Analysis and Design of random variable of speech detection as an indicator for depressed person

Navin Khambhala
Department of Electronics and communication engineering
PIET
Limda
Baroda
newnavin009@gmail.com

A. C. Suthar
Department of Electronics and communication engineering
LCIT
Bhandu
Mahesana
acsuthar@yahoo.co.in

Hardik Mewada
Department of Electronics and communication engineering
PIET
Limda
Baroda
sjbv.hardik@gmail.com

***Abstract*- In recent years, the problem of automatic detection of depressed state from the speech signal has gained some initial interest; however questions remaining include how speech segments should be selected, what features provide good discrimination and what benefits feature normalization might bring given the speaker-specific nature of depression. In this research, these questions are addressed empirically using classifies configurations employed in emotion recognition from speech, evaluated on depressed/neutral speech database. The PSD was extracted from the collected voiced speech samples using the method of the spectrum analysis. Results demonstrate that detailed spectra features are well suited to the task; speaker normalization provides benefits mainly for less detailed features and dynamic information appears to provide little benefit.**

Keywords- Power Spectrum Density (PSD), Depression, Classification, voiced speech.

I. INTRODUCTION

According to the World Health Organization, India has one of the highest suicide rates worldwide due to depression. The country's health ministry estimates that up to 120,000 people kill themselves every year and almost 40 percent of them are under the age of 30. Bullying by college seniors, post-examination or the death of a relative have all been named as reasons for the recent depression wave in India [9].

Several studies have been conducted since 1984 on the effects of emotional arousal on speech production based on indication of speech rate, voice articulation and respiration. Some methods based on acoustic features that were investigated are estimation of fundamental phonation or pitch, features based on a nonlinear model of speech production, relation of vocal tract features to emotional speech and speech energy estimation. Research findings show evidence of specific vocal characteristics among patients at the level of near term suicide. The investigation of correlation between vocal characteristic and psychological suicidal state was proposed by Drs. Stephan and Marilyn Silverman [10].

More recent work has seen the first steps towards automatic analysis of speech. Speech production cues such as pitch and formant measures are useful due in part to the effects of increased tension in the vocal tract associated with depression. Spectral and energy based measures are also useful in classifiers, as depressive speech can contain more information in the higher energy bands when compared with neutral speech. Spectral centroid based methods including the sub-band spectral centroid features have recently shown promise in other applications, and other work shows that these newer measures potentially include information useful in the classification of depression. Although systems for the classification of depressed speech have been proposed, there is considerable further research to be conducted, particularly in light of the extensive speech-based emotion recognition literature, from which insight can be gained towards the classification of depressed speech. In this research, we examine the effect of segment selection and the choice of different speech characterization methods on the automatic classification of depressed and neutral speech [4].

II. METHODOLOGY

A. Database and Preprocessing

The database consists of patients who were categorized as either depressed or normal by a clinician who was not involved in the research project. All speech recordings were made on mono channel. Previous voice recordings used in [7] were sampled at 10 kHz but in this study, all voice recordings were digitized with a higher sampling rate of 44.1 kHz for better sound quality. These recordings were then edited using the NCH wave pad audio digital editor by removing the interviewer ′s voice, removing long pauses that are present for more than 0.5 second and removing background noises such as door slams, sneezing and paper rustling sounds.

We combined zero crossings rate and energy calculation for segment selection [3, 7]. Zero-crossing rate is an important parameter for voiced/unvoiced classification. It is also often used as a part of the front-end processing in automatic speech recognition system. The zero crossing count is an indicator of the

frequency at which the energy is centred in the signal spectrum. We expect to find that voiced segments that provide the most effective discrimination between normal and depressed person. So that using this parameter to combine voiced speech in one row.

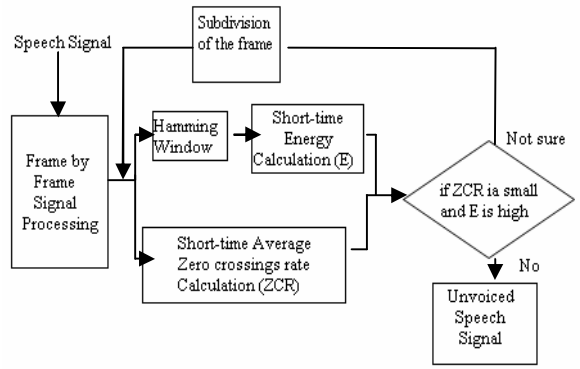

Fig.1: Block diagram of the voiced/unvoiced classification

B. Feature Extraction

The PSD was extracted from the collected voiced speech samples using the method of the spectrum analysis. Each segment was analyzed using a non-overlapping window of voice. A simple Fast Fourier Transform based power spectrum estimation was applied on every frame to obtain the PSD [1, 2].

Eight equal 250 Hz bands of energy spectrum as shown in Table were extracted from a frequency range of 0-2,000 Hz using the trapezoidal numerical integration [2]. PSD8 (1,750-2,000 Hz) was removed due to the fact that the contained information that is linear dependent on the other seven spectral energy bands and contained only a very small energy. PSD for a full frequency range of 0-2,000 Hz (PSD total) was also kept for further investigation.

Each estimated PSD band (PSD1, PSD2, PSD3, PSD4, PSD5, PSD6 and PSD7) and the calculated PSD total obtained from all 500 frames were summed up. Ratios for individual spectral energy bands were calculated by dividing the total summed energy in each 250 Hz band over the total summed energy in 0-2000 Hz. A collection of features to represent each 20-seconds segments voiced speech samples were stored in a single row vector comprised of seven spectral energy ratios. We calculate mean and standard deviation for each of eight spectral bands [5]. This parameter provides the effective discrimination between normal and depressed person.

PSD Band	Frequency range (Hz)
PSD1	0-250
PSD2	250-500
PSD3	500-750
PSD4	750-1000
PSD5	1000-1250
PSD6	1250-1500
PSD7	1500-1700
PSD total	0-2000

Table 1.1: division of frequency band

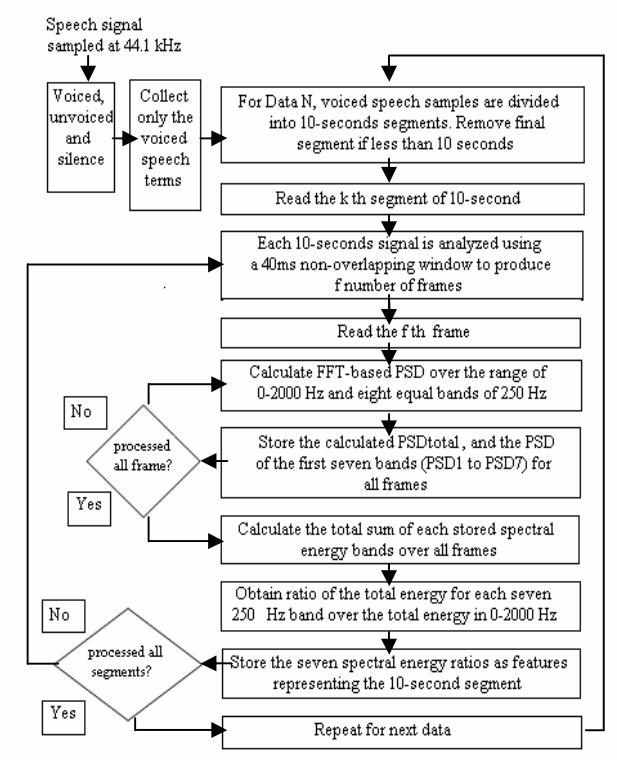

Fig.2: Flow chart for the PSD feature extraction

C. Modeling of Depressed Speech

Following the approach of many paralinguistic speech classification systems based on acoustic features, including some focused on depressed speech recognition, we employ models using PSD feature to model depressed and neutral speech [4]. In contemporary systems, researchers often make use of multiple subsystems and score-level fusion, to combine the benefits of individual systems and advance the state of the art. The overall system described in above section is summarized in Fig. 3

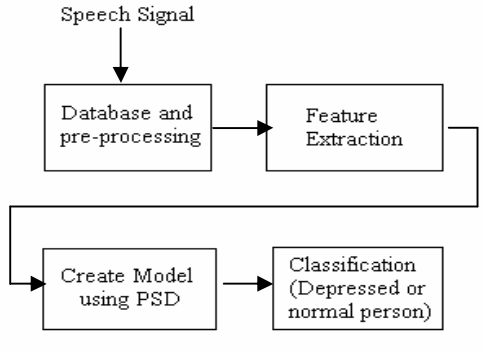

Fig. 3: Depression detection system configuration

III. RESULT AND DISCUSSION

Table 1.2 summarizes the mean and standard deviation for each group collected from speech data.

For all three groups, most of the energies contained inside PSD1, PSD2 and PSD3. The depressed state exhibits a decreasing trend of mean energy ratios as the frequency sub-bands get higher. The normal groups on the other hand, revealed a higher energy ratio in PSD2 as compared to PSD1. Overall, the high standard deviation shows the results were not clustered near the mean but were more scattered around the mean.

PSD Band	Depressed	Normal
PSD1 Ratio	0.383 ± 0.111	0.285 ± 0.112
PSD2 Ratio	0.360 ± 0.097	0.395 ± 0.119
PSD3 Ratio	0.185 ± 0.105	0.243 ± 0.068
PSD4 Ratio	0.027 ± 0.022	0.031 ± 0.015
PSD5 Ratio	0.020 ± 0.019	0.016 ± 0.009
PSD6 Ratio	0.013 ± 0.008	0.013 ± 0.008
PSD7 Ratio	0.006 ± 0.004	0.010 ± 0.005

Table 1.2: Mean and standard deviation collected from the Spectral energy ratios in speech

Compare depressed to normal, the mean energy ratios in PSD1 and PSD5 exhibit a decreasing trend, PSD6 remains the same and other sub-bands reveal an increment in mean energy ratios. The result illustrates that depressed groups exhibit a larger mean value compared to the normal group.

IV. CONCLUSION

During the data pre-processing stage, speech recordings were sampled at 44.1 kHz compared to previous publications where the speech recordings were sampled at 10 kHz. Several insights have been gained from this study of depressed/neutral classification. Voiced speech segments appear to be mildly preferable for the purpose; however segment selection is not critical. Detailed spectral features are very well suited to the speaker-dependent problem, but are also the feature of choice for the speaker-independent case. Feature warping needs to be used with care; however its behaviour in this investigation is similar to that for emotion recognition, despite the differences in the structure of the respective recognition problems. Temporal feature evolution appears not to provide any benefit for depressed/normal discrimination. For future work, the effects of noisy speech recordings on classification can be used.

ACKNOWLEDGEMENT

The authors would like to acknowledge A.C. Suthar and Hardik N. Mewada for their contribution to this work and for reviewing the paper.

REFERENCES

[1] Daniel J. France, Richard G. Shiavi, Stephen Silverman, Marilyn Silverman, and D. Mitchell Wilkes, "Acoustical Properties of Speech as Indicators of Depression and Suicidal Risk" IEEE transactions on biomedical engineering, vol. 47, NO. 7, July 2000.

[2] Hande Kaymaz Keskinpala, Thaweesak Yingthawornsuk, D. Mitch Wilkes, Richard G. Shiavi, and Ronald M. Salomon, "Screening For High Risk Suicidal States Using Mel-Cepstral Coefficients And Energy In Frequency Bands" 15th European Signal Processing Conference (EUSIPCO 2007), Poznan, Poland, September 3-7, 2007.

[3] Bachu R.G., Kopparthi S., Adapa B., Barkana B.D. "Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal" , Department of Electrical Engineering at the University of Bridgeport, Bridgeport, CT.

[4] Nicholas Cummins1, Julien Epps, Michael Breakspear and Roland Goecke. "An Investigation of Depressed Speech Detection: Features and Normalization" School of Elec. Eng. and Telecomm, The University of New South Wales, Sydney, Australia.

[5] Asli Ozdas, Richard G. Shiavi, Stephen E. Silverman, Marilyn K. Silverman, and D. Mitchell Wilkes, "Investigation of Vocal Jitter and Glottal Flow Spectrum as Possible Cues for Depression and Near-Term Suicidal Risk" IEEE transactions on biomedical engineering, vol. 51, no. 9, september 2004.

[6] Thaweesak Yingthawornsuk and Richard G. Shiavi, "Distinguishing Depression and Suicidal Risk in Men Using GMM Based Frequency Contents of Affective Vocal Tract Response" International Conference on Control, Automation and Systems , Oct. 14-17, 2008.

[7] Mark Greenwood, Andrew Kinghorn, "Automatic Silence /Unvoiced/Voiced Classification Of Speech" University of Sheffield Regent Court, 211Portobello St, Sheffield S14DP, UK

[8] A. Ozdas, R. G. Shiavi, D. M. Wilkes, M. K. Silverman, S. E. Silverman, "Analysis of Vocal Tract Characteristics for Near-term Suicidal Risk Assessment" Vanderbilt University, Nashville, TN, USA, 2004

[9] Surge in suicide rates among Indian youth Asia Deutsche Welle Feb-18, 2010, website : http://www.dw-world.de/dw/

[10] Wikipedia, Human Voice, http://en.wikipedia.org/wiki/Human_Voice.