpretation of the response center concept is that it reflects a subject’s expected sensation in response to a stimulus (e.g., a prototype or “best exemplar” of the subject’s phoneme category). When a percept (generated by the internal noise model) falls in the response region corresponding to a given phoneme (or, in other words, when the percept is closer to

FIG. 2. Summary of the two-step iteration involved in a three-dimensional F1F2F3 MPI model for vowel identification. Internal noise model generates a percept by adding noise (proportional to input JNDs) to the formant locations of a given vowel. Decision model selects response center (i.e., best exemplar of a given vowel) with formant locations closest to those of percept.

the response center of that phoneme than to any other response center), then the decision model predicts that the subject will select that phoneme as the one that she/he heard. The ideal experienced listener would have response centers that are equal to the *stimulus centers*, which we define as the average location of tokens for a particular phoneme in the perceptual space. In other words, this listener's expectations match the actual physical stimuli. When this is not the case, one can implement a bias parameter to accommodate for differences between stimulus and response centers. In the present study, all listeners are treated as ideal experienced listeners so that stimulus and response centers are equal.

Using a Monte Carlo algorithm that implements each component of the MPI model, one can simulate vowel identifications to any desired number of iterations, and compile the results into a confusion matrix. Each iteration can be summarized as a two-step process. First, one uses the internal noise model to generate a sample percept for a given phoneme. Second, one uses the decision model to select the phoneme that has the response center closest to the percept. Figure 2 illustrates a block diagram of the two-step iteration involved in a three-dimensional MPI model for vowel identification, where the three dimensions are the average locations along the electrode array stimulated in response to the first three formants: F1, F2, and F3.

Mathematical formulation. The Gaussian distribution that underlies the internal noise model for the F1F2F3 perceptual dimension combination can be described as follows. Let E_i represent the i th vowel out of the nine possible vowels used in the present study. Let E_{ij} represent the j th token of E_i , out of the five possible tokens used for this vowel in the present study. Each token is described as a point in the three dimensional F1F2F3 perceptual space. Let this point T be described by the set $T = \{T_{F1}, T_{F2}, T_{F3}\}$, so that $T_{F2}(E_{ij})$ represents the F2 value of the vowel token E_{ij} . Let $J = \{J_{F1}, J_{F2}, J_{F3}\}$ represent the subject's set of JNDs across perceptual dimensions so that J_{F2} represents the JND along the F2 dimension. Now let $X = \{x_{F1}, x_{F2}, x_{F3}\}$ be a set of random variables across perceptual dimensions, so that x_{F2} is a random

variable describing any possible location along the F2 dimension. Since perceptual dimensions are assumed to be independent, the normal probability density describing the likelihood of the location of a percept that arises from vowel token E_{ij} can be defined as $P(X|E_{ij})$ where

$$P(X|E_{ij}) = \frac{1}{J_{F1}J_{F2}J_{F3}(\sqrt{2\pi})^3} e^{-(x_{F1} - T_{F1}(E_{ij}))^2/2J_{F1}^2} \times e^{-(x_{F2} - T_{F2}(E_{ij}))^2/2J_{F2}^2} e^{-(x_{F3} - T_{F3}(E_{ij}))^2/2J_{F3}^2}. \quad (1)$$

Each presentation of E_{ij} results in a sensation that is modeled as a point that varies stochastically in the three dimensional F1F2F3 space following the Gaussian distribution $P(X|E_{ij})$. This point, or "percept," can be defined as $X' = \{x'_{F1}, x'_{F2}, x'_{F3}\}$, where x'_{F2} is the coordinate of X' along the F2 dimension. The prime script is used here to distinguish X' as a point in X . The stochastic variation of X' arises from a combination of "sensation noise," which is a measure of the observer's sensitivity to stimulus differences along the relevant dimension, and "memory noise," which is related to uncertainty in the observer's internal representation of the phonemes within the experimental context.

In the decision model, the percept X' is categorized by finding the closest response center. Let $R(E_k) = \{R_{F1}(E_k), R_{F2}(E_k), R_{F3}(E_k)\}$ be the location of the *response center* for the k th vowel so that $R_{F2}(E_k)$ represents the location of the response center for this vowel along the F2 perceptual dimension. For vowel E_k , the stimulus center can be represented as $S(E_k) = \{S_{F1}(E_k), S_{F2}(E_k), S_{F3}(E_k)\}$, where $S_{F2}(E_k)$ is the location of the stimulus center for vowel E_k along the F2 perceptual dimension. $S_{F2}(E_k)$ is equal to the average F2 value across the five tokens of E_k [i.e., the average of $T_{F2}(E_{kj})$ for $j=1, \dots, 5$]. When a listener's expected sensation in response to a given phoneme is unbiased, then we say that the response center is equal to the stimulus center; i.e., $R(E_k) = S(E_k)$. Conversely, if the listener's expectations (represented by the response centers) are not in line with the physical characteristics of the stimulus (represented by the stimulus centers), then we say that the listener is a biased observer. In the present study, all listeners are treated as unbiased observers so that response centers are equal to stimulus centers.

The closest response center to the percept X' can be determined by comparing X' with all response centers $R(E_z)$ for $z=1, \dots, n$ using the Euclidean measure

$$D_z = \sqrt{\left(\frac{x'_{F1} - R_{F1}(E_z)}{J_{F1}}\right)^2 + \left(\frac{x'_{F2} - R_{F2}(E_z)}{J_{F2}}\right)^2 + \left(\frac{x'_{F3} - R_{F3}(E_z)}{J_{F3}}\right)^2}. \quad (2)$$

If $R(E_k)$ is the closest response center to the percept X' (in other words, if D_z is minimized when $z=k$), then the phoneme that gave rise to the percept (i.e., E_i) was identified as phoneme E_k and one can update Cell_{ik} in the confusion matrix accordingly. Using a Monte Carlo algorithm, the process of generating a percept with Eq. (1) and categorizing this percept using Eq. (2) can be continued for all vowel tokens to any desired number of iterations. It is important to note that the JNDs that appear in the denominator of Eq. (2) are

used to ensure that all distances are measured as multiples of the relevant just-noticeable-difference along each perceptual dimension.

B. Stimulus measurements

Electrograms of the vowel tokens used in the present study were obtained for two types of Nucleus device and one type of Advanced Bionics device using specialized hardware and software. In both cases, vowel tokens were presented over loudspeaker to the device's external microphone in a sound attenuated room. The microphone was placed approximately 1 m from the loudspeaker and stimuli were presented at 70 dB C-weighted sound pressure level (SPL) as measured next to the speech processor's microphone.

Depending on the experiment conducted in the present study, measurements were obtained from either a standard Nucleus-22 device with a Spectra body-worn processor or a standard Nucleus-24 device with a Sprint body-worn processor. In either case, the radio frequency (RF) information transmitted by the processor (through its transmitter coil) was sent to a Nucleus dual-processor interface (DPI). The DPI, which was connected to a PC, captured and decoded the RF signal, which was then read by a software package called *sCILab* (Bögli *et al.*, 1995; Wai *et al.*, 2003). The speech processor was programmed with the spectral peak (SPEAK) stimulation strategy where the thresholds and maximum stimulation levels were fixed to 100 and 200 clinical units, respectively. Depending on the experiment, the frequency allocation table was set to FAT No. 7 and/or FAT No. 9.

For the Advanced Bionics device, electrograms were obtained by measuring current amplitude and pulse duration directly from the electrode array of an eight-channel Clarion 1.2 "implant-in-a-box" connected to an external speech processor (provided by Advanced Bionics Corporation, Valencia, CA, USA). The processor was programmed with the continuous interleaved sampling (CIS) stimulation strategy and with the standard frequency-to-electrode assigned by the processor's programming software. For each electrode, the signal was passed through a resistor and recorded to PC by one channel of an eight-channel IOtech WaveBook/512H Data Acquisition System [12-bit analogue to digital (A/D) conversion sampled at 1 MHz].

C. Comparing predicted and observed confusion matrices

Two measures were used to assess the ability of the MPI model to generate a matrix that best predicted a listener's observed vowel confusion matrix. The first method provides a global measure of how a model matrix generated with the MPI model differs from an experimental matrix. The second method examines how the MPI model accounts for the specific error patterns observed in the experimental matrix. For both measures, matrix elements are expressed in units of percentage so that each row sums to 100%.

1. Root-mean-square difference

The first measure is the root-mean-square (rms) difference between the predicted and observed matrices. With this

TABLE I. Example of a 2×2 comparison table comparing the vowel pairs confused more than a certain percentage of the time (5% in this case) by the subjects, to the vowel pairs that the model predicted would be confused.

Threshold=5%	Predicted $\geq 5\%$	Predicted $< 5\%$
Observed $\geq 5\%$	A=5	B=1
Observed $< 5\%$	C=1	D=29

measure, the differences between each element of the observed matrix and each element of the predicted matrix are squared and summed. The sum is divided by the total number of elements in the matrix (e.g., $9 \times 9 = 81$) to give the mean-square, and its square-root the rms difference in units of percent. With this measure, the predicted matrix that minimized rms was defined as the best-fit to the observed matrix.

2. Error patterns

The second measure examines the extent to which the MPI model predicts the pattern of vowel pairs that were confused (or not confused) more frequently than a predefined percentage of the time. Vowel pairs were analyzed without making a distinction as to the direction of the confusion within a pair, e.g., "had" confused with "head" vs "head" confused with "had." That is, in a given confusion matrix, the percentage of time the i th and j th vowel pair was confused is equal to $(\text{Cell}_{ij} + \text{Cell}_{ji})/2$. This approach was adopted to simplify the fitting criteria between observed and predicted matrices and should not be taken to mean that confusions within a vowel pair are assumed to be symmetric. In fact, there is considerable evidence that vowel confusion matrices are not symmetric either for normal hearing listeners (Phatak and Allen, 2007), or for the CI users in the present study.

After calculating the percentage of vowel pair confusions in both the observed and predicted matrices, a 2×2 contingency table can be constructed based on a threshold percentage. Table I shows an example of such a contingency table using a threshold of 5%. Out of 36 possible vowel pair confusions, cell A (upper left) is the number of true positives, i.e., confusions ($\geq 5\%$) made by the subject and predicted by the model. Cell B (upper right) is the number of false negatives, i.e., confusions ($\geq 5\%$) made by the subject but not predicted by the model. Cell C (lower left) is the number of false positives, i.e., confusions ($\geq 5\%$) predicted to occur by the model but not made by the subject. Lastly, cell D (lower right) is the number of true negatives, i.e., confusions not made by the subject ($< 5\%$) and also predicted not to occur by the model ($< 5\%$). With this method of matching error patterns, a best-fit predicted matrix was defined as one that predicted as many of the vowel pairs that were either confused or not confused by a given listener as possible while minimizing false positives and false negatives. That is, best-fit 2×2 comparison matrices were selected so that the maximum value of B and C was minimized. Of these, the comparison matrix for which the value $2A - B - C$ was maximized was then selected. When more than one value for JND produced the same maximum, the JND that also yielded the lowest rms out of the group was selected. Best-fit 2×2 com-

TABLE II. Demographics of CI users tested for this study: 7 users of the Advanced Bionics device (C) and 18 users of the Nucleus device (N). Age at implantation and experience with implant are stated in years. Speech processing strategies are CIS, ACE (Advanced Combination Encoder), and SPEAK.

Subject	Implanted age	Implant experience	Implanted device	Strategy	No. of channels
C1	66	3.4	Clarion 1.2	CIS	8
C2	32	3.4	Clarion 1.2	CIS	8
C3	61	5.9	Clarion 1.2	CIS	8
C4	23	5.5	Clarion 1.2	CIS	8
C5	53	6.1	Clarion 1.2	CIS	5
C6	39	2.7	Clarion 1.2	CIS	6
C7	43	2.2	Clarion 1.2	CIS	8
N1	31	5.2	Nucleus CI22M	SPEAK	18
N2	59	11.2	Nucleus CI22M	SPEAK	13
N3	71	3	Nucleus CI22M	SPEAK	14
N4	67	2.9	Nucleus CI22M	SPEAK	19
N5	45	3.9	Nucleus CI22M	SPEAK	20
N6	48	9.1	Nucleus CI22M	SPEAK	16
N7	16	4.6	Nucleus CI22M	SPEAK	18
N8	66	2.3	Nucleus CI22M	SPEAK	18
N9	48	1.7	Nucleus CI24M	ACE	20
N10	42	2.3	Nucleus CI24M	SPEAK	16
N11	44	3.1	Nucleus CI24M	SPEAK	20
N12	75	1.7	Nucleus CI24M	SPEAK	19
N13	65	2.2	Nucleus CI24M	SPEAK	20
N14	53	1.9	Nucleus CI24M	SPEAK	20
N15	45	4.2	Nucleus CI24M	SPEAK	20
N16	45	3.2	Nucleus CI24M	SPEAK	20
N17	37	0.9	Nucleus CI24M	SPEAK	20
N18	68	1.2	Nucleus CI24M	SPEAK	20

parison matrices were obtained at three values for threshold: 3%, 5%, and 10%. Different thresholds were necessary to assess errors made by subjects with very different performance levels. A best-fit 2×2 comparison matrix was labeled “satisfactory” if both A and D were greater than (or at least equal to) B and C. According to this definition a satisfactory comparison matrix is one where the model was able to predict at least one-half of the vowel pairs confused by an individual listener, and do so with a number of false positives no greater than the number of true positives (vowel pairs accurately predicted to be confused by the individual).

III. EXPERIMENT 1: VOWEL IDENTIFICATION

A. Methods

1. CI listeners

Twenty-five postlingually deafened adult users of CIs were recruited for this study. Participants were compensated for their time and provided informed consent. All participants were over 18 years of age at the time of testing, and the mean age at implantation was 50 years ranging from 16 to 75 years. Participants were profoundly deaf ($PTA > 90$ dB) and had at least 1 year of experience with their implant before testing, with the exception of N17 who had 11 months of post-implant experience when tested. The demographics for this group at time of testing are presented in Table II, includ-

ing age at implantation, duration of post-implant experience, type of CI device and speech processing strategy, as well as number of active channels.

2. Stimuli and general procedures

Vowel stimuli consisted of nine vowels in /hVd/context, i.e., heed, hawed, heard, hood, who’d, hid, hud, had, and head. Stimuli included three tokens of each vowel recorded from the same male speaker. Vowel tokens would be presented over loudspeaker to CI subjects seated 1 m away in a sound attenuated room. The speaker was calibrated before each experimental session so that stimuli would register a value of 70 dB C-weighted SPL on a sound level meter placed at the approximate location of a user’s ear-level microphone. In a given session listeners would be presented with one to three lists of the same 45 stimuli (i.e., up to 135 presentations) where each list comprised a different randomization of presentation order. In each list, two tokens of each vowel were presented twice and one token was presented once. Before the testing session, listeners were presented with each vowel token at least once knowing in advance the vowel to be presented for practice. During the testing session, no feedback was provided. All three lists were presented on the same day, and a listener was allowed a break between lists if required.

3. Application of the MPI model

Step 1. All seven possible combinations of one, two, or three dimensions consisting of mean locations of formant energies F1, F2, and F3 along the electrode array were tested.

Step 2. Mean locations of formant energies along the electrode array were obtained from electrodiagrams of each vowel token that was presented to CI subjects. A set of formant location measurements was obtained for each CI listener. Obtaining these measurements directly from each subject’s external device would have been optimal, but time consuming. Instead, four generic sets of formant location measurements were obtained. One set was obtained for the Nucleus-24 spectra body-worn processor with the SPEAK stimulation strategy using FAT No. 9, and three sets were obtained for the Clarion 1.2 processor with the CIS stimulation strategy using the standard FAT imposed by the device’s fitting software. The three sets of formant locations for Clarion users were obtained with the speech processor programmed using eight, six, and five channels. One Clarion subject had five active channels in his FAT, another one had six channels, and the remaining five had all eight channels activated. Two out of 18 of the Nucleus subjects and 4 out of 7 of the Clarion subjects used these standard FATs, whereas the other subjects used other FATs with slight modifications. For example, a Nucleus subject may have used FAT No. 7 instead of FAT No. 9, or one or more electrodes may have been turned off, or a Clarion subject may have used extended frequency boundaries for the lowest or the highest frequency channels. For these other subjects, each generic set of formant location measurements that we obtained was then modified to generate a unique set of measurements. Using

linear interpolation, the generic data set was first transformed into hertz using the generic set's frequency allocation table and then transformed back into millimeters from the most basal electrode using the frequency allocation table that was programmed into a given subject's speech processor at the time of testing. This method provided a unique set of formant location measurements even for those subjects with one or more electrodes shut off, typically to avoid facial twitch and/or dizziness.

Step 3. Using a CI listener's set of formant location measurements for a given perceptual dimension combination, MPI model-predicted matrices were generated while JND was varied using one degree of freedom from 0.03 to 6 mm in steps of 0.005 mm (i.e., a total of 1195 predicted matrices). The lower bound of 0.03 mm was selected as it represents a reasonable estimate of the lowest JND for place of stimulation in the cochlea achievable with present day CI devices (Firszt *et al.*, 2007; Kwon and van den Honert, 2006). Each predicted matrix (one for each value of JND) consisted of 5 000 iterations per vowel token, i.e., 225 000 entries in total. Predicted matrices were compared with the listener's observed vowel confusion matrix to obtain the JND that provided the best-fit between predicted matrices and the CI listener's observed vowel matrix. A best-fit JND value and predicted matrix was obtained for each CI listener, for each of the seven perceptual dimension combinations, both in terms of the lowest rms difference and in terms of the best 2 × 2 comparison matrix using thresholds of 3%, 5%, and 10%. The combination of perceptual dimensions that provided the best-fit to the data was then examined, both from the point of view of rms difference and of error patterns.

B. Results

Vowel identification percent correct scores for the CI listeners tested in the present study are listed in the second column of Table III. The scores ranged from near chance to near perfect.

1. rms differences between observed and predicted matrices

Also listed in Table III are the minimum rms differences between predicted and observed matrices as a function of seven possible perceptual dimension combinations. The perceptual dimension combination that produced the lowest minimum rms is highlighted in bold, and rms values greater than 1% above the lowest minimum rms have been omitted. As one can observe, the perceptual dimension combination that produced the lowest minimum rms was F1F2F3 for 15 out of 25 listeners. For eight of the remaining ten listeners, the F1F2F3 perceptual dimension combination provided a fit that was not the best, but was within 1% of the best-fit. Of these remaining ten listeners, six were best fitted by the F1F2 combination, three by the F2 combination, and one by the F1F3 combination.

The third column of Table III contains the rms difference between each listener's observed vowel confusion matrix and a purely random matrix, i.e., one where all matrix elements are equal. Any good model should yield a rms dif-

TABLE III. Minimum rms difference between CI users' observed and predicted vowel confusion matrices for seven perceptual dimension combinations comprising F1, F2, and/or F3. The lowest rms values across perceptual dimensions are highlighted in bold and only values within 1% of this minimum were reported. The second and third columns list observed vowel percent correct and the rms difference between observed matrices and a purely random matrix.

CI User	Vowel (%)	rms							
		Random	F1F2F3	F1F2	F1F3	F2F3	F1	F2	F3
C1	72.6	25.2	9.9	10.0	...	10.1	...	...	...
C2	98.5	31.0	5.2	5.4	...	16.0	...	...	...
C3	94.1	29.7	6.3	6.7	...	...	...	...	...
C4	80.0	26.3	9.1	9.5	...	...	...	...	...
C5	21.5	11.0	14.9	15.0	14.5	...	...	...	15.5
C6	43.7	16.5	10.8	11.1	11.4	...	...	...	...
C7	83.7	27.0	6.0	6.1	...	...	...	...	...
N1	80.0	28.2	14.9	15.3	...	15.7	...	...	...
N2	22.2	11.5	...	13.8	...	...	...	14.1	14.7
N3	73.3	24.6	8.0	...	...	8.1	...	...	...
N4	70.4	26.7	13.3	...	...	13.3	...	12.7	...
N5	95.6	30.0	5.4	4.4	...	...	...	...	...
N6	81.7	27.2	11.4	12.0	...	12.4	...	...	...
N7	72.6	23.5	...	10.4	...	...	...	...	...
N8	26.1	11.6	11.9	11.6	...	12.2	...	12.4	...
N9	80.0	26.7	9.0	...	...	...	...	...	...
N10	81.5	26.3	10.7	10.1	...	...	...	...	...
N11	85.0	27.9	10.2	...	...	...	...	...	...
N12	42.2	16.4	11.9	12.7	...	12.1	...	12.5	...
N13	79.3	25.4	8.4	9.2	...	...	...	...	...
N14	81.5	26.9	10.0	...	...	...	...	...	...
N15	91.1	29.5	9.7	9.2	...	...	...	...	...
N16	59.3	24.7	15.3	...	...	15.8	...	14.8	...
N17	71.1	24.3	10.2	...	...	...	...	9.8	...
N18	66.7	24.2	12.1	...	...	13.0	...	...	...
Mean	70.1	24.1	10.5	11.1	14.9	12.7	17.7	13.7	19.7
		No. of best rms	15	6	1	0	0	3	0

ference that is much smaller than the values that appear in this column. Indeed, this is true for 20 out of 25 CI users for which the lowest minimum rms values achieved with the MPI model (highlighted in bold) are at least 10% lower than those for a purely random matrix (i.e., third column of Table III). The remaining five CI users (C5, C6, N2, N8, and N12) had the lowest vowel identification scores in the group (between 21% and 44% correct). For these subjects, the MPI model does not do much better than a purely random matrix, especially for the three subjects whose scores were only about twice chance levels.

A repeated measures analysis of variance (ANOVA) on ranks was conducted on the rms values we obtained for all subjects. Perceptual dimension combinations, as well as the random matrix comparison, were considered as different treatment groups applied to the same CI subjects. A significant difference was found across treatment groups ($p < 0.001$). Using the Student-Newman-Keuls method for multiple post-hoc comparisons, the following significant group differences were found at $p < 0.01$: F1F2F3 rms

TABLE IV. Best 2×2 comparison matrices between observed vowel confusion matrices from CI users and those predicted from MPI model. Key for best comparison matrices is on bottom: dim=perceptual dimension combination, thr=threshold at which best comparison matrix was obtained, and p -value=result of Fisher exact test; A, B, C, and D, as in Table I. Bottom right, average best 2×2 comparison matrix.

C1	F1F2F3	C2	F1F2F3	C3	F1F2F3	C4	F1F2F3	C5	F1F2F3
5%	<0.001	5%	1.00	10%	0.024	5%	0.003	5%	0.026
7	0	0	0	3	2	4	2	23	4
1	28	2	34	3	28	2	28	4	5
C6	F1F2F3	C7	F1F2F3	N1	F2	N2	F1F2F3	N3	F1F2F3
5%	0.002	5%	0.013	10%	0.027	10%	0.015	5%	0.003
12	3	3	2	2	1	11	7	4	2
5	16	2	29	2	31	3	15	2	28
N4	F1F2F3	N5	F1F2	N6	F2F3	N7	F1F2F3	N8	F1F2F3
5%	<0.001	3%	0.005	10%	0.027	10%	0.013	5%	0.041
4	2	4	1	2	1	3	2	16	5
0	30	4	27	2	31	2	29	6	9
N9	F1F2F3	N10	F1F2F3	N11	F1F2F3	N12	F1F2F3	N13	F1F2F3
5%	0.010	3%	0.024	10%	0.010	5%	<0.001	5%	0.030
3	1	5	5	2	0	14	4	4	4
3	29	3	23	2	32	2	16	3	25
N14	F1F2F3	N15	F1F2F3	N16	F1F2F3	N17	F1F2F3	N18	F1F2F3
10%	0.027	10%	0.010	5%	0.026	3%	0.002	5%	0.003
2	1	2	0	5	4	11	4	9	4
2	31	2	32	4	23	4	17	4	19

Key			
Subject	Dim		
thr	p -value		
A	B		
C	D		

Average	
6.20	2.44
2.76	24.60

<F1F2 rms<F2F3 rms<F2 rms<F1F3 rms<F1, F3 and random rms. No significant differences were found between F1, F3, and the random case.

2. Prediction of error patterns

Table IV shows the extent to which the MPI model can fit the patterns of vowel confusions made by individual CI users. The table lists one example of a best 2×2 comparison matrix for each subject. At the bottom of Table IV is a key that identifies where to find for each comparison matrix the subject identifier, the perceptual dimension from which the best comparison matrix was selected, the threshold (3%, 5%, or 10%), the p -value obtained from a Fisher exact test, and elements A–D of the comparison matrix as outlined in Table I of Sec. II. The following criteria were used for selecting the matrices listed in Table IV: (1) a satisfactory 2×2 comparison matrix with F1F2F3 at the 5% threshold, (2) a satisfactory matrix with F1F2F3 at any threshold, and (3) a satisfactory matrix at any perceptual dimension. Under these criteria, satisfactory matrices were obtained for 24 out of 25 subjects. The only exception was subject C2 who confused very few vowel pairs and for whom a satisfactory comparison matrix could not be obtained. On the lower right of Table IV is an average of elements A–D for all 25 exemplars listed in Table IV. On average, the MPI model predicted the pattern of vowel confusions in 31 out of 36 possible vowel pair

confusions. As for the Fisher exact tests, the comparison matrices in Table IV were significant at $p < 0.05$ for 24 out of 25 subjects (again subject C2 was the exception), half of which were significant at $p \leq 0.01$.

Table V shows the number of satisfactory best-fit 2×2 comparison matrices obtained for each listener at each perceptual dimension combination. As comparison matrices were obtained at thresholds of 3%, 5%, and 10%, the maximum number of satisfactory comparison matrices at each perceptual dimension combination is 3. The bottom row of Table V lists the total number of satisfactory comparison matrices at each perceptual dimension combination. As one can observe, the F1F2F3 combination produced the largest number of satisfactory best-fit 2×2 comparison matrices, corroborating the result obtained with the best-fit rms criteria.

C. Discussion

It is not surprising that a model based on the ability to discriminate formant center frequencies can explain at least some aspects of vowel *identification*. Rather, what is novel about the results of the present study is that the MPI model produced *confusion matrices* that closely matched CI users' vowel confusion matrices, including the general pattern of errors between vowels, despite differences in age at implantation, implant experience, device and simulation strategy

TABLE V. Number of “satisfactory” 2×2 comparison matrices at thresholds of 3%, 5%, and 10% for each perceptual dimension.

Subject	F1F2F3	F1F2	F1F3	F2F3	F1	F2	F3
C1	3	3	0	3	0	3	0
C2	0	0	0	0	0	0	0
C3	1	0	0	0	0	0	0
C4	2	2	0	2	0	2	0
C5	1	1	1	0	1	0	0
C6	3	3	3	3	3	3	0
C7	3	2	0	1	0	1	0
N1	0	0	0	0	0	1	0
N2	2	3	1	3	1	3	2
N3	2	2	0	2	0	2	0
N4	3	3	0	3	0	3	0
N5	0	1	0	0	0	0	0
N6	0	0	0	1	0	0	0
N7	2	3	0	2	0	3	0
N8	2	1	1	2	1	2	0
N9	3	0	0	2	0	1	0
N10	1	2	0	1	0	2	0
N11	1	0	0	1	0	0	0
N12	3	3	3	3	3	3	3
N13	2	2	0	2	0	2	0
N14	2	0	0	1	0	0	0
N15	1	0	0	1	0	0	0
N16	3	3	0	2	0	3	0
N17	1	3	1	2	0	3	1
N18	3	2	2	3	0	3	0
Total	44	39	12	40	9	40	6

used (Table II), as well as overall vowel identification level (Table III). It is important to stress that these results were achieved with only one degree of freedom. The ability to demonstrate how a model accounts for experimental data is strengthened when the model can capture the general trend of the data while using fewer instead of more degrees of freedom (Pitt and Navarro, 2005). With one degree of freedom, when a model with F1F2F3 does better than a model with F1F2, or when a model with F1F2 does better than a model with F2 alone, one can interpret the value of an added perceptual dimension without having to account for the possibility that the improvement was due to an added fitting parameter.

Whether in terms of rms differences (Table III) or prediction of error patterns (Table V) it is clear that F1F2F3 was the most successful formant combination in accounting for CI users’ vowel identification. Upon inspection of the other formant dimension combinations, both Tables III and V suggest that models that included the F2 dimension tended to do better than models without F2, and Table III suggests that the F1F2 combination was a close second to the F1F2F3 combination. The implication may be that F2, and perhaps F1, are important for identifying vowels in most listeners, whereas F3 may be an important cue for some implanted listeners, particularly for r-colored vowels such as heard, but perhaps not for others (Skinner *et al.*, 1996).

The model was able to explain most of the confusions made by most of the individual listeners, while making few

false positive predictions. This is an important result because one degree of freedom is always sufficient to fit one independent variable, such as percent correct, but it is not sufficient to predict a data set that includes 36 pairs of vowels. It should come as no surprise that percent correct scores in a predicted vowel matrix drop as the JND parameter is increased. Any model that employs a parameter to move data away from the main diagonal would accomplish the same result. However, the MPI model succeeds in the sense that increasing the JND moves data away from the main diagonal toward a specific vowel confusion pattern determined by the set of perceptual dimensions proposed. Although the fit between predicted and observed data was not perfect, it was strong enough to suggest that the proposed model captures some of the mechanisms CI users employ to identify vowels.

IV. EXPERIMENT 2: F1 IDENTIFICATION

A. Methods

One of the premises underlying the MPI model of vowel identification by CI users in the present study is that a relationship exists between these listeners’ ability to identify vowels and their ability to identify steady-state formant frequencies. To test this premise, 18 of the 25 CI users tested for our vowel identification task were also tested for first-formant (F1) identification.

1. Stimuli and general procedures

The testing conditions for this experiment were the same as for the vowel identification experiment in Sec. III A 2, differing only in the type and number of stimuli to identify. For F1 identification, stimuli were seven synthetic three-formant steady-state vowels created with the Klatt 88 speech synthesizer (Klatt and Klatt, 1990). The synthetic vowels differed from each other only in steady-state first-formant center frequencies, which ranged between 250 and 850 Hz in increments of 100 Hz. The fundamental, second, and third formant frequencies were fixed at 100, 1500, and 2500 Hz, respectively. Steady-state F1 values were verified with an acoustic waveform editor. The spectral envelope was obtained from the middle portion of each stimulus, and the frequency value of the F1 spectral peak was confirmed. Each stimulus was 1 s in duration and the onset and offset of the vowel envelope occurred over a 10 ms interval, this transition being linear in dB. The stimuli were digitally stored using a sampling rate of 11 025 Hz at 16 bits of resolution. Listeners were tested using a seven-alternative, one interval forced choice absolute identification task. During each block of testing stimuli were presented ten times in random order (i.e., 70 presentations per block). Prior to testing, participants would familiarize themselves with each stimulus (numbered 1–7) using an interactive software interface. During testing, participants would cue the interface to play a stimulus and then select the most appropriate stimulus number. After each selection, feedback about the correct response was displayed on the computer monitor before moving on to the next stimulus. Subjects completed seven to ten testing blocks (with the exception of listeners N6 and N7 who completed six and five testing blocks, respectively). This number of testing blocks was chosen as it was typically sufficient for most listeners to

provide at least two runs representative of asymptotic, or best, performance.

2. Cumulative- d' (Δ') analysis

For each block of testing a sensitivity index d' (Durlach and Braida, 1969) was calculated for each pair of adjacent stimuli (1 vs 2, 2 vs 3, ..., 6 vs 7) and then summed to obtain the total sensitivity, i.e., Δ' , which is the cumulative- d' across the range of first-formant frequencies between 250 and 850 Hz (i.e., from stimuli 1 to 7). For a given pair of adjacent stimuli, d' was calculated by subtracting the mean responses for the two stimuli and dividing by the average standard deviation of the responses to the two stimuli. For each CI user, the two highest Δ' among all testing blocks were averaged to arrive at the final score for this task. The average of the highest two Δ' scores represents an estimate of asymptotic performance, i.e., failure to improve Δ' . Asymptotic performance was sought as it provides a measure of sensory discrimination performance after factoring in learning effects and factoring out fatigue. As is customary for Δ' calculations, any d' score greater than 3 was set to $d' = 3$ (Tong and Clark, 1985). We defined the JND as occurring at $d' = 1$, so that Δ' equals the number of JNDs across the range of first-formant frequencies between 250 and 850 Hz. We then divided this range (i.e., 600 Hz) by Δ' to obtain the average JND in Hz.

To test the premise that a relationship exists between CI listeners' ability to identify vowels and their ability to discriminate steady-state formant frequencies, two correlation analyses were made using the average JNDs (in hertz) measured in the F1 identification task. One comparison was between JNDs (in hertz) and vowel identification percent correct scores. The other comparison was between JNDs (in hertz) and the F1F2F3 MPI model input JNDs (in millimeters) that yielded best-fit predicted matrices in terms of low-est rms difference.

B. Results

Listed in Table VI are CI subjects' observed percent correct scores for vowel identification and observed average JNDs (in hertz) for first-formant identification (F1 ID). Also listed in Table VI are CI subjects' predicted vowel identification percent correct and input JNDs (in millimeters) that provided best-fit model matrices using the F1F2F3 MPI model. Comparing the observed scores, a scatter plot of vowel scores and JNDs for the 18 CI users tested on both tasks (Fig. 3, top panel) yields a correlation of $r = -0.654$ ($p = 0.003$). This result suggests that in our group of CI users, the ability to correctly identify vowels was significantly correlated with the ability to identify first-formant frequency. Furthermore, for the same 18 CI users, a scatter plot of the MPI model input JNDs in millimeters against the observed JNDs in hertz from F1 identification (Fig. 3, bottom panel) yields a correlation of $r = 0.635$, $p = 0.005$ (without the data point with the highest predicted JND in millimeters, $r = 0.576$ and $p = 0.016$). Hence, a significant correlation exists between the JNDs obtained from first-formant identification and the JNDs obtained indirectly by optimizing model matrices to fit the vowel identification matrices obtained from

TABLE VI. Observed percent correct scores for vowel identification and average JNDs (in hertz) for first-formant identification, and F1F2F3 MPI model-predicted vowel percent correct scores and input JNDs that minimized rms difference between predicted and observed vowel confusion matrices for CI users tested in this study (NA=not available).

Subject	Observed		Predicted (F1F2F3)	
	Vowel (%)	JND (Hz)	Vowel (%)	JND (mm)
C1	72.6	279	72.6	0.095
C2	98.5	144	91.6	0.040
C3	94.1	138	89.5	0.040
C4	80.0	NA	77.8	0.080
C5	21.5	359	24.1	0.685
C6	43.7	111	45.9	0.125
C7	83.7	88	84.9	0.060
N1	80.0	NA	70.9	0.280
N2	22.2	NA	28.8	1.575
N3	73.3	141	71.6	0.230
N4	70.4	247	70.6	0.280
N5	95.6	NA	91.8	0.070
N6	81.7	131	75.5	0.225
N7	72.6	123	80.7	0.150
N8	26.1	324	29.0	1.725
N9	80.0	NA	76.9	0.270
N10	81.5	NA	72.6	0.175
N11	85.0	159	80.8	0.220
N12	42.2	224	45.8	0.820
N13	79.3	116	80.4	0.225
N14	81.5	138	79.4	0.235
N15	91.1	NA	87.3	0.140
N16	59.3	185	52.8	0.645
N17	71.1	141	72.7	0.315
N18	66.7	311	64.1	0.430

the same listeners. That is, fitting the MPI model to one data set (vowel identification) produced JNDs that are consistent with JNDs obtained with the same listeners from a completely independent data set (F1 identification).

C. Discussion

The significant correlations in Fig. 3 lend support to the hypothesis that CI users' ability to discriminate the locations of steady-state mean formant energies along the electrode array contributes to vowel identification, and also provides a degree of validation for the manner in which the MPI model of the present study connects these two variables. Nevertheless, the correlations were not very large, accounting for approximately 40% of the variability observed in the scatter plots. One important difference between identification of vowels and identification of formant center frequencies is that the former involves the assignment of lexically meaningful labels stored in long-term memory whereas the latter does not. Hence, if a CI user has very good formant center frequency discrimination, their ability to identify vowels could still be poor if their vowel labels are not sufficiently resolved in long-term memory. That is, good formant center frequency discrimination is necessary but not sufficient for good vowel identification.

As a side note, the observed JNDs in Table VI were larger than those reported by Fitzgerald *et al.* (2007).

FIG. 3. Top panel: scatter plot of vowel identification percent correct scores against observed JND (in hertz) from first-formant identification obtained from 18 CI users ($r = -0.654$, $p = 0.003$). Bottom panel: scatter plot of F1F2F3 MPI model's input JNDs (in millimeters) that produced best-fit to subjects' observed vowel matrices (minimized rms) against these subjects' observed JND (in hertz) from first-formant identification ($r = 0.635$ and $p = 0.005$).

However, this is to be expected as their F1 discrimination task measured the JND above an F1 center frequency of 250 Hz, whereas our measure represented the average JND for F1 center frequencies between 250 and 850 Hz.

V. EXPERIMENT 3: FREQUENCY ALLOCATION TABLES

A. Methods

Skinner *et al.* (1995) examined the effect of FAT Nos. 7 and 9 on speech perception with seven postlingually deafened adult users of the Nucleus-22 device and SPEAK stimulation strategy. Although FAT No. 9 was the default clinical map, Skinner *et al.* (1995) found that their listeners' speech perception improved with FAT No. 7. The speech battery they used included a vowel identification task with 19 medial vowels in /hVd/context, 3 tokens each, comprising 9 pure vowels, 5 *r*-colored vowels, and 5 diphthongs. The vowel confusion matrices they obtained (and recordings of the stimuli they used) were provided to us for the present study.

1. Application of MPI model

The MPI model was applied to the vowel identification data of Skinner *et al.* (1995) in order to test the model's ability to explain the improvement in performance that occurred when listener's used FAT No. 7 instead of FAT No. 9.

As a demonstration of how the MPI model can be used to explore the vast number of possible settings for a given CI fitting parameter in a very short amount of time, the MPI model was also used to provide a projection of vowel percent correct scores as a function of ten different frequency allocation tables and JND.

Step 1. One perceptual dimension combination was used to model the data of Skinner *et al.* (1995) and to generate predictions at other FATs. Namely, mean locations of formant energies along the electrode array for the first three formants combined, i.e., F1F2F3, in units of millimeters from the most basal electrode.

Step 2. Because our MPI model predicts identification of and confusions among vowels based on CI users' discrimination of mean formant energy locations, only ten of the vowels used by Skinner *et al.* (1995) were used in our model; i.e., the nine purely monophthongal vowels and the *r*-colored vowel heard. Using the original vowel recordings used by Skinner *et al.* (1995) and sCILab software (Bögli *et al.*, 1995; Wai *et al.*, 2003), two sets of formant location measurements were obtained from a Nucleus-22 spectra body-worn processor programmed with the SPEAK stimulation strategy. One set of measurements was obtained while the processor was programmed with FAT No. 7, and the other while the processor was programmed with FAT No. 9. Both sets of measurements were used for fitting Skinner *et al.*'s (1995) data, and for the MPI model's projection of vowel percent correct as a function of JND. For the model's projection at other FATs, formant location measurements were obtained using linear interpolation from FAT No. 9. The other frequency allocation tables explored in this projection were FAT Nos. 1, 2, and 6–13.

Step 3. For Skinner *et al.*'s (1995) data, the MPI model was run while allowing JND to vary as a free parameter until model matrices were obtained that best-fit the observed group vowel confusion matrices at FAT Nos. 7 and 9. The JND parameter was varied from 0.1 to 1 mm of electrode distance in increments of 0.01 mm using one degree of freedom; i.e., JND was the same for each perceptual dimension. Only one value of JND was used to find a best-fit to both sets of observed matrices in terms of minimum rms combined for both matrices. For the MPI model's projection of vowel identification as a function of the various FATs, model matrices were obtained for JND values of 0.1, 0.2, 0.4, 0.8, and 1.0 mm of electrode distance, where JND was assumed to be the same for each perceptual dimension. Percent correct scores were then calculated from the resulting model matrices. In all of the above simulations, the MPI model was run using 5000 iterations per vowel token.

B. Results

1. Application of MPI model to Skinner *et al.* (1995)

For the ten vowels we included in our modeling, the average vowel identification percent correct scores for the group of listeners tested by Skinner *et al.* (1995) were 84.9% with FAT No. 7 and 77.5% with FAT No. 9. For the MPI model of Skinner *et al.*'s (1995) data, a JND of 0.24 mm produced best-fit model matrices. The rms differences between observed and predicted matrices were 4.3% for FAT

TABLE VII. 2×2 comparison matrices for MPI model matrices produced with JND=0.24 mm and Skinner *et al.*'s (1995) vowel matrices obtained with FAT Nos. 7 and 9. The data follow the key at the bottom of Table IV.

FAT No. 7	F1F2F3	FAT No. 9	F1F2F3
3%	$p < 0.001$	3%	$p < 0.001$
6	3	6	5
2	34	1	33

No. 7 and 6.2% for FAT No. 9. The predicted matrices had percent correct scores equal to 85.1% with FAT No. 7 and 79.4% with FAT No. 9. Thus, the model predicted that FAT No. 7 should result in better vowel identification (which was true for all JND values between 0.1 and 1 mm) and it also predicted the size of the improvement. The 2×2 comparison matrices that demonstrate the extent to which model matrices account for the error pattern in Skinner *et al.*'s (1995) matrices are presented in Table VII. The comparison matrices were compiled using a threshold of 3%. With one degree of freedom, the MPI model produced model matrices that account for 40 out of 45 vowel pair confusions in the case of FAT No. 7 and 39 out of 45 vowel pair confusions in the case of FAT No. 9. For both comparison matrices, a Fisher's exact test yields $p < 0.001$.

2. MPI model projection at various FATs

The FAT determines the frequency band assigned to a given electrode. The ten FATs used to produce MPI model projections of vowel percent correct scores are summarized in Table VIII, which depicts the FAT number (1, 2, and 6–13), channel number (starting from the most apically stimulating electrode), and the lower frequency boundary (in hertz) assigned to a given channel (the upper frequency boundary for a given channel is equal to the lower frequency boundary of the next highest channel number, and the upper boundary for the highest channel number is provided in the bottom row). The percent correct scores obtained from MPI model matrices at each FAT, and as a function of JND are summarized in Fig. 4. Two observations are worth noting. First, a lower JND for a given frequency map results in a higher predicted percent correct score. That is, a lower JND would provide better discrimination between formant values and hence a smaller chance of confusing formant values belonging to different vowels. Second, for a fixed JND, percent correct scores begin to gradually decrease as the FAT number is increased to higher FAT numbers beyond FAT No. 7, with the exception of JND=0.1 mm where a ceiling effect is observed. As FAT number increases from No. 1 to No. 9, a larger frequency range is assigned to the same set of

TABLE VIII. Frequency allocation table numbers (FAT No.) 1, 2, and 6–13 for the Nucleus-22 device. Channel numbers begin with the most apically stimulated electrode and indicate the lower frequency boundary (in hertz) assigned to a given electrode. Bottom row indicates upper frequency boundary for highest frequency channel. Approximate range of formant frequency regions indicated by text in bold: F1 (300–1000 Hz), F2 (1000–2000 Hz), and F3 (2000–3000 Hz).

Channel	FAT No.									
	1	2	6	7	8	9	10	11	12	13
1	75	80	109	120	133	150	171	200	240	150
2	175	186	254	280	311	350	400	466	560	300
3	275	293	400	440	488	550	628	733	880	700
4	375	400	545	600	666	750	857	1 000	1 200	1100
5	475	506	690	760	844	950	1 085	1 266	1 520	1500
6	575	613	836	920	1022	1 150	1 314	1 533	1 840	1900
7	675	720	981	1080	1200	1 350	1 542	1 800	2 160	2300
8	775	826	1127	1240	1377	1 550	1 771	2 066	2 480	2700
9	884	942	1285	1414	1571	1 768	2 020	2 357	2 828	3100
10	1015	1083	1477	1624	1805	2 031	2 321	2 708	3 249	3536
11	1166	1244	1696	1866	2073	2 333	2 666	3 110	3 732	4062
12	1340	1429	1949	2144	2382	2 680	3 062	3 573	4 288	4666
13	1539	1642	2239	2463	2736	3 079	3 518	4 105	4 926	5360
14	1785	1904	2597	2856	3174	3 571	4 081	4 761	5 713	6158
15	2092	2231	3042	3347	3719	4 184	4 781	5 578	6 694	7142
16	2451	2614	3565	3922	4358	4 903	5 603	6 537	7 844	8368
17	2872	3063	4177	4595	5105	5 744	6 564	7 658	9 190	...
18	3365	3589	4894	5384	5982	6 730	7 691	8 973	...	...
19	3942	4205	5734	6308	7008	7 885	9 011	...	...	...
20	4619	4926	6718	7390	8211	9 238	...	...	...	...
Upper	5411	5772	7871	8658	9620	10 823	10 557	10 513	10 768	9806

FIG. 4. F1F2F3 MPI model prediction of vowel identification percent correct scores as a function of FAT No. and JND (in millimeters). Filled circles: Skinner *et al.*'s (1995) mean group data when CI subjects' used FAT Nos. 7 and 9.

electrodes. For FAT Nos. 10–13, the relatively large frequency span is maintained while the number of electrodes assigned is gradually reduced. Hence, the MPI model predicts that vowel identification will be deleteriously affected by assigning too large of a frequency span to the CI electrodes. In Fig. 4, the two filled circles joined by a solid line represent the vowel identification percent correct scores obtained by Skinner *et al.* (1995) for the ten vowel tokens we included in our modeling.

C. Discussion

The very first thing to point out is the economy with which the MPI model can be used to project estimates of CI users' performance. The simulation routine implementing the MPI model produced all of the outputs in Fig. 4 in a matter of minutes. Contrast this with the time and resources required to obtain data such as that of Skinner *et al.* (1995), which amounts to two data points in Fig. 4. It would be financially and practically impossible to obtain these data experimentally for all the frequency maps available with a given cochlear implant, let alone for the theoretically infinite number of possible frequency maps.

Without altering any model assumptions, the model predicts the increase in percent correct vowel identification attributable to changing the frequency map from FAT No. 9 to FAT No. 7 with the Nucleus-22 device. In retrospect, Skinner *et al.* (1995) hypothesized that FAT No. 7 might result in improved speech perception because it encodes a more restricted frequency range onto the electrodes of the implanted array. Encoding a larger frequency range onto the array involves a tradeoff: The locations of mean formant energies for different vowels are squeezed closer together. With less space between mean formant energies, the vowels become more difficult to discriminate, at least in terms of this particular set of perceptual dimensions, resulting in a lower percent correct score.

How does this concept apply to the MPI model projections at different FATs displayed in Fig. 4? The effect of different FAT frequency ranges on mean formant locations along the electrode array is depicted in Table VIII where approximate formant regions are indicated in bold. The frequency boundaries defined for each formant are 300–1000 Hz for F1, 1000–2000 Hz for F2, and 2000–3000 Hz for F3.

Under this definition of formant regions, five or more electrodes are available for each of F1 and F2 for all maps up to FAT No. 8, and progressively decrease for higher map numbers. In Fig. 4, percent correct changes very little between FAT Nos. 1 and 8, suggesting that F1 and F2 are sufficiently resolved, and then drops progressively for higher map numbers. Indeed, FAT No. 9 has one less electrode available for F2 in comparison to FAT No. 7, which may explain the small but significant drop in percent correct scores with FAT No. 9 observed by Skinner *et al.* (1995).

Apparently, the changes in the span of electrodes for mean formant energies in FAT Nos. 7 and 9 are of a magnitude that will not contribute to large differences in vowel percent correct score for JND values that are very small (less than 0.2 mm) or very high (more than 0.8 mm), but are relevant for JND values that are in between these two extremes.

Although the prediction of the MPI model in Fig. 4 suggests that there is not much to be gained (or lost, for that matter) by shifting the frequency map from FAT No. 7 to FAT No. 1, there is strong evidence to suggest that such a change could be detrimental. Fu *et al.* (2002) found a significant drop in vowel identification scores in three postlingually deafened subjects tested with FAT No. 1 in comparison to their clinically assigned maps (FAT Nos. 7 and 9), even after these subjects used FAT No. 1 continuously for three months. Out of all the maps in Table VIII, FAT No. 1 encodes the lowest frequency range to the electrode array, and potentially has the largest frequency mismatch to the characteristic frequency of the neurons stimulated by the implanted electrodes; particularly for postlingually deafened adults who retained the tonotopic organization of the cochlea before they lost their hearing. The results of Fu *et al.* (2002) suggest that the use of FAT No. 1 in postlingually deafened adults results in an excessive amount of frequency shift, i.e., an amount of frequency mismatch that precludes complete adaptation. In Fig. 4, response bias was assumed to be zero (see Sec. IIA2) so that no mismatch occurred between percepts elicited by stimuli and the expected locations of those percepts. The contribution of a nonzero response bias to lowering vowel percent correct scores for the type of frequency mismatch imposed by FAT No. 1 is addressed in Sagi *et al.*, (2010) wherein the MPI model was applied to the vowel data of Fu *et al.* (2002).

VI. EXPERIMENT 4: ELECTRICAL DYNAMIC RANGE REDUCTION

A. Methods

The electrical dynamic range is the range between the minimum stimulation level for a given channel, typically set at threshold, and the maximum stimulation level, typically set at the maximum comfortable loudness. Zeng and Galvin (1999) systematically decreased the electrical dynamic range of four adult users of the Nucleus-22 device with SPEAK stimulation strategy from 100% to 25% and then to 1% of the original dynamic range. In the 25% condition, dynamic range was set from 75% to 100% of the original dynamic range. In the 1% condition, dynamic range was set from 75% to 76% of the original dynamic range. CI users were then

tested on several speech perception tasks including vowel identification in quiet. One result of Zeng and Galvin (1999) was that even though the electrical dynamic range was reduced to almost zero, the average percent correct score for identification of vowels in quiet dropped by only 9%. We sought to determine if the MPI model could explain this result by assessing the effect of dynamic range reduction on formant location measurements. If reducing the dynamic range has a small effect on formant location measurements, then the MPI model would predict a small change in vowel percent correct scores.

1. Application of MPI model

Step 1. One perceptual dimension combination was used to model the data of Zeng and Galvin (1999). Namely, mean locations of formant energies along the electrode array for the first three formants, i.e., F1F2F3, in units of millimeters from the most basal electrode.

Step 2. Three sets of formant location measurements were obtained, one for each dynamic range condition. For the 100% dynamic range condition, sCILab recordings were obtained for the vowel tokens used in experiment 1 of the present study, using a Nucleus-22 spectra body-worn processor programmed with the SPEAK stimulation strategy and FAT No. 9. The minimum and maximum stimulation levels in the output of the speech processor were set to 100 and 200 clinical units, respectively, for each electrode. For the other two dynamic range conditions, the stimulation levels in these sCILab recordings were adjusted in proportion to the desired dynamic range. That is, the charge amplitude of stimulation pulses, which spanned from 100 to 200 clinical units in the original recordings, was proportionally mapped to 175–200 clinical units for the 25% dynamic range condition, and to 175–176 clinical units for the 1% dynamic range condition. Formant locations were then obtained from electrodograms of the original and modified sCILab recordings.

Step 3. In Zeng and Galvin, 1999, the average vowel identification score in quiet for the 25% dynamic range condition was 69% correct. Using the formant measurements for this condition, the MPI model was run while varying JND, until a JND was found that produced a model matrix with percent correct equal to 69%. This value of JND was then used to run the MPI model with the other two sets of formant measurements for the 100% and 1% dynamic range conditions. In each case, the MPI model was run with 5000 iterations per vowel token, and the percent correct of the resulting model matrices was compared with the scores observed in Zeng and Galvin, 1999.

B. Results

With the MPI model, a JND of 0.27 mm provided a vowel percent correct score of 69% using the formant measurements obtained for the 25% dynamic range condition. With the same value of JND, the formant measurements obtained for the 100% and 1% dynamic range conditions yielded vowel matrices with 71% and 68% correct, i.e., a drop of 3%. The observed scores obtained by Zeng and Galvin (1999) for these two conditions were 76% and 67%, respectively, i.e., a drop of 9%. On one hand, the MPI model

employed here explains how a large reduction in electrical dynamic range results in a small drop in the identification of vowels under quiet listening conditions. On the other hand, the MPI model underestimated the magnitude of the drop observed by Zeng and Galvin (1999).

C. Discussion

It should not come as a surprise that the F1F2F3 MPI model employed here predicts that a large reduction in the output dynamic range would have a negligible effect on vowel identification scores in quiet. After all, reducing the output dynamic range (even 100-fold) causes a negligible shift in the location of mean formant energy along the electrode array. More importantly, why did this model underestimate the observed results of Zeng and Galvin (1999)? One explanation may be that the model does not account for the relative amplitudes of formant energies, which can affect percepts arising from F1 and F2 center frequencies in close proximity (Chistovich and Lublinskaya, 1979). Reducing the output dynamic range can affect the relative amplitudes of formant energies without changing their locations along the electrode array. This effect may explain why Zeng and Galvin (1999) found a larger drop in vowel identification scores than those predicted by the MPI model. Hence, the MPI model employed in the present study may be sufficient to explain the vowel identification data of experiments 1 and 3, but may need to be modified to more accurately predict the data of Zeng and Galvin (1999).

Of course, the prediction that reducing the dynamic range will not largely affect vowel identification scores in quiet only applies to users of stimulation strategies such as SPEAK, ACE, and n-of-m. This effect would be completely different for a stimulation strategy like CIS, where all electrodes are activated in cycles, and the magnitude of each stimulation pulse is determined in proportion to the electric dynamic range. For example, in a CI user with CIS, the 1% dynamic range condition used by Zeng and Galvin (1999) would result in continuous activation of all electrodes at the same level regardless of input, thus obliterating all spectral information about vowel identity.

VII. CONCLUSIONS

A very simple model predicts most of the patterns of vowel confusions made by users of different cochlear implant devices (Nucleus and Clarion) who use different stimulation strategies (CIS or SPEAK), who show widely different levels of speech perception (from near chance to near perfect), and who vary widely in age of implantation and implant experience (Tables II and III). The model's accuracy in predicting confusion patterns for an individual listener is surprisingly robust to these variations despite the use of a single degree of freedom. Furthermore, the model can predict some important results from the literature, such as Skinner *et al.*'s (1995) frequency mapping study, and the general trend (but not the size of the effect) in the vowel results of Zeng and Galvin's (1999) studies of output electrical dynamic range reduction.

The implementation of the model presented here is specific to vowel identification by CI users, dependent on

discrimination of mean formant energy along the electrode array. However, the framework of the model is general. Alternative models of vowel identification within the MPI framework could use dynamic measures of formant frequency (i.e., formant trajectories and co-articulation), or other perceptual dimensions such as formant amplitude or vowel duration. One alternative to the MPI framework might involve the comparison of phonemes based on time-averaged electrode activation across the implanted array, treated as a single object rather than breaking it down into specific “cues” or perceptual dimensions (cf. Green and Birdsall, 1958; Müsch and Buus, 2001). Regardless of the specific form they might take, computational models like the one presented here can be useful for advancing our understanding about speech perception in hearing impaired populations, and for providing a guide for clinical research and clinical practice.

ACKNOWLEDGMENTS

Norbert Dillier from ETH (Zurich) provided us with his sCilab computer program, which we used to record stimulation patterns generated by the Nucleus speech processors. Advanced Bionics Corporation provided an implant-in-a-box so we could monitor stimulation patterns generated by their implant. Margo Skinner (may she rest in peace) provided the original vowel tokens used in her study as well as the confusion matrices from that study. This study was supported by NIH-NIDCD Grant Nos. R01-DC03937 (P.I.: Mario Svirsky) and T32-DC00012 (PI: David B. Pisoni) as well as by grants from the Deafness Research Foundation and the National Organization for Hearing Research.

- Bögli, H., Dillier, N., Lai, W. K., Rohner, M., and Zillus, B. A. (1995). Swiss Cochlear Implant Laboratory (Version 1.4) ([computer software]), Zürich, Switzerland.
- Braida, L. D. (1991). “Crossmodal integration in the identification of consonant segments,” *Q. J. Exp. Psychol.* **43A**, 647–677.
- Braida, L. D., and Durlach, N. I. (1972). “Intensity perception. II. Resolution in one-interval paradigms,” *J. Acoust. Soc. Am.* **51**, 483–502.
- Chatterjee, M., and Peng, S. C. (2008). “Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition,” *Hear. Res.* **235**, 143–156.
- Chistovich, L. A., and Lublinskaya, V. V. (1979). “The ‘center of gravity’ effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli,” *Hear. Res.* **1**, 185–195.
- Durlach, N. I., and Braida, L. D. (1969). “Intensity perception. I. Preliminary theory of intensity resolution,” *J. Acoust. Soc. Am.* **46**, 372–383.
- Firszt, J. B., Koch, D. B., Downing, M., and Litvak, L. (2007). “Current steering creates additional pitch percepts in adult cochlear implant recipients,” *Otol. Neurotol.* **28**, 629–636.
- Fitzgerald, M. B., Shapiro, W. H., McDonald, P. D., Neuburger, H. S., Ashburn-Reed, S., Immerman, S., Jethanamest, D., Roland, J. T., and Svirsky, M. A. (2007). “The effect of perimodiolar placement on speech perception and frequency discrimination by cochlear implant users,” *Acta Oto-Laryngol.* **127**, 378–383.
- Fu, Q. J., Shannon, R. V., and Galvin, J. J., III (2002). “Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant,” *J. Acoust. Soc. Am.* **112**, 1664–1674.
- Green, D. M., and Birdsall, T. G. (1958). “The effect of vocabulary size on articulation score,” Technical Memorandum No. 81 and Technical Note No. AFCRC-TR-57-58, University of Michigan, Electronic Defense Group.
- 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.
- Hood, L. J., Svirsky, M. A., and Cullen, J. K. (1987). “Discrimination of complex speech-related signals with a multichannel electronic cochlear implant as measured by adaptive procedures,” *Ann. Otol. Rhinol. Laryngol.* **96**, 38–41.
- Iverson, P., Smith, C. A., and Evans, B. G. (2006). “Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement and duration,” *J. Acoust. Soc. Am.* **120**, 3998–4006.
- Jenkins, J. J., Strange, W., and Edman, T. R. (1983). “Identification of vowels in ‘vowelless’ syllables,” *Percept. Psychophys.* **34**, 441–450.
- Kewley-Port, D., and Watson, C. S. (1994). “Formant-frequency discrimination for isolated English vowels,” *J. Acoust. Soc. Am.* **95**, 485–496.
- Kirk, K. I., Tye-Murray, N., and Hurtig, R. R. (1992). “The use of static and dynamic vowel cues by multichannel cochlear implant users,” *J. Acoust. Soc. Am.* **91**, 3487–3497.
- Klatt, D. H., and Klatt, L. C. (1990). “Analysis, synthesis, and perception of voice quality variations among female and male talkers,” *J. Acoust. Soc. Am.* **87**, 820–857.
- Kwon, B. J., and van den Honert, C. (2006). “Dual-electrode pitch discrimination with sequential interleaved stimulation by cochlear implant users,” *J. Acoust. Soc. Am.* **120**, EL1–EL6.
- Müsch, H., and Buus, S. (2001). “Using statistical decision theory to predict speech intelligibility. I. Model structure,” *J. Acoust. Soc. Am.* **109**, 2896–2909.
- Peterson, G. E., and Barney, H. L. (1952). “Control methods used in a study of the vowels,” *J. Acoust. Soc. Am.* **24**, 175–184.
- Phatak, S. A., and Allen, J. B. (2007). “Consonant and vowel confusions in speech-weighted noise,” *J. Acoust. Soc. Am.* **121**, 2312–2326.
- Pitt, M. A., and Navarro, D. J. (2005). In *Twenty-First Century Psycholinguistics: Four Cornerstones*, edited by A. Cutler (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 347–362.
- Ronan, D., Dix, A. K., Shah, P., and Braida, L. D. (2004). “Integration across frequency bands for consonant identification,” *J. Acoust. Soc. Am.* **116**, 1749–1762.
- Sagi, E., Fu, Q.-J., Galvin, J. J., III, and Svirsky, M. A. (2010). “A model of incomplete adaptation to a severely shifted frequency-to-electrode mapping by cochlear implant users,” *J. Assoc. Res. Otolaryngol.* (in press).
- Shannon, R. V. (1993). In *Cochlear Implants: Audiological Foundations*, edited by R. S. Tyler (Singular, San Diego, CA), pp. 357–388.
- Skinner, M. W., Arndt, P. L., and Staller, S. J. (2002). “Nucleus 24 advanced encoder conversion study: Performance versus preference,” *Ear Hear.* **23**, 2S–17S.
- Skinner, M. W., Fourakis, M. S., Holden, T. A., Holden, L. K., and Demorest, M. E. (1996). “Identification of speech by cochlear implant recipients with the multiplex (MPEAK) and spectral peak (SPEAK) speech coding strategies I. vowels,” *Ear Hear.* **17**, 182–197.
- Skinner, M. W., Holden, L. K., and Holden, T. A. (1995). “Effect of frequency boundary assignment on speech recognition with the SPEAK speech-coding strategy,” *Ann. Otol. Rhinol. Laryngol.* **104**, (Suppl. 166), 307–311.
- Svirsky, M. A. (2000). “Mathematical modeling of vowel perception by users of analog multichannel cochlear implants: Temporal and channel-amplitude cues,” *J. Acoust. Soc. Am.* **107**, 1521–1529.
- Svirsky, M. A. (2002). In *Etudes et Travaux*, edited by W. Serniclaes (Institut de Phonétique et des Langues Vivantes of the ULB, Brussels), Vol. **5**, pp. 143–186.
- Syrdal, A. K., and Gopal, H. S. (1986). “A perceptual model of vowel recognition based on the auditory representation of American English vowels,” *J. Acoust. Soc. Am.* **79**, 1086–1100.
- Teoh, S. W., Neuburger, H. S., and Svirsky, M. A. (2003). “Acoustic and electrical pattern analysis of consonant perceptual cues used by cochlear implant users,” *Audiol. Neuro-Otol.* **8**, 269–285.
- Thurstone, L. L. (1927a). “A law of comparative judgment,” *Psychol. Rev.* **34**, 273–286.
- Thurstone, L. L. (1927b). “Psychophysical analysis,” *Am. J. Psychol.* **38**, 368–389.
- Tong, Y. C., and Clark, G. M. (1985). “Absolute identification of electric pulse rates and electrode positions by cochlear implant subjects,” *J. Acoust. Soc. Am.* **77**, 1881–1888.
- Wai, K. L., Bögli, H., and Dillier, N. (2003). “A software tool for analyzing multichannel cochlear implant signals,” *Ear Hear.* **24**, 380–391.
- Zahorian, S. A., and Jagharghi, A. J. (1993). “Spectral-shape features versus formants as acoustic correlates for vowels,” *J. Acoust. Soc. Am.* **94**, 1966–1982.
- Zeng, F. G., and Galvin, J. J., III (1999). “Amplitude mapping and phoneme recognition in cochlear implant listeners,” *Ear Hear.* **20**, 60–74.