Published in final edited form as:

Hear Res. 2018 December ; 370: 316–328. doi:10.1016/j.heares.2018.10.014.

DEACTIVATING COCHLEAR IMPLANT ELECTRODES TO IMPROVE SPEECH PERCEPTION: A COMPUTATIONAL APPROACH

Elad Sagi and Mario A. Svirsky

New York University School of Medicine, New York, NY, USA

Abstract

A potential bottleneck to improving speech perception performance in cochlear implant (CI) users is that some of their electrodes may poorly encode speech information. Several studies have examined the effect of deactivating poorly encoding electrodes on speech perception with mixed results. Many of these studies focused on identifying poorly encoding electrodes by some measure (e.g. electrode discrimination, pitch ordering, threshold, CT-guided, masked modulation detection), but provide inconsistent criteria about which electrodes, and how many, should be deactivated, and without considering how speech information becomes distributed across the electrode array. The present simulation study addresses this issue using computational approaches. Previously validated models were used to generate predictions of speech scores as a function of all possible combinations of active electrodes in a 22-electrode array in three groups of hypothetical subjects representative of relatively better, moderate, and poorer performing CI users. Using high-performance computing, over 500 million predictions were generated. Although deactivation of the poorest encoding electrodes sometimes resulted in predicted benefit, this benefit was significantly less relative to predictions resulting from model-optimized deactivations. This trend persisted when using novel stimuli (i.e. other than those used for optimization) and when using different processing strategies. Optimum electrode deactivation patterns produced an average predicted increase in word scores of 10% with some scores increasing by more than 20%. Optimum electrode deactivation patterns typically included 11 to 19 (out of 22) active electrodes, depending on the performance group. Optimal active electrode combinations were those that maximized discrimination of speech cues, maintaining 80% to 100% of the physical span of the array. The present study demonstrates the potential for further improving CI users' speech scores with appropriate selection of active electrodes.

Keywords

Cochlear implants; electrode deactivation; computational modeling

Corresponding Author: Elad Sagi, PhD, Department of Otolaryngology, NBV-5E5, New York University School of Medicine, 550 First Ave., New York, NY, USA 10016, Elad.Sagi@nyumc.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. INTRODUCTION

Cochlear implants (CIs) are implantable auditory prosthetic devices that have successfully restored functional hearing to hundreds of thousands of individuals with severe-to-profound hearing impairment world-wide. Since their introduction as a clinical intervention several decades ago, the combination of relaxed patient selection criteria, improved clinical management of hearing loss, and regular advances in technology and surgical approaches have brought regular improvements in speech perception outcomes at the group level (Blamey et al., 2013). Nevertheless, individual performance remains highly variable, e.g. open-set word scores in quiet with modern-day CIs can range from near zero to near perfect (Holden et al., 2013; Kraaijenga et al., 2016). Despite these differences, apart from a few programming parameters, CI device settings are typically kept at default settings provided by the manufacturer. This is not a criticism of current clinical practice. There are millions of possible CI device setting combinations that can be programmed. There is also very little guidance for clinicians on which combination of device settings will maximize benefit for an individual CI user (Govaerts et al., 2010).

Closing the gap between better and poorer performing CI users remains one of the major goals in the CI field. One issue thought to prevent CI users with poorer speech scores from performing better is that a subset of their intracochlear electrodes may poorly convey speech information. Multichannel CIs include an array of electrodes (currently from 12 to 22 depending on manufacturer) inserted along the length of the cochlea, where more apical electrodes tend to produce lower pitch percepts and more basal electrodes tend to produce higher pitch percepts. Aside from deactivating nonfunctional electrodes, or those that induce discomfort (non-auditory or auditory), the large majority of CIs are programmed clinically with all viable electrodes active (Zeitler et al., 2008). Presumably, this is the best way to map the full range of important auditory frequencies in tonotopic fashion onto the cochlear length covered by the electrode array. However, some electrodes may be poorly suited to convey speech information. For example, in some CI users, neighboring electrodes may not be discriminable from each other and some electrodes may even produce anti-tonotopic pitch sensations (that is, a given electrode may sound lower pitched rather than higher pitched than its more apical neighbor). Having discriminable electrodes well-ordered in pitch is thought to contribute to better speech perception (Collins et al., 1997; Donaldson and Nelson, 2000; Henshall and McKay, 2001; Kenway et al., 2015). Furthermore, stimulation of some electrodes can interact and interfere with neural encoding of spectral and temporal information provided by other electrodes (e.g. forward masking studies reviewed in McKay, 2012).

A body of research has developed to mitigate some of these effects through selective deactivation of stimulating electrodes with the goal of improving CI users' speech perception. This approach is necessarily selective because it is practically impossible to try all combinations of stimulating electrodes. For example, mathematically speaking, the total number of active electrode combinations in a 22-channel CI device, from 1 through 22 active electrodes is $2^{22} - 1 = 4,194,303$ possible combinations. Instead, the more principled approach has been to assess the encoding abilities of CI users' electrodes by some measure, and to deactivate the poorer-encoding electrodes. These assessments of electrode encoding

abilities include measures of electrode discrimination (Zwolan et al., 1997; Debruyne et al., 2017), pitch ranking (Saleh et al., 2013; Vickers et al., 2016) and tonotopicity (Henshall and McKay, 2001); measures that aim to assess the integrity of the electrode-neural interface by using absolute thresholds (Bierer and Litvak, 2016; Zhou, 2016) and CT-based anatomical measurements (Noble et al., 2014; Labadie et al., 2016); and measures of masked modulation detection thresholds (Zhou and Pfingst, 2012; Garadat et al., 2013). In terms of improving speech perception outcomes with respect to the baseline condition (where all viable electrodes are activated), results have been mixed. Three of these studies reported little to no improvement in speech scores for subjects using maps with electrodes selectively deactivated (Henshall and McKay, 2001; Vickers et al., 2016; Debruyne et al., 2017). Three show significant improvements in speech scores for some CI subjects, though not at the group level (Zwolan et al., 1997; Bierer and Litvak, 2016; Labadie et al., 2016). Five studies demonstrated significant improvements in CI users' speech scores at the group level (Zhou and Pfingst, 2012; Garadat et al., 2013; Saleh et al., 2013; Noble et al., 2014; Zhou, 2016).

In addition to outcomes, these electrode-deactivation studies vary in the criteria used for removing poorer encoding electrodes. For example, Zwolan et al. (1997) and Debruyne et al. (2017) retained mutually discriminable electrodes without restriction resulting in a large range in the number of electrodes deactivated across subjects (from 0 to 9 in the former study, and from 4 to 14 in the latter study). Henshall and McKay (2001) deactivated 3 to 4 non-tonotopic electrodes in their CI subjects. Bierer and Litvak (2016) deactivated high-threshold electrodes with the restrictions that no more than two adjacent electrodes in apical, middle, and basal frequency regions of the array could be deactivated, and that at least 8 electrodes were to be retained across the array. Zhou (2016) deactivated five non-adjacent electrodes with highest thresholds. Noble et al. (2014) and Labadie et al. (2016) deactivated electrodes that were farther away from the modiolus than their neighboring electrodes, while retaining at least 8 active electrodes. Garadat et al. (2013) partitioned the array into 5 regions and deactivated the electrode in each region with poorest masked modulation detection threshold. In cases where one region had electrodes with better overall thresholds than other regions, no more than two adjacent electrodes were deactivated in the other regions.

All these electrode selection criteria are entirely logical but they do not consider how speech information is distributed across active electrodes. It is known that some frequencies are more relevant for understanding speech than others (Healy et al., 2013; Bosen and Chatterjee, 2016; Lee and Mendel, 2017). Although deactivating a poorly encoding electrode may sound beneficial, in practice the deactivation unavoidably results in a number of changes for each remaining electrode. For example, deactivating a poorly encoding electrode and reallocating its frequency range to a better encoding electrode may be undesirable if the frequency range in question has limited importance for understanding speech. In other words, speech perception may not improve if deactivation of an electrode doesn't improve a listener's overall encoding of speech cues. Another factor to consider is that deactivating electrodes may cause an important range of frequencies to be concentrated over a smaller cochlear distance, reducing the listener's ability to differentiate speech cues within that range of important frequencies.

Another issue with all the above listed electrode selection criteria is that, by practical necessity, they must resort to unproven heuristic rules. For example, different studies decided to retain at least one half of all available electrodes, or to deactivate exactly five electrodes for all subjects. However, and beyond the issue of which specific electrodes should be removed, it is impossible to determine a-priori that five (or any other number) is the optimal number in general, or whether the number of electrodes to be removed should be a function of some psychophysical or anatomical measure, or whether it even makes a difference if one removes any number of electrodes. These and many other questions cannot be answered without an exhaustive exploration of a huge parametric space. For example, if a patient has 22 active channels and we want to use a map that includes 6 to 21 channels, there are more than 4 million possible combinations, and it is just not possible to test all those combinations experimentally.

Thus, in the present study, computational modeling is used to generate predictions of CI users' phoneme and word scores as a function of all possible electrode deactivation combinations in a 22-electrode CI array. Hundreds of millions of predictions are generated for three groups of hypothetical CI subjects with relatively poorer, moderate, and better electrode discrimination values (randomly generated) along the array. The modeling approach herein assesses how speech cues represented by a CI are affected by the encoding abilities of active electrodes, thereby producing predictions of phoneme and word scores. Optimal predicted speech scores (and their associated active electrode combinations) are compared with predicted speech scores obtained when electrodes associated with poorest encoding abilities are deactivated. The computational modeling approach of the present study can be used to inform site selection criteria when considering electrode deactivation studies, and when programming individual CI users.

Admittedly, the significance of the present study hinges on the validity of the underlying computational models. We do not claim that those models provide a perfect explanation of speech perception by CI users. Instead, we are interested in seeing what insights might be provided by a comprehensive analysis of all possible electrode combinations when using computational models that have provided explanations of several different speech perception results in previous studies of speech perception by cochlear implant users. For example, the computational models that underlie those employed in the present study have been successful at explaining the following: patterns of integration of waveform and channel-amplitude cues in CI users (Svirsky, 2000), vowel perception in CI users based on their ability to discriminate formant cues as well as CI users' vowel scores with changes to their frequency-allocation-map and output dynamic range (Sagi et al., 2010b), consonant perception in CI users based on their ability to discriminate a subset of acoustic-phonetic cues (Svirsky et al., 2011), perceptual adaptation to frequency mismatch in CI users (Sagi et al., 2010a; Svirsky et al., 2015), and CI users' vowel perception in noise (Sagi and Svirsky 2017). These models have been able to explain not just percent correct scores but also most of the vowel pairs that would or would not be confused by individual CI users, and the same can be said for consonants. These results were obtained for users of different cochlear implants and stimulation strategies over a wide range of speech perception performance. These models are not perfect, and their ability to predict specific confusions between phoneme pairs are more accurate for perception of vowels in quiet than for consonants in

quiet or vowels in noise. Despite the models' shortcomings, we believe there is value in using them to investigate how speech perception by CI users may be affected when removing any give subset of electrodes. The present study aims to complement existing studies of electrode deactivation by using a novel computational approach that may contribute to the discussion.

2. METHODS

2.1. Simulated subjects

The present study employs models validated previously on CI users to generate predictions of vowel, consonant, and word scores in three groups of twenty hypothetical subjects (60 in total) representative of relatively better, moderate, and poorer performing CI users. Simulated subjects are assigned random electrode discrimination, or just-noticeable-difference (JND), values to each of their 22 electrodes that vary uniformly over a range dependent on the performance group, i.e. from near 0 to 1 electrode for the better performing group, from near 0 to 2 electrodes for the moderately performing group, and from near 0 to 4 electrodes for the poorer performing group. To avoid performance group differences due to randomization, the same twenty randomizations were used for each group but scaled according to the electrode discrimination range assigned.

2.2. Stimuli

Speech cue measurements of vowel and consonant stimuli (Section 2.3.1) are used as inputs to the MPI model (Section 2.3.2). The vowel stimuli used for model optimization ("training stimuli") were recordings of nine vowels in /hVd/ context, i.e. /æ, ɔ, e, ɜ, i, ʊ, ʌ, and u/ (Cochlear Corporation and the University of Iowa, 1995). These recordings comprised several utterances of each vowel recorded from the same male talker, a native speaker of American English. The consonant stimuli used for model optimization were recordings of 16 consonants in /aCa/ context. These recordings comprised several utterances of each consonant recorded from the same female talker, who was a native speaker of American English (Tyler et al., 1987). The consonants were /b, d, f, g, k, l, m, n, p, r, s, ʃ, t, ð, v, and z/.

Modeling predictions obtained with these training stimuli were further validated (c.f. Section 2.5) for their generalizability to "novel" stimuli from contrasting talkers, i.e. female instead of male vowels, male instead of female consonants. The vowel stimuli used for validation were recordings from 5 female speakers of American English (taken from Hillenbrand et al., 1995), one utterance each of the 9 vowels used for model optimization. The consonant stimuli used for validation were recordings from one male speaker (Tyler et al., 1987), three utterances each of the 16 consonants used for model optimization.

2.3. Modeling

In the present study, vowel and consonant scores were generated using the multidimensional phoneme identification (MPI) model (Svirsky, 2000, 2002) which has been successfully employed in predicting vowel and consonant identification by cochlear implant (CI) users (Sagi et al., 2010, 2010b, 2017; Svirsky et al., 2011, 2015). The MPI model predicts a listener's phoneme identification based, in part, on their ability to discriminate a postulated

set of relevant speech cues. The modeling assumptions and speech cues employed in the present study closely follow Sagi et al. (2017) for generating vowel scores and Svirsky et al. (2011) for generating consonant scores. By way of overview, in the present study, speech cues are measured from emulations of the stimulation patterns resulting from CI processing of the vowels and consonants used in this study (Section 2.3.1). These speech cues are obtained for every possible combination of active electrodes in a hypothetical 22-electrode cochlear implant, i.e. $2^{22}-1 = 4,194,303$ possible combinations. Speech cue measurements and simulated subject JNDs for a given active electrode combination are used as inputs to the MPI model to generate vowel and consonant percent correct scores specific to that electrode combination. Using high-performance-computing, phoneme scores are generated for all possible combinations of active electrodes, for each simulated subject JND randomization, i.e. just over 500 million phoneme score predictions (Section 2.3.2). Predicted vowel and consonant scores are then combined to produce predicted monosyllabic word scores (Section 2.3.3) following the model of Rabinowitz et al. (1992). Details of these steps are provided in the sections that follow.

2.3.1. Speech cue measurements—In the present study, speech cues are measured from electrodiagrams of the speech stimuli described in Section 2.2. An electrodiagram is a representation of the pulse sequence generated by a CI speech processor in response to an acoustic stimulus (akin to a rough spectrogram), i.e. a record of which electrode is stimulated at a given point in time, and at what magnitude. Much like how acoustic speech cues can be measured from spectrograms, electric speech cues can be measured from electrodiagrams. Electrodiagrams are typically obtained by decoding the output of a speech processor, or by measuring the output at the electrode contacts of an ‘implant-in-a-box’. Although somewhat less realistic, electrodiagrams can also be obtained with software emulation that mimics the signal processing cascade implemented by a CI speech processor. Because of the large number of electrodiagrams required to obtain speech cue measurements for all stimuli as a function of all possible active electrode combinations, electrodiagram emulation was used in the present study.

Electrodiagram emulation was achieved with the Nucleus Matlab Toolbox (NMT, provided by Cochlear Americas, Denver, USA). The toolbox includes a default 22-channel ACE speech processor map. Map parameters within this default map were modified as follows. Channel stimulation rate was set to 900 pulses per second, with the number of maxima set to 8. The latter was true for all maps with 8 or more active electrodes. For maps with less than 8 electrodes, the number of maxima was set to the number of active electrodes. Threshold and Comfort levels were set to 100 and 200 current level units, respectively. The frequency allocation table (FAT) was set to the default range of 188 – 7938 Hz. This frequency range was fixed regardless of the number of active electrodes. The specific frequency range allocated to each electrode was automatically determined by the number of emulated active electrodes (i.e. number of bands), which was systematically varied from 1 through 22 electrodes. The resulting FATs as a function of the number of active electrodes were confirmed to be equivalent to those found in the clinical fitting software, i.e. Custom Sound 4.2 (Cochlear Ltd, Lane Cove, Australia). For a given number of active electrodes, one can also specify the set of active electrodes by modifying the ‘electrodes’ parameter. With these

map settings, one can apply processing to any sound file (.wav format) and produce electrodograms. To ensure accurate processing, it is necessary to ensure sound files are resampled at 16 kHz. To ensure that emulated output matches that typically obtained from a clinical speech processor (which includes front-end pre-emphasis), the pre-emphasis function provided in NMT was applied with default values to vowel and consonant sound files before being processed with map settings. Examples of electrodograms of the vowel in “hid” and consonant in “ada” for two combinations of 11 active electrodes are shown in Figure 1.

The speech cues employed for vowels are the mean locations of stimulation along the electrode array in response to the first two formants (F1 and F2), averaged separately for the first and second halves of the vowel segment (Sagi and Svirsky, 2017). These four speech cues, i.e. mean locations of F1 and F2 in the nucleus and offglide portions of the vowel, were shown to be highly predictive of CI listener’s vowel identification percent correct scores when coupled with these listeners’ JNDs for these cues. Examples of F1 and F2 obtained from electrodograms are highlighted in the vowels of Figure 1 (circles and squares, respectively, in both panels A and B). The time points marking vowel segments within sound files were measured prior to electrodogram processing with Praat (Boersma, 2001). These time markers were used to isolate vowel segments in emulated electrodograms from which vowel formant cues were obtained. Details of how these cues are measured from electrodograms are provided in Sagi and Svirsky (2017).

The speech cues employed for consonant stimuli were the same as those used in Svirsky et al. (2011). They are five speech cues in total. Three comprise the mean locations of stimulation along the electrode array in response to the first three formants (F1, F2, and F3), averaged over the duration of the consonant segment including small portions of the preceding and following vowel /a/. These cues provide consonant place of articulation information. Another cue was the amplitude ratio of the first formant energy relative to the total energy in all three formants. This cue provides information about voicing in that voiced consonants have a relatively larger low-frequency amplitude ratio than unvoiced consonants. It also provides limited information about manner of articulation. That is, nasals, liquids and glides would have larger amplitude ratios than fricatives and affricates. The other cue employed was duration of a silent gap within the consonant segment. This cue provides information about manner of articulation (e.g. stop consonants vs other consonants) as well as voicing information (e.g. voiced vs. unvoiced stop consonants). When used with the MPI model, this combination of spectral, temporal, and amplitude cues was shown to account for many aspects of CI listeners’ individual patterns of consonant confusions and correlated strongly with these listeners’ transmission of voicing, place, and manner features (Svirsky et al., 2011).

Some of the details for obtaining these consonant speech cues from electrodograms differ from Svirsky et al. (2011). As in the previous study, formant values were obtained by evaluating the mean electrode location of stimulation energy (summed across the duration of the consonant segment) in the subset of electrodes that encode a particular formant. For example, if the range of acoustic F2 values for all consonant segments was between 1000 Hz and 2000 Hz, then the subset of electrodes to which these frequencies were allocated was

used to evaluate the mean F2 electrode location for each consonant from their electrodiagram. What differs in the present study are the frequency values used to determine the boundaries between formant ranges. These were determined by analyzing the consonant stimuli of the present study using Praat software. A different set of ranges were used for the fricative consonants /f/, /s/, and /ʃ/ because their formant frequency ranges tended to be higher than the other consonants. In particular, the F1/F2 and F2/F3 boundaries for these fricative consonants were 1620 Hz and 2830 Hz, respectively. For all other consonants, the F1/F2 and F2/F3 boundaries were 1100 Hz and 2050 Hz, respectively. Another difference in the present study was in the evaluation of the silent gap duration. In the present study, silent gap duration was defined as the largest duration within the consonant segment without stimulation pulses, when evaluated across all electrodes encoding frequencies above 600 Hz. Examples of formant, amplitude ratio, and silent gap duration measurements from electrodiagrams of the consonant in “ada” are highlighted in Figure 1 (panels C and D).

Evaluation of vowel and consonant cues from electrodiagrams was automated using MATLAB (MathWorks, Natick, USA). With each change in map parameters, whether a change in the number of active electrodes or in the combination of active electrodes, a new set of speech cues were evaluated from digitized stimuli using this automated routine. Each set of speech cues served as input to the MPI model, from which vowel and consonant percent correct scores were generated.

2.3.2. MPI implementation—In the MPI model, speech cues postulated to contribute to phoneme identification are combined to form a multidimensional perceptual space where measured speech cue values serve as coordinates that define the locations of phonemes within the space. Neighboring phonemes within that space are predicted to be confused more often than relatively more distant phonemes. A well-known example of this concept for vowel identification is that of Peterson and Barney (1952). Different vowels that were closest in a 2-dimensional plot of their average F1 and F2 values were most often confused. In the present study, the postulated speech cues used to model vowel and consonant identification were described in Section 2.4.1. A block-diagram schematic of the MPI model subcomponents is shown in Figure 2 for the generalized case of m-dimensions.

The MPI consists of two model subcomponents. In the internal noise model subcomponent (left side of Figure 2), simulated percepts are generated by adding noise to phoneme locations in the multidimensional perceptual space. The noise follows a Gaussian distribution centered about a given phoneme location. The magnitude of the noise along a given dimension is gauged by a listener’s discrimination ability (proportional to their just-noticeable-difference, or JND) for the speech cue represented by that dimension. Hence, more poorly discriminated speech cues (higher JND) tend to produce more ambiguous percepts that have a higher likelihood of being misclassified. In the internal noise model, the ‘JND parameters’ are the standard-deviations of the noise distribution that control the amount of perceptual noise along each speech cue dimension.

In the decision model subcomponent, percepts are classified based on the closest response center within the perceptual space. Response centers are best exemplars (centroids) of a listener’s representation of phoneme categories, stored in their long-term memory. If a

listener's response centers are unbiased, then they should coincide with the actual mean locations of phonemes within the space as defined by their associated speech cues. The latter are referred to as stimulus centers. In other words, response centers represent listener expectations and stimulus centers represent physical reality. It's possible for response centers to be biased, i.e. unequal to stimulus centers. This can occur when CI users do not adapt (or incompletely adapt) to their speech processor map settings (e.g. Sagi et al., 2010b; Svirsky et al., 2015). Alternatively, some CI users are capable of completely adapting to their speech processor settings and one can assume they have no response bias, i.e. equal stimulus and response centers (Sagi et al., 2010, 2017; Svirsky et al., 2011). In the present study, zero response bias was assumed, and response centers were calculated from the average speech cue values for each phoneme category. Multiple iterations of the MPI model subcomponents for all phonemes of interest results in a series of stimulus and response values that can be accumulated in a confusion matrix, from which one can calculate a percent correct score.

In previous applications of the MPI model, a JND parameter was assumed constant for its associated speech cue irrespective of the range of speech cue values. For example, JND for F2 was fixed even though F2 values could range over several electrodes, each with differing amounts of encoding abilities. This assumption is equivalent to saying that the JND for F2 is represented by the average encoding ability across those electrodes encoding F2. In the present study, a central question is the effect of removing electrodes on encoding of speech cues, and thereby on speech perception. We assume that the JND associated with a given speech cue results from some combination of the JNDs associated with all electrodes that are employed to represent that cue. The following formulation is intended primarily for relating JND for place of stimulation to place-based speech cues (e.g. location of mean formant energy along the array). It is assumed that the JND for a place-based speech cue is equal to a weighted sum of the JNDs for all electrodes, weighted by the inverse square distance between the speech cue location along the array and every electrode. In particular,

$$JND_S = \sqrt{\frac{\sum_E w_E \times JND_E^2}{\sum_E w_E}} \quad \text{where} \quad w_E = \frac{1}{(L_S - L_E)^2 + 1} \quad (1)$$

where JND_S represents the JND associated with speech cue S and JND_E represents the JND associated with electrode E. The equation relating JND_S and JND_E is a standard formulation for a weighted mean with weighting w_E , where L_S and L_E represent the locations of speech cue S and electrode E along the array in electrode units. That latter units are a simplified way of representing distance along the array without knowing the exact distance between electrodes in millimeters, which can vary along the array depending on manufacturer. Distance along the array is assumed to be a continuous variable. In Equation 1, the additional factor of 1 added to the denominator of the weighting factor is to ensure that weighting is equal to 1 when L_S and L_E are equal, and less than one otherwise. Equation 1 provides a unique JND value for each speech cue that may vary for each stimulus. This is in contrast to previous applications of the MPI model where JND values for speech cues were constant across stimuli. The major advantage of our current approach is that one can

evaluate how deactivating an electrode (with its associated JND) affects the JND for a given speech cue, and hence overall speech perception.

In the present study, Equation 1 was only applied to the formant speech cues in the modeling of vowels and consonants. That is, each formant speech cue measurement was assigned a unique JND value based on Equation 1 that depended on the combination of active electrodes and their associated JNDs. In the consonant modeling, the amplitude ratio and silent gap duration speech cues were assigned fixed JNDs throughout all simulations. The JND for amplitude ratio was 0.2, and the JND for silent gap duration was 10 ms. These values were selected based on Svirsky et al. (2011).

Using the MPI model, vowel and consonant scores were generated for every possible active electrode combination, for every JND randomization. For a given combination of active electrodes, a set of speech cue measurements were generated from electrodiagram emulation of vowel and consonant stimuli which were used as input to the MPI model. JND for formant speech cues were calculated with Equation 1 using these speech cue measurements and JNDs for each active electrode (as specified by the JND randomization). With these speech cue measurements and associated JND values, vowel and consonant confusion matrices were generated from which percent correct scores were calculated. The numbers of iterations for each percent correct score generated were 2500 per vowel and 1500 per consonant.

2.3.3. Monosyllabic word model—In the present study, word score predictions are generated from the vowel and consonant scores obtained with the MPI model using a relationship proposed by Rabinowitz et al. (1992). Following a power-law model proposed by Boothroyd and Nittrouer (1988), Rabinowitz et al. demonstrated how CI users' monosyllabic word recognition scores can be predicted based on their vowel and consonant recognition scores. For meaningless words comprising consonant-vowel-consonant (CNC) segments, one could expect that the probability of correct word recognition would be a product of the probabilities of correctly identifying each phoneme within that word independently. However, for meaningful CNC words, contextual relationships among phonemes could allow listeners to correctly identify a word after correctly perceiving less than the total number of phonemes within the word. Therefore, Rabinowitz et al. suggested a model where word scores are predicted by the product of phoneme scores raised to a power equal to the fraction of phonemes in the word necessary for correct identification. Using vowel, consonant, and word recognition scores from 20 users of older generation CI devices, they found that word scores predicted by this relationship correlated strongly ($r = 0.9$, $p < 0.001$) with observed word scores when using an exponent equal to 0.89 (i.e. 2.67 out of 3 phonemes).

For the present study, we used the same method as Rabinowitz et al., but based on a relatively newer dataset. Most of this data was previously published in Sagi et al. (2009, 2010b) and Svirsky et al. (2011) which consist of vowel, consonant, and word scores from many of the same CI subjects. It was found that the exponent reported by Rabinowitz et al. consistently underestimated observed CNC scores in our data set. Instead, we found that the same formula provided a very good fit to the data ($r = 0.9$, $p < 0.001$), when using an

exponent of 0.5 (equal to 1.5 phonemes out of 3) instead of the 0.89 reported by Rabinowitz et al.

Therefore, the formula used in the present study to generate word score predictions from vowel and consonant scores generated from the MPI model is the following:

$$W = (C \times V \times C)^{0.5} \quad (2)$$

Where W is the proportion of words correct, C is the proportion of consonants correct, and V is the proportion of vowels correct. In the present study, word scores were generated for every possible active electrode combination, for every JND randomization. In particular, Equation 2 was applied to the predicted vowel and consonant scores obtained for the same combination of active electrodes.

2.4. Group comparisons

From the very large number of generated phoneme and word scores, those obtained for four electrode-combination conditions were isolated and compared. One condition was the electrode combination that yielded the maximum possible percent correct score predicted by the model at each number of active electrodes when using training stimuli (“model-optimized”). Another condition consisted of phoneme and word score predictions obtained when removing the poorest encoding electrodes at each number of active electrodes (“best-electrodes”). This condition reflects the current approach used in the literature of evaluating encoding abilities of electrodes and deactivating the poorest ones, without considering how speech information becomes distributed across the remaining active electrodes. The third condition consisted of phoneme and word score predictions obtained when removing the best encoding electrodes at each number of active electrodes (“worst-electrodes”). This group represents a baseline control condition that should be consistently worse than the other two groups. The fourth condition consisted of phoneme and word score predictions with all electrodes active (i.e. “full-array”). The phoneme and word scores resulting from these electrode combination conditions were compared for each JND subject group (i.e. indicative of poorer, moderate, and better performing CI users) using 95% confidence intervals (i.e. ± 2 standard errors of the mean). The ranges of phoneme and word scores for the model-optimized and best-electrodes conditions, relative to the full-array condition, were also compared.

2.5. Model Validation

2.5.1. Novel Stimuli—To assess whether differences in phoneme and word scores across electrode combination conditions generalized beyond the specific training stimuli used for optimization (i.e. 9 vowels from one male talker, and 16 consonants from one female talker), phoneme predictions were generated using novel stimuli from talkers who were different than those who produced the training stimuli (Section 2.2). Predicted scores with novel stimuli were not obtained for all possible active-electrode combinations. Rather, they were only obtained for the full-array conditions and the numbers of active electrodes that produced maximum model-optimized scores with training stimuli for each JND subject

group. At these numbers of electrodes, predictions with novel stimuli were obtained with best-electrode and model-optimized (via training stimuli) electrode combinations.

Differences among active-electrode conditions (i.e. model-optimized, best-electrode, and full array) within each JND subject group were assessed using repeated-measures analysis of variance (RMANOVA), and follow up post-hoc comparisons. Standard RMANOVA and Holm-Sidak post-hoc tests were applied to data sets that passed tests of Normality. Otherwise, RMANOVA on ranks and post-hoc Tukey tests were applied. For post-hoc testing, significance was set for a family-wise alpha equal to 0.05.

2.5.2. ACE vs CIS—To assess whether differences among electrode-combination conditions were unique to the ACE processing strategy employed during CI emulation, model predictions were also generated for a limited data set using the CIS processing strategy. In the ACE processing strategy, a subset of active electrodes with maximum energy are selected for stimulation every FFT analysis cycle. In the present study, these number of maxima were fixed at 8 channels for 8 or more active electrodes (Section 2.3.1). With the CIS processing strategy, all active electrodes are selected for stimulation every FFT analysis cycle. The procedures for obtaining model predictions with CIS are nearly identical as those with ACE. The only exception is in emulation of electrograms from which speech cues were measured. Instead of using the default ACE map provided with NMT software, we implemented the default CIS map ensuring that the number of channels selected per cycle was equal to the number of active electrodes.

Comparison of model predictions with ACE vs CIS were done only for vowels, and only for two numbers of active electrodes corresponding to those that yielded maximum model-optimized vowel scores for the poorer and moderate JND subject groups. Using CIS and the vowel stimuli used for training (Section 2.2), predicted vowel scores were generated for all possible combinations of active electrodes in order to determine the maximum vowel score (i.e. model-optimized electrode combination for CIS). Predicted vowel scores corresponding to the best-electrode condition across subjects were also generated, and the differences between the two conditions was obtained (i.e. model-optimized scores minus best-electrode scores). At question is whether the difference between model-optimized and best-electrode conditions are the same when using ACE and CIS processing strategies. Paired t-tests, or signed-rank tests where normality test failed, were used to assess differences in model-optimized and best-electrodes scores between ACE and CIS generated predictions.

2.6. Cumulative d-prime analysis

One premise of the present study is that an electrode deactivation strategy can improve predicted speech scores if that strategy results in better discrimination of speech cues relative to the range over which those speech cues span. To test this premise cumulative d-prime analysis was applied to the modeling results of the present study. Cumulative d-prime (d'), also referred to as total sensitivity, is an index of discrimination across a range of values (Durlach and Braida, 1969; Braida and Durlach, 1972). It can also be described as the total number of JNDs across the range. One can estimate d' as the ratio of the range of interest divided by the average JND across the range (assuming JND is the change in stimulus

continuum corresponding to a d-prime of 1). For example, if the average JND for the F1 formant dimension for vowels was 0.5 electrode units, and F1 across all vowels spanned 2.5 electrodes, then $d' = 2.5/0.5 = 5$. In other words, the range of F1 values comprises 5 JNDs in total.

In the present study, d' was only calculated for the numbers of active electrodes that yielded maximum vowel and consonant scores in the model-optimized conditions across JND subject groups using training stimuli. At these numbers of active electrodes, d' was evaluated for model-optimized and best-electrode combinations for each simulated subject within each group. For a given electrode combination, d' was calculated as follows. For vowels, mean electrode locations of F1 and F2 were calculated over the entire vowel duration and Equation 1 was used to transform JNDs per active electrodes (as determined by a particular simulated subject's JND randomization) into JNDs per formant cue. The range for F1 was specified by the difference between minimum and maximum F1 electrode location values across all vowel tokens. d' for F1 was calculated as the ratio of the range of F1 divided by the average JND across F1 values. A similar approach was used to calculate d' for F2 and the two values of d' were added. For consonants, d' was calculated in a similar manner as that for vowels, summing d' across the F1, F2, and F3 speech cue dimensions. In order to maintain a similar scale for d' between vowels and consonants, summed values of d' were divided by the number of formant dimensions, i.e. 2 for vowels and 3 for consonants.

Correlation analysis was used to compare d' and predicted percent correct scores, separately for vowels and consonants. A high correlation would suggest that differences in predicted scores between model-optimized and best-electrode conditions are related to differences in d' . That is, performance with different active electrode conditions is largely determined by how an electrode combination affects JNDs for speech cues relative to the range of electrodes over which those speech cues span (i.e. total number of JNDs across the range)

3. RESULTS

3.1. Model Optimization with Training Stimuli

Model-optimized phoneme and word scores as a function of number of active electrodes for relatively poorer, moderate, and better JND subject groups are presented in Figure 3 (filled circles). Also plotted are predicted scores resulting when electrodes with poorest encoding JNDs are deactivated, i.e. best electrodes remaining for a given number of active electrodes (unfilled circles), as well as predicted scores resulting when electrodes with best encoding JNDs are deactivated, i.e. worst electrodes remaining for a given number of active electrodes (unfilled squares). Vowel, consonant, and word scores are in the top, middle, and bottom panels of Figure 3, respectively. Relatively poorer, moderate, and better JND subject groups (20 randomizations per group) are in left, middle, and right columns of Figure 3. Scores are plotted as means $\pm$ 2 standard errors. In each of the nine graphs in Figure 3, scores for each electrode combination condition (model-optimized, best-electrodes, and worst-electrodes) converge at the right-most point which is the mean score of the full array condition (22 active electrodes). This result occurs because there is only one possible model prediction for

a given JND randomization when all electrodes are active. In our model implementation, the full array condition produced the following average scores as a function of JND subject group (horizontal lines in Figure 3): 38% for vowels, 40% for consonants, and 25% for words in the poorer JND subject group; 61% for vowels, 53% for consonants, and 42% for words in the moderate JND subject group; and, 87% for vowels, 69% for consonants, and 64% for words in the better JND subject group.

As one might expect when including only the poorest encoding electrodes (worst-electrodes), scores consistently produced lower phoneme and word scores on average than either model-optimized or best-electrodes conditions, in every JND subject group. Furthermore, in every JND subject group, worst-electrodes scores tend to decrease monotonically as the number of active electrodes are reduced. The reason for including this condition was to provide a baseline to compare the other two electrode combination conditions, as well as to confirm the conceptual notion (a sanity check of sorts) that selecting the poorest encoding electrodes should produce lower phoneme and word scores as the number of active electrodes is decreased.

The opposite case, including best encoding electrodes, does not always produce the highest speech scores. In fact, the model-optimized predictions were consistently and significantly higher (based on confidence intervals) than best-electrodes scores for vowels, consonants, and words for many numbers of active electrodes. This result is not entirely surprising in that, when using training stimuli, model-optimized predictions represent the maximum possible prediction. Hence, by definition, the best-electrodes score could never be higher than the model-optimized score in the case of predictions obtained with training stimuli.

One motivation for studies exploring electrode deactivation is the potential for improving speech scores relative to the full array condition. Results for the worst-electrodes condition were as expected. Scores obtained when only retaining electrodes with worst encoding abilities (unfilled squares in Figure 3) never produced better scores than the full array condition. Results for the best-electrodes condition (unfilled circles) were consistent with the findings of several studies listed in the Introduction of the present study. Namely, removing the poorest encoding electrodes can produce improvements in speech scores, but with limited results. In our modeling, largest improvements for the best-electrodes condition were in vowel scores for poorer and moderate JND subject groups, where improvements were approximately 10% on average after deactivating nearly half the array (unfilled circles in upper left and upper middle panels of Figure 3). Improvements in vowel scores for the better JND subject group (upper right panel), and improvements in consonant and word scores in all JND subject groups (remaining panels in Figure 3) were modest on average (never more than a few percentage points) for the best-electrodes condition.

For the model-optimized condition (filled circles in Figure 3), improvements in scores relative to the full array were found at many possible numbers of active electrodes. For vowel scores, the largest improvements were found for poorer and moderate JND subject groups (upper left and upper middle panels in Figure 3) when deactivating approximately half the array (albeit using a different combination of electrodes than the best-electrodes condition). The sizes of these improvements were approximately 30% and 20% higher on

average than the full array for the poorer and moderate JND subject groups, respectively. For the better JND subject group, the model-optimized condition produced more modest improvements in vowel scores relative to the full array (upper right panel of Figure 3). For consonant scores, the largest improvements were found for moderate and better JND subject groups when using between 10 and 16 active electrodes (center middle and center right panels in Figure 3). In these cases, consonant scores were approximately 10% higher on average than consonant scores with the full array. For the poorer JND subject group, the model-optimized condition produced more modest improvements in consonant scores relative to the full array (center left panel in Figure 3). For word scores (bottom row), improvements were found for all JND subject groups. The model-optimized condition produced 10% improvement in word scores, on average, relative to the full array when using between 10 and 16 active electrodes (more towards 10 electrodes for the poorer JND subject group, i.e. lower left panel in Figure 3, and more towards 16 electrodes for the better JND subject group, i.e. lower right panel in Figure 3).

In addition to examining the average improvement in predicted performance achieved by the model-optimized and best-electrodes conditions over the full array (Figure 3), it is also informative to assess the range of improvement in scores across the 20 JND randomizations in each JND subject group. This is shown in Figure 4 where phoneme and word scores are displayed as box plots showing 10th, 25th, 50th, 75th, and 90th percentiles. For ease of presentation, only data for 9 through 21 active electrodes are included in Figure 4. Although the average improvements in speech scores over the full array for the best-electrodes condition was relatively modest at specific numbers of active electrodes (Figure 3), some JND randomizations yielded improvements in scores that exceeded twice the average improvement, and many JND randomizations resulted in decreased scores relative to the full array. This result is consistent with some of the inter-individual differences in studies that show improvement in speech scores after deactivating electrodes with poorer encoding abilities in some CI subjects, but not for others (see studies cited in the Introduction of the present study). In contrast, at numbers of active electrodes that produced the largest average improvements in speech scores for the model-optimized condition, positive increments in speech scores were achieved for all JND randomizations. Furthermore, some JND randomizations achieved more than twice the average improvements in speech scores. For example, the maximum improvements in word scores were around 25% in poorer and moderate JND subject groups, and around 20% in the better JND subject group.

3.2. Model Validation: Novel Stimuli

With training stimuli, the number of active electrodes that produced maximum vowel and consonant scores for poorer, moderate, and better JND subject groups were 11, 13, and 19 active electrodes, respectively, in the case of vowels and 11, 16, and 16 electrodes, respectively, in the case of consonants. The model-optimized and best-electrode combinations at those numbers of electrodes for each JND subject group obtained with training stimuli were used to produce phoneme predictions with novel stimuli (Section 2.2). The results of these predictions with novel stimuli are shown in the box-plots of Figure 5 (showing 10th, 25th, 50th, 75th, and 90th percentiles), which also includes full-array predictions with the novel stimuli.

Median vowel scores (Figure 5, top panel) for model-optimized, best-electrode, and full-array conditions were, respectively, as follows: 44%, 36%, and 33% for poorer JND subjects; 58%, 54%, and 51% for moderate JND subjects; and, 80%, 76%, and 72% for better JND subjects. For each JND subject group, vowel scores across electrode combination conditions were significantly different (RMANOVA, $p < 0.001$). In each case, model-optimized scores were significantly greater than best-electrode and full-array conditions (Holm-Sidak, $p < 0.05$). Best-electrode scores were significantly greater than full-array conditions in the poorer and better JND subject groups (Holm-Sidak, $p < 0.05$), but not in the moderate JND group.

Median consonant scores (Figure 5, bottom panel) for model-optimized, best-electrode, and full-array conditions were, respectively, as follows: 44%, 41%, and 40% for poorer JND subjects; 55%, 53%, and 50% for moderate JND subjects; and, 69%, 69%, and 65% for better JND subjects. For each JND subject group, consonant scores across electrode combination conditions were significantly different (RMANOVA on ranks for poorer JND group, $p < 0.001$, and parametric RMANOVA for moderate and better JND groups, $p < 0.001$). For all JND groups, model-optimized scores were significantly greater than the full array condition (Tukey test, $p < 0.05$, for poorer JND group; Holm-Sidak, $p < 0.05$, for moderate and better JND groups). Model-optimized scores were significantly better than best-electrode scores in the poorer JND group (Tukey test, $p < 0.05$) and in the moderate JND group (Holm-Sidak, $p < 0.05$), but not in the better JND group. Best-electrode scores were significantly better than full-array scores in the moderate and better JND groups (Holm-Sidak, $p < 0.05$), but not in the poorer JND group.

In summary, the trend where model-optimized scores were larger than best-electrode and full-array scores found with training stimuli generalized to the novel stimuli, although to a lesser extent particularly for novel consonants. It is important to note that model-optimized scores in Figure 5 obtained with novel stimuli do not represent the maximum possible predicted scores because no optimization was done with novel stimuli.

3.3. Model Validation: ACE vs CIS

Maximum vowel scores with training stimuli and the ACE processing strategy were obtained with 11 electrodes for the poorer JND subject group, and with 13 electrodes for the moderate JND subject group. Model-optimized and best-electrode vowel predictions with training stimuli were also obtained at these numbers of active electrodes using the CIS processing strategy. Difference in percent scores between model-optimized and best-electrode conditions (i.e. model-optimized score minus best-electrode score) are shown in the box-plots of Figure 6 (showing 10th, 25th, 50th, 75th, and 90th percentiles) for ACE and CIS processing strategies. Model-optimized vowel scores were significantly larger than best-electrode scores for both ACE and CIS processing strategies, in poorer and moderate JND groups (paired t-test, $p < 0.001$). The median increase in vowel scores for model-optimized over best-electrode conditions for ACE and CIS, respectively, were as follows: 19% and 16% for the poorer JND group; and, 13% and 12% for the moderate JND group. Improvements with the model-optimized condition over the best-electrode conditions were significantly larger with the ACE strategy than with the CIS strategy (Signed-rank test, $p <$

0.001, for the poorer JND group; paired t-test, $p < 0.001$, for the moderate JND group), but the size of the effect was not large. More generally, the finding that model-optimized scores were significantly larger than best-electrode scores was not unique to the ACE processing strategy, it also generalized to the CIS processing strategy.

3.4. Cumulative d-prime Analysis

Cumulative d-prime (d') analysis was applied to the best-electrode and model-optimized electrode combinations at the numbers of electrodes that yielded maximum model-optimized phoneme scores for poorer, moderate, and better JND subject groups (i.e. 11, 13, and 19 active electrodes, respectively, for vowels; and, 11, 11, and 16 active electrodes, respectively, for consonants). Figure 7 shows predicted vowel and consonant scores (top and bottom panel) and associated d' for best-electrode and model-optimized conditions (unfilled and filled circles) at these numbers of active electrodes. For vowels, d' values were summed for F1 and F2 dimensions, and then divided by 2. For consonants, d' values were summed for F1, F2, and F3 dimensions, and then divided by 3. For both vowels and consonants, predicted scores and associated d' values are highly correlated ($r = 0.915$, $p < 0.001$ for vowels; $r = 0.951$, $p < 0.001$ for consonants). This result means that differences between model-optimized and best-electrode scores are largely related to differences in d' that these electrode combinations produce. That is, electrode combinations that provide more JNDs for speech cues over their respective ranges tend to produce higher phoneme scores.

4. DISCUSSION

Although a numeric exercise, the present study provides valuable information that is impossible to obtain in a clinic, or through experimental testing alone. After examining hundreds of millions of generated speech scores, the present study demonstrated that there may be untapped potential for improving speech scores in CI users by appropriate deactivation of electrodes. By relating encoding abilities of individual electrodes to how speech information becomes distributed over electrodes, the present study found consistent improvements in speech scores in comparison to deactivating only the poorest encoding electrodes. That is, optimal speech scores were not typically obtained by only including electrodes with best encoding abilities (i.e. lowest JNDs). This result was true for different levels of encoding abilities, representative of poorer to relatively better performing CI users.

There are several caveats to the modeling used in the present study. Encoding ability of electrodes was represented as JND for place of stimulation along the electrode array (e.g. electrode discrimination) so the reported trends apply primarily to deactivating electrodes based on place coding abilities, in contrast to temporal encoding abilities. In particular, JNDs associated with each electrode affected the JND for location of mean formant energy along the array, but not the other speech cue dimensions. Another caveat is that speech cue measurements were obtained from a relatively small sample of vowel and consonant recordings. A larger speech corpus may improve generalizability of results. Additionally, when examining all possible active electrode combinations, it was assumed that CI users are capable of completely adapting to the speech information presented to every combination of active electrodes. When this is not the case, predicted improvements may overestimate

observed outcomes. Another caveat is that reported predictions in the present study were obtained using JND randomizations representative of individual CI subjects. The values of JND for each electrode were independent and uncorrelated across electrodes. Some CI users may have electrode encoding abilities that are correlated to cochlear regions of the array. Furthermore, the range of JND values used did not include a range indicative of better than average CI users. Indeed, the better JND subject group produced average word scores of 64%, which represents average performance in users of present-day CI devices (Holden et al., 2013; Kraaijenga et al., 2016). Lastly, speech score predictions were based on CI processing with a 22-electrode Nucleus device.

These caveats notwithstanding, there were several trends in the present study that may be more generalizable to electrode deactivation studies. For example, although predictions were obtained using a small sample of vowel and consonant recordings (training stimuli), the general trend of the results generalized to novel stimuli. Model-optimized electrode combinations achieved with training stimuli tended to produce larger scores than either the best-electrode or full-array conditions when using novel stimuli. This was true even though the novel stimuli were from different speakers of different genders than those used for training, and even though no optimization was carried out with the novel stimuli.

Furthermore, the results of the present study were not specific to processing strategy and produced similar trends when using either the ACE or CIS strategies. This result contrasts with Vickers et al. (2016) who found no improvements in speech scores when selectively deactivating poorest encoding electrodes based on pitch ranking in users of the ACE strategy. They suggested the lack of improvement may have been due to the ACE strategy. In contrast our results suggest that the lack of improvement in their study may have been because their best-electrode condition may not have been the optimal selection, irrespective of stimulation strategy. Indeed, we found the difference between model-optimized and best-electrode conditions to be larger with the ACE strategy than with the CIS strategy. This may be because ACE magnifies the difference between important and unimportant channels. However, the size of the difference in speech scores between ACE and CIS, although significant, was small (a few percentage points at best).

More generally, the results of the present study suggest that poorer performing CI subjects may benefit from maps with relatively fewer active electrodes, and better performing CI subjects may benefit from relatively more active electrodes. In our projections, maximum improvements were found with 11 out of 22 active electrodes for the poorer JND subject group, and with 16 to 19 out of 22 active electrodes for the better JND subject group. For the moderate JND subject group (i.e. less than average performing CI users), similar amounts of improvement were found using 11 through 16 active electrodes. There also may be a tradeoff between improvements in vowels versus consonants for poorer versus better performing CI users. In our model-optimized projections, the poorer JND subject group displayed greater improvements in vowel scores, but relatively modest improvements in consonant scores. In contrast, the better JND subject group displayed relatively modest improvements in vowel scores, but greater improvements in consonant scores. However, this tradeoff balanced out in words scores, for which the model-optimized projections produced relatively equal improvements for all JND subject groups.

So why was it that better predicted speech scores were obtained by choosing a set of active electrodes that included at least some relatively poorer encoding electrodes, in comparison to a set of active electrodes that included only the best encoding electrodes? One reason has to do with how encoding abilities at each electrode (i.e. the JND associated with each electrode) relate to encoding of relevant speech cues, e.g. JND for formants (i.e. Equation 1). Another reason has to do with the physical span of electrodes over which speech cues vary. That is, deactivating an electrode may inadvertently reduce the cochlear distance available to encode place-based cues like formants. The ratio of these two values, i.e. range of electrodes for a formant cue and the corresponding JND for this formant, gives the total number of JNDs, i.e. cumulative d-prime (d'), available for this formant cue. We propose that deactivating electrodes that maximize the number of JNDs available for speech cues over their respective ranges will tend to yield higher speech scores (Figure 7).

To buttress the importance of maintaining the physical span of electrodes when selecting an optimal set of active electrodes, it's interesting to examine the model-optimized electrode combinations with 9 or more active electrodes that produced maximum phoneme and word scores in the present study. In particular, the proportion of these electrode combinations that included the most apical and basal electrodes (a span of 21 electrode distances) was 83% for the poorer JND subject group, 90% for the moderate JND subject group, and 95% for the better JND subject group. All of the remaining combinations retained a span of 18 (out of 21) electrode distances or more.

The present study has implications for individualized fitting of cochlear implants. The results of the present study suggest untapped potential for improving CI users' speech scores with the appropriate selection of active electrodes. A common theme in other electrode deactivation studies is evaluating the encoding abilities of individual electrodes, and then deactivating the poorest encoding electrodes along the array (or subsections thereof). Beyond logical guessing, there is no clear guideline as to which, and how many, electrodes need to be deactivated in order to achieve optimal results. The approach of the present study can be paired with these efforts to isolate an optimal combination of active electrodes best suited for an individual CI user. That is, one would evaluate electrode encoding abilities of a CI user using one or more of the approaches recommended by other studies and use the present modeling approach, or cumulative d-prime analysis of speech cues, to find the best combination of active electrodes. Future efforts include prospective studies with actual CI users, relating the JND measures of the present study to electrode encoding abilities as measured by other studies, as well as improving the accuracy of model predictions using a more robust corpus of vowel and consonant stimuli, and by using emulation of CI processing for other device manufacturers.

ACKNOWLEDGEMENTS

This research was supported by a grant from NIH/NIDCD grant number R01-DC03937 (PI: Mario Svirsky), and by a research contract from Cochlear Americas (PI: J. Thomas Roland Jr.). This work has utilized computing resources at the High-Performance Computing Facility at NYU Langone Medical Center.

REFERENCES

- Blamey P, Artieres F, Baskent D, Bergeron F, Beynon A, Burke E, ..., and Lazard DS (2013). "Factors affecting auditory performance of postlingually deaf adults using cochlear implants: an update with 2251 patients." *Audiology and Neurotology*, 18, 36–47. [PubMed: 23095305]
- Bierer JA and Litvak L (2016). "Cochlear implant programming may improve speech perception: current focusing and channel deactivation." *Trends in hearing*, 20, 1–12.
- Boersma P (2001). "Praat, a system for doing phonetics by computer." *Glott International*, 5, 341–345.
- Boothroyd A and Nitttrouer S (1988). "Mathematical treatment of context effects in phoneme and word recognition." *Journal of the Acoustical Society of America*, 84, 101–114. [PubMed: 3411038]
- Bosen AK, and Chatterjee M (2016). "Band importance functions of listeners with cochlear implants using clinical maps." *Journal of the Acoustical Society of America*, 140, 3718–3727.
- Braida LD, and Durlach NI (1972). "Intensity perception. II. Resolution in one-interval paradigms," *Journal of the Acoustical Society of America* 51, 483–502.
- Cochlear Corporation and the University of Iowa (1995). "Cochlear Corporation/the University of Iowa Revised Cochlear Implant Test Battery," Englewood, Colo, USA.
- Collins LM, Zwolan TA, and Wakefield GH (1997). "Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects." *Journal of the Acoustical Society of America*, 101, 440–455.
- Debruyne JA, Francart T, Janssen MLJ, Douma K and Brokx JPL (2017). "Fitting prelingual deafened adult cochlear implant users based on electrode discrimination performance." *International journal of audiology*, 56, 174–185. [PubMed: 27758152]
- Durlach NI, and Braida LD (1969). "Intensity perception. I. Preliminary theory of intensity resolution," *Journal of the Acoustical Society of America* 46, 372–383. [PubMed: 5804107]
- Donaldson GS and Nelson DA (2000). "Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies." *Journal of the Acoustical Society of America*, 107, 1645–1658. [PubMed: 10738818]
- Garadat SN, Zwolan TA and Pfingst BE (2013). "Using temporal modulation sensitivity to select stimulation sites for processor maps in cochlear implant listeners." *Audiology and Neurotology*, 18, 247–260. [PubMed: 23881208]
- Govaerts PJ, Vaerenberg B, De Ceulaer G, Daemers K, De Beukelaer C, and Schauwers K (2010). "Development of a software tool using deterministic logic for the optimization of cochlear implant processor programming." *Otology & Neurotology*, 31, 908–918. [PubMed: 20418791]
- Healy EW, Yoho SE and Apoux F (2013). "B and importance for sentences and words reexamined." *Journal of the Acoustical Society of America*, 133, 463–473. [PubMed: 23297918]
- Henshall KR and McKay CM (2001). "Optimizing electrode and filter selection in cochlear implant speech processor maps." *Journal of the American Academy of Audiology*, 12, 478 – 489.
- Hillenbrand J, Getty LA, Clark MJ, and Wheeler K (1995). "Acoustic characteristics of American English vowels," *Journal of the Acoustical Society of America* 97, 3099–3111. [PubMed: 7759650]
- Holden LK, Finely CC, Firszt JB, Holden TA, Brenner C, Potts LG, ... Skinner MW (2013). "Factors affecting open-set word recognition in adults with cochlear implants." *Ear and Hearing*, 34(3), 342–360. [PubMed: 23348845]
- Kenway B, Tam YC, Zebunnisa V, Harris F, Gray R, Birchall J, Carlyon R, and Axon P (2015). "Pitch discrimination: an independent factor in cochlear implant performance outcomes." *Otology and Neurotology*, 36, 1472–1479. [PubMed: 26375968]
- Kraaijenga VJC, Smit AL, Stegeman I, Smilde JJM, van Zanten GA, and Grolman W (2016). "Factors that influence outcomes in cochlear implantation in adults, based on patient-related characteristics – a retrospective study." *Clinical Otolaryngology*, 41, 585–592. [PubMed: 26493690]
- Labadie RF, Noble JH, Hedley-Williams AJ, Sunderhaus LW, Dawant BM, and Gifford RH (2016). "Results of postoperative, CT-based, electrode deactivation on hearing in prelingually deafened adult cochlear implant recipients." *Otology and Neurotology*, 37, 137–145. [PubMed: 26719955]

- Lee S and Mendel LL (2017). "Derivation of frequency importance functions for the AzBio sentences." *Journal of the Acoustical Society of America*, 142, 3416–3427. [PubMed: 29289102]
- Noble JH, Gifford RH, Hedley-Williams AJ, Dawant BM and Labadie RF (2014). "Clinical evaluation of an image guided cochlear implant programming strategy." *Audiology and Neurotology*, 19, 400–411. [PubMed: 25402603]
- Peterson GE, and Barney HL (1952). "Control methods used in a study of the vowels," *Journal of the Acoustical Society of America* 24, 175–184.
- Rabinowitz WM, Eddington DK, Delhorne LA and Cuneo PA (1992). "Relations among different measures of speech reception in subjects using a cochlear implant." *Journal of the Acoustical Society of America*, 92, 1869–1881. [PubMed: 1401531]
- Sagi E, Fu Q-J, Galvin JJ, III, & Svirsky MA (2010). "A model of incomplete adaptation to a severely shifted frequency-to-electrode mapping by cochlear implant users," *Journal of the Association for Research in Otolaryngology* 11, 69–78. [PubMed: 19774412]
- Sagi E, Kaiser AR, Meyer TA, and Svirsky MA (2009). "The effect of temporal gap identification on speech perception by users of cochlear implants," *Journal of Speech, Language, and Hearing Research* 52, 385–395.
- Sagi E, Meyer TA, Kaiser AR, Teoh SW, & Svirsky MA (2010b). "A mathematical model of vowel identification by users of cochlear implants," *Journal of the Acoustical Society of America* 127, 1069–1083. [PubMed: 20136228]
- Sagi E and Svirsky MA (2017). "Contribution of formant frequency information to vowel perception in steady-state noise by cochlear implant users." *Journal of the Acoustical Society of America*, 141, 1027–1038. [PubMed: 28253672]
- Saleh SM, Saeed SR, Meerton L, Moore DR and Vickers DA (2013). "Clinical use of electrode differentiation to enhance programming of cochlear implants." *Cochlear Implants International*, 14, 16–18.
- Svirsky MA (2000). "Mathematical modeling of vowel perception by users of analog multichannel cochlear implants: temporal and channel-amplitude cues," *Journal of the Acoustical Society of America* 107, 1521–1529. [PubMed: 10738806]
- Svirsky MA (2002). "The multidimensional phone identification (MPI) model: a new quantitative framework to explain the perception of speech sounds by cochlear implant users," in: *Etudes et Travaux Vol. 5*, edited by Serniclaes W (Institut de Phonétique et des Langues Vivantes of the ULB, Brussels), pp. 143–186.
- Svirsky MA, Fitzgerald MB, Sagi E, & Glassman EK (2015). "Bilateral cochlear implants with large asymmetries in electrode insertion depth: implications for the study of auditory plasticity," *Acta Oto-Laryngologica* 135, 354–363. [PubMed: 25719506]
- Svirsky MA, Sagi E, Meyer TA, Kaiser AR, & Teoh SW (2011). "A mathematical model of medial consonant identification by cochlear implant users," *Journal of the Acoustical Society of America* 129, 2191–2200. [PubMed: 21476674]
- Tyler RS, Preece JP, and Lowder MW (1987). "The Iowa audiovisual speech perception laser video disk." *Laser Videodisk and Laboratory Report*, University of Iowa at Iowa City, Department of Otolaryngology-Head and Neck Surgery.
- Vickers DA, Degun A, Canas A, Stainsby T and Vanpoucke F (2016). "Deactivating Cochlear Implant Electrodes Based on Pitch Information for Users of the ACE Strategy." In: *Advances in Experimental Medicine and Biology*, Vol. 894, pp.115–123. New York: Springer.
- Zeitler DM, Lalwani AK, Roland JT, Jr., Habib MG, Gudis D, and Waltzman SB (2008). "The effects of cochlear implant electrode deactivation on speech perception and in predicting device failure." *Otology and Neurotology*, 30, 7–13.
- Zhou N (2016). "Monopolar detection thresholds predict spatial selectivity of neural excitation in cochlear implants: implications for speech recognition." *PLoS ONE*, 11(10), 1–18.
- Zhou N and Pfingst BE (2012). "Psychophysically based site selection coupled with dichotic stimulation improves speech recognition in noise with bilateral cochlear implants." *Journal of the Acoustical Society of America*, 132, 994–1008 [PubMed: 22894220]

Zwolan TA, Collins LM and Wakefield GH (1997). "Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects." *Journal of the Acoustical Society of America*, 102, 3673–3685. [PubMed: 9407659]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

- Deactivating poorest cochlear implant electrodes may not improve speech perception
- Computational approach examined millions of electrode deactivation combinations
- Model used also considers how speech information distributes along electrode array
- Model-optimized combinations improved predicted word scores 10% on average
- Improvement persisted in relatively poorer, moderate, and better performance groups

Figure 1:
Examples of electrodograms of the vowel in “hid” (panels A and B) and consonant in “ada” (C and D) for two different combinations of 11 active electrodes (i.e. “best-electrodes” in panels A and C, and “model-optimized” in panels B and D; see Discussion section for more details).

INTERNAL NOISE MODEL

Figure 2:

Schematic for an m-dimensional MPI model showing internal noise model and decision model subcomponents. Internal noise proportional to JNDs along each dimension is added to phoneme locations to generate “percepts” which are classified by the decision model based on the closest response center.

PREDICTED SCORES (± 2 standard errors)

Figure 3:

Mean predicted phoneme and word scores (± 2 standard errors) as a function of the number of active electrodes for “model-optimized” (filled circles), “best-electrodes” (unfilled circles), and “worst-electrodes” (squares) electrode combination conditions, and for poorer, moderate, and better JND subject groups (left, middle, and right columns). Horizontal line in each panel indicates predicted scores with full-array condition.

PREDICTED SCORES MINUS FULL ARRAY (range of scores)

Figure 4:

Range of phoneme and word scores (10th, 25th, median, 75th, and 90th percentiles) within each JND subject group (left, middle, and right columns). Scores are shown as a function of the number of active electrodes (between 9 and 21) minus scores with the full array for “model-optimized” (filled bars) and “best-electrodes” (unfilled bars) electrode combination conditions.

Figure 5:

Evaluation of active electrode conditions (model-optimized, best electrodes, and full array) using novel stimuli. Number of electrodes chosen for model-optimized and best electrodes conditions were those that yielded maximum vowel and consonant scores using training stimuli for poorer, moderate, and better JND subject groups.

Difference in Vowel Scores Model-optimized vs Best-electrodes

Figure 6:

Difference in predicted vowel scores between active electrode conditions (model-optimized less best-electrodes) using ACE and CIS processing strategies for poorer and moderate JND subject groups. Number of electrodes chosen were those that yielded maximum model-optimized scores using training stimuli.

Figure 7:

Predicted vowel and consonant scores vs cumulative d-prime measures for formant speech cues (F1F2 for vowels, and F1F2F3 for consonants) for model-optimized and best electrode conditions measured across simulated subject groups. Cumulative d-prime measures evaluated for each speech cue, summed across speech cues, and then divided by the number of speech cues, i.e. two for vowels, and three for consonants (Section 2.6).