2012 International Conference on Conference on Signal Processing Systems (ICSPS 2012)
IPCSIT vol. 58 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V58.2

Analysis of Timing Pattern of Speech as Possible Indicator for Near-Term Suicidal Risk and Depression in Male Patients

Nik Wahidah Hashim [1+], Mitch Wilkes [1], Ronald Salomon [2] and Jared Meggs [2]

[1]Department of Electrical Engineering, Vanderbilt University, Nashville, TN 37235 USA
[2] Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, TN 37212 USA

Abstract. Patients who are diagnosed with depression without appropriate clinical recognition of their hidden suicidal tendencies are at elevated risk of making a suicide attempts. An important clinical problem remains the differentiation between non-suicidal and more lethal episodes of depression. In an effort to find a reliable method that could assist clinicians in risk assessment, information in the speech signal has been found to contain characteristic changes associated with high risk suicidal states. Among numerous characteristics available, this paper addresses the question of which speech signal contribute most to the discrimination between high risk suicidal and depressed patients. Analysis is based on features related to the timing patterns of speech (voiced, unvoiced and silence), specifically the Transition Parameters and Interval Probability Density Functions (PDF). Automatic speech data sets were collected from readings of a standard "rainbow passage" essay. Linear and quadratic classifiers were used to obtain the decision boundary for the pairwise classification and the classifier performance was estimated using the method of Equal-Test-Train data, K-cross validation and Jackknife. Use of the jackknife procedure as a means to measure a classifier performance for all-data classification, within the first data set, revealed a single Transition Parameter to be a significant discriminator with 74% correct classification. Certain combinations of interval pdf for voiced and silence yielded 75%-100% correct classification. Results achieved 83% correct classification for a single Transition Parameter and 94% overall correct classification with 100% high risk prediction when using a single Transition Parameter combined with a single bin from voiced or silence interval pdf.

Keywords: classification, suicide, depression, speech pauses, prediction, voiced

1. Introduction

Suicide continues to be a major concern to public health worldwide. To view the importance of this issue, on average one suicide occurs every 14.2 minutes in the United States. On average, one young person dies in suicide every two hours. Non-completed suicide attempts numbered 922,725 during this interval, translating to an average of one attempt every 34 seconds. Male exhibit a greater risk of death from suicide as a gender wise analysis reported a ratio of 3.7 male to 1 female by suicide [1]. Despite decades of research, accurate prediction of suicide and imminent suicide attempts still remains elusive. Inaccurate assessment tools may mislead clinicians to believe that patients who are actually at imminent risk of committing suicide are experiencing a less severe depressive disorder. Identification of an imminent suicidal risk at an early stage may allow the patient to receive proper hospitalization and treatment. Commonly used suicide risk assessment tools comprise a series of questionnaires and checklists with rating scales that can be evaluated with reliability by trained clinicians [2]. However, even clinicians with advanced psychiatric training may benefit from information from a second source that can give quantitative results and yield a better detection of imminent risk.

Speech contains implicitly hidden information that reflects psychological states, including affective states or the presence of diseases such as Parkinson's [3]-[9]. Previous studies have suggested that

+ Corresponding author. Tel.: +1 615 335 5668
E-mail address: nik.nur.wahidah.nik.hashim@vanderbilt.edu

depression is associated with distinctive speech patterns. Among the characteristics are decreases in intonation, phonation stress, loudness, inflection and intensity, increase in duration of speech, sluggishness in articulation, narrow pitch range, monotonous, and lack in vitality [10]-[12]. The previous studies relating to investigation of the vocal cues for depression and imminent suicidal risk detection often revolves around spectrum-based measures of the voice signal [3]-[6].

Several studies have observed the correlation between different characteristics of prosody and speech rate with major depression. According to Monrad-Krohn [13], the definition of prosody consists of the normal variation of pitch, stress and rhythm which includes silent intervals of pauses. Alpert [14] separated speech productivity and pausing under the term fluency and defined prosody as emphasis and inflection. Speech rate comprises a combination of phonation length (voiced), frequency of short pauses and the duration of pauses. The study reported herein focuses on the use of certain features related to the rhythm, fluency of speech and speech rate in an attempt to capture information related to voiced and silent pauses and quantify these features as an indicator of depressed and suicidal speech.

In an earlier study conducted by Szabadi[15], both phonation and pause time were extracted through a patient's counts from 1 to 10 in a non-spontaneous 'automatic' speech. Depressed patients showed no change in phonation time but exhibited a decrease in pause time during the period of improvement compared to a constant length of pause time in the control and healthy patients. Nilsonne [16] used a collection of recordings from patients reading a standardized text at their own speed. Three relevant parameters collected were the time between the first and the last vocalization, summation of vocalization and summation of pauses. Most patients were reported to read more slowly during depression and displayed a decrease in pause time as patients became less depressed. Prolonged pauses that occurred in between a series of questions and answers were examined by Alpert [14]. Depressed patients demonstrated longer pause duration in between interviewer's questions and patient's answers. Mundt [17] investigated both automatic and free speech. Results indicated that total pause time in automatic speech had better correlation with depression as opposed to using free speech, but a reverse effect was observed in pause variability and voice-to-pause ratio.

In this paper, a new features based on the timing patterns of speech that consists of voiced/silence transition probabilities and the distribution of interval lengths of voiced and silent pauses are proposed for the analysis of vocal cues for depressed (DP) and high risk suicidal (HR) detection. The previous research mostly studied the average or total length of pauses. Instead, this study looks at the interval histograms of voiced, unvoiced and silence pauses which may reveal more information by their shape or certain characteristics within their histograms. Some information in the Transition Parameters and Interval pdf are related to the pause lengths, but we are looking at the information in a different way.

2. Database Collection

All recording sessions were conducted at the Vanderbilt Psychiatric Hospital in a nearly closed-room environment to minimize the disturbance of background noise. Patients were asked to read from a standardized "rainbow passage" which contains every sound in the English language and is considered to be phonetically balanced with the ratios of assorted phonemes similar to the ones in normal speech [18]. Variations in phonemes and articulation can almost be eliminated because each patient was reading the same passage.

Table 1: The number of patients (male and female) reading sessions

Database A	**HR**	**DP**
Total number of patients	7	12
Database B	**HR**	**DP**
Total number of recordings	18	
Total number of patients (total number of HR)	8	
Total number of patients with 3 sessions	4	
Total number of patients with 2 sessions	2	
Total number of patients with 1 session	2	

Two types of databases were used for this study. All speech samples were digitized at 44.1 kHz sampling rate for both databases. Table 1 shows the number of reading sessions. In the first database (Database A), audio acquisitions were made using a high-quality Audix SCX-one cardioid microphone. The second

database (Database B) was collected from each patient for at most three distinct sessions over an interval of days after receiving treatments. All patients were labeled as high risk suicidal during their first session and the state of the patient during the next recording session were made blind to the researcher and categorized as *others*. In this case, *others*indicate that patients are no longer considered high risk. Audio acquisitions were made using a portable high-quality field recorder, a TASCAM DR-1.

In the pre-processing stage, recordings were edited using Audacity 2.0.1 to remove any identifying information, interviewer's voice and background noises. Pauses are important information that needed to be preserved thus, they were kept unedited. The sampled signals were divided into 40ms non-overlapping frames and a voiced/unvoiced/silence decision was made for each frame based on the method in [4].

3. Methodology

3.1. Feature extraction

3.1.1 Transition parameters

Consider a sampled signal that contains a combination of voiced, unvoiced and silence frames which can be labeled as three different states of t = 1, 2, and 3 respectively. The probabilities were estimated with a method of an observable discrete-time Markov process [19] implemented using the statistics toolbox available in MATLAB. The state and sequence are initially known with the emission probabilities set to be the matrix identity. One of the output parameters given is the estimated transition matrix which in this case is a three-by-three matrix$t_{ij}\{i = 1,2,3\ and\ j = 1,2,3\}$. Each row i is a conditional probability density function given that you are in state i and column j is the next possible state. The nine features were concatenated into a row vector representing each patient as $\{t_{11}, t_{12}, t_{13}, t_{21}, t_{22}, t_{23}, t_{31}, t_{32}, t_{33}\}$.

3.1.2 Interval length pdf

The pdf is estimated by counting the number of occurrence for a consecutive number of 40ms frames per interval, that belongs to voiced, unvoiced or silence within a sampled signal. The implementation procedure to obtain the Interval Length pdf for a sampled signal is as follows:

1. For the Interval Length pdf of voiced intervals, find all the voiced intervals in the signal.
2. Count all the intervals of length one (40ms) and divide by the total number of voiced intervals for normalization.
3. Do the same for voiced interval of lengths two (80ms) through 24 (0.96 sec) and normalize.
4. Count all the intervals of length 25 (one sec or longer) and normalize. At this point, you have a vector of interval length percentages, i.e., a histogram.
5. Repeat step 1-4 for unvoiced (labeled '2') and silence (labeled '3') for a maximum of 0.24s (six frames per interval) and 2.0s (50 frames per interval) respectively.

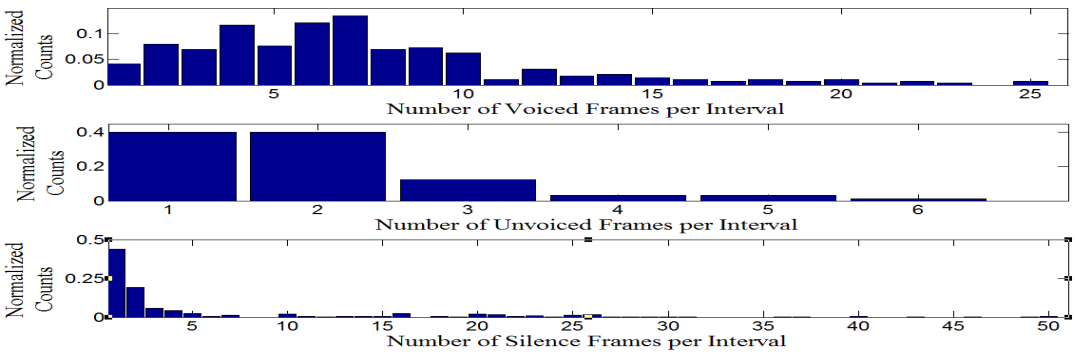


Fig. 1: Examples of the voiced, unvoiced and silence interval pdf distributions

Examples of the resulting pdfs are shown in Fig. 1. Each bin is treated as a feature. For the silence interval distribution, every five consecutive interval ratios were combined in order to reduce the number of features from 50 to 10

3.2 Classification

The decision boundaries for the two-class classification of high risk suicidal and depressed were obtained using a quadratic and linear classifier. The resampling methods that were adopted in this research were Equal Test-Train, Jackknife (Leave-One-Out) and Cross-Validation. Using Equal Test-Train method, all data in the training set are also used for testing to show whether the data can be separated. The use of the jackknife is to show whether information obtained from or within a subpopulation can predict the behavior of the unknown individual. The procedure involves leaving out one patient's data from the data set and develops a training data set with the remaining *N*-1 patients. Cross-validation is an effective resampling method without replacement for the problem of small data sets. The data sets are partitioned into two samples sets of randomly chosen 30% testing data and70% training data. This method was performed iteratively for 100 runs and the averages for all outputs were computed.

The analysis was divided into two stages. In the first stage, classification analysis was performed within Database A on a single and multiple combinations of features using the three methods of resampling within each feature category (i.e., Transition Parameters, voiced, unvoiced and silence interval pdf). Features that yielded the best classification result within each category were then combined

For the second stage, features that were recognized to perform well according to analysis in stage 1 were then used to identify high risk recordings in Database B. Classification of suicidal/others were implemented by treating all patients in Database A as the training data and each one recording in Database B as the test data. This method will determine whether it is possible to classify patients from Database B with prior knowledge from a different subpopulation (Database A).

4 Results and Discussion

4.1 Stage 1: Classification of High Risk Suicidal and Depressed Speech in Male Reading

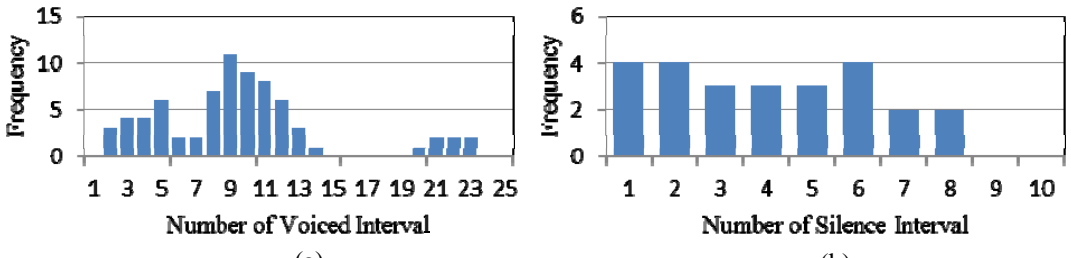
4.1.1 Transition parameter

Table 2 presents results obtained from an analysis of a single feature Silence-to-Voiced (t_{31}) from Transition Parameters that provided the best high risk and depressed classification. Classification using linear and quadratic classifiers yielded nearly similar results. All-Data percentage indicates vectors that are correctly classified over both groups. High risk and depressed percentages denote the percentage of vectors that are correctly classified within each group respectively.

Table 2: Results for high risk and depressed male automatic speech classification using Transition Parameters

Transition Parameter	Feature: **Silence-to-Voiced (t_{31})**		
	All-Data %	**High Risk %**	**Depressed %**
Equal-Test-Train	74	71	75
Jackknife	74	71	75
Cross-Validation	73	73	72

Results indicate that the classifier performed equally well in classifying both high risk and depressedfor all methods of resampling. Approximately five of the seven (~70%) high risk suicidal patients were correctly classified as suicidal and about nine out of 12 (~75%) depressed patients were correctly classified as depressed using all methods of resampling.



(a) (b)

Fig. 2: Histogram of the individual (a) 25 voiced interval ratios and (b) 10 silence interval ratios that contributed 75% to 100% correct jackknife classification using a single and/or combination of features within male high risk and depressed.

Analysis of Interval pdfs were divided into three; voiced, unvoiced and silence. Classification on unvoiced features did not yield good results. For voiced and silence intervals, classification was performed with a single feature (i.e., a histogram bin) from the collection of features in the 25 and 10 bins respectively and also any possible combinations of two and three features within each group. The number of occurrences that a single bin contributes to a single and/or combination of features classification within the range of 75% to 100% correct classification for high risk and depressed using the jackknife are represented in Fig. 2(a) and Fig. 2(b) for voiced and silence intervals.

The most discriminating information occurred when patients hold their vowels for a range of time intervals from 0.16s (eight consecutive frames) to 0.48s (12 consecutive frames) with a peak at an interval of 0.36s (nine consecutive frames). On the other hand, silence pauses that occurred within an approximately 40ms (one frame) to 1.2s (30 consecutive frames) time interval contained most of the information relating to the variability characteristics between high risk and depressed.

4.1.2 Combined feature set

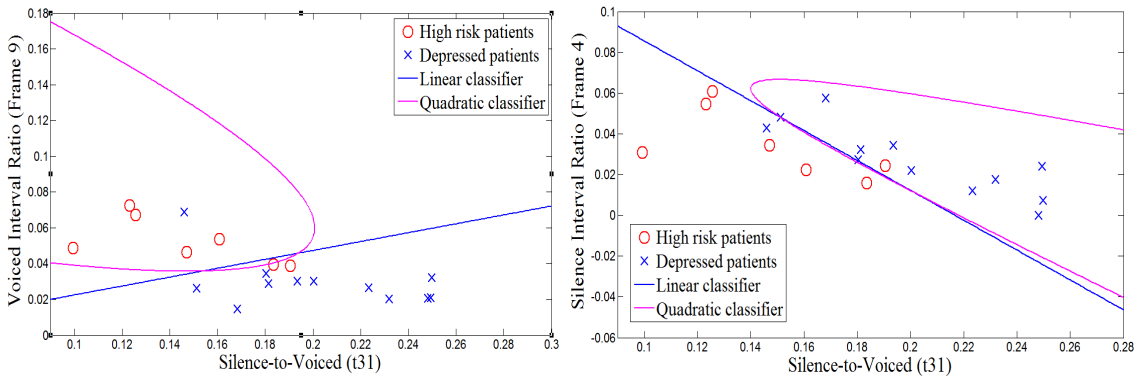
The variability that exists within each feature can either complement or nullify each other.Classification analysis was performed on the combination of Silence-to-Voiced (t_{31}) with each single feature of eight to 12 voiced frames per interval and one to six silence frames per interval.

Table 3: Results of the combined feature sets classification for high risk and depressed male automatic speech

	Feature: t_{31} + **Voiced9**			
	All-Data %	**High Risk %**	**Depressed %**	**Classifier**
Equal-Test-Train	84	71	92	Linear/Quadratic
Jackknife	84	71	92	Linear
Cross-Validation	79	71	88	Linear
	Feature: t_{31} + **Silence4**			
	All-Data %	**High Risk %**	**Depressed %**	**Classifier**
Equal-Test-Train	89	86	92	Linear/Quadratic
Jackknife	74	71	75	Linear
Cross-Validation	79	72	85	Linear

As shown in Table 3, the ninth voiced interval (Voiced9) and fourth silence interval (Silence4) produced the best classification when combined with Silence-to-Voiced (t_{31}). The overall results showed that the classifier performed better on depressed compared to high risk suicidal.

Figure 3(a) and 3(b) plot the distribution of high risk and depressed patients using the combined feature set. By observation, the distributions of high risk and depressed were distinct from each other and vectors that are misclassified were fairly close to the boundary except for one of the depressed patients as shown in figure 3(a).



(a) (b)

Fig. 3: Plot of the high risk and depressed patient distribution for the combined feature set of Voiced-to-Silence (t31) with (a) voiced interval ratios in frame 9 and with (b) silence interval ratios in frame 4 using linear and quadratic discriminant classiier.

4.2 Stage 2: Analysis of Classification between Two Populations

Classifiers using features obtained from analysis in stage 1 that were trained on Database A, were then tested on Database B for identifying high risk suicidal recordings. The result of the classification is shown in Table 4. Using the feature of Silence-to-Voiced (t_{31}) alone, seven out of eight recordings that were labeled as high risk were successfully identified. The results improve significantly to a perfect classification of high risk suicidal when Silence-to-Voiced (t_{31}) was combined with Voiced9 and Silence4. Remarkably, the classifier that was trained using only Voiced9 produced the same result as the combined features.

Table 4: Results of the tested classifier for the identification of high risk suicidal recordings in male patient database B

Feature Combination	All Data %	High Risk %	Others %	Classifier
t_{31}	83	86	82	Quadratic
Voiced9	94	100	91	Linear/Quadratic
t_{31} + Voiced9	94	100	91	Linear/Quadratic
t_{31} + Silence4	89	100	82	Quadratic

5 Conclusion

Two features of Transition Parameters and Interval PDF that relates to the timing pattern of speech were investigated and discussed in this report. The features are not affected by the acoustic content of the speech since it is derived from a pattern-based speech. The results of this investigation correlate with the previous findings where it was shown that features relating to voice and silence from Transition Parameters and Interval pdf provided prominent results in classification of high risk suicidal and depressed patients. Thefact that only one or two parameters were able to produce the quality of discrimination and generate such strong performance across two datasets recorded with different devices strengthens the arguments that the features are robust and that these results are not coming from over modeling. The advantage of the simple and small features worked well with the small amount of data set.It is a strong indication that there are significant information within these parameters.

6 References

[1] J. L. McIntosh, U.S.A Suicide: 2009 Official Final Data, *American Association of Suicidology* (2010), http://www.suicidology.org. Accessed 26 July 2012

[2] C. M. Perlman, E. Neufeld, L. Martin, M. Goy, and J. P. Hirdes, Suicide Risk Assessment Inventory: A Resource Guide for Canadian Health care Organizations, Toronto, *ON: Ontario Hospital Association and Canadian Patient Safety Institute*, 2011.

[3] D. J. France, Acoustical properties of speech as indicators of depression and suicidal risk, Ph.D., Thesis, Vanderbilt University, 1997.

[4] A. Ozdas, Analysis of Paralinguistic Properties of Speech for Near-term Suicidal Risk Assessment, Ph.D., Thesis, Vanderbilt University, 2001.

[5] T. Yingthawornsuk, "Acoustic Analysis of Vocal Output Characteristics for Suicidal Risk Assessment", Ph.D., Thesis, Vanderbilt University, 2007.

[6] H. K. Keskinpala, Analysis of Spectral Properties of Speech for Detecting Suicide Risk and Impact of Gender Specific Differences, PhD Thesis, Vanderbilt University, 2011

[7] K. R. Scherer, Vocal affect expression: A review and model for the future research, *Psychological Bulletin*, vol. 99, pp. 143-145 (1986)

[8] D. Ververidis, C. Kotropoulos, Emotional Speech Recognition: Resources, Features and Methods, *Speech Communication*, vol. 48, pp. 1162-1181, 2006.

[9] H. K. Rouzbahani, M. R. Daliri, Diagnosis of Parkinson's Disease in Human using Voice Signal, *Basic and Clinical Neuroscience*, 2(3): 12-20, 2011.

[10] A. Askenfelt, S. Nilsonne, Voice Analysis in Depressed Patients: Rate of Change of Fundamental Frequency Related to Mental State, *Quarterly Progress and Status Report*, vol. 21, pp. 2-3, 1980.

[11] H. Ellgring, K. R. Scherer, Vocal Indication of Mood Change in Depression, *Journal of Nonverbal Behavior*, 20(2): 83-110, 1996.

[12] J. K. Darby, H. Hollien, Vocal and Speech Patterns of Depressive Patients, *International Journal of Phoniatrics, Speech Therapy and Communication Pathology*, 29(4): 279-91, 1997.

[13] G. H. Monrad-Krohn, The Third Element of Speech: Prosody in the Neuro-Psychiatric Clinic, *The British Journal of Psychiatry*, vol. 103, pp. 326-331, 1957.

[14] M. Alpert, E. R. Pouget, R. R. Silva, Reflections of depression in acoustic measures of the patient's speech, *Journal of Affective Disorders*, vol. 66, pp. 59-69, 2001.

[15] E. Szabadi, C. M. Bradshaw, J. A. O. Besson, Elongation of Pause-Time in Speech: A simple, Objective Measure of Motor Retardation in Depression, *The British Journal of Psychiatry*, vol. 129, pp. 592-597, 1976.

[16] A. Nilsonne, Acoustic Analysis of Speech Variables during Depression and After Improvement, *Acta. Psychiatr. Scand*, 76(3): 235-45, 1987.

[17] J. C. Mundt, P. J. Snyder, M. S. Cannizzaro, K. Chappie, D. S. Geralts, Voice Acoustic Measures of Depression Severity and Treatment Response Collected Via Interactive Voice Response (IVR) Technology, *Journal of Neurolinguistic*, vol. 20, pp. 50-64, 2007.

[18] International Phonetic Association, Phonetic description and the IPA chart, Handbook of the International Phonetic Association: a guide to the use of international phonetic alphabet, (*Cambridge University Press*, 1999 in press)

[19] L. R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceedings of the IEEE*, 77(2): pp. 257-286, 1989.

Nik Nur Wahidah Nik Hashim received the B.S. and M.S. degree in electrical engineering from Vanderbilt University, Nashville, TN in 2009 and 2011 respectively. She is currently working towards her Ph.D. degree in electrical engineering at Vanderbilt University, Nashville, TN. She has been assigned as a Graduate Teaching Assistant and Graduate Research Assistant while at Vanderbilt. Her research interests include speech processing, digital signal processing and pattern recognition.

Don Mitchell Wilkes (S'79–M'87) received the B.S.E.E. degree from Florida Atlantic University, and the M.S.E.E. and Ph.D. degrees from the Georgia Institute of Technology, Atlanta, in 1981, 1984, and 1987, respectively. From 1983 to 1984, he was a Graduate Teaching Assistant at Georgia Institute of Technology, and from 1984 to 1987, he was a Graduate Research Assistant. From August 1987 to June 1994, he was an Assistant Professor of electrical engineering, and from June 1994 to the present, he has been an Associate Professor at Vanderbilt University, Nashville, TN. His research interests include intelligent service robotics, image processing and computer vision, digital signal processing and signal modelling.

Ronald Murray Salomon received the B.S. degree from Massachusetts Institute of Technology, and the M.D. degree from Universite De L'Etat A Liege, in 1976 and 1983, respectively. He is currently an Associate Professor of Psychiatry in Vanderbilt University School of Medicine. He has been a member of the faculty since 1995. He directs the Psychiatry Clerkship for the Medical School, the Grand Rounds Program, and the Adult Division. He is also a member of the Vanderbilt Institute for Clinical and Translational Research Scientific Advisory Committee, and investigator in the Vanderbilt Kennedy Center for Research on Human Development and also in the Department of Psychiatry's Psychiatric Neuroimaging Program.

Jared Scott Meggs received the B.S. degree from Bethel College and the M.S. degree from Lipscomb University in 2008 and 2011, respectively. Since, he has received training in basic science and clinical research at Vanderbilt University Medical Center, where he currently provides research support for the department of psychiatry. His research interests include neurobiology of depression, neuromodulation for psychiatric disorders, and suicidality.