IDENTIFYING IMPAIRED COCHLEAR IMPLANT CHANNELS VIA SPEECH-TOKEN CONFUSION MATRIX ANALYSIS

Jeremiah J. Remus and Leslie M. Collins

Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708
{jjr6, lcollins}@ee.duke.edu

ABSTRACT

Cochlear implant patients exhibit a wide range of performance on speech recognition tasks. One potential explanation for such variability is the existence of psychophysically observed phenomena that might indicate the presence of anomalous percepts associated with certain electrical stimuli, which in turn could limit the transmission of important auditory cues. Exhaustive psychophysical testing to detect all such psychophysical anomalies is time prohibitive; however, the search for anomalous channels could be expedited with prior information identifying channels potentially containing an anomaly. This study proposes a method of analyzing confusion matrices from speech token recognition tasks with the intent of identifying impaired channels. Results using both normal-hearing subjects tested with impaired acoustic models and cochlear implant subjects suggest that the proposed methods are providing information about the probability of impairment on each channel.

Index Terms— Cochlear Implants, Correlational Method, Confusion Matrix, Impairment, Hidden Markov Models

1. INTRODUCTION

Cochlear implants have been shown to successfully restore at least some sense of hearing in severely deafened individuals. However, performance for speech recognition tasks can vary significantly from patient to patient (e.g. [1]). One possible explanation for the variability in performance is the possible presence of channel-specific anomalies resulting in unexpected percepts associated with certain electrical stimuli, which could further indicate limitations in the auditory information capacity of different channels or electrodes in the cochlear implant. The presence of such impairments has been observed in numerous studies through psychophysical testing (e.g. [2]) suggesting that in some cochlear implant patients the transmitted stimulus does not result in the expected percept.

Several studies have linked psychophysically-observed impairments to speech recognition ability (e.g. electrode discrimination [3]), suggesting that the observed impairments deleteriously affect the transmission of cues important for speech perception. In one study of electrode discrimination and speech recognition, Zwolan et al. [3] also illustrated that tuning the presentation of auditory information based on psychophysically-observed impairments can potentially improve speech recognition, suggesting potential gains in speech recognition performance through psychophysically-inspired tuning of the cochlear implant speech processor parameters.

However, due to the number of electrodes in the cochlear implant array, the number of stimulus parameters, and interaction of different variables, exhaustive psychophysical testing is time prohibitive. This study proposes performing analysis on confusion matrices from tests of speech token recognition to gather information about the location of potential impaired channels. Such techniques would require minimal time investment to collect speech token recognition data. By focusing the search for impaired channels on channels identified through preliminary analysis to be most likely to contain impairments, it may be possible to expedite the identification of the impaired channels through directed testing.

The correlational method (Lutfi [4]) is the basis for the methods of confusion matrix analysis used in this study for measuring information about the presence of impaired channels. It has been used previously (e.g. [5]) to estimate channel and electrode weights in both cochlear implant subjects and normal-hearing listeners using acoustic models of cochlear implants. In this study, the correlational method is modified for the task of identifying impaired channels via confusion matrix analysis. Section 2 describes the correlational method and the necessary task of specifying how auditory cues are distributed across the channels. Section 3 discusses the two sets of listening experiment data, one from normal-hearing subjects using impaired acoustic models of cochlear implants and another using cochlear implant subject data, used to test the performance of the proposed techniques. Section 4 presents a discussion of the outcomes of the current study.

2. CORRELATIONAL METHOD FOR CONFUSION MATRIX ANALYSIS

The correlational method is a potentially useful technique in cochlear implant research for analyzing the information-bearing capacity of individual channels based on the results of a token identification task. In the implementation used by [5] for estimating channel weights, a token identification task was performed with each channel having a different signal-to-noise

Supported by NIH grant 1-R01-DC007994-01

ratio (SNR). The correlation between the percentage of tokens correctly identified and SNR was calculated for each channel. Intuitively, as SNR decreases on "more important" channels, performance should drop more significantly than when SNR is decreased on less important channels. Hence, a higher correlation between channel SNR and percentage of tokens correctly identified suggests that the subject is relying more heavily on that channel for token identification.

To apply the correlational method to the task of identifying potentially impaired channels, the correlation between the percentage of tokens correctly identified and the presence of cues is calculated for each channel. A channel *C* will contain varying amounts of information necessary for correctly identifying each token. If channel *C* is impaired, i.e. does not properly transmit the auditory cues necessary for token identification, then there may be an observed correlation between the amount of information channel *C* contains about token *T* and the correct identification rate for token *T*. Specifically, it would be expected that the recognition rate of token *T* would decrease as the amount of information about token *T* contained on an impaired channel increases. Hence, higher correlations between the amount of token cue information and the identification rates of those tokens should identify channels more likely to contain limitations on presenting auditory cues.

Calculating the correlational method metrics *Y* for each channel requires a n_T by 1 vector *X* of token identification rates, equal to the diagonal of the confusion matrix, and a n_T by n_C classification matrix *M*, where n_C is the number of channels and entry (*T*,*C*) in *M* specifies the presence or amount of information the C^{th} channel contains about the T^{th} token. Letting M_C equal the C^{th} column of *M*, a linear regression can be calculated between M_C and *X*. To characterize the relationship between information content and token identification rates, the metrics are calculated as the slope coefficient of the linear regression using Equation 1. Higher calculated metrics indicate a stronger relationship between the presence of information and decreasing token recognition rates, suggesting that the channel is more likely to contain an impairment to transmitting auditory cues.

$$Y(C) = \frac{n_T X^T M_c - \left(\sum X\right)\left(\sum M_c\right)}{n_T X^T X - \left(\sum X\right)\left(\sum X\right)} \qquad (1)$$

2.1. Measures of Channel-Specific Information

An important step in implementing the correlational method for confusion matrix analysis to identify impaired channels is specifying the amount of token cue information contained on each channel. Two methods for defining the token cues were considered in this study to specify the presence/quantity of token cue information on each channel in the form of the classification matrix. The first method utilized vowel token formants as the token discriminating cues, and the vowel formant frequencies were used to construct the classification matrix. For each token *T*, the channels containing F1 and F2 were labeled important and all others were considered to not contribute to identification of that token. Thus there are only two levels of information defined in this case for a channel to contain: zero or one.

The second method used a hidden Markov model to calculate a quantitative measure of the information for each token. The hidden Markov model (HMM) was trained using speech tokens processed through an acoustic model of a cochlear implant. The log-likelihood was calculated for the HMM producing a speech token processed through an *impaired* acoustic model of a cochlear implant; thus it provides a quantitative measure of the distance between the HMM of the token and an impaired instance of that same speech token. An n_T by n_T matrix of log-likelihoods was constructed, one column for each HMM and one row for each token. A total of n_C matrices of log-likelihoods were constructed, one matrix for impairments modeled on each of the n_C different channels. Each column of the classification matrix *M* can be constructed from one log-likelihood matrix by averaging the off-diagonal values in each row. The average of the off-diagonal values corresponds to the distance to potential incorrect responses. Higher average distance implies greater discrimination of the token given the information contained on that channel; hence the average off-diagonal average provides an estimate of the amount of information contained on each channel that is useful for discriminating each token.

3. PERFORMANCE OF CONFUSION MATRIX ANALYSIS

The performance of the proposed application of the correlational method to identifying impaired channels was tested on both normal-hearing subjects using a cochlear implant acoustic model and cochlear implant subject data. Each data set has certain benefits. Using acoustic models provides significant control over the experiment, allowing the implementation of a specific impairment and avoiding patient-specific factors that can influence performance. However, the variability intrinsic to cochlear implant subjects (e.g. number of electrodes in array, speech recognition ability, psychophysical and physiological variation) is a useful tool for gauging the robustness of any proposed technique to the realities of clinical testing. Therefore, testing using both sets of data provides an appropriate set of test cases for both preliminary investigation and more rigorous evaluation.

3.1. Data sets

Two sets of data were analyzed to evaluate the proposed confusion matrix analysis technique based on the correlational method. The first set of data was collected using twenty-four normal-hearing subjects from the population of students and staff at Duke University. Normal-hearing subjects listened to tokens processed using an eight-channel acoustic model based on the CIS cochlear implant speech processor [6]. The acoustic model mimics the temporal and spectral information presented by the CIS speech processor. Subjects were tested using nine vowel tokens (had, hawed, head, heard, heed, hid, hood, hud, who'd) and fourteen consonant tokens (b, d, f, g, j, k, m, n, p, s, sh, t, v, z) presented in /aCa/ context. Subjects were trained on two randomly ordered repetitions of the set of tokens, and tested on an additional five repetitions of the token set.

Eight impaired acoustic models were created by removing individual channels, an impairment implemented by setting the envelope of the signal on a single channel to zero. Subjects were tested using four of the eight impaired models, as well as the baseline unimpaired acoustic model. The impaired models are identified as imp$I$ with I the channel number (1 through 8) impaired, i.e. removed.

The second set of data came from the study by Zwolan et al. [3] investigating the relationship between electrode discrimination and speech recognition. From the data collected by Zwolan et al., vowel and consonant confusion matrices and electrode discrimination data were obtained for six subjects. Subjects were tested with five repetitions of the same nine vowels and fourteen consonants used in the normal-hearing subject study. Four of the subjects (S2, S3, S4, and S5) improved on a majority of speech recognition tasks when tested with an experimental map that removed indiscriminable electrodes. Thus, it can be inferred that these impaired electrodes have a deleterious effect on speech recognition. Additional experimental details are available in [3].

3.1. Results

The metric used in this study to quantify the information about the presence of impaired channels measured by the proposed analyses was the percentage of possible false alarms ($\%_{FA}$), i.e. the percentage of non-impaired channels incorrectly identified as impaired channels. This was determined to be a more useful metric than other measures such as percent error since it illustrates when the methods are not performing perfectly but are still better than chance.

Fig. 1 shows the average $\%_{FA}$ for the analysis of the normal-hearing confusion matrices. Since the impaired acoustic models contained a single impaired channel, these results correspond to the false alarm rate at 100% probability of detection. Results are plotted averaged across all subjects for each impaired model, and lower values indicate better performance. Using the vowel formant frequencies in the correlational method restricted testing for impairments to channels containing formant frequencies. In the eight-channel acoustic model used in this study, channels 1, 7, and 8 are absent any formants, thus the data set is reduced to cases testing for impairments on channels 2-6 and no results are shown for the correlational method using formants on channels 1, 7, and 8. The number of possible false alarms is also reduced (from seven to four). Since the motivation for the proposed technique is to analyze token identification rates to identify deleterious impairments, the results in Fig. 1 only include data from subjects whose token identification rates fall below perfect performance (100% correct) by a statistically significant margin ($p < 0.05$).

Results show that performance using both measures of token cue information is best for impairments on channels 3 through 6. This coincides with the region of the frequency spectrum containing the most critical speech information, so it

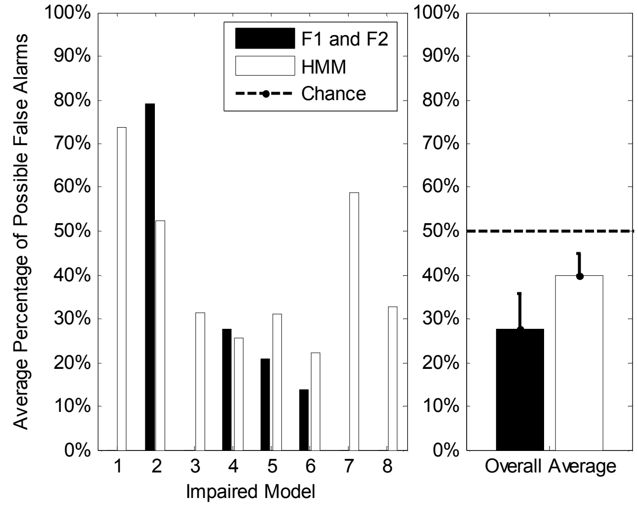

Fig. 1. Average percentage of possible false alarms ($\%_{FA}$) for each impaired model (left) and averaged across all models (right) for the correlational method using vowel formants (black bars) and the HMM (white bars). Chance level performance is 50%; error bars on overall average performance are 95% confidence intervals. Note: Average $\%_{FA}$ for imp3 with F1 and F2 equals 0%.

may be easier to quantify the token cues. The HMM also had some success for impairments on channel 8, likely making use of fricative consonant discrimination. The results averaged across all models and subjects are shown on the right. Error bars indicate 95% confidence levels and the dashed line at 50% indicates chance level performance. The correlational method using both definitions of token cue information clearly outperforms chance by a statistically significant margin.

The cochlear implant subject results were analyzed using the same metric ($\%_{FA}$). However, the number of indiscriminable electrodes varied from subject to subject (S2, S3, S4, S5, S6, and S7 had one, one, seven, three, four, and seven indiscriminable electrodes, respectively). Therefore, it is appropriate to consider $\%_{FA}$ for two separate circumstances. In the first case, $\%_{FA}$ was calculated for identifying a single impaired channel (if the patient has C_{IMP} impaired channels, this equates to performance at a probability of detection equaling $1/C_{IMP}$). This measure of performance corresponds to the process of identifying impairments being considered sequentially, where remediation is performed first on the most dominant impairment. The second case for measuring performance was $\%_{FA}$ to find all impaired channels, i.e. 100% probability of detection. This is analogous to the electrode discrimination task performed by Zwolan et al. However, if information about the set of indiscriminable electrodes can be provided through analysis of the results of a token identification task it would be significantly faster than psychophysical testing.

Fig. 2 shows the results for analysis of the cochlear implant subject data. The top row of plots is $\%_{FA}$ for each subject for the first case (finding the first indiscriminable electrode) for both measures of token cue information. The bottom row of plots is for the second case (finding all

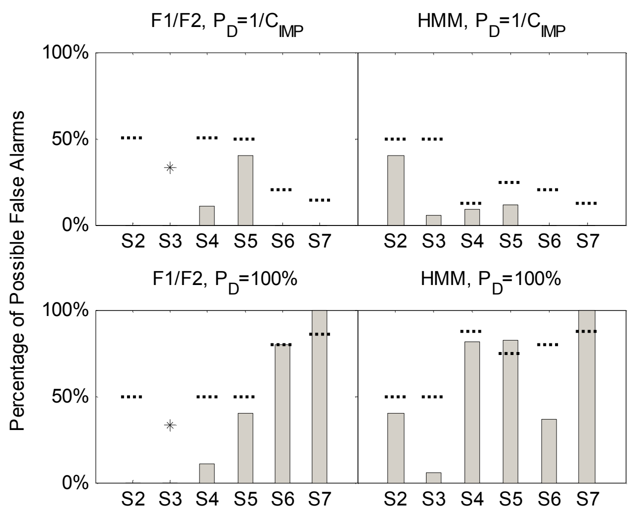

Fig. 2. Percentage of possible false alarms ($\%_{FA}$) for six cochlear implant subjects. The left column plots use the vowel formants as token cues; right column plots use the HMM. The top row of plots show $\%_{FA}$ for finding the first indiscriminable electrode; the bottom row shows $\%_{FA}$ for finding all indiscriminable electrodes. Dashed lines are chance performance. Note: No vowel formants were present on the indiscriminable electrode for S3.

indiscriminable electrodes). The dashed lines for each subject indicate the chance level of performance, which is dependent on the number of indiscriminable electrodes, total number of electrodes, and probability of detection.

The level of performance using the two measures of token cue information is approximately equivalent, based on the number of subjects with low, near chance, and above chance $\%_{FA}$. Using the vowel formants restricts testing for indiscriminable electrodes to channels containing formants; thus, usually only a subset of the indiscriminable electrodes were included in this task. For S3, the single indiscriminable electrode did not contain any formants and could not be tested. Therefore, one apparent benefit of the HMM measure is the ability to test for impairments on all electrodes.

4. DISCUSSION

This study investigated the application of the correlational method, previously used for estimating channel weights, to identifying impaired electrodes based on patterns of vowel and consonant token identification. Two measures of token cue information were considered: the vowel formant frequencies and an HMM-based distance metric. Results were generated using confusion matrices from both normal-hearing subjects tested with acoustic models and cochlear implant subjects. The results of this study suggest that the proposed confusion matrix analyses are providing information about the location of impaired channels.

This study considered vowel and consonant token confusions; however, previous studies (e.g. [5]) used nonsense syllables for channel weight estimation. The ideal set of stimuli would contain channel-specific cues rich in temporal information that would measure the full information transmission capabilities of each channel. One benefit of using vowel tokens or other speech stimuli may be the ability to identify standard token cues such as formant frequencies. However, quantitative measures of information such as the HMM-based technique in this study may allow the use of tokens without requiring extensive knowledge contributed by the investigator.

Indeed, one goal of this study was the comparison of two methods for defining the importance of each channel for the identification of each token. Using vowel formants requires the investigator to input the vowel formant frequencies, whose identification can be a subjective process. The HMM-based measure automates the generation of the classification matrix, removing the investigator from the loop. For more complex stimuli, an automated method for defining the token cues may be particularly appealing.

It is also possible that the outcomes of the proposed analyses could be used as prior information in Bayesian adaptive psychometric procedures such as QUEST [7]. Psychophysical testing is typically conducted in a straightforward fashion that does not include prior information; however, the Bayesian framework of methods like QUEST could utilize informative priors. The results of this study, below chance by a statistically significant margin in the case of normal-hearing subjects, indicate that information about the presence of impairments is provided by the proposed confusion matrix analyses. Using that information in directed psychophysical testing or a more sophisticated Bayesian framework could potentially expedite the identification of channel-specific impairments.

5. REFERENCES

[1] R. C. Dowell, D. J. Mecklenburg, and G. M. Clark, "Speech Recognition for 40 Patients Receiving Multichannel Cochlear Implants," *Archives of Otolaryngology - Head & Neck Surgery*, vol. 112, pp. 1054-1059, 1986.

[2] R. V. Shannon, "Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction," *Hearing Research*, vol. 12, pp. 1-16, 1983.

[3] T. A. Zwolan, L. M. Collins, and G. H. Wakefield, "Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects," *Journal of the Acoustical Society of America*, vol. 102, pp. 3673-3685, 1997.

[4] R. A. Lutfi, "Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks," *Journal of the Acoustical Society of America*, vol. 97, pp. 1333-1334, 1995.

[5] M. A. Mehr, C. W. Turner, and A. Parkinson, "Channel weights for speech recognition in cochlear implant users," *Journal of the Acoustical Society of America*, vol. 109, pp. 359-366, 2001.

[6] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz, "Better speech recognition with cochlear implants," *Nature*, vol. 352, pp. 236-238, 1991.

[7] A. B. Watson and D. G. Pelli, "QUEST: A Bayesian adaptive psychometric method," *Perception & Psychophysics*, vol. 33, pp. 113-120, 1983.