Introduction

- Suicide is the eighth leading cause of death in the United States and depression is the most common precursor to suicide. Almost 50%-75% of all suicides are linked to depression.
- In previous work, researchers have looked at features such as pitch, amplitude modulation characteristics, vocal jitter, glottal flow spectrum, degree of formant movement, and formant bandwidths, power spectral density, Mel-frequency Cepstral Coefficients (MFCCs), and Teager energy among others [1], [2], [3].
- The motivation for the current study comes from the fact that neuro-physiological changes associated with depression affect motor coordination and therefore the disruption of articulatory control and kinematics [1].
- In this study, we want to understand if our aperiodicity/periodicity/pitch (APP) measure is sensitive enough to capture changes in vocal cord coordination. Specifically, we look at jitter, shimmer and breathiness. Researchers have looked at jitter and shimmer based first differences associated with depression affect motor coordination and therefore the disruption of articulatory energy among others [1], [2], [3].

Database

- Analysis was performed using the Mundi database [4] which consists of data collected from 35 physician-referred patients undergoing treatment for depression. The patients were assessed weekly over a period of 6 weeks.
- The speech signals consist of 4 vowels (a, i, u, ae) and 3 utterances where the subjects talked about their emotional & physical states, and ability to function.
- Subject sessions were labeled as depressed if their Hamilton Depression (HAMD) score was 17 or greater, and as non-depressed if their score was 7 or lower, with scores of 8 to 16 excluded because their depression status is ambiguous [1].
- Only 6 patients showed the transition from depressed to non-depressed (see Table 1).

Modulation Rate

- Normally, the peak modulation frequency which is a measure of syllable rate is between 3 and 4 Hz.
- People who are depressed tend to speak at a slower rate and hence their peak modulation frequency may be lower compared to that on days when they are feeling better.
- The peak modulation rate was calculated using the Neural Systems Lab toolbox [5]. An auditory spectrogram of the speech signal is computed which mimics the performance of the cochlea. From this analysis, we get the rate (rate of frequency change in Hz) and scale (inter frequency spacing) of the modulation profile for the vowels, showing prominent clusters of peaks in vowels (left), an unvoiced fricative (center) and a breathy vowel (right) are shown below. Note that the dip in the spectrogram occurs outside a certain tolerance region of other channels whose center frequency was <2000 Hz.

Jitter, Shimmer and Breathiness

- To quantify these parameters, we used the Aperiodicity/Periodicity/Pitch detector [6] that is based on the Average Magnitude Difference Function (AMDF)

\[
\gamma_n(k) = \sum_{m=0}^{n} (x(m+n)e(m) - x(m+n-k)e(m-k))
\]

- For a periodic portion of speech, when k is equal to a pitch period, \(\gamma(k)\) will be close to 0. The resulting AMDF dip will have a value close to 1. We divided the speech spectrum into 60 channels and computed the AMDF for each channel. To measure jitter and shimmer we only consider those channels whose center frequency was < 2000 Hz.
- The AMDF dip profile for a frame from a vowel (left), an unvoiced fricative (center) and a breathy vowel (right) are shown below. Note that the dip profile for the vowels shows prominent clusters of dips at \(T_o\), 2\(T_o\) and 3\(T_o\) where \(T_o\) is the fundamental period of the vowel. The dip profile for the unvoiced fricative, on the other hand, shows no such clusters since the fricative is aperiodic.

Table 2: List of Patients and the days on which they were depressed or not depressed based on HAMD scores

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Depressed</th>
<th>Not-Depressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>Day-14</td>
<td>Day-42</td>
</tr>
<tr>
<td>111</td>
<td>Day-00</td>
<td>Day 14, 28, 47</td>
</tr>
<tr>
<td>123</td>
<td>Day-00</td>
<td>Day-31</td>
</tr>
<tr>
<td>127</td>
<td>Day-00</td>
<td>Day-42</td>
</tr>
<tr>
<td>128</td>
<td>Day-00, 13</td>
<td>Day-27, 41</td>
</tr>
</tbody>
</table>

Table 2 shows that 4 of the 6 speakers show an increase in modulation rate on the days they feel better.

Future Work

- The range of values of jitter and shimmer are speaker dependent. We are currently exploring ways to normalize these values so that we can build a speaker-independent classifier that can distinguish between depressed and not-depressed states based on speech. Our current method normalizes the jitter values by the fundamental pitch period.
- In the next phase of this research, we will look at the patients on the days when their HAMD values lie between 7-17 to understand the variation in our measures on the days when they are not considered to be depressed or not depressed.
- We plan to determine how to measure jitter, shimmer and breathiness in free flowing speech where there may be considerably more acoustic variability due to coarticulation.
- We plan to analyze speech from a wide range of students who are not depressed to get baseline measures for jitter, shimmer, breathiness and syllabic rate.

References