Objective Estimation of Suicidal Risk using Vocal Output Characteristics

Yingthawornsuk T[1], Kaymaz Keskinpala H[1], France D[3], Wilkes DM[1], Shiavi RG[1,2], Salomon RM[4].

[1]Department of Electrical and Computer Engineering, Vanderbilt University, Nashville, TN, USA
[2]Department of Biomedical Engineering, Vanderbilt University, Nashville, TN, USA
[3]Department of Anesthesiology, Vanderbilt University Hospital, Nashville, TN, USA
[4]Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, TN, USA

Abstract

Vocal output characteristics of speech have previously been identified as possible cues to the assessment of suicide risk, and there is evidence that certain vocal parameters may be used as suicidal discriminators. The acoustic properties of male speech samples comprised of individuals carrying diagnoses of depression, suicide risk, and remission were analyzed and statistically compared.

The male sample contained 10 high-risk suicidal patients, 13 depressed patients, and 9 remitted patients. Acoustic analyses of voiced power distribution were performed on speech samples extracted from audio recordings collected from the patients during clinical interviews. Features derived from the power spectral densities were found to be powerful discriminators of class membership in both studies of interview and reading passage sessions. The results support theories that identify psychomotor disturbances as central elements in depression and suicidality.

Index Terms: suicidal speech, depression, power spectral density

1. Introduction

Suicide is a common outcome in persons with serious mental disorders. However, it remains a phenomenon that is underresearched and poorly understood. Moreover, methods to help to identify persons who are at elevated risk are sorely needed in clinical practice. This study represents an attempt to identify characteristic vocal patterns in persons with imminent suicidal potential which could lead to the development of new technology to aid in the assessment of suicidal potential. This project brings together investigators from the divergent disciplines of Psychiatry and Biomedical Engineering to study vocal acoustic properties in suicidal states. We will contrast three study groups: near-term suicidal, depressed, and remitted.

The present study of vocal acoustic properties in suicidal states will use tightly controlled recruitment and recording conditions to replicate and extend findings from recordings made in previous and ongoing studies in both uncontrolled and acoustically controlled clinical interview settings.

In published pilot studies [2], [4], [5], analytical techniques have been developed to determine if subjects were in one of three mental states: healthy control, non-suicidal depressed, or high-risk suicidal. The initial sets of recordings used for these published analyses were made in a wide variety of clinical and technical conditions, without the advantages of an acoustically controlled environment or modern high fidelity equipment. Most were recorded in the 1960's through '80's by a clinical practitioner (the late Dr. S. Silverman), who routinely taped his therapy sessions. He assembled this set of tapes for just such studies of acoustical characteristics of suicidal speech, which he strongly believed could produce a clinical tool for detection of high-risk individuals. Each selected tape predated a known subsequent suicide attempt with high lethality or completed suicide. Comparison subject speech was taken from the same tapes (e.g. the interviewer's voice) or from recordings made later in more controlled environments. The comparison subjects were clinically diagnosed and assigned to groups of either healthy controls or non-suicidal depression. In the early studies using these clinical tapes, analysis focused on segments in the high-risk recordings selected by Dr. Silverman as evocative of suicidal speech sounds. The comparison tapes were sampled at random. With this method, diagnostic groupings were successfully separable using parameters of vocal acoustics.

Vocal cues have been used as indicators in diagnosing the syndrome underlying a person's abnormal behavior or emotional state by experienced clinicians [2, 6], but these skills are not in widespread clinical use. Considerable evidence suggests that emotional arousal produces changes in the speech production scheme by affecting the respiratory, phonatory, and articulatory processes that in turn are encoded in the acoustic signal [7]. Emotional arousal produces a tonic activation of striated musculature, and the sympathetic, and parasympathetic nervous systems [8]. Changes in heart rate, blood pressure, respiratory patterns, muscle tension, and motor activity transiently alter respiratory, phonatory, and articulatory functions in speech production [9] in an *acutely* state-related fashion, directly tied to emotions. Consequently, emotional disturbances can be expected to cause measurable changes in speech parameters. Certain changes in speech parameters may be specific to near-term suicidal states.

Emotional content of the voice can be associated with acoustical variables such as the level, range, contour, and perturbation of the fundamental frequency, the vocal energy, the distribution of energy in the frequency spectrum, the location, bandwidth and intensity of the formant frequencies, and a variety of temporal measures [10]. Research has shown that depression has a major effect on the acoustic characteristics of voice as compared to normal controls. Prosody is slower and the energy in the speech is distributed differently over the frequency range between 0 and 2,000 Hz. Recently some

research has been reported to show that suicide also has some effect on vocal characteristics.

This paper is organized as follows: Section 2 provides a detailed description of database, the patient populations involved, the acoustical feature extraction, and statistical analyses performed on the processed data. Section 3 presents the results of the two session studies and compares the acoustical properties of suicidal, depressed, and remitted speech. Sections 4 and 5 discuss and conclude some of the salient features of the results.

2. Methodology

2.1. Database

In this research, the recordings were obtained from three different patient groups; high–risk suicidal patients, depressed patients, remitted patients. Each study subject from each patient group has two types of speech samples recorded. They are speech samples from the interviews with a therapist and the speech samples from reading a predetermined part of a book. The unedited speech was randomly extracted from the interview session and reading passage session to represent each subject. During the reading session, each patient read the standardized text, the "Rainbow Passage" [3], which is used in speech science since it contains all of the normal sounds in spoken English and it is phonetically balanced.

A portable audio data acquisition system which was used for this study consists of a Sony VAIO laptop computer containing Pentium IV 2 GHz CPU, 512 Mb memory, 60 GB hard drive, 20X CD/DVD read/write unit, 250 Gb external hard drive, Windows XP OS, and ProTools LE digital audio editor; Digital Audio Mbox for audio signal acquisition; and Audix SCX-one cartiod microphone.

The recordings of the 13 male depressed patients, 10 male suicidal patients, and 9 male remitted patients were obtained from ongoing study supported by American Foundation for Suicide Prevention. The ages of the patients were between 25 and 65 years. All speech signals were digitized by using a 16-bit analog to digital converter with a sampling rate of 10 kHz with an anti-aliasing filter (i.e., 5-kHz low-pass). The background noise and the voices other than the patient's voice were removed by using the GoldWave v.5.08 audio editor. This software was also used to remove the silences which were longer than 0.5 seconds for getting a continuous speech record. The preprocessing is finished by dividing the edited continuous speech into 20-seconds segments. For minimizing the introduction of spurious frequency effects resulting from abrupt transitions in the edited speech, the segmentation points were selected at zero crossings or at the beginning of the pauses in the edited continuous speech.

Two steps of preprocessing were used. First all speech segments were tested for voicing and only voiced speech samples were kept for further analysis. Second, all speech signals were detrended and normalized to have a variance of 1 before analysis to compensate for possible differences in recording level among subjects. For each patient the length of the voiced interview speech was approximately 8 minutes and the reading speech was approximately 2 minutes.

2.2. Feature Extraction

Power spectral densities (PSD's) of the voiced speech were obtained by using the classical method of PSD estimation based on Welch method with non–overlapping 100–point Hamming windows. The algorithm was written in MATLAB with using 1024-point fast Fourier transforms (FFT) to estimate the spectra with 40-ms windowing over each 20-second segment [2]. Six features were calculated. Four were the power in the four different frequency ranges: from 0 Hz to 500 Hz, 500 Hz to 1000 Hz, 1000 Hz to 1500 Hz, and finally from 1500 Hz to 2000 Hz. The other two were the value of peak power and the frequency of the peak power. For each of the 500 Hz sub-bands (PSD_1, PSD_2, PSD_3, PSD_4), the percentages of the total power were calculated and stored.

2.3. Comparative Statistical Classification of Class Features

All estimated acoustical features were arranged and stored in matrix form for statistical analysis. Each of the five acoustical features (i.e., peak power, peak location, PSD_1, PSD_2, PSD_3) formed an output vector of features representing a 20-second segment for each subject. The PSD_4 were not taken into this comparative statistical classification due to the property of linear dependency among sub-bands. Each output matrix contained N rows and M columns (N x M matrix), where N was the number of means obtained from each 20-second segment of voiced speech and M was the number of acoustical features. Mathematically, the suicidal, depressed, and remitted classes were defined into three matrices. The parameter matrix representing each class were imported into the On-Line Pattern Analysis System (PcOLPARS, PAR Government Systems, La Jolla, CA), and the SYSTAT (SPSS Inc., Chicago, IL) statistical package for feature analyses and discriminant analyses. Projection analyses and quantile-quantile (Q-Q) plots were used to check the assumption that the three classes of data were normally distributed. Coordinate, eigenvalue, and Fisher Pair-wise projection algorithms were employed in PcOLPARS to verify that each data set exhibited the elliptical unimodal scatter characteristic of multivariate normal distributions. Q-Q plots were performed in SYSTAT to test the marginal normality of each univariate feature.

For this study, pair-wise, (i.e., suicidal-depressed, depressed-remitted, remitted-suicidal) statistical analyses were performed on set of all vocal parameter vectors. All features were combined to design a classifier. Classification score and performance (i.e., measures of sensitivity, specificity, positive predictive (PPV), negative predictive (NPV) values) were calculated. The pair-wise statistical analyses included the calculation and comparison of class covariance matrices, comparison of class vocal parameter features using analysis of variance (ANOVA), application of quadratic discriminators with using the hold-one-out method, and 95% confidence interval were used in all statistical analyses. The *hold-one-out* or *Jackknife* method was used in this discriminant analysis to compensate for the small (N<30) class sizes. All discriminant analyses were performed in SYSTAT.

3. Results

Means and standard deviations of the set of features, peak power, peak location (Hz), and the percentages of total power from each categorized group of patients during the interview were summarized in Table I. Suicidal speech was characterized by peak location which was significantly lower frequency when compared to remitted speech. These can be also seen for the reading passage session summarized in Table II.

Table I. *PSD Statistics for Male (Interview Session)*

	Suicidal	Depressed	Remitted
Peak Power	(20.88, 2.70)	(20.63, 3.34)	(20.92, 1.70)
Peak Location (Hz)	(284.47, 84.89)	(292.02, 58.55)	(331.17, 67.98)
% Power in PSD_1	(0.79, 0.08)	(0.79, 0.08)	(0.74, 0.05)
% Power in PSD_2	(0.19, 0.07)	(0.18, 0.08)	(0.23, 0.04)
% Power in PSD_3	(0.01, 0.01)	(0.03, 0.02)	(0.02, 0.01)

Mean and standard deviation values are presented

Table II. *PSD Statistics for Male (Reading Session)*

	Suicidal	Depressed	Remitted
Peak Power	(21.52, 2.22)	(20.80, 1.59)	(21.53, 2.08)
Peak Location (Hz)	(298.83, 112.84)	(296.10, 66.11)	(351.65, 74.24)
% Power in PSD_1	(0.78, 0.08)	(0.82, 0.06)	(0.75, 0.09)
% Power in PSD_2	(0.19, 0.08)	(0.16, 0.05)	(0.23, 0.09)
% Power in PSD_3	(0.01, 0.01)	(0.02, 0.01)	(0.01, 0.01)

Mean and standard deviation values are presented

The depressed speech exhibited elevated PSD_1, PSD_3 and reduced PSD_2 for interview session when compared to the remitted speech. This trend can be also noticed for the reading passage session. For the results of the suicidal-remitted pair-wise study, the percentage of the total power in the 0 to 500-Hz sub-band (PSD_1) was reduced for remitted speech while the percentage of power in the higher sub-bands (PSD_2 and PSD_3) increased. These can be also noticed for the reading passage session except PSD_3. It indicated no significant difference for the percentage of total power between groups.

Table III.

Sensitivity, Specificity, Positive Predictive (PPV), and Negative Predictive (NPV) Values for Male Pair-wise (Interview Session) and Classification Analyses

Pair-wise Groups	%Classification	Sensitivity	Specificity	PPV	NPV
Suicidal/Depressed	77	0.89	0.63	0.76	0.80
Depressed/Remitted	94	0.94	0.94	0.94	0.94
Suicidal/Remitted	85	0.91	0.76	0.84	0.86

The results of pair-wise discriminant analyses performed on the male study populations were summarized in Tables III and IV for interview and reading passage session, respectively. For interview speech session the depressed patients were fairly well differentiated from suicidal patients (77%). However, the depressed patients and suicidal patients were effectively (i.e. 94%, 85%) differentiated from remitted patients. These discriminant analyses were primarily on the basis of PSD features. Table III also summarized the cumulative classification scores obtained for each statistical analysis using the discriminating features and the performance characteristics of each discriminant function with scores of sensitivity, specificity, PPV, and NPV. To calculate a measure of sensitivity, the clinical measurement of suicidal speech from the confusion matrix of classification was selected as the conditional parameter for suicidal-depressed pair-wise study. In opposite direction, when a measure of specificity was calculated, the clinical measurement of depressed speech was selected as the conditional parameter.

Table IV summarized all statistical scores and performances of classification for reading passage session. Interestingly, the classification results from the reading showed just the opposite trends. The differentiation between suicidal speech and depressed speech was correctly classified at 82%. Whereas the correct classification scores among the other comparisons were lower.

The comparison spectrograms representing each patient voiced speech randomly selected from each diagnostic group are depicted in Figure 1. Some difference on energy distribution along frequency span can be obviously noticed among patients. The dark blue area represents the floor energy level.

Table IV.

Sensitivity, Specificity, Positive Predictive (PPV), and Negative Predictive (NPV) Values for Male Pair-wise (Reading Session) and Classification Analyses

Pair-wise Groups	%Classification	Sensitivity	Specificity	PPV	NPV
Suicidal/Depressed	82	0.73	0.88	0.82	0.82
Depressed/Remitted	73	0.85	0.58	0.71	0.76
Suicidal/Remitted	75	0.73	0.76	0.73	0.76

4. Discussion

In general, the performance measures summarized in Table III indicated that the combined acoustical features based on the power spectral density analysis provided the powerful discriminator to effectively separate the depressed speech from the remitted speech for interview session. These features also expressed their power of speech discrimination between suicidal and depressed for reading classification as shown in Table IV.

The highest classification score for suicidal-depressed pair-wise study from reading speech was obviously observed that it was less than score of depressed-remitted pair-wise study from interview speech and the overall score was also less. These may suggest that the text-dependent speech recording may have troublesome in speech production or recording quality during reading session. Patients may have different postures during audio recording of reading session. These might be the source of lower percentage of obtained correct classification. The audio recordings in the further study will be more carefully controlled and adjusted to improve the quality of recordings.

From Table III, an average jackknifed classification score of 94% was obtained using the quadratic classifier with the *hold-one-out method* for the depressed-remitted pair-wise study. This percentage of classification was the highest score found among pair-wise studies for interview study. And the 85% correct classification was also nearly as effective in separating suicidal speech from remitted speech.

For reading study with using the same type of classifier and procedure method, the highest classification score of 82% was obtained for suicidal-depressed pair-wise study. As compared between sessions of recording on the same pair-wise study, this score was higher than that from interview study. As observed the measure value of specificity (0.63) from interview study was

less than from reading study. This can be interpreted that measurement of depressed speech was misclassified as suicidal speech greater than that from case of reading. The text-dependent approach of speech recording expressed that the combined features of PSD was more effective and capable of being a powerful discriminator to separate depressed speech from suicidal speech

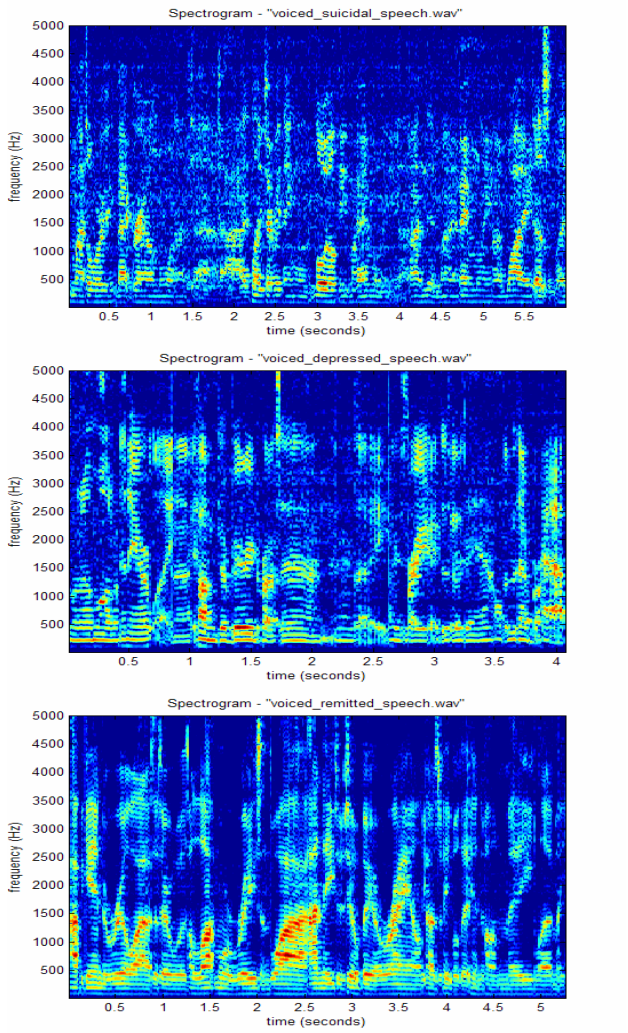

Figure 1 *Spectrograms of voiced speech: suicidal (upper), depressed (middle), and remitted (bottom).*

For further investigation of feature analysis, the discriminant rank method was used. By means of predictive validity, we performed the ranking method in PcOLPARS and found all sub-bands (PSD_1, PSD_2, and PSD_3) to be the specified features with higher discriminating power than value of peak power and frequency of peak power. With the small size of sample, the statistical power of classification was reduced and the wide confidence intervals of estimated parameters were occurred. On larger samples the important differences of vocal output characteristics between groups of patients may be precisely identified. Classification performance measures such as sensitivity, specificity, PPV, and NPV reveal how accuracy and effectiveness of vocal features are as discriminators of psychological state. Improvement of classification can be increased by multi-parameter classifiers [1]. The classifier design would be our further investigation associated with the assessment of suicide risk.

5. Conclusions

The studied PSD features of vocal output characteristics were most effective in differentiating remitted speech from depressed and suicidal speech for interview session and suicidal speech from depressed speech for reading session with high correct classification score. These suggested that the power spectral density analysis can be used as the acoustical features to assess the suicide risk.

Results from discriminating analysis for reading study also suggested that text-dependent classification can be another powerful discriminating approach to design a classifier for the assessment of suicide risk. These findings illustrated that this pioneering approach of vocal output characteristics should be further investigated in effectiveness of classification.

6. References

[1] H. Stassen, "Modeling affect in terms of speech parameters", Psychopathol., Vol. 21, p 83–88, 1988.

[2] D.J. France, R.G. Shiavi, S. Silverman, M. Silverman, and D.M. Wilkes, "Acoustical properties of speech as indicators of depression and suicidal risk", IEEE Trans. Biomed. Eng., 47(7):829-837, 2000.

[3] G. Fairbanks, Voice and Articulation Drillbook, Harper & Row, New York, 1960.

[4] A. Ozdas, R. G. Shiavi, D. M. Wilkes, M. K. Silverman and S. E. Silverman, "Analysis of Vocal Tract Characteristics for Near-term Suicidal Risk Assessment", Methods of Information in Medicine, Vol. 43, p 36-38, 2004.

[5] A. Ozdas, R. G. Shiavi, D. M. Wilkes, M. K. Silverman and S. E. Silverman, "Investigation of Vocal Jitter and Glottal Flow Spectrum as Possible Cues for Depression and Near-Term Suicidal Risk", IEEE Trans. Biomed. Eng., Vol. 51, p 1530-1540, 2004.

[6] K. Scherer, Nonlinguistic Vocal Indicators of Emotion and Psychopathology, in C. E. Izard, ed., Emotions in Personality and Psychopathology, Plenum Press, New York, 1979, p 493-529.

[7] K. R. Scherer, Vocal correlates of emotional arousal and affective disturbance, in H. Wagner and A. Manstead, eds., Handbook of social psychophysiology, Wiley, New York, 1989.

[8] J. K. Darby, Speech and voice studies in psychiatric populations, in J. K. Darby, ed., Speech Evaluation in Psychiatry, Grune & Stratton, Inc., New York, 1981.

[9] K. R. Scherer, Speech and emotional states, in J. K. Darby, ed., Speech Evaluation in Psychiatry, Grune and Stratton, Inc., New York, 1981.

[10] K. R. Scherer, Vocal affect expression: A review and a model for future research, Psychological Bulletin, Vol. 99, p 143-165, 1986.