Acoustical properties of speech as indicators of depression and suicidal risk

Michael Landau
School of Engineering, Vanderbilt University

Acoustical properties of speech have been shown to be related to mental states such as remission and depression. The objective of this project was to relate the energy in frequency bands with the severity of the mental state using the Beck Depression Inventory (BDI). Recorded speech was obtained from male and female subjects with mental states of remission, depression, and suicidal risk. These subjects had recorded automated and spontaneous speech samples. Multiple regression analysis was used to relate the independent energy band ratio variables with the dependent BDI scores, and thus allow the determination of equitable BDI scores for future patients. For the male group, the square of the 3rd energy band and the cross-product of the 2nd and 3rd energy band were prominent in both the reading and interviewed groups. Therefore the equation with the 2nd lowest Akaike Information Criterion (AIC) score was chosen for the reading male group, and the 1st lowest AIC score was chosen for the interviewed male group. For the female group, the square and cross-product of the 1st and 2nd energy bands were prominent in both the reading and interviewed groups. Therefore the 2nd lowest AIC score was chosen for the reading female group, and the 1st lowest AIC score was chosen for the interviewed female group. The clinician could thus determine the patient’s mood or state of mind by comparing the estimated BDI score with the ranges of total BDI scores: remitted 0 – 20, depressed 15 – 38, suicidal 38 – 46.

Introduction

Each year, 11 out of 100,000 people take their own lives and anywhere from 240,000 to 600,000 attempts are made in the United States alone, making suicide the eighth leading cause of death across the nation [1]. Moreover, national suicidal rates universally show an upward long-term trend, proving that something must be done in order to assess and prevent possible high-risk candidates from acting upon their tendencies.

Current methods utilize clinician and psychologist evaluations of a patient’s mental state based on histories, interviews, self-reports, and reports by others. However, these methods of elucidation prove to be time consuming and inaccurate based on the clinicians experience. Similarly, not every person with suicidal tendencies will be forthright with their mental state and seek help from a clinician or doctor. More often than not, they will visit their doctor with a completely unrelated problem in an attempt to contingently cry for help. Finally, in the case of a town with a small population and little to no access to a doctor with experience in psychoanalysis, there may be no way of telling if a person is actually depressed or suicidal. Therefore, a quick and easy method to predict someone’s mental state during a short visit is needed to augment clinical evaluation.

2. Hypothesis

The ultimate goal of this project is to do just that: provide a quick and simple method to evaluate a patient’s mental state and determine if they might be at risk for suicide. Figure 1 depicts the legacy Human Oriented Process (HOP) and the new Computer Oriented Process (COP) that...
is under development. In order to facilitate COP development, the current study attempts to relate the frequency content in speech to the mental state of persons in three study groups: remitted, near-term suicidal and depressed.

The frequency content is evaluated through the use of a MATLAB program developed by Thaweesak Yingthawornsuk, a Vanderbilt engineering graduate student. His program extracts the energy band ratios from an edited audio file from a patient’s recording of spontaneous and/or automated speech. The spontaneous speech comes from an interview with a clinician, and includes questions regarding the patient’s normal daily activities and behaviors. The automatic speech is obtained from the reading of the “Rainbow Passage”, an articulation exercise that is phonetically balanced for English. For a copy of the Rainbow Passage, please refer to Appendix 3 [2]. The patient’s mental state will be assessed based on a total score from the Beck Depression Inventory – II (BDI-II), a self-scored patient survey containing questions about the person’s feelings for the week previous to when the questionnaire was administered [3].

The goal of my research is to relate these two separate variables – the energy band ratios of the interview (independent) and the total BDI score (dependent) – with a multiple regression equation. The purpose of this is to predict how depressed or high-risk the person is based solely upon the highly accurate energy band ratio estimations. The best regression model will be determined by minimizing an information theoretic score using the Akaike’s (AIC) and Bayesian’s (BIC) information criterion and penalty factors [4]. This offline process is depicted in figure 2 and will be described in section 3.3.

Figure 2: Offline Analysis for the COP

<table>
<thead>
<tr>
<th>Model Hypotheses / Parameter Estimation</th>
<th>Ensemble of BDI / EBR (3bands) quadruplets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questionnaire</td>
<td>Patient Response</td>
</tr>
<tr>
<td>Interview</td>
<td>Psychologist</td>
</tr>
<tr>
<td>Signal Processing</td>
<td>Audio Editor</td>
</tr>
<tr>
<td>- EBR Extractor</td>
<td>Measured EBRs</td>
</tr>
</tbody>
</table>

3. Materials and Methods

3.1. Database

Each audio recording was obtained from an ongoing study supported by the American Foundation for Suicide Prevention, and can be divided into four populations, each with three different classifications. There were 27 male reading patients, 19 male interview patients, 30 female reading patients, and 31 female interview patients, and all were between the ages of 25 and 65 years. The patients were either interviewed by a clinician and/or required to read the “Rainbow Passage” as mentioned above. Depending on the clinician’s professional opinion, these patients were then separated into three mood classes, including suicidal, depressed, or remitted (those who were previously high risk or depressed, but have mitigated their symptoms).

Each patient also filled out a survey known as the Beck Depression Inventory – II (BDI-II), adapted from the original version created by Dr. Aaron T. Beck in 1961. The most current version of the questionnaire, published in 1996, regards mental as well as physical depression related symptoms. The responses to these 21 questions provide a numerical score ranging from 0 to 64, where the higher scores relate to a more high-risk or suicidal mental state [3]. For a copy of the questionnaire please refer to Appendix 3.

There is also a fourth mood class, the ideational suicide patients, in which the subjects often talk about committing suicide, but will most likely not act upon their thoughts or tendencies. However, this group tends to fall in a wide range of total BDI score, falling anywhere between a score of 0 – 58. They were therefore not included in the model hypotheses determination.

3.2. Feature Extraction

Important tools that were utilized are the MATLAB scripts that extract the interviewer and unvoiced portions of the audio files. All speech signals are first digitized by using a 16-bit analog to digital converter with a sampling rate of 10 kHz. The program then prompts the user to select a portion of the interview to distinguish the interviewer and patient. It then separates the voiced and unvoiced portions of the speech, discarding all silences and sounds that don’t involve the vocal cords (e.g. the sounds ‘s’ and ‘th’ make). For a flowchart of the training and separation programs,
please refer to Appendix 2.

Additionally, perhaps the most important tool used in this project was Thaweesak Yingthaworn-suk’s MATLAB script that estimates the Power Spectral Density (PSD) of the edited audio files of a patient’s voiced speech from both an interview and a reading of the rainbow passage. The PSD estimate was determined using the Welch method with non-overlapping 100-point Hamming windows [5]. Four features were then calculated within a 0-2000 Hz frequency range, where the percentages of the total energy (PSD1, PSD2, PSD3, PSD4) fall in each of the four 500 Hz sub-bands. These four subsets subsequently produce the energy band ratios for each level.

3.3. Regression Analysis

In a previous project I conducted with this data, I created various single variable linear regression models for each of the first three PSD sub-bands, and concluded that the first energy band was the most significant in estimating the total BDI score. However, these models excluded the relation of the BDI score to each of the sub-bands in a single, easily managed equation. Therefore, a multiple regression model is necessary in order to relate the independent energy band ratio variables with the dependent BDI scores, and thus allow the determination of equitable BDI scores for future patients.

The multiple regression model, first used by Pearson in 1908, is a tool used to learn more about the relationship between several predictor variables and a criterion variable, producing a model similar to equation 1 [6]:

\[ y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 \] (eqn. 1)

In the case of this project, the y or criterion variable will represent the total BDI score, the x or predictor variables will represent the energy band ratios, and the beta coefficients represent the weighting independent contributions of each independent energy band ratio:

\[ b_{diest} = b_0 + b_1 \text{psd}_1 + b_2 \text{psd}_2 + b_3 \text{psd}_3 \] (eqn. 2)

Since some variables, i.e. the energy band ratios, may affect the goodness of fit (R^2) value more than others, quadratic variables will be introduced into the model. In order to determine the R^2 value, beta coefficients, expected dependent variable, and residuals of each model, the dependent and independent values of each patient in each of the four groups will be entered into a MATLAB equation designed by G. Anthony Reina (reference?). The R^2 value (eqn. 3), a.k.a. the coefficient of determination, is a number with variability from 0 to 1 that reveals how closely the estimated values for the statistical model correspond to the actual data [6]. In this definition, the term “variability” is defined as the sum of squares.

\[
R^2 = 1 - \frac{\sum (y_i - \text{exp})^2}{\sum (y_i - \text{avg})^2} \] (eqn. 3)

The variable ‘yi’ is an actual BDI measurement, ‘y_{exp}’ is the respective expected value, and ‘y_{avg}’ is the average of the yi values. The residual is defined as the error in estimating the dependent variable, or the actual y value minus the expected y value. Since there are 64 different combinations of the quadratic variables \(x_1^2\), \(x_2^2\), \(x_3^2\), \(x_1x_2\), \(x_1x_3\), \(x_2x_3\), there will be 64 models to choose from, each including all three linear variables \((x_1, x_2, x_3)\).

In order to choose the most appropriate model, Akaike’s (AIC, eqn. 4) and Bayesian’s (BIC, eqn. 5) information criterion were utilized [4]. The minimization of these criterions can indicate the “best” model. AIC and BIC formulae measure of the goodness of fit of the estimated model, and their derivations are grounded in the concept of entropy.

\[ AIC = 2k + n \ln \left( \frac{RSS}{n} \right) \] (eqn. 4)

\[ BIC = k \ln(n) + n \ln \left( \frac{RSS}{n} \right) \] (eqn. 5)

Note that simply minimizing the Residual Sum of Squares (RSS) will not work since the RSS is a decreasing function of the number of free model parameters (k) and becomes zero when the number of free parameters equals the number of measurement samples (n). The AIC penalizes free parameters, the order of the model hypothesis, less strongly than does the Bayesian criterion and therefore is associated with models with higher coefficients of determination.

4. Results and Discussion
The data in each of the four groups of patients were first organized into graphs in order to find the range of energy band ratios and total BDI scores. Figure 3 is an example of how the reading male patients relate to each other; for the rest of the graphs, please refer to Appendix 1. The graph plots the BDI score against each separate independent variable (energy band ratios 1, 2, and 3) for each patient. The colors represent different classes of patients, where ‘sc’ are the suicidal patients, ‘dp’ are the depressed patients, ‘rm’ are the remitted patients, and ‘id’ are the ideational patients.

Table I displays the relationship between the total BDI score and the energy band ratios for each patient in each of the four groups, where ‘sc’ are the suicidal patients, ‘dp’ are the depressed patients, ‘rm’ are the remitted patients, and ‘id’ are the ideational patients.

Table I. Range of Data Values

<table>
<thead>
<tr>
<th>Gender</th>
<th>Session</th>
<th># Patients</th>
<th>BDI Score (sc)</th>
<th>BDI Score (dp)</th>
<th>BDI Score (rm)</th>
<th>PSD1</th>
<th>PSD2</th>
<th>PSD3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Reading</td>
<td>27</td>
<td>20 - 57</td>
<td>9 - 34</td>
<td>0 - 16</td>
<td>0.60 - 0.94</td>
<td>0.054 - 0.468</td>
<td>0.003 - 0.044</td>
</tr>
<tr>
<td>Male</td>
<td>Interview</td>
<td>19</td>
<td>40 - 57</td>
<td>9 - 30</td>
<td>0 - 16</td>
<td>0.60 - 0.94</td>
<td>0.058 - 0.384</td>
<td>0.002 - 0.054</td>
</tr>
<tr>
<td>Female</td>
<td>Reading</td>
<td>30</td>
<td>28 - 51</td>
<td>18 - 38</td>
<td>19 - 28</td>
<td>0.65 - 0.96</td>
<td>0.038 - 0.335</td>
<td>0.001 - 0.044</td>
</tr>
<tr>
<td>Female</td>
<td>Interview</td>
<td>31</td>
<td>34 - 51</td>
<td>18 - 38</td>
<td>0 - 21</td>
<td>0.61 - 0.97</td>
<td>0.031 - 0.361</td>
<td>0.001 - 0.051</td>
</tr>
</tbody>
</table>

The data in each of the four groups of patients were first organized into graphs in order to find the range of energy band ratios and total BDI scores. Figure 3 is an example of how the reading male patients relate to each other; for the rest of the graphs, please refer to Appendix 1. The graph plots the BDI score against each separate independent variable (energy band ratios 1, 2, and 3) for each patient. The colors represent different classes of patients, where ‘sc’ are the suicidal patients, ‘dp’ are the depressed patients, ‘rm’ are the remitted patients, and ‘id’ are the ideational patients.

Table I displays the relationship between the total BDI score and the energy band ratios for each patient in each of the four groups, where ‘sc’ are the suicidal patients, ‘dp’ are the depressed patients, ‘rm’ are the remitted patients, and ‘id’ are the ideational patients.

From Table I, it shows that there is a definitive range for each energy band ratio. The 1st energy band ratio seems to fall in the range of 1 - .546, the 2nd energy band’s ratio is between .546 - .054, and the 3rd energy band ratio will approximately fall in the range of .054 - .001. As expected, the suicidal patients have a higher BDI score, whereas the depressed patients have mid-range BDI scores, and remitted patients have the lowest total BDI scores. However, the remitted and depressed patient’s BDI scores overlap somewhat, therefore a rough estimate will have to be made in order to classify a new patient’s mood (see Table VII). In the overlap region (roughly 15 – 20), there’s ambiguity between the remitted and depressed mood classes, and more information about the patient’s history may be needed.

After regressing all of the models, the AIC and BIC scores were calculated and included in a chart (eg: figure 4) for each of the 64 model hypotheses in each group (Appendix 1).

These scores were normalized in order to fit them on one chart. Also, the residual sum of squares (RSS) was included in the chart, to demonstrate that the RSS decreases with more parameters and therefore is not the proper cost function.
to minimize. In this figure, ‘MP’ stands for model parameters or number of variables used in each model hypothesis.

The lowest AIC, BIC, and RSS scores were then organized into Tables II, III, IV, and V (Tables III, IV, and V are located in Appendix 1).

In the LN(r) column, “r” represents the normalized sum of squares of the residuals (RSS/N), and “var” represents the number of variables or parameters in the model, and x1, x2, and x3 correlate with energy band ratios 1, 2, and 3 respectively.

The cubic variables (ie: x1^3, x2^3, x3^3) were also tested in the model hypotheses. However, they only increased the goodness of fit about 5-10% for the male reading and female patients. Moreover, these models did not produce the lowest AIC and BIC scores. Since adding a cubic variable tends to over-fit the data and exhibit several nearby maxima and minima next to each other, they are therefore not included in the model hypotheses. But it should be noted for the male interview group, adding the x1^3, x2^3, and x3^3 variables tended to increase the R^2 value by about 30%, and therefore more research is needed to test for the significance of these parameters.

5. Conclusion

Table VI. Final Model Hypotheses for Four Groups

<table>
<thead>
<tr>
<th>Gender</th>
<th>Session</th>
<th>Model Hypothesis (Multiple Linear Equation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Reading</td>
<td>y = 21 - 36x1 + 13x2 + 4250x3 - 80406x3^2 - 3946x2x3</td>
</tr>
<tr>
<td>Male</td>
<td>Interview</td>
<td>y = -4030 + 4050x1 + 4120x2 + 187970x1x3 - 192000x2x3 - 237010x3^2</td>
</tr>
<tr>
<td>Female</td>
<td>Reading</td>
<td>y = 8926 - 196261x1 - 15468x2 + 2705x3 + 10743x1^2 + 6710x2^2 + 17286x1x2 - 13213x2x3</td>
</tr>
<tr>
<td>Female</td>
<td>Interview</td>
<td>y = 26240 - 57148x1 - 38456x2 - 35499x3 + 30935x1^2 + 12338x2^2 + 43308x1x2 + 43956x1x3</td>
</tr>
</tbody>
</table>
The information theoretic criterion I chose for the best model prediction is AIC (ie: not BIC) because they coincide with the models that provide better coefficients of determination (R2) for the equation fits. For the male group, the square 3rd energy band (x3) and the cross-product of the 2nd (x2) and 3rd energy band were prominent in both the reading and interviewed groups. Therefore the equation with the 2nd lowest AIC score was chosen for the reading male group (eqn. 19), where the gain in error is roughly 10%, and the 1st lowest AIC score was chosen for the interviewed male group (eqn. 42). For the female group, the square and cross-product of the 1st (x1) and 2nd (x2) energy bands were prominent in both the reading and interviewed groups. Therefore the 2nd lowest AIC score was chosen for the reading female group (eqn. 47), and the 1st lowest AIC score was chosen for the interviewed female group (eqn. 46).

Table VII. Final Range for Three Mood Classes

<table>
<thead>
<tr>
<th>Mood Class</th>
<th>Total BDI Score Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suicidal</td>
<td>38 - 64</td>
</tr>
<tr>
<td>Depressed</td>
<td>15 - 38</td>
</tr>
<tr>
<td>Remitted</td>
<td>0 - 20</td>
</tr>
</tbody>
</table>

The final range of total BDI scores for each mood class was determined based on data from Table I. For any future patient, the power spectral densities can be extracted from an audio recording of their interview or reading of the Rainbow Passage. The first three power spectral densities (energy band ratios) could then be integrated into the model hypothesis of the patient’s respective group, and the total BDI score would be estimated. The clinician could thus determine the patient’s mood or state of mind by comparing the estimated BDI score with Table VII.

References

Appendix 1: Regression analysis, graphs, charts, and tables
### Table II. Model Hypotheses for Reading Males (yellow: 1st lowest score, blue: 2nd lowest score)

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
<th>LN R2</th>
<th>Model Hypothesis (Equation Fit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>6618</td>
<td>2814</td>
<td>5.23</td>
<td>$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$</td>
</tr>
<tr>
<td>19</td>
<td>6640</td>
<td>2825</td>
<td>5.22</td>
<td>$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$</td>
</tr>
</tbody>
</table>

### Table III. Model Hypotheses for Interviewed Males (yellow: 1st lowest score, blue: 2nd lowest score)

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
<th>LN R2</th>
<th>Model Hypothesis (Equation Fit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>5735</td>
<td>2875</td>
<td>5.54</td>
<td>$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$</td>
</tr>
<tr>
<td>51</td>
<td>5808</td>
<td>2925</td>
<td>5.77</td>
<td>$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7$</td>
</tr>
</tbody>
</table>

### Table IV. Model Hypotheses for Reading Females (yellow: 1st lowest score, blue: 2nd lowest score)

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
<th>LN R2</th>
<th>Model Hypothesis (Equation Fit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>1678</td>
<td>2173</td>
<td>4.49</td>
<td>$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7$</td>
</tr>
<tr>
<td>47</td>
<td>1617</td>
<td>2145</td>
<td>4.53</td>
<td>$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7$</td>
</tr>
</tbody>
</table>

### Table V. Model Hypotheses for Interviewed Females (yellow: 1st lowest score, blue: 2nd lowest score)

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
<th>LN R2</th>
<th>Model Hypothesis (Equation Fit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>4079</td>
<td>4365</td>
<td>5.39</td>
<td>$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7$</td>
</tr>
<tr>
<td>28</td>
<td>4303</td>
<td>5195</td>
<td>5.44</td>
<td>$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7$</td>
</tr>
</tbody>
</table>

---

Spring 2008 | Volume 4 | Number 1
Appendix 2: Block Diagrams

Audio File Editor Block Diagram: Training

Audio File Editor Block Diagram: Separation

Determination of Best Model Hypothesis Block Diagram

Model Hypotheses / Parameter Estimation

\[ AIC = 2k + n \ln \left( \frac{RSS}{n} \right) \]

\[ BIC = k \ln(n) + n \ln \left( \frac{RSS}{n} \right) \]

Multiple linear regression analysis

- R² value, F-value, p-value, beta coefficients, expected dependent variable, residuals
- Residual sum of squares (RSS) and AIC & BIC equation analysis
- Lowest AIC and BIC scores
- Best model (chosen based on lowest AIC score)

Offline Process
The Rainbow Passage

(Articulation exercise including all the normal sounds of spoken English)
AV3F English Public Speaking (Hopkins)
Department of Translation Studies, University of Tampere

When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow.

Throughout the centuries people have explained the rainbow in various ways. Some have accepted it as a miracle without physical explanation. To the Hebrews it was a token that there would be no more universal floods. The Greeks used to imagine that it was a sign from the gods to foretell war or heavy rain. The Norsemen considered the rainbow as a bridge over which the gods passed from earth to their home in the sky. Others have tried to explain the phenomenon physically. Aristotle thought that the rainbow was caused by reflection of the sun’s rays by the rain.

Since then physicists have found that it is not reflection, but refraction by the raindrops which causes the rainbows. Many complicated ideas about the rainbow have been formed. The difference in the rainbow depends considerably upon the size of the drops, and the width of the colored band increases as the size of the drops increases. The actual primary rainbow observed is said to be the effect of super-imposition of a number of bows. If the red of the second bow falls upon the green of the first, the result is to give a bow with an abnormally wide yellow band, since red and green light, when mixed, form yellow. This is a very common type of bow, one showing mainly red and yellow, with little or no green or blue.


Last Updated 14 November 2001