DISTINGUISHING HIGH RISK SUICIDAL SUBJECTS AMONG DEPRESSED SUBJECTS USING MEL–FREQUENCY CEPSTRUM COEFFICIENTS AND CROSS VALIDATION TECHNIQUE

H. Kaymaz Keskinpala¹, T. Yingthawornsuk¹, D. M. Wilkes¹, R. G. Shiavi¹,², R. M. Salomon³
¹Department of Electrical Engineering and Computer Science, Vanderbilt University, TN, USA
²Department of Biomedical Engineering, Vanderbilt University, TN, USA
³Department of Psychiatry, Vanderbilt University School of Medicine, TN, USA

Abstract: This paper describes a way to distinguish high risk suicidal patients among depressed patients using mel-frequency cepstrum coefficients. Distinguishing high risk suicidal patients among depressed patients is an important problem; a practical solution to this would prevent the loss of many lives. In this study, the vocal characteristics of female and male patients' speech samples were analyzed and a small subset of the first ten mel-frequency cepstrum coefficients were used to classify high risk suicidal patients and depressed patients. Cross validation was used to observe classification performance. There were two different types of speech samples from both male and female patients. One of them was the speech sampled during a clinical interview and the other was speech sampled during a text-reading session.

Keywords: Speech, MFCC, suicide, depression, cross validation

I. INTRODUCTION

It is reported [1] that mental disorders are very common in the United States and internationally. Twenty-six percent of Americans, 18 or older, carried a mental disorder in 2005. In the same year, major depressive disorder affected 6.7 percent of the U.S. population. [1] More than 90 percent of the people who committed suicide had a diagnosable mental disorder, most commonly a depressive disorder [2]; so, there is an important relationship between depression and suicide.

As can be seen from these statistics, suicide is an important public health problem and has a strong relationship with depression. Therefore, it is very important to evaluate a depressed patient's risk of committing suicide. Psychiatrists evaluate this risk using clinical interviews and rating scales, such as the Hamilton depression rating scale. [3] Additionally, it is known that psychological states affect a person's speech production system. It was proposed by S. E. Silverman that vocal parameters of human speech could assist in recognizing and then assessing suicide risk. [4]

Some researchers have studied the relationship between vocal tract characteristics and suicidal risk. Tolkmitt et al. compared the formant information of vowels that occurs in the identical phonetic context during the patient's recovery period. [5] France et al. observed long term averages of the formant information and found that they were able to distinguish high risk suicidal patients from depressed and control patient groups. [6] Yingthawornsuk et al. used the percentages of the total power, its highest peak value and its frequency location to distinguish between high risk suicidal, depressed and remitted (had been depressed previously but recovered) groups. [7] In another study, Yingthawornsuk et al. used the spectral energy and the GMM based feature of the vocal tract system response for separating two groups of female patients carrying a diagnosis of depression and suicidal risk. [8] Kaymaz Keskinpala et al. used both energy in frequency bands, and first eight mel-cepstral coefficients to distinguish between high risk suicidal and depressed patients. [9] Ozdas used lower order mel-cepstral coefficients to distinguish high risk suicidal patients from non-suicidal ones using Gaussian mixture models and unimodal Gaussian models. [10]

Mel-frequency cepstral coefficients are useful parameters that have been used in many speech processing systems, such as in [10]. Logan proposed using mel-frequency cepstral coefficients for modeling music. [11] Godino-Llorente et al. used short term mel-cepstral parameters for pathological voice quality assessment. [12] Choi worked on compensating the mel-frequency cepstral coefficients for speech recognition in noisy environments. [13]

This paper presents work on distinguishing high risk suicidal patients from depressed patients using a small subset of the first ten mel-frequency cepstral coefficients for female and male patients. Cross validation was used to estimate the classification performance. The optimal mel-frequency cepstrum coefficients are found for female and male patients and for both the reading and interview sessions of each gender.

II. METHODOLOGY

A. Database

A.1. Information about the Database

The database for this research is obtained from an ongoing study within the Department of Psychiatry at
Vanderbilt University School of Medicine and supported by the American Foundation for Suicide Prevention. The study and consent process was developed in collaboration with, and approved by the Vanderbilt University Institutional Review Board. The database is composed of recordings from male and female subjects whose ages are between 25 and 65 years of age. Psychiatric clinicians, not involved in this study, categorized these patients as depressed, and with or without high risk of suicide, and referred them to research personnel for consent procedures, diagnostic confirmation and a brief recording. The number of the female patients and male patients that are used in this study is shown in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Female and Male Patient Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female / Male</td>
</tr>
<tr>
<td>Depressed</td>
</tr>
<tr>
<td>High-Risk</td>
</tr>
</tbody>
</table>

The database contains two different types of speech samples. One sample type was recorded while the patient was interviewed by a physician or highly-trained research assistant. This type of speech sample was named the "Interview Session". The other one is named the "Reading Session" and was recorded while the patient read predetermined part of a book. Quiet, closed rooms in clinical settings provided the recording environments.

A.2. Preprocessing

All speech signals were digitized by using a 16-bit analog to digital converter at a sampling rate of 10 kHz with an anti-aliasing filter. GoldWave v.5.08 audio editor was used to remove the silences which are longer than 0.5 seconds and the voices that is not belong to the patient. In this study, 76 seconds of each female patient's continuous speech from both interview and reading sessions were stored for analysis. For male patients, 66 seconds of continuous speech were stored. All stored speech signals underwent detection for voiced and unvoiced speech segments. Only voiced segments were used for subsequent analysis.

B. Feature Extraction

The features used for the analysis were a small subset of the first ten mel-frequency cepstrum coefficients in each patient's speech sample. Each speech signal was divided into 512 points of voiced segments. For each voiced speech segment the log-magnitude spectrum was computed from discrete Fourier transform (DFT). The spectrum was then filtered by a series of 16 triangular band-pass filters. The bandwidths and center frequencies of these filters are chosen according to the mel-scale.

The human ear is more sensitive to changes in the low frequency portion of the frequency spectrum. Thus, the mel-scale was formulated for the sampling of the frequency spectrum based on this property of human auditory perception. The linear frequency spectrum was mapped based on the human auditory perception with mapping approximately linear on the 0 – 1 kHz range and logarithmic above 1 kHz. The following formula is the suggested formula that models this relationship in which $F_{\text{mel}}$ is the perceived frequency and $F_{\text{Hz}}$ is the actual frequency.

$$F_{\text{mel}} = 2595 \log_{10} \left[ 1 + \frac{F_{\text{Hz}}}{700} \right]$$

Vocal tract length normalization was performed for each patient. The bandwidths and center frequencies of the filters in the mel-scale filter bank were then adjusted according to this normalization factor. The last step is to calculate the inverse discrete Fourier transform (IDFT) to obtain the mel-frequency cepstrum coefficients. The procedure is shown in Fig. 1 below.

After the first ten mel-frequency cepstrum coefficients were calculated for each frame, the values in all frames are averaged to have one value for each mel-frequency cepstrum coefficient for each patient.

C. Cross-Validation Classification

The k-fold cross validation technique [17] with quadratic discriminant function was performed on the mel-frequency cepstrum coefficients data. The data files were split randomly into two subsets. One set is for training the data and the other is for testing the data. Sixty-five percent of the data was used to train the data for estimating the quadratic classification function. Then using this quadratic classification function, 35% of the
data was tested by performing the classification. The variance of the performance estimates was reduced by averaging the results from 10 different runs of cross validation.

A simple approach is used to seek sub-optimal combinations of one, two, and three coefficients for classifications. The cross validation procedure is performed for each mel–frequency cepstrum coefficient separately. The cepstral coefficient that gives the maximum classification result is determined first. Next, this cepstral coefficient is paired with all the other cepstral coefficients and cross validation classification is performed again. The resulting pair of cepstral coefficients that gave the maximum classification result is determined. The same process is repeated for three cepstral coefficients that gave the maximum classification. Three classification performances (one coefficient performance, two coefficients performance, and three coefficients performance) are then compared and then the set giving the best performance is assigned as the optimal coefficients.

This performance testing is performed for three criteria: determining only the maximum depressed classification, and then only for the maximum high risk suicidal classification, and finally for the maximum total classification of depressed and high risk suicidal classification.

### III. RESULTS

The depressed–high risk suicidal pairwise classification using k-fold cross validation technique was performed for finding the optimal coefficient(s) that gave the maximum classification performance. The results for the male interview and reading sessions are shown in Table 3 and Table 4, respectively.

**Table 3. Male Interview Session’s Classification Results**

<table>
<thead>
<tr>
<th>Only Depressed</th>
<th>Coefficients 1 and 4</th>
<th>78.60%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only High Risk Suicidal</td>
<td>Coefficient 3</td>
<td>86.00%</td>
</tr>
<tr>
<td>Total Classification</td>
<td>Coefficient 3</td>
<td>77.20%</td>
</tr>
</tbody>
</table>

**Table 4. Male Reading Session’s Classification Results**

<table>
<thead>
<tr>
<th>Only Depressed</th>
<th>Coefficients 2, 9 and 1</th>
<th>89.80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only High Risk Suicidal</td>
<td>Coefficient 2</td>
<td>93.00%</td>
</tr>
<tr>
<td>Total Classification</td>
<td>Coefficient 2</td>
<td>78.00%</td>
</tr>
</tbody>
</table>

The optimal features for depressed classification are coefficients 1 and 4 for the interview session with a classification performance of 78.60%; on the other hand optimal features are coefficients 2, 9 and 1 for the reading session with a classification performance of 89.80%.

Coefficient 3 is the optimal feature for both high risk suicidal classification and total classification of depressed–high risk suicidal with a classification performance of 86% and 77.20% respectively in the interview data.

The optimal feature for both high risk suicidal classification and total classification of depressed–high risk suicidal classification of the reading session is coefficient 2. The classification performance was 93% for the high risk suicidal classification and 78% for the total classification of depressed–high risk suicidal classification.

**Table 5. Female Interview Session’s Classification Results**

<table>
<thead>
<tr>
<th>Only Depressed</th>
<th>Coefficients 1, 5, and 7</th>
<th>78.90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only High Risk Suicidal</td>
<td>Coefficient 9</td>
<td>70.10%</td>
</tr>
<tr>
<td>Total Classification</td>
<td>Coefficient 9</td>
<td>66.40%</td>
</tr>
</tbody>
</table>

Table 5 shows the results for the female interview session. The optimal features for depressed classification are coefficient 1, 5 and 7 with a classification performance of 78.90%; on the other hand the optimal feature is coefficient 9 for both high risk suicidal classification and total classification of depressed–high risk suicidal classification with a classification performance of 70.10% and 66.40% respectively.

**Table 6. Female Reading Session’s Classification Results**

<table>
<thead>
<tr>
<th>Only Depressed</th>
<th>Coefficients 3 and 2</th>
<th>70.10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only High Risk Suicidal</td>
<td>Coefficient 8</td>
<td>71.10%</td>
</tr>
<tr>
<td>Total Classification</td>
<td>Coefficient 9</td>
<td>63.90%</td>
</tr>
</tbody>
</table>

Table 6 presents the female reading session classification results and optimal features. The optimal features for depressed classification are coefficients 3 and 2 with a classification performance of 70.10%. For high risk suicidal classification, the optimal feature is coefficient 8 with a classification performance of 71.10%. Coefficient 9 is the optimal coefficient for the total classification of depressed–high risk suicidal classification.
classification with a classification performance of 63.90%.

IV. DISCUSSION AND CONCLUSIONS

This paper demonstrates that mel-frequency cepstrum coefficients are a good indicator for discriminating between depressed patients at high- and low-risk of suicidal behavior. Male and female patients were analyzed separately. The mel-frequency cepstrum coefficients discriminated among the depressed patients, with matching of the vocal to clinical assessment with a performance better than 70%.

The controlled text-reading tended to give better results for male subjects especially for high risk suicidal classification and depressed classification. The total classification was about the same for both reading session and interview sessions. The maximum classification results that are obtained from the male subjects are noticeably better than the female subjects’ results.

The findings may be limited by several factors, including the imperfections of the recording environments, the reliance on clinical assessments (by non-research as well as research diagnosticians) for a reference standard, and the variable timing of recordings relative to peak intensities of suicidal risk.

Never-the-less, the findings may ultimately be applicable to the development of clinically practical instruments for detecting vocal stress that could indicate a need for increased attention to suicidal risk assessment. The results for male subjects especially for high risk suicidal performance better than 70%.

The controlled text-reading tended to give better results for male subjects especially for high risk suicidal classification and depressed classification. The total classification was about the same for both reading session and interview sessions. The maximum classification results that are obtained from the male subjects are noticeably better than the female subjects’ results.

The findings may be limited by several factors, including the imperfections of the recording environments, the reliance on clinical assessments (by non-research as well as research diagnosticians) for a reference standard, and the variable timing of recordings relative to peak intensities of suicidal risk.

Never-the-less, the findings may ultimately be applicable to the development of clinically practical instruments for detecting vocal stress that could indicate a need for increased attention to suicidal risk assessment. These findings, along with other findings in the literature, indicate that feedback and feed-forward regulatory pathways for speech production are impaired in depression. Identifiable and quantifiable alterations in these pathways may provide needed paradigms for the study of the pathophysiology of depression.

REFERENCES


