Characterization of ΔMFCC in Depressed Speech Sample as Assessment of Suicidal Risk

Pramote Anunvrapong, and Thaweesak Yingthawornsuk

***Abstract*—**Former research reports have shown for quantitative information related to depression that associates with variation in speech quality of speaker. This association has been identified between the acoustical characteristics of vocal outcome in term of Filter frequency response (known as Cepstral based estimation corresponded to the vocal-tract response) and degree level of symptom related to emotional states of a speaker. Especially, in case of speakers who have been long-term depressed and even been elevated to be high-risk in which speakers tend to plan for committing a suicide themselves if experienced severe depression.

In this work we have investigated and concluded from our experiment and findings obtained from the designed comparative study on MFCC and Delta-MFCC (ΔMFCC) features extracted from three different categorized speech databases recorded from severe depressed, recovered (remitted) and suicidal volunteers. The ΔMFCCs corresponded to the sixteen consecutive MFCCs have been further extracted, analyzed, statistically tested and classified between different speech sample groups in pairwise manner with specifically selected classifiers, ML and LMS. The correct score of classification can be determined as high as 95% in case of Depressed/Suicidal pairwise testing phase in classification. Result has suggested that the focused acoustical features reveal some promisingly property as a symptom assessment tended to achieve in class separation based on its vocal characteristics which depends highly on the reliability of speech processing techniques and procedure applied in this study.

***Keywords*—** Speech, ΔMFCC, MFCC, Depression, ML

I. Introduction

OUR world nowadays has an increasing population growth and still climbing up every year, but inversely all natural resources are declining due to high consumption. It can affect our daily healthy living. In some serious situation it can make our life at risk increased without our notification. That risk has been known as clinical depressed, or even in some severe case suicidality could happen resulted from long-term suffering depression.

Pramote Anunvrapong is with Division of Electronics & Telecommunication Engineering, Faculty of Engineering, Rajamangala University of Technology Krungthep, Thailand.

Thaweesak Yingthawornsuk is with Department of Media Technology, Faculty of Industrial Education and Technology, King Mongkut's University of Technology Thonburi, Thailand.

It has been able to acknowledge from many research reports regarding such suicidal risk in person. Suicide is very popular public health problem associated with high population. There are many increase rates appeared such as hotline call-in, which is simultaneously monitored by physicians to first-hand assess in those callers who may or may not be depressed. It may not be the best way to save life in a program of suicide prevention but it is very necessary procedure that needs in such program. The main issue on this case of remote assessment for analyzing and evaluating on voice of speaker who calls is in that how accurate the diagnosis made by physician could possibly impact on speaker's mental healthiness and lethal risk to his/her life if he/she is planning to commit suicide due to elevating of depression symptom at that moment. If the psychiatrist can diagnose the symptom of depression or even suicidal risk correctly and rapidly, it can help save life of that patient and assign a proper treatment from the beginning of admission to hospital.

The formerly research reports [1-3, 5-12] have been proposed on utilizing of the acoustical parameters in speech that associate with the emotional affection on recognizing pattern and assessment of the degree level of suicidal severity in depressive speakers. The most common methods to assess, if patients were at severe state of depression or even at elevated risk of suicide, are self-scored patient survey, report by other, clinical interviews and rating scales [4]. Diagnosis and decision making on clinical categories patients belong to are clinical procedure with time consuming in which practitioners have to get involved in several steps such as information gathering, background profile checking, hospital visiting and admission records, diagnosing with simultaneous response in judging if patient were psychologically safe from risk of committing suicide or clinically identified and pinpoint for one of the symptom categories, dramatically necessitates for physician to conclude the diagnosing result with the correct decision making on admission and treatment for that patient.

As reported in the published studies [5-10], several analytical techniques have been proposed for achievement of measuring the particular changes, as a result of affection from the underlying symptom of depression, in acoustics of speech of depressed patients. It has been concluded that the suicidal speech in severely depressed speaker is very similar to that of common depressive one, but the tonal quality of speech significantly changes when the symptom of near-term suicidal risk highly strikes at the moment.

All following sections are organized and detailed as follows:

Section II describes on method, database, feature extraction, PCA and classification. Section III deals with experimental result and discussion. Section IV provides conclusion and future research direction at the end.

II. Methodology

A. Speech Database

The database consists of speech samples recorded from interviewing session with psychiatrist. It is categorized into three groups of 10 remitted, depressed and high-risk suicidal female subjects. The pre-processing is carried out by first digitizing all speech signals through a 16-bit analog to digital converter at a sampling rate 10 KHz via a 5 KHz anti-aliasing low-pass filer. Prior to detection of voiced, unvoiced, silent segment in speech files, the monitor and screening on any sound artifact possibly appeared during interviewing are offline implemented by using the Goldwave, including the silences longer than 0.5 seconds are manually removed. All speech signals of remitted, depressed and high-risk suicidal speakers are carefully processed under the same condition of pre-processing and the similar acoustical environment control is made during the period of recording speech sample in interviewing conversation.

B. Speech Segmentation

Based on the exploiting fact that the unvoiced segments of speech signal are very high frequency component compared to the voiced speech which is low frequency and quasi-periodic. To classify which segments of speech signal based on their energy and then weighted using the Dyadic Wavelet Transform (DWT) of speech samples were computed in each segment of 256 samples/frame. The unvoiced speech segments can be readily detected by comparing the energies of DWTs at the lowest scale $\delta_1 = 2^1$ and the highest energy level is $\delta_5 = 2^5$. Any segment of speech signal with its largest energy level estimated at scale $\delta_1 = 2^1$ is favorably classified as an unvoiced segment, otherwise found voiced segments. The following equation is the energy threshold defined as unvoiced segment;

$$UV = (n|\delta_i = 2^1) \quad ; \quad n = 1, \ldots\ldots\ldots, N \tag{1}$$

Where *uv* is speech segment classified as unvoiced at which the *n* segment with energy at scale δ_1 maximized.

C. MFCC & ΔMFCC Extraction

Voiced segments of all speech signals in database are processed for Mel-Scale Frequency Cepstral Coefficients (MFCC) [7-8,10]. The estimation procedure of studied energy parameter is described.

- Windowing each concatenated voiced-segment into 25.6 ms-length frames
- Computing the logarithm of the discrete Fourier Transform (DFT) for all windowed frames of voiced speech

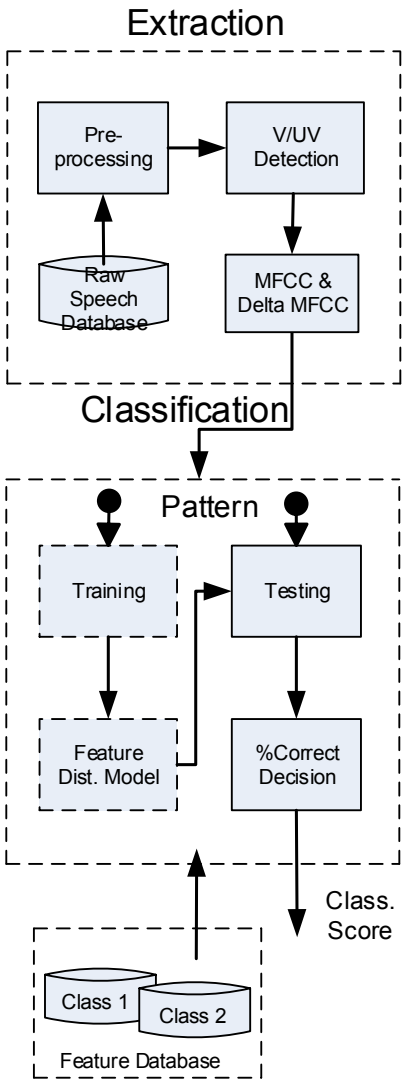
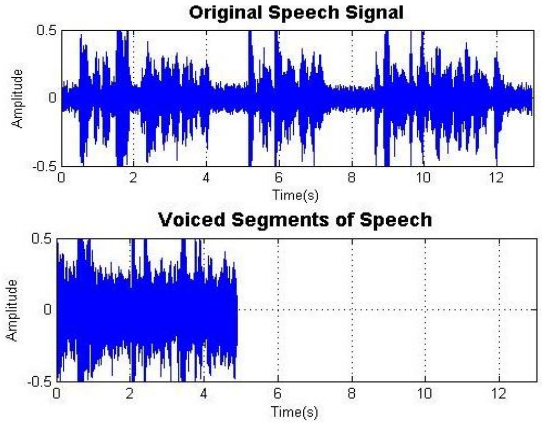


Fig. 1 Overall feature extraction and classification procedure



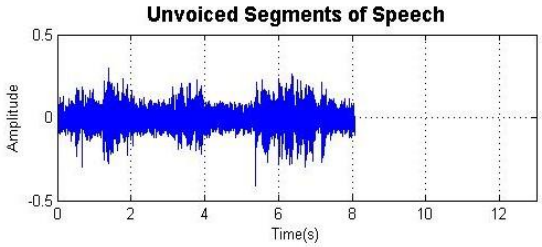


Fig. 2 Original speech waveform (upper), voiced segment (middle) and unvoiced segment (lower)

- Applying the log-magnitude spectrum through the 16 triangular bandpass filter bank with center frequencies corresponding to Mel-frequency scale
- Computing the inverse discrete Fourier Transform (IDFT), then calculate the 16-order cepstral coefficients
- Computing the delta coefficients with frame size of two
- Analyzing all extracted features from dataset with two dimensional PCA and then classifying with ML and LMS classifiers

The purpose of Mel-frequency scale is to map between linear to logarithmic scale for frequencies of speech signal higher than 1 kHz. The characteristics of spectral frequency will correspond to human auditory perception. The Mel-scale frequency mapping is defined [11]:

$$f_{mel} = 2595 * LOG_{10}\left[1 + \frac{f_{lin}}{700}\right] \quad (2)$$

in which f_{mel} is the perceived frequency and f_{lin} is the real linear frequency in speech signal.

In filtering phase, a series of the 16 triangular bandpass filters, N_s = 16 is used for a filter bank whose center frequencies and bandwidths are selected according to the Mel-scale. Once the center frequencies and bandwidths of the filter are obtained, the log-energy output of each filter *i is* computed and encoded to the MFCC by performing a Discrete Cosine Transform (DCT) defined as follow:

$$C_n = \frac{2}{N'}\sum_{i=1}^{N_f} x_k cos\left(k_i \frac{2\pi}{N'} n\right) \quad ;1 \leq n \leq p \quad (3)$$

Regarding less complexity, the factor $\frac{2}{N'}$ in equation 3 is discarded from algorithm computation.

Delta-MFCC (ΔMFCC) features were reported in [12] to gain more information on dynamic characteristics to the static MFCC features. They improve accuracy by adding a characterization of temporal dependencies to pattern classification, which is nominally assumed to be statistically independent of one another. For a short-time MFCC coefficient, the ΔMFCC features are typically defined as

$$D_n = C_{n+m} - C_{n-m} \quad (4)$$

where n is the index of the analysis frames and practically m is approximately number of 2.

D. Principal Component Analysis

The PCA technique has been applied to MFCC features to extract the most significant components of feature. This technique helps reduce multi-dimension of dataset down to two dimensions which is adequate for training and testing phases in classification.

E. Pairwise Classification

Several classifiers such as Maximum Likelihood (ML) and Least Mean Squares (LMS) are selected to train and test on two dimensional **Δ**MFCC dataset and compare among three different subject groups for performances of individual classification. In this study three groups of extracted **Δ**MFCC samples are arranged into pairwise manners which are RMT/DPR, RMT/SUI and DPR/SUI. First, **Δ**MFCC samples are randomly selected for 20% from sample dataset, and then used to train classifier, and other 35%, 50% from same dataset for training the same classifier. The reason of doing these is to compare the performances of classification among categorized subject groups, which might be affected from sizes of sample. Several trials on random selection of samples for training and testing approximately hundred times are proceeded to find the average performance of classification.

III. Experimental Results And Discussion

Original speech waveform, voiced and unvoiced segments of speech signal are plotted in Figure (2). The difference in amplitude and time interval can be obviously notified between voiced and unvoiced segments. Averaged errors in classification are tabulated in categorized pairwise groups versus types of classifier listed in Table 1&2 for all cases of various selected training samples, summarized averages from DPR/SUI classification found to have the least pairwise error, when the 20% of randomizedly selected samples trained and consecutively tested with Least Mean Squares classifier.

The comparative errors obtained from several trials on selections of **Δ**MFCC sample in classification are graphically depicted in Figures for cases of training and testing with LMS and ML classifiers. As seen in box-and-whisker diagrams, sampled **Δ**MFCC represented in Figures 5-7 for any groups vs suicidal risk group provided very less error of classification approximately 0.1 for all 20%, 35% and 50% of training and testing samples for mostly LMS classifier. The greater errors can be seen as well for mostly ML classification in cases of randomizedly selected percentages of **Δ**MFCC sample.

Based on the extracted **Δ**MFCC, the fairly high correct classifying scores can be obtained in this study, which are likely productive for its class discriminative property beneficial to emotional disorder assessment. More various acoustical parameters are suggested into the same account with studied **Δ**MFCC proposed here for obtained more accurate classification and improvement of research result toward the objective commitment of our research work.

TABLE I
CLASSIFICATION ERRORS BETWEEN
DEPRESSED AND HIGH-RISK SUICIDAL GROUPS

Classification	Percent of sample in testing overall classifier		
	20%	35%	50%
LMS	0.391	0.390	0.394
ML	0.551	0.551	0.551

TABLE II
CLASSIFICATION ERRORS BETWEEN
REMITTED AND HIGH-RISK SUICIDAL GROUPS

Classification	Percent of sample in testing overall classifier		
	20%	35%	50%
LMS	0.461	0.461	0.460
ML	0.470	0.470	0.470

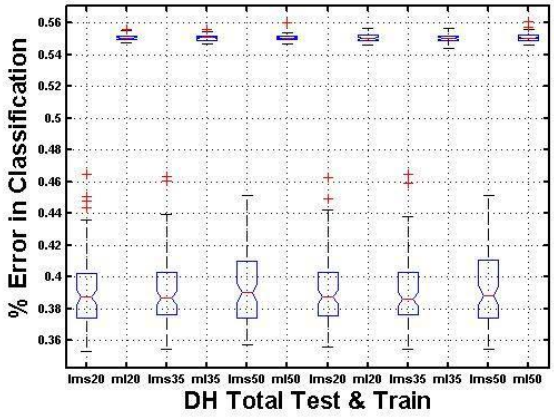


Fig. 3 Comparison of Box plots between LMS and ML classification overall with 20%, 35% and 50% of samples from depressed and suicidal groups

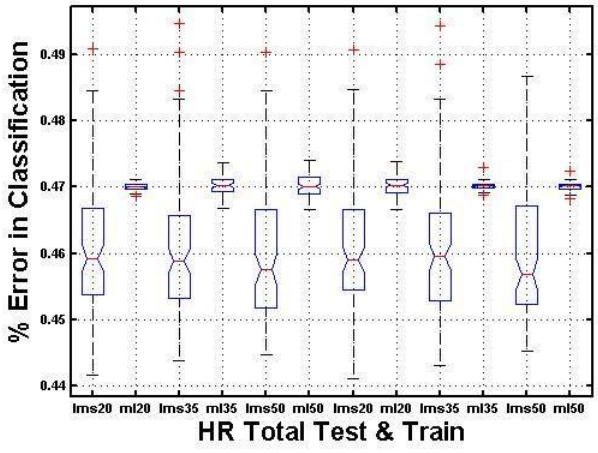


Fig. 4 Comparison of Box plots between LMS and ML classification overall with 20%, 35% and 50% of samples from suicidal and remitted groups

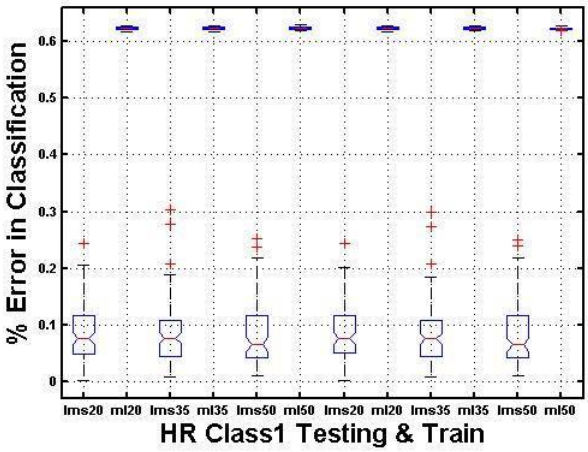


Fig. 5 Comparison of Box plots between LMS and ML classification on class 1 with 20%, 35% and 50% of samples from suicidal and remitted groups

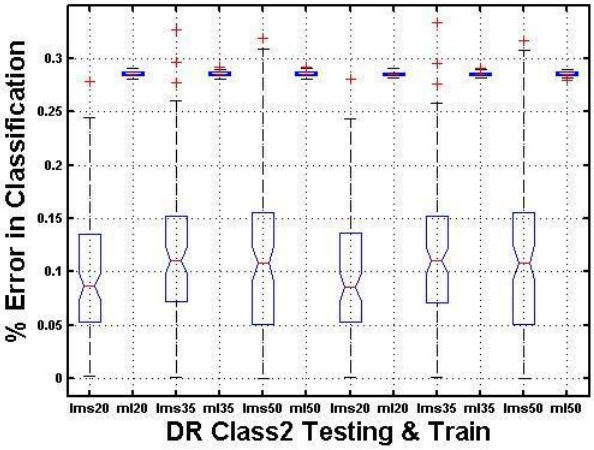


Fig. 6 Comparison of Box plots between LMS and ML classification on class 2 with 20%, 35% and 50% of samples from depressed and remitted groups

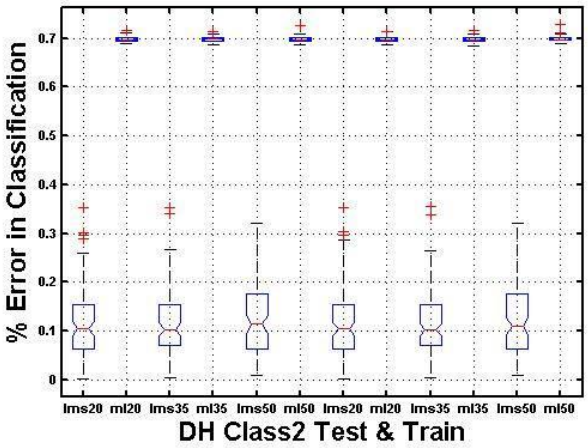


Fig. 7 Comparison of Box plots between LMS and ML classification on class 2 with 20%, 35% and 50% of samples from depressed and suicidal groups

IV. Conclusion

The **Δ**MFCC's property is able to indicate speaker's psychiatric state, especially in class separation between depressed and suicidal speaker groups. Different sampling percentages investigated in this study show no affection on classification scores in both selected classifiers. The ongoing direction of this research will focus on more effective acoustics that can be used as additional to the currently studied **Δ**MFCC in class separation with highly significantly statistical difference, and larger size of speech sample database required.

V. Acknowledgement

This work has been financially granted by National Research Council of Thailand.

References

[1] T.Yingthawornsuk, "Comparative Study on Vocal Cepstral Emission of Clinical Depressed & Normal Speaker", Int'L Conf. On Control Automation & Systems, Korea, Oct. 26 -29, 2011.

[2] T.Yingthawornsuk & et. al, "Comparative Study of Pairwise Classification by ML & NN on Unvoiced Segments in Speech Sample", Int'L Conf. On System & Electronic Engineering (ICSEE' 2012), Phuket, Thailand, Dec. 18 -19, 2012.

[3] T.Yingthawornsuk, "Classification of Depressed Speakers Based on MFCC in Speech Sample", Int'L Conf. On Advances in Electrical & Electronics Engineering, Pattaya, Thailand, April 13 – 15, 2012.

[4] M. Hamilton, "A rating scale for depression", Journal of Neurology, Neurosurgery and Psychiatry, Vol. 23, pp. 56-62, 1960. http://dx.doi.org/10.1136/jnnp.23.1.56

[5] France, D.J., et al., "Acoustical properties of speech as indicators of depression and suicide", IEEE transactions on BME, 2000. 47:p 829-837.

[6] F. Tolkmitt, H. Helfrich, R. Standke, K.R. Scherer, "Vocal Indicators of Psychiatric Treatment Effects in Depressives and Schizophrenics", J.Communication Disorders, Vol.15, pp.209-222, 1982. http://dx.doi.org/10.1016/0021-9924(82)90034-X

[7] Godino-Llorente J.I., Gomez-Vilda P., and Blanco-Velasco M., "Dimensionality Reduction of a pathological Voice Quality Assessment System Based on Gaussian Mixture Models and Short Term Cepstral Parameters", IEEE Transaction on Biomedical Engineering, 53(10):1943-1953, 2006. http://dx.doi.org/10.1109/TBME.2006.871883

[8] Lu-Shih Alex Low, et al., " Content Based Clinical Depression Detection in Adolescents", 17th EUSIPCO 2009, Scotland Aug. 24-28, 2009.

[9] T. Yingthawornsuk, R.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 System 2008, Seoul, Korea, 2008.

[10] Ozdas, A., Shiavi, R.G., Wilkes, D.M., Silverman, M., Silverman, S.,"Analysis of Vocal Tract Characteristics for Near-term Suicidal Risk Assessment", Meth. Info. Med., vol. 43, pp 36-38, 2004.

[11] Koeing, W., "A new frequency scale for acoustic measurements", Bell Telephone Laboratory Record", Vol. 27, pp. 299-301, 1949.

[12] S. Furui, "Speaker-independent isolated word recognition based on emphasized spectral dynamics," Proc. ICASSP, 1986.