INVESTIGATION OF GLOTTAL FLOW SPECTRAL SLOPE AS POSSIBLE CUE FOR DEPRESSION AND NEAR-TERM SUICIDAL RISK

Asli Ozdas¹, Hasmila Omar³, Richard G. Shiavi², Stephen E. Silverman⁴, Marilyn K. Silverman⁴, and D. Mitchell Wilkes³

¹Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee, USA
²Department of Biomedical Engineering, Vanderbilt University, Nashville, Tennessee, USA
³Department of Electrical and Computer Engineering, Vanderbilt University, Nashville, Tennessee, USA
⁴Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut, USA

Abstract: When reviewing his clinical experience in treating suicidal patients, one of the authors observed that successful predictions of suicidality were often based on the patient’s voice independent of content. In this study we investigated the discriminating power of an excitation-based speech parameter, the glottal flow spectrum. There were two sets of subjects, male and female. Each set consisted of 10 high-risk near-term suicidal patients, 10 major depressed patients, and 10 non-depressed control subjects. As a result of two sample statistical analyses, the slope of the glottal flow spectrum, was a significant discriminator in five of six comparisons (p<0.05). A maximum likelihood classifier, developed by combining the a posteriori probabilities of two features, yielded correct classification scores between 60 and 95%.

Keywords: Speech, glottal flow spectrum, suicide, depression, classification

I. INTRODUCTION

Identification of individuals at imminent suicidal risk is often one of the most important judgments that clinicians must make. This task requires gathering and weighing of a variety of information and data from numerous sources by experienced clinicians [1]. These methods help in categorizing individual patients as “high risk”, but they are not sufficient to determine if a patient is at imminent risk. Stephen and Marilyn Silverman describe suicidal speech as similar to depressed speech but exhibiting significant perceptual changes in its qualities when a patient becomes near-term suicidal. The exhibition of these qualities was often a decisive factor in alerting the clinicians to the need to take preventative action [2]. These clinical findings together with the literature on the clinical importance of a patient’s voice in psychiatry led to the hypothesis that near-term suicidality may be associated with changes in speech production and articulation that differ from non-suicidal persons. Our own studies are showing this [3][4].

Many studies have been done using the fundamental frequency. However, the fundamental frequency provides information only about the duration of the glottal cycle. Besides fundamental frequency, glottal flow waveform was also reported to be altered as a result of excessive tension or lack of coordination in the laryngeal musculature under emotional stress [5]. Investigation of this phenomenon showed an increase in the amount of high frequency energy in the glottal pulses under emotional stress. In this paper, we explore the significance of the slope of the glottal flow spectrum (spectral tilt) as an indicator of near-term suicidal risk.

II. DATABASE FORMULATION

Glottal flow spectral analyses were performed on sets of audio recordings for males and females. Each set contained 10 near-term suicidal patients, 10 depressed patients, and 10 non-depressed control subjects collected from existing databases. All the patients used in this research were white Caucasians between the ages of 25 and 65. Because of the inability to record psychiatric speech in controlled settings, all of the speech samples were recorded during real-life situations (i.e., therapy sessions, suicide notes left on tapes, etc with various tape recorders at various recording environments). A high-risk, near-term suicidal patient was defined as one who has committed suicide or attempted suicide and failed within minutes to weeks from the time of their voice recordings. The audio recordings of the depressed and control groups were extracted from the database of an ongoing study in the Vanderbilt University Department of Psychiatry. The control group was comprised of depressed individuals who, after receiving cognitive therapy or pharmacotherapy, were judged to be no longer depressed and not in need of further treatment. The selected non-depressed control subjects met the following criteria: 1) a Hamilton rating scale (17 item version) for a depression score of 7 or less [6]; 2) a Beck depression score of 7 or less [7]. The depressed patients met the following criteria: 1) major depressive disorder as defined by the research diagnostic criteria [8]; 2) a Beck depression score of 20 or greater; 3) a Hamilton rating scale for depression score 14 or greater.

All of the selected audio recordings were digitized using a sixteen-bit analog to digital converter. The sampling rate was 10 KHz, with an anti-aliasing filter (i.e.,
5KHz low-pass) precisely matched to the sampling rate. The digitized speech waveforms were then imported into a MicroSound Editor where silence pauses exceeding 0.5 seconds were removed to obtain a record of continuous speech. Thirty seconds of continuous speech from each subject were stored for analyses.

III. METHODS

A. Glottal Spectral Slope Feature Extraction

Vocal tract effects were removed from the speech spectrum while estimating glottal flow spectrum. It was assumed that the frequency response of the vocal tract shapes the speech spectrum for different vowels and glottal flow spectrum stays the same for all vowels. Therefore, the glottal flow spectrum can be estimated if energy normalized frames from voiced speech spectra are averaged to remove the effects of vocal tract shaping. The averaged vocal tract response will have an all pass characteristics if a wide variety of vowel spectra are used, and the average energy normalized frames will yield the glottal flow spectrum. This approach provides a representation that reflects the properties of glottal flow waveform.

A1. Estimation of Glottal Spectrum

a) The patient speech is broken into segments containing 256 samples.
b) Voiced and unvoiced speech detection is performed on each segment. However, only voiced segments are retained for analysis. The method used is based on wavelets and developed by Ozdas [4].
c) The periodogram for each voiced segment is calculated using the discrete Fourier transform.
d) Each periodogram is normalized by its energy.
e) All normalized periodograms are then averaged to remove the effects of varying vocal tract response.
f) The average energy of all voiced segments is then used to scale the average normalized periodogram back to its original amplitude. This is the glottal flow spectrum estimate.

A2 Estimation of Glottal Spectral Slope

The spectral slope is calculated using a least squares line fit on a log-log scale is performed over 300-3000 Hz frequency band of the glottal flow spectrum. The slope given by the least square error approximation gives the glottal spectral slope for each patient. Fig. 1 shows an example of the estimation procedure.

B. Comparative Statistical Analyses and Classification

B1. Statistical Tests

Two-sample (i.e., control-depressed, control-suicidal, and depressed-suicidal) t-tests were performed separately on glottal spectral slope estimates to determine any statistically significant differences in means [9].

B2. Maximum Likelihood Classifier

In order to evaluate the discriminating power of the slope among groups, a Maximum Likelihood (ML) classifier was developed for each parameter. The ML classifier employs the Probability Density Function (PDF) of each class to make a decision as to which class PDF results in the closest match for a test data sample. The PDFs of the class distributions were assumed to be unimodal Gaussian and were generated by using the means and variances estimated from the training samples. Given the trained class model, classification of the test samples was accomplished according to Bayes' decision rule, where a test subject was assigned to the class for which it had the maximum a posteriori probability for its set of observations.

Ideally, this procedure is conducted by splitting the total data set into a training set and a test set. Because of the limited number of patients in this case, Lachenbruch's holdout procedure was employed [10]. This procedure is very useful for small data class sizes because it makes it possible to use the same subject for both training and testing rather than using only half of the data for each part.
IV. RESULTS

A. Magnitude of Glottal Slopes

The estimated magnitudes of the slopes of the glottal flow spectra for each subject are given in Figs. 2 and 3.

Fig. 2: Magnitude of Glottal Slopes for Males

Notice that the controls have the highest values and the depressed subjects have the lowest values.

B1. Statistical Test

The p-values for pair-wise comparison of the means are shown in Table 1.

Table 1. P-Values for Mean Comparisons of Spectral Slope Estimates

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control/Depressed</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Depressed/Suicidal</td>
<td>0.0005</td>
<td>0.0035</td>
</tr>
<tr>
<td>Control/Suicidal</td>
<td>0.1156</td>
<td>0.0266</td>
</tr>
</tbody>
</table>

The means of the groups of spectral slope values are significantly different (p < 0.05) for five of the six comparisons. The only comparison that is not different is the one between control and suicidal females.

B2. Maximum Likelihood Classifier

The ML classification results for glottal flow spectral slope are presented in Table 2.

Table 2. ML Pairwise Classification Results (%) For Spectral Slope Estimates

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control/Depressed</td>
<td>95</td>
<td>90</td>
</tr>
<tr>
<td>Depressed/Suicidal</td>
<td>85</td>
<td>75</td>
</tr>
<tr>
<td>Control/Suicidal</td>
<td>60</td>
<td>60</td>
</tr>
</tbody>
</table>

The ML classifier yielded overall classification scores between 60% and 95%. The highest between depressed and control classes and the lowest between suicide and control classes. This was consistent in the male and female populations.

V. DISCUSSION AND CONCLUSION

Analyses of glottal spectral slope measurements indicated that both near-term suicidal and depressed patients exhibit significantly higher energies in the upper frequency bands of the glottal flow spectrum compared to healthy controls. These shifts are significantly different among most of the comparisons. The spectral content of the glottal spectra is more similar between controls and suicidal subjects while those for depressed subjects have a broader bandwidth. In addition it is possible to use the spectral slope to classify subjects as belonging to one of three groups. Evidence for similar energy shifts in long-term energy spectra during depression and near-term suicidal states have been reported by various researchers [11]. Most of the studies that investigated this phenomenon have revealed that the speech of patients who suffer from major depressive illness contains more energy at higher frequency bands, which was shifted toward lower frequencies after treatment. Here, it is important to note that it is not possible to collect speech samples from suicidal persons shortly before their suicide attempts in a
systematic manner. Therefore expanding the database requires a considerable amount of time.

REFERENCES