IDENTIFYING CHANNEL-SPECIFIC IMPAIRMENTS IN COCHLEAR IMPLANT PATIENTS VIA PARTIAL LEAST SQUARES DISCRIMINANT ANALYSIS OF SPEECH-TOKEN CONFUSION MATRICES

Jeremiah J. Remus and Leslie M. Collins

1Dept. of Electrical and Computer Engineering
Clarkson University, Potsdam, NY 13699
jremus@clarkson.edu

2Dept. of Electrical and Computer Engineering
Duke University, Durham, NC 27708
lcollins@ee.duke.edu

ABSTRACT

It is not uncommon for cochlear implant patients to have individual electrodes that produce anomalous percepts that impair or prevent the effective transmission of auditory information. Exhaustive psychophysical testing to detect all such information channels is time prohibitive; however, if impaired channels could be identified quickly then the application of remediation strategies becomes more cost-effective. Previous studies have suggested that the missing speech information could produce a predictable pattern of errors in speech-token identification tasks [1, 2]. This study investigates the application of partial least squares discriminant analysis to identifying the presence of channel-specific impairments based on confusion matrices generated from vowel and consonant token identification tasks. The results of this study, using normal-hearing subjects tested with acoustic models, suggest that the partial least squares discriminant analysis can successfully distinguish impaired and unimpaired models, as well as identify channel-specific impairments, without requiring a significant amount of labeled training data.

Index Terms— Cochlear implants, Partial least squares discriminant analysis, Confusion matrix, Impairment

1. INTRODUCTION

Cochlear implants restore hearing in severely deafened individuals; however, the degree of benefit provided by cochlear implants can vary greatly on an individual basis. Some patients perform very well in post-operative tests of speech perception, while others do not achieve the same level of benefit from their cochlear implant [3]. One possible contributor to the poorer performance of some implant recipients could be the presence of channel-specific anomalies that cause abnormal percepts for certain stimuli, thus preventing some channels from transmitting all possible speech cues. The existence of such psychophysically-observed anomalies has been documented in numerous studies (e.g. [4, 5]), suggesting that for some implant patients, certain stimuli do not produce the expected percept. Several studies have also linked these psychophysical anomalies to the speech perception ability of cochlear implant patients [5]. These results provide support for the hypothesis that such channel-specific impairments can deleteriously impact speech perception by preventing effective transmission of the necessary speech cues. Thus, the term "channel-specific anomalies" is used in this study as a general reference to any physiological or perceptual phenomenon that 1) can be measured psychophysically and 2) may interfere with the transmission of speech cues.

Previous studies have shown that patient-specific tuning based on psychophysical data can improve speech recognition scores for some individuals (e.g. [5]). These results motivate the identification of channel-specific impairments in individual patients, allowing for tuning and compensation based on patient-specific psychophysical data. However, tuning based on psychophysics requires a time-consuming collection of psychophysical data that is impractical in a clinical setting. The number of channels in the implant, the number of stimulus parameters, and potential interaction of different variables renders a comprehensive data collection time-prohibitive.

Several studies have considered methods to expedite the collection of the information necessary for tuning the implant via analysis of speech-token confusion matrices [1, 2]. These techniques are of interest because they process speech recognition data that can be gathered more expeditiously. The previous studies by the authors have investigated techniques using formant frequencies and model-generated features for predicting confusions under certain impaired conditions, and using these models to identify impairments in experimental data. These techniques show promise for identifying impaired channels, and suggest that the patterns of token confusions may provide information about missing channels due to the poor transmission of discriminating cues.

There are a significant number of classification algorithms in the pattern recognition and machine learning literature; however, these techniques have not yet been easily transitioned to speech-based confusion matrix analysis. This task suffers from the classic “large p, small n” problem: there is a limited amount of labeled training data and the potentially large number of input features, which presents a problem for many popular pattern recognition algorithms such as classification trees [6] and the support vector machine [7]. However, data

Supported by NIH grant 1-R01-DC007994-01
sets with similar characteristics (i.e. large p, small n) occur frequently in chemometrics. Partial least-squares discriminant analysis (PLSDA) [8] is a commonly-used technique in chemometrics for identifying compositional elements based on spectroscopy data. In this study, PLSDA has been used to identify channel-specific impairments based on speech-token confusion matrices generated from speech recognition studies performed by several normal-hearing subjects using acoustic models of cochlear implant speech processors. Section 2 provides some background and description of the PLSDA algorithm. Section 3 discusses the listening study and data set used to train the PLSDA algorithm for impaired channel identification. Sections 4 and 5 discuss the results and conclusions of the study.

2. PARTIAL LEAST SQUARES DISCRIMINANT ANALYSIS (PLSDA)

Partial least squares discriminant analysis (PLSDA) was originally developed within the chemometrics community, but has since been applied to a variety of topics including bioinformatics (e.g. [9]) and medical diagnosis (e.g. [10]). It is most frequently applied as a regression method to model the relationship between a set of independent variables (i.e. features) X and dependent variables Y. PLSDA is particularly appropriate for the confusion matrix analysis task; while there are many relevant algorithms in the pattern recognition literature, many of them either encounter numerical issues or over-fit the training data when applied to a “large p, small n” task that is considered here. For example, if the individual entries in the confusion matrix are used as inputs to the classification algorithm, it is possible to have hundreds of possible features, which exceeds the number of training samples available in this study by a factor of two. Due to the small size of the labeled training data set, it is also impractical to apply one of the many available feature selection techniques [11]. However, the PLSDA algorithm has characteristics that make it appropriate for the confusion matrix analysis task.

The first stage in PLSDA is a partial least squares decomposition, which performs a linear projection to a lower-dimensional subspace. This first stage allows the use of high-dimensional data sets without running into the large p, small n problem. One advantage of partial least squares over other linear subspaces projection methods such as principal component analysis is that partial least squares method utilizes the data labels. Therefore, the resulting lower-dimensional subspace is more likely to maintain separability between classes, by using a criterion that seeks linear projections w and q that maximize the covariance between the independent and dependent variables X and Y, respectively, in the lower-dimensional projection space.

$$\max_{w, q} \text{cov}(Xw, Yq)$$

(1)

This contrasts with PCA, which maximizes the variance of the data under the constraint of a unit-norm weight vector, ignoring any available class labels for the training data. Once the lower dimensional subspace has been found in PLSDA, a set of regression weights are calculated to explain the projected dependent variables using the projected independent variables. The PLS linear projection and regression weights can be combined into a single p by m weight matrix $B_{PLS}$:

$$\hat{Y}_0 = X_0 B_{PLS}$$

(2)

$$B_{PLS} = R \text{diag}(b) Q^T$$

where m is the number of classes, p is the dimensionality of the data (i.e. number of columns of $X_0$), $\hat{Y}_0$ are the estimated labels for centered test samples $X_0$, and the p by m matrix $B_{PLS}$ is calculated using the vector of regression weights b, the matrix R containing the loadings that project the centered data into the partial least squares subspace, and the matrix Q of weights that decompose the dependent variables.

3. EXPERIMENTAL METHOD AND SETUP

3.1. Listening study data collection

Listening studies were conducted using normal-hearing subjects recruited from the student and staff populations at Duke University. The test materials consisted on vowel and consonant tokens processed using an eight-channel acoustic model that mimics the temporal and spectral information presented by the CIS speech processor [12]. Thirty-five subjects were tested using nine vowel tokens (had, hawed, head, heard, heed, hid, hood, hud, who’d) and fourteen consonant tokens (b, d, f, 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. The distribution of token recognition scores were similar enough across the unimpaired and eight impaired models that the absolute level of speech recognition should not bias performance; successful identification will require identification of patterns in the token misidentification and stimulus-response confusion pairs.

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. Each subject was tested using four of the eight impaired models, as well as the baseline unimpaired acoustic model. The impaired models are identified as impl with “i” indicating the channel number (1 through 8) that was impaired, i.e. removed. Each utilization of each acoustic model by a study participant was considered a separate sample for the purposes of the experiment. Thus, the available data set contains 129 samples, with 12 to 17 repetitions of each model.

3.2. Confusion matrix analysis: training and testing

The inputs to the PLSDA algorithm consisted of each value from the vowel and consonant confusion matrices. The confusion matrix for the nine vowel tokens produces 81 features; the confusion matrix for the fourteen consonant tokens produces an additional 196 features. The PLSDA algorithm was trained and tested using a fifteen-fold cross-validation paradigm. There were nine acoustic models used in...
the listening study (one unimpaired, eight impaired), resulting in a nine-class (m-ary) classification task.

The PLSDA algorithm was also compared to a previously-used technique for confusion matrix analysis based on a correlation method. As in the previous study, the performance metric in this comparison was the false positive rate at true positive rate equal to unity, i.e. the proportion of incorrect acoustic model classifications prior to detection of the true model condition. In the previous study, this was determined to be a more useful metric than other performance measures such as percent error since it indicates levels of performance that are less than perfect but still better than chance.

4. RESULTS

The PLSDA technique requires specification of the number of components to use in the lower-dimensional projection. Fig. 1 shows percent correct identification of the acoustic model as a function of the number of components used in the PLSDA algorithm. Three feature sets were considered for comparison: vowels only, consonants only, and both vowels and consonants together. Performance as a function of number of components indicates a critical number of required components, below which performance has not yet stabilized. Performance for any of the three feature sets is relatively consistent when using between 10 and 20 PLSDA components. As the specified models become overly-complex (i.e. large numbers of PLSDA components are used), performance begins to degrade as the models become over-trained. Impairment identification is substantially improved when using both vowels and consonants as features versus use of either individual token set.

Fig. 2 shows confusion matrices associated with each of the three different feature sets (generated with the PLSDA classifier using 15 components), allowing for a careful analysis of the classification of the individual models. The trends in identification of the individual acoustic models are consistent with knowledge of the vowel and consonant cues. For example, the eighth channel contains little information that is useful for discriminating vowel tokens, but high-frequency information is necessary to distinguish different fricative consonants. The inability to distinguish these fricatives in indicative of a lack of information on the eighth channel, which PLSDA is capable of learning, and identification of impairments on the eighth channel based on the consonant confusion patterns is a relatively easy task. The acoustic model with an impairment applied to the first channel ("imp1") was the most difficult to correctly classify. A large number of confusions occurred between the unimpaired acoustic model ("none") and the acoustic model with an impairment on channel #1 ("imp1"). The first channel in the acoustic model is assigned speech information in the frequency range from 150Hz to 240Hz. This low frequency information does not contain a substantial number of discriminating cues for vowels or consonants, supported by the fact that average speech recognition scores for all subject using acoustic model imp1 were higher than scores with most of the other impaired models (only model imp8 had higher average speech recognition scores). Therefore, there may be a lack of strong, consistent patterns in the token confusions with acoustic model imp1 which results in a more difficult classification task.

To better evaluate the performance of PLSDA in the context of previous confusion matrix analysis results, the PLSDA approach used in this study was applied to the same data set used in Remus and Collins [1]. A comparison was made between PLSDA and two implementations of the correlational method approach that utilized different sets of features: formant frequencies and a model-based distance measure. From the previously published results, the techniques CorrMeth and HMM CorrMeth resulted in false positive probabilities of 28.8% and 37.2%, respectively. When evaluating PLSDA on the same test samples, the impaired channels were identified with a 15.4% probability of false positive. This is a statistically significant improvement over the results achieved with either of the correlational method approaches.

5. DISCUSSION

This study investigated the application of partial least squares discriminant analysis (PLSDA) to identify impaired perceptual conditions based on patterns of vowel and consonant confusions. The PLSDA algorithm is particularly appropriate for this task because it is able to manage a large number of input features, a property that impedes the use of many common pattern recognition algorithms. The limited training data, and time required for collecting training data, also complicate the use of feature selection methods. Thus, the level of performance achieved using PLSDA, a supervised technique that is capable of operating on a high-dimensional data set with limited training data, is an encouraging result. Comparison to previous confusion matrix analysis techniques
based on the correlational method indicates PLSDA provides a significant improvement in false positive probability. It was previously argued [1, 2] that one advantage of using vowel tokens in the listening task is that standard cues, such as formant frequencies, can be readily identified and linked to individual channels to identify poor transmission of cues on those channels if the corresponding vowel tokens are poorly recognized. However, PLSDA does not require such detailed analysis of the token discriminating cues since it operates directly on the confusion matrices. Therefore, it is possible to design listening tasks for follow-on confusion matrix analysis that utilize other types of tokens such as nonsense syllables, which have been used previously for channel weight estimation [13]. 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. Identification of a more optimal set of speech tokens could further improve the information that can be gathered from confusion matrix analysis.

The results of this study provide additional support for the use of speech-token confusion matrix analysis to quickly assess the transmission of speech information. This is particularly relevant in the assessment and fitting of cochlear implant patients, where tuning to address anomalous channels has been previously shown to improve speech recognition scores. Further development of methods for confusion matrix analysis will hopefully lead to additional measures for quickly and efficiently collecting information that will be useful for evaluating and assessing perceptual conditions in patients, leading to effective remediation strategies.

6. REFERENCES