MOOD STATE PREDICTION FROM SPEECH OF VARYING ACOUSTIC QUALITY FOR INDIVIDUALS WITH BIPOLAR DISORDER

John Gideon* *Emily Mower Provost** *Melvin McInnis*†

Departments of: Computer Science and Engineering* and Psychiatry†, University of Michigan

ABSTRACT

Speech contains patterns that can be altered by the mood of an individual. There is an increasing focus on automated methods to monitor speech in patients suffering from mental disorders. However, audio recordings can be modulated in unintended ways when different devices of varying acoustic quality are used, making them difficult to analyze. In order to expand these technologies to more individuals it is necessary to overcome these limitations. This paper explores speech collected from phone recordings for analysis of mood in individuals with bipolar disorder. Two different phones with varying amounts of clipping, loudness, and noise are employed. We describe methodologies for use during preprocessing, feature extraction, and data modeling to correct these differences and make the devices more comparable. Our system performs significantly better than one without these pipeline modifications, making it feasible to further expand distribution of mobile health using speech.

Index Terms— Bipolar Disorder, Mood Modeling, Mobile Health, Speech Analysis

1. INTRODUCTION

Bipolar disorder (BP) is characterized by swings in mood between mania, or heightened mood, and depression, or lowered mood. BP is pervasive, affecting 4% of people in the United States [1]. Both mania and depression profoundly impact the behavior of affected individuals, resulting in potentially devastating economic, social, and professional consequences. The current treatment paradigm involves routine monitoring of individuals through regular clinical visits. However, there are insufficient resources to ensure that all individuals with BP have access to this type of care [2]. This scarcity of available care points to the need for novel approaches to regular mood monitoring and the potential of computational approaches to serve as auxiliary methods. In this paper, we present an investigation into automatic speech analysis using mobile phone conversations as a way to predict mood, as well as the complications that arise from the diversity of real world recordings.

Research has demonstrated that speech patterns are affected by mood and contribute to accurate clinical assessments [3]. For example, both the Hamilton Depression Scale (HAMD) [4] and Young Mania Rating Scale (YMRS) [5] use clinical observations of speech to determine the severity of depression or mania [4, 5]. There is an opportunity to discover how speech cues can be automatically processed to augment objective measures available in clinical assessments. Mobile phones provide an effective platform for naturally monitoring these speech cues and have shown promise for BP [6, 7, 8]. However, changes in recording quality between different types of phones can severely decrease the predictive capabilities of a system. These include clipping, loudness variations, and different levels of background noise.

Much mood speech research has been centered around identifying speech features for recognizing depression. Among these, are pitch, energy, rhythm, and formants [9, 10, 11, 12, 13, 14]. Short pauses and increased pitch have been correlated with mania [10, 12, 14, 15, 16]. However, much of the work in identifying speech associated with mania has focused on differentiating it from schizophrenia and cannot be directly applied [17, 18]. Many mood related studies collected their speech from controlled environments [10, 12, 13] or used a single type of recording device [7, 8, 19] and do not necessarily reflect the variations in background noise and microphone quality present in real world recordings. As such, their models would be difficult to apply to a widely distributed mobile health system.

In this paper, we focus on one of the challenges associated with real-world distributed mood recognition: variability in recording. We examine the differences between the two phones used on this study and analyze preprocessing and modeling methods that allow us to build models of mood across the database as a whole. These methods include declipping [20], noise-robust segmentation [21], feature normalization [13], and multi-task learning [22]. We provide evidence that mood-related changes in speech are captured in this model using the structured assessment calls captured from different phone types. Please see Figure 1 for a system overview.

The novelty of our approach is the investigation into acoustic variations caused by recording with different types of phones and the preprocessing and modeling changes that are necessary to detect mood under these conditions. Our results suggest that this pipeline of methods from the preprocessing and feature extraction up through data modeling can effectively increase the performance of these types of mixed device systems. Considering the baseline AUCs of 0.57±0.25 for manic and 0.64±0.14 for depressed and the significant increase to 0.72±0.20 and 0.75±0.14, respectively, it is clear that there are benefits to carefully improving how data is handled from devices with different acoustics.

2. PRIORI DATASET

The PRIORI Dataset is a collection of smartphone conversational data (reviewed and approved by the Institutional Review Board of the University of Michigan, HUM00052163). The participants were recruited from the HC Prechter Longitudinal Study of Bipolar Disorder at the University of Michigan [23]. The inclusion criteria include a diagnosis of rapid-cycling BP, type I or II. Individuals with substance abuse or neurological illnesses were excluded. Recruitment and participation in the study is ongoing. Participants are enrolled for six to twelve months and are provided with an Android smartphone with the secure recording application (*PRIORI app*) installed. The app runs in the background and turns on whenever a phone call is made, recording only the participant's side of the dialog. The speech data are encrypted in real-time and stored on the phone. The encrypted files are then uploaded to a HIPAA-compliant server.

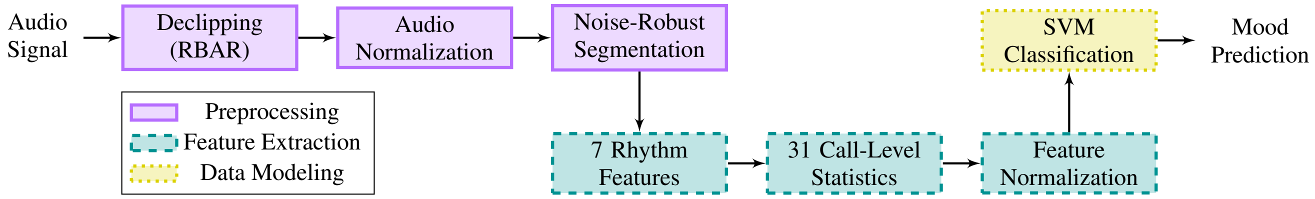

Fig. 1: Audio pipeline divided into three stages of preprocessing (Section 3), feature extraction (Section 4), and data modeling (Section 5).

Phone	#Subjects	#Assess.	%Clipped	RMS	SNR$_{dB}$
S3	18	456	2.74%	0.397	21.2
S5	17	287	0.02%	0.066	25.1
Both	35	743	1.69%	0.269	23.1

Table 1: Differences in data amounts and acoustics between the Galaxy S3 and S5. The percent clipped assessments (Assess.) and the mean percent of samples per call clipped are shown. Root mean square (RMS) values are calculated to show the loudness or each device microphone. Signal to noise ratio (SNR) is calculated as the relative power in the speech verses silence regions in decibels (dB).

Mood	Total	# Per Subject	% Per Subject
Euthymic	275	7.9±7.7	30%
Manic	107	3.1±4.0	12%
Depressed	247	7.1±7.5	28%
Mixed	95	2.7±3.6	13%
Excluded	175	5.0±4.7	17%

Table 2: Distribution of assessment classes of mood. The total number of observations of each mood class is given. The mean and standard deviation of observations for each class per subject is shown, along with the average percentage of each.

2.1. Data Description

The recorded calls are designated into one of two groups: assessment and personal. Participants take part in weekly calls with our study clinicians in which the HAMD and YMRS interviews are conducted. The assessment calls establish a ground truth for the participant's mood over the previous week. The remainder of the data are referred to as personal calls. The personal calls represent all calls that take place outside of the clinical context. These calls are not annotated to ensure patient privacy and are not used in this study.

The PRIORI Bipolar Dataset currently contains 37 participants who have made 34,830 calls over 2,436 hours. Each participant has been on the study for an average of 29.2 weeks with a standard deviation of 16.4 weeks. Additionally, there have been 780 recorded weekly clinical assessments. Only these structured calls are used in this study. 23 of these assessments were transcribed with speech and silence locations to aid in the development of segmentation.

2.2. Phone Model Differences

The Samsung Galaxy series of phones, including the S3, S4, and S5 are used by participants. Only two of the participants were given S4s and their data are excluded from this study. The distribution of subjects with S3s and S5s can be seen in Table 1. The two models of phone include model-specific microphones and processing. One of the effects of this recording and processing is clipping. Clipping occurs most often in the S3, with an average of 2.74% of speech samples at maximum range. This sensitivity is also demonstrated by the average root mean square value of 0.397 for the S3. Additionally, the noise is much more pronounced, as seen in the lower signal to noise ratio of 21.2 dB for the S3.

2.3. Label Assignment

The HAMD and YMRS scales are continuous measures of mood, ranging from a score of 0 (not symptomatic) to 34 (highly symptomatic). In this paper, we treat the prediction problem as classification, binning the HAMD and YMRS into categories of symptomatic (depressed or manic, respectively) and asymptomatic (euthymic). Scores under a threshold of 6 on both scales are assigned a label of euthymic. Scores above 10 on the HAMD and below 6 on YMRS are assigned a label of depressed. Scores above 10 on the YMRS and below 6 on the HAMD are assigned a label of manic. Data in six-ten range on either scale and data with labels above 10 on both scales are excluded. Table 2 shows the class distribution.

The large standard deviations seen in Table 2 demonstrate the widely varying amounts of mood episodes between individuals with BP. Additionally, some individuals have disparities among the proportions of times spent in each mood. For example, one participant experienced 27 weeks of euthymia and two weeks of mania. This imbalance is rectified in the methodology through various weighting schemes used in feature ranking and validation, explained later.

3. PREPROCESSING

As seen in Section 2.2, the two phones used in this study have different acoustic properties. The S3 has more clipping, higher volume, and a sensitivity to background noise. Because of this, it is necessary to carefully preprocess the data before feature extraction using declipping, audio normalization, and noise-robust segmentation in order to make calls from different devices more comparable.

Declipping: The declipping algorithm *Regularlized Blind Amplitude Reconstruction* (RBAR) [20] was used to approximate the original signal. This is a closed form solution approximation of an algorithm called *Constrained Blind Amplitude Reconstruction* (CBAR) [24]. Each algorithm extrapolates the clipped sections of audio beyond their original values, while minimizing the second derivative of the signal, and have been shown to improve the performance of automatic speech recognition [20, 24]. Both algorithms also ignore any unclipped regions, so it should work well in a system suited for audio of variable clipping amounts, as seen in Table 1.

Audio Normalization: The audio signal is scaled by dividing by the maximum absolute value. This ensures that the signal ranges from at most -1 to 1 and is necessary after running declipping, as it extrapolates the signal beyond these bounds. It also ensures that the varying loudness between the two phone types, as seen in Table 1, becomes more comparable.

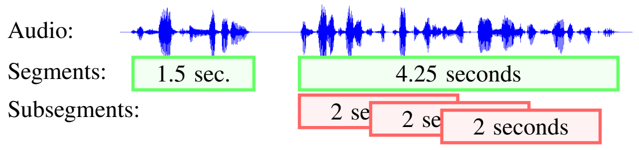

Fig. 2: Segments of speech are found. Segments of 2 seconds or longer are divided into subsegments of 2 seconds in 1 second steps.

Segmentation: Each call is segmented using an algorithm based on [21]. This was selected to be robust to noise variations. This is necessary, given the differences in SNR between the phones (Table 1). The algorithm extracts five signals representative of speech likelihood, including: harmonicity, clarity, prediction gain, periodicity, and perceptual spectral flux. These are then combined using principal component analysis (PCA). The final signal is the largest eigenvalue. It is smoothed by a Hanning window of 25ms and normalized by subtracting by the 5th percentile over the call and dividing by the standard deviation. This ensures that signals from different calls all share a similar baseline for silence. Segments of 25ms are created wherever the combo signal exceeds a 1.8 threshold. Overlapping segments are then merged and any silences less than 700ms are removed. The threshold and minimum silence amounts were found by validating over the transcribed assessments. The segments are further divided into subsegments of 2s with 1s overlap. Segments less than 2s are discarded. Constant window sizes are used to ensure that variations in the features are not caused by changes in segment size [25]. The full segmentation process is shown in Figure 2.

4. FEATURE EXTRACTION

Rhythm Features: Individuals in manic or depressed episodes exhibit changes in the rhythm of their speech [26]. Rhythm features are calculated for each subsegment by first extracting the voicing envelope. The envelope is used to calculate the spectral power ratio and spectral centroid. The envelope is decomposed into two intrinsic mode functions (IMF) using empirical mode decomposition [27]. Tilsen and Arvaniti [25] empirically demonstrated that the extracted IMFs are reflective of syllable- and word-level fluctuations. The IMFs are used to extract five segment-level features: the power ratio between the two IMFs and the mean and standard deviation of the instantaneous frequencies associated with each IMF.

Call-Level Statistics: The seven rhythm features are transformed into call-level features by taking the mean, standard deviation, skewness, kurtosis, minimum, maximum, range, and 1^{st}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}, and 99^{th} percentiles of the subsegment measures. Additionally, the differences between the 50^{th} and 25^{th}, 75^{th} and 50^{th}, 75^{th} and 25^{th}, 90^{th} and 10^{th}, and 99^{th} and 1^{st} percentiles are included. This set is augmented with the percentage of the call that is above 10%, 25%, 50%, 75%, and 90% of the range. Finally, the call-level feature trend is captured by fitting a linear regression model to the features extracted over each segment (R^2, mean error, and mean squared error). This results in a total of 217 features.

Feature Normalization: Call-level features are Z-normalized either (1) globally, using the mean and standard deviation of all training data, or (2) by subject, using the mean and standard deviation of each subject's own data. Previous research has shown that normalization by subject can reduce the disparity between subject feature distributions caused by speaker differences and aid in the detection of mood [13]. This method may also help reduce some of the differences in subject feature distributions due to differences in phones.

5. DATA MODELING

The classification goal is to identify if a given call is (1) from a manic or euthymic episode or (2) from a depressed or euthymic episode. Subjects are only included in analysis if they have at least six total assessments in order to ensure enough data to process features by subject. Additionally, subjects must contain at least two euthymic calls and two manic/depressed calls. This ensures that there is enough data to measure test performance. With these restrictions, 15 subjects are used when considering mania (12 S3s and 3 S5s) and 18 subjects are used when considering depression (11 S3s and 7 S5s).

Support Vector Machines (SVM) [28] are used to classify the speech. SVMs learn a decision boundary between two classes of data with an explicit goal of identifying a boundary that maximally separates the two classes. The classifiers are implemented using both linear and radial basis function (RBF) kernels. Euthymic samples are given a weight equal to the number of manic/depressed samples divided by the number of euthymic samples. Manic/depressed samples are given a weight of one. This ensures that there is no bias towards the mood with more samples by increasing the penalty for misclassification of minority labels. Multi-task SVMs [22] are also used for certain experiments. This algorithm weights the kernel function using a parameter *rho* in order to decrease the importance of data from a different task. In this case, the task is considered to be the phone type. On one extreme, rho can be selected to behave as a normal SVM and consider the tasks to be equal. On the other extreme, the selected rho can consider the tasks to be completely independent.

The models are trained using leave-one-subject-out cross-validation, ensuring that there is no overlap between the speakers used to train and test the system. The model parameters include: kernel type (RBF vs. linear), gamma (RBF only), number of features with respect to a ranked list, cost parameter (C), and rho (multi-task only). The parameter combination is chosen to optimize leave-one-training-subject-out cross-validation, where the contribution of each training subject is proportional to his/her amount of data.

Features are ranked using a heuristic of Weighted Information Gain (WIG). The heuristic was chosen due to the observed subject-specific label imbalance, which may result in the identification of features that are tied to subject identity, rather than mood. This can result in a classifier learning to associate all instances with a single mood state from a biased subject. WIG allows for each sample to be ascribed an importance that ensures both classes contribute equally from each subject. This is implemented using the weighted entropy functions described in [29]. Each sample is given a weight equal to the total number of samples in its subject divided by the number of occurrences of its label in its subject. This ensures that minority and majority samples are given equal weight over each subject, while subjects are given weight proportional to their number of samples.

The system performance was measured using Area Under the Receiver Operating Characteristic Curve (AUC). AUC assesses the ability of a system to correctly rank pairs of instances from opposing classes. It has a chance rating of 0.5 and ideal rating of 1.

6. RESULTS AND DISCUSSION

In this section we demonstrate the ability to differentiate between euthymic and symptomatic moods, despite using two types of mobile phones with different acoustics. The results are presented in Table 3. In addition to reporting the combined test performance of both phone types, test results are broken down into individual types. However, all phones from both types are always used to train models. A paired t-test with a significance of 0.05 is used to compare results to the

Model	Manic AUC	Depressed AUC
S3	0.52±0.22	0.66±0.17
S5	0.78±0.31	0.62±0.09
Both	0.57±0.25	0.64±0.14

(a) No Declipping and Global Normalization (Baseline)

Model	Manic AUC	Depressed AUC
S3	0.68±0.16	0.62±0.14
S5	0.79±0.21	0.69±0.18
Both	**0.70±0.17***	0.65±0.15

(b) RBAR Declipping and Global Normalization

Model	Manic AUC	Depressed AUC
S3	0.73±0.22	0.74±0.10
S5	0.79±0.37	0.80±0.21
Both	**0.74±0.24***	**0.77±0.15***

(c) No Speech Segmentation (Silence Included)

Model	Manic AUC	Depressed AUC
S3	0.66±0.15	0.73±0.15
S5	0.71±0.35	0.78±0.10
Both	**0.67±0.19***	**0.75±0.14***

(d) No Declipping and Subject Normalization

Model	Manic AUC	Depressed AUC
S3	0.67±0.20	0.67±0.21
S5	0.72±0.41	0.65±0.11
Both	**0.68±0.23***	0.66±0.18

(e) Multi-Task SVM Using Baseline Preprocessing

Model	Manic AUC	Depressed AUC
S3	0.71±0.19	0.66±0.14
S5	0.78±0.23	0.79±0.13
Both	**0.72±0.20***	0.71±0.15

(f) Multi-Task SVM Using Best Preprocessing

Table 3: Classification results using various methods. **Bolded*** AUCs denote results significantly better than baseline (paired t-test, p=0.05).

baseline performance and a significant difference is marked with an asterisk and bolded. Significance tests are not performed on individual phone type results, as there are too few samples individually.

Baseline Performance: The baseline system consists of no declipping algorithm and global normalization. The results in Table 3a show an AUC of 0.64±0.14 for depressed and a near chance performance of 0.57±0.25 AUC for manic. However, the three S5s performed better than the S3s in the manic test with 0.78±0.31 AUC. This could indicate that even though the S5 only makes up 20% of the phones, its higher quality recordings allow for it to perform well in testing. Alternately, the speaker population that makes up those subjects using the S5s could be more homogeneous. The S5 continues to outperform the S3 in the rest of the manic experiments.

Evaluation of Declipping: Table 3b shows the results of declipping when using global normalization. While the performance of the depressed tests remain mostly unaffected, the manic test increases significantly to an AUC of 0.70±0.17. This is due to the improvement in the S3, where larger amounts of clipping occurred, as seen in Table 1. We hypothesize that the stronger improvement in manic tests, compared with depressed tests, is due to the fact that manic S3 calls have significantly more clipping than euthymic and depressed S3 calls (unpaired t-test, p=0.05). The percent of clipping in euthymic, manic, and depressed S3 calls are 2.73±1.25%, 3.21±1.13%, and 2.41±1.07%, respectively.

Evaluation of Segmentation: The effect of segmentation was studied in an experiment wherein the segments were no longer identified using the method outlined in Section 3. Instead, the 2 second subsegments were taken over the entire call - silences included. It performed the best of all tests with significant increases from the baselines for both moods (Table 3c). However, we hypothesize that this is actually due to the rhythm features indirectly capturing information about the assessment structure. For example, an individual who is euthymic would have more silence due to their brief interview answers. This highlights one of the potential pitfalls to avoid when working with structured calls to train a model to recognize acoustic aspects of mood. For this reason, it is necessary to use accurate segmentation to avoid these misleading results.

Evaluation of Feature Normalization: Normalization by subject significantly increased the performance of both manic and depressed tests from baseline, as shown in Table 3d. This method has the ability to correct for different feature distributions among speakers, as explained in [13]. These results demonstrate that this correction can also benefit systems with variable recording devices of different quality.

Multi-task SVM Analysis: The use of a multi-task SVM can also rectify some of the effects of different device types. Table 3e shows a significant improvement in manic from baseline by selecting a low value for rho and treating data from different phone types as less informative. Depression does not see much improvement, as a high rho value is selected, indicating that the data is already comparable without preprocessing. This gives further evidence to the reason preprocessing works well for manic but has little effect on depressed. Another multi-task experiment was run using the preprocessing methods that worked best for each mood - RBAR declipping and subject normalization for manic and subject normalization for depressed. These results can be seen in Table 3f, with the highest manic AUC of 0.72±0.20, which is significantly better than baseline. This implies that multi-task learning can be helpful even after preprocessing devices for compatibility, depending on the similarity.

7. CONCLUSION

This paper presents a set of methods to improve the comparability of data collected from across devices of different acoustics. This is essential for any mobile health system using speech that aims to be widely distributed, as the prospect of varying audio quality is unavoidable. Our results demonstrate that through a combination of preprocessing, feature extraction, and data modeling techniques it is possible to mitigate the effects of differing amounts of clipping, loudness, and noise. This is best shown by the increase in performance from the baseline AUCs of 0.57±0.25 for manic and 0.64±0.14 for depressed to the significantly higher AUCs of 0.72±0.20 and 0.75±0.14, respectively. However, there was not a comprehensive solution for both mood types, which indicates the need for careful consideration of all steps along any pipeline.

The ultimate goal will be for the system to be totally passive, requiring no active input from the BP patient or the clinic. Current methods using structured assessments are not enough, as they require weekly interview calls. However, the transition to personal calls will require solutions to many problems, including how to control for the confounding factors of variations in subject symptomatology, episode patterns, and conversational styles. The refinement of techniques developed in this study to increase device comparability may be adaptable to these issues. In particular, it will be necessary to determine how to adapt the system to particular individuals and determine which features are indicative of mood and not some other misleading factor. Although, if effective, it will greatly assist in the way that mental health care is managed.

8. REFERENCES

[1] Ronald C Kessler, Patricia Berglund, Olga Demler, Robert Jin, Kathleen R Merikangas, and Ellen E Walters, “Lifetime prevalence and age-of-onset distributions of dsm-iv disorders in the national comorbidity survey replication,” *Archives of general psychiatry*, vol. 62, no. 6, pp. 593–602, 2005.

[2] Jules Angst, Robert Sellaro, and Felix Angst, “Long-term outcome and mortality of treated versus untreated bipolar and depressed patients: a preliminary report,” *International Journal of Psychiatry in Clinical Practice*, vol. 2, no. 2, pp. 115–119, 1998.

[3] National Institute of Mental Health, “Bipolar disorder in adults,” http://www.nimh.nih.gov/health/publications/bipolar-disorder-in-adults/Bipolar_Disorder_Adults_CL508_144295.pdf, Accessed: September - 2015.

[4] M Hamilton, “Hamilton depression scale,” *ECDEU Assessment Manual For Psychopharmacology, Revised Edition. Rockville, MD: National Institute of Mental Health*, pp. 179–92, 1976.

[5] RC Young, JT Biggs, VE Ziegler, and DA Meyer, “A rating scale for mania: reliability, validity and sensitivity.,” *The British Journal of Psychiatry*, vol. 133, no. 5, pp. 429–435, 1978.

[6] Zahi N Karam, Emily Mower Provost, Satinder Singh, Jennifer Montgomery, Christopher Archer, Gloria Harrington, and Melvin G Mcinnis, “Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech,” .

[7] Robert LiKamWa, Yunxin Liu, Nicholas D Lane, and Lin Zhong, “Moodscope: building a mood sensor from smartphone usage patterns,” in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 2013, pp. 389–402.

[8] Venet Osmani, Alban Maxhuni, Agnes Grünerbl, Paul Lukowicz, Christian Haring, and Oscar Mayora, “Monitoring activity of patients with bipolar disorder using smart phones,” in *Proceedings of International Conference on Advances in Mobile Computing & Multimedia*. ACM, 2013, p. 85.

[9] Daniel Joseph France, Richard G Shiavi, Stephen Silverman, Marilyn Silverman, and D Mitchell Wilkes, “Acoustical properties of speech as indicators of depression and suicidal risk,” *Biomedical Engineering, IEEE Transactions on*, vol. 47, no. 7, pp. 829–837, 2000.

[10] Nicola Vanello, Andrea Guidi, Claudio Gentili, Sandra Werner, Gilles Bertschy, Gaetano Valenza, Antonio Lanata, and Enzo Pasquale Scilingo, “Speech analysis for mood state characterization in bipolar patients,” in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 2104–2107.

[11] Thomas F Quatieri and Nicolas Malyska, “Vocal-source biomarkers for depression: A link to psychomotor activity.,” in *Interspeech*, 2012.

[12] A Guidi, N Vanello, G Bertschy, C Gentili, L Landini, and EP Scilingo, “Automatic analysis of speech f0 contour for the characterization of mood changes in bipolar patients,” *Biomedical Signal Processing and Control*, 2014.

[13] Nicholas Cummins, Julien Epps, Michael Breakspear, and Roland Goecke, “An investigation of depressed speech detection: Features and normalization.,” in *Interspeech*, 2011, pp. 2997–3000.

[14] Ernest H Friedman and Gary G Sanders, “Speech timing of mood disorders,” *Computers in Human Services*, vol. 8, no. 3-4, pp. 121–142, 1991.

[15] Heidi S Resnick and Thomas F Oltmanns, “Hesitation patterns in the speech of thought-disordered schizophrenic and manic patients.,” *Journal of abnormal psychology*, vol. 93, no. 1, pp. 80, 1984.

[16] Michael F Pogue-Geile and Thomas F Oltmanns, “Sentence perception and distractibility in schizophrenic, manic, and depressed patients.,” *Journal of abnormal psychology*, vol. 89, no. 2, pp. 115, 1980.

[17] Michael Alan Taylor, Robyn Reed, and Sheri Berenbaum, “Patterns of speech disorders in schizophrenia and mania.,” *The Journal of nervous and mental disease*, vol. 182, no. 6, pp. 319–326, 1994.

[18] Ralph E Hoffman, Susan Stopek, and Nancy C Andreasen, “A comparative study of manic vs schizophrenic speech disorganization,” *Archives of General Psychiatry*, vol. 43, no. 9, pp. 831–838, 1986.

[19] A Grunerbl, Amir Muaremi, Venet Osmani, Gernot Bahle, Stefan Ohler, Gerhard Tröster, Oscar Mayora, Christian Haring, and Paul Lukowicz, “Smart-phone based recognition of states and state changes in bipolar disorder patients,” 2014.

[20] Mark J Harvilla and Richard M Stern, “Efficient audio declipping using regularized least squares,” .

[21] Seyed Omid Sadjadi and John HL Hansen, “Unsupervised speech activity detection using voicing measures and perceptual spectral flux,” *Signal Processing Letters, IEEE*, vol. 20, no. 3, pp. 197–200, 2013.

[22] Theodoros Evgeniou and Massimiliano Pontil, “Regularized multi–task learning,” in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004, pp. 109–117.

[23] Scott A Langenecker, Erika FH Saunders, Allison M Kade, Michael T Ransom, and Melvin G McInnis, “Intermediate: cognitive phenotypes in bipolar disorder,” *Journal of affective disorders*, vol. 122, no. 3, pp. 285–293, 2010.

[24] Mark J Harvilla and Richard M Stern, “Least squares signal declipping for robust speech recognition,” in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.

[25] Sam Tilsen and Amalia Arvaniti, “Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages,” *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. 628–639, 2013.

[26] Frederick K Goodwin and Kay Redfield Jamison, *Manic-depressive illness: bipolar disorders and recurrent depression*, Oxford University Press, 2007.

[27] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu, “The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[28] Corinna Cortes and Vladimir Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[29] Marek Śmieja, “Weighted approach to general entropy function,” *IMA Journal of Mathematical Control and Information*, p. dnt044, 2014.