RESOLUTION OF POLE MODEL TO AUTHENTICATE AGITATION FROM VOICE SIGNAL

Deepak Kumar Singh,
Assistant Professor,
Department of Electronics & Communication Engineering
Bharat Institute of Technology, Meerut
Affiliated to UPTU Lucknow; Uttar Pradesh, India

Abstract— *Voice is the vocalized form of communication. Voice is produced by turn the tongue, to give the voiced or unvoiced airflow. Agitation authentication results from voice signals, looking on the extraction of agitation characteristics. Stretching of a spoken articulation can be affected by the agitation. In mankind-computer or mankind-mankind interaction systems, Agitation authentication systems could provide users with services by being suited to their agitation. This research paper is to find the reflection coefficient, Gain, Denominator of Pole Model and then authenticate the agitations from the voice. Major steps are Pitch Conjecture, Characteristics Extraction, Statistical Resolution, Classification of the agitation from the voice signal. The voice signal is obtained from the wave file. A single line is considered and the Pitch of the line can be mark by the Pitch Marker Algorithm. The statistical Classification of the agitation can be done by the Discriminant Resolution and K-NN classifiers to get a better success rate.*

Keywords—*Agitation Authentication, Pole Model, Average Magnitude Difference Function, Discriminant Classifiers, K-NN classifiers*

I. INTRODUCTION

Voice is primary to the objectives of an agitation authentication system, as are facial expressions and gestures. A Scientific challenge is how to close the gap between voice signals and human agitation. The agitational Science World is an important component of information process. Automatic agitational voice authentication system can be characterized by the selected characteristics and the agitational categories. Agitation Detection from Voice has many potential applications. Voice characteristics that are commonly used in agitation authentication can be grouped as pitch and pitch-derived measures which are related to voiced voice generation mechanism and vocal tract formation. The second group contains various energy descriptors such as mean or standard deviation of energy of an articulation that are related to related to voice production processes. The third group comprises temporal characteristics, such as articulation duration, pauses related to behavioural voice production processes.

Recognizing human agitation in voice signal has attracted much attention and plays an important role in affect computing, artificial intelligence and signal processing areas. Most traditional agitation authentication systems have focused on the modeling of spectral characteristics or prosodic characteristics. The automatic authentication of agitations in voice has made much use of prosody characteristics such as pitch, intensity and duration, which are easier to handle. Voice quality characteristics changed when different articulation s occurred due to the different phonation types such as breathy, creaky, harsh etc. They include Harmonics-to-noise ratio, jitter, Shimmer**.** The Performance of the temporal and Statistical Characteristics is adopted in the recent Research Topics.

There are seven characteristics selection filtering methods such as CFS, Chisquare, Consistency, Gain Ratio, Info Gain, Relief and Symmetrical Uncertainty. Among these methods, Relief Method yielded an accuracy of 72%. The methods include Group wise and component wise selection. Component-wise characteristics selection methods rely on searching through all possible subsets of characteristics and fall into either wrapper or filter categories. The aim of Group wise selection is to find a combination of characteristics subgroups which maximizes agitation classification accuracy. These methods are applied to perform acoustic characteristics selection for agitation classification. In order to extract articulatory characteristics of an agitational voice, a set of articulatory classifiers are trained to learn the mapping between the acoustic signals and the articulation states. The spectral and temporal characteristics include formants, LPC, 12 Mel-frequency Cepstral coefficients.The combined information of the prosodic, spectral, statistic characteristics provide higher accuracy with SVM classifications.The class level spectral characteristics yielded a higher accuracy rate of 100%. Formant frequencies are properties of the vocal tract system and they need to be inferred from the voice signal rather than just measured. The Mel Frequency Cepstral co-efficient (MFCC) is a choice for voice authentication systems. The MFCC is good for noise free data. In MFCC, the frequency bands are equally spaced on the Mel scale. MFCCs are used as characteristics in voice authentication systems which can automatically authenticate numbers spoken into a telephone

Narayanan and Chul Min in the year 2005 have discussed about the Fold stratified cross validations to estimate how accurately a predictive model will perform .Classification errors were noted and reliable results were attained. Along with this, PCA and LDA methods were also employed and their results showed that PCA was useful for dimensionality reduction.

Chen in the year 2008 implemented a new idea regarding the ranked voting method. The weighted Voting Algorithm

is the improved algorithm which makes the voted agitations attached by different weights. Agitations are summed up and the agitation which has maximum value is determined as the result. The Classifier fusion system using relative characteristics vector can achieve better authentication result. The authors Shashidhar G. Koolagudi, Nitin Kumar and K. Sreenivasa Rao designed SVM in 2011 for two-class pattern classification. Multiclass pattern classification problems can be solved using a combination of binary support vector machines. One-against-the-rest approach can be used in two-class classification problems. Gaussian Mixture Models (GMMs) are among the most statistically mature methods for clustering. They model the probability density function of observed data points using a multivariate Gaussian mixture density. GMM refines the weights of each distribution for the given set of inputs through expectation-maximization algorithm.

The paper is organized as follows. Section II presents the methodology to extract the characteristics for the voice signal and the classification techniques involved for testing and training of samples. Section III discusses the experimental results. Finally conclusion in Section IV.

II. METHODOLOGY

Agitation Authentication is useful in the Automatic Voice Resolution. This work presents us the detection of agitation from the voice signal by analyzing the Pole model with its coefficients such as AR, Gain, and Reflection. The samples are collected from the wave file, and the peak samples are determined from the frame of samples. The samples between any two peaks can be considered as a single frame. The coefficients of an Pole model can be obtained for each frame.

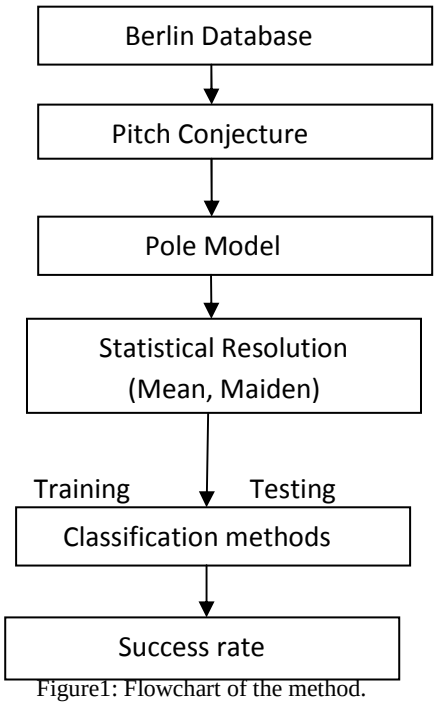

Figure1: Flowchart of the method.

Fig 1 shows the flow diagram for the proposed method. The voice signals are obtained from the database and further the peak of the samples are identified.

A. DATABASE

The database used here is Berlin Database. It is a German agitational database comprising seven agitations including anger, happiness, sad, fear, disgust, boredom and neutral voice. Ten articulation s were produced by ten speakers in seven different agitations. The recordings were made using a Sennheiser MKH 40 P 48 MICROPHONE and a TASCAM DA P1 Portable DAT-Recorder in an anechoic chamber of the Technical University Berlin, Department of Technical Acoustics. The length of the articulations is high. The articulation s whose agitation authentication is low can be eliminated from the database so as to have a better authentication rate.

B. PITCH CONJECTURE

The pitch period can be marked by the Pitch Marker Algorithm .
Steps Followed:

a. Identification of the 1st and 2nd order peaks.
b. Finding Average Magnitude Difference Function (AMDF) of the 2nd order peaks.
c. Normalization of AMDF.
d. Proper selection of Threshold voltage.

The positive peaks are the 1st order peaks which are derived from the original voice samples. The 2nd order peaks are derived from the 1st order peaks. The 2nd order peaks are used for the experimental purposes due to important reasons. One reason is that the number of the 2nd order peaks within a given frame is less than that of the first order peak . So the calculations can be much reduced. Another reason is that the 2nd order peaks can provide better reference points for pitch conjecture. The normalization of AMDF is to map the values of AMDF into the interval [0, 1]. The threshold value should be close to the minimum of AMDF. The average values of all AMDF's can be given by

$$\text{Ave AMDF} = \frac{1}{i}\sum_{j=1}^{i} AMDF\ (Kj) \quad \ldots\ldots\ldots\ldots\ldots(1)$$

that can be selected as the value of threshold to improve the accuracy and for easier implementation.

C. POLE MODEL

Extraction of characteristics is an important form of dimensionality reduction. When the input data to an algorithm is too large to be processed, then the input data will be transformed into a reduced representation set of characteristics called characteristics vectors. The transformation of the input data into the set of characteristics is called characteristics extraction.
Linear prediction (LP) is a method for signal source modeling dominant in voice signal processing. Linear prediction can be used as an alternative to spectral resolution by Fourier transform. The object of linear prediction is to estimate a set of coefficients which describes the behavior of an LTI system when its design is not available. LP calculates a set of coefficients which provides an estimate or *a* prediction for a forthcoming output sample given knowledge of previous input and/or output samples. Pole models are mathematically expressed as AR model which is Auto Regressive. The Pole model has the coefficients of AR Denominator Coefficients, Gain and Reflection Coefficients. The composite spectrum effects of radiation, vocal tract and glottal

excitation are represented by a time varying digital filter .The transfer function *S (z)* of the signal is to be modeled by an all pole filter.

$$S(z) = \frac{G}{H(Z)} = \sum_{k=1}^{p} x^{k} akz^{-k} \ldots\ldots\ldots\ldots\ldots(2)$$

$$\text{Where } H(z) = 1 - \sum_{k=1}^{p} x^{k} akz^{-k} \ldots\ldots\ldots(3)$$

G is the gain factor computed, k is the reflector coefficient and *p* is the number of poles in the model spectrum. Pole model is a natural representation of the non nasal voiced sounds. When the order p is high enough, the model provides a good representation for almost all the sounds of voice. The main advantage of this method is the Gain parameter G and the coefficients { } can be estimated in a straight forward and calculationally efficient manner. The input signal is proportional to the error signal with the constant of proportionality being the Gain constant G can be given by ,k

$$E(n) = G\, u(n) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

The Gain calculation for voiced and unvoiced voice given by

$$G^2 = R(0) - \sum_{k=1}^{p} ak\, R(k) \ldots\ldots\ldots\ldots\ldots\ldots\ldots(5)$$

Here R (0) refers the autocorrelation function with zero lag, refers the denominator coefficients. The autocorrelation method of linear prediction is formulated in time, autocorrelation and spectral domains. It approximates the short time signal power spectrum by an all-pole-spectrum. It provides Non iterative Calculation. The error measure leads to better Spectral Envelope Fit for an All - Pole Spectrum. It can be also less sensitive to the effects of high pitch on the spectrum.

(AR) model to the windowed input data by minimizing the forward prediction error. This formulation leads to the Yule-Walker equations, which are solved by the Levinson-Durbin recursion. The outputs are always nonsingular. The AR Yule method uses the Levinson Durbin Recursion of the sample Autocorrelation sequence to compute the coefficients.

1. Liinear Discriminant Resolution

LDA is expressed as a linear combination of characteristics. The assumption of the LDA method is that it is not reasonable to consider that the independent variables are normally distributed.

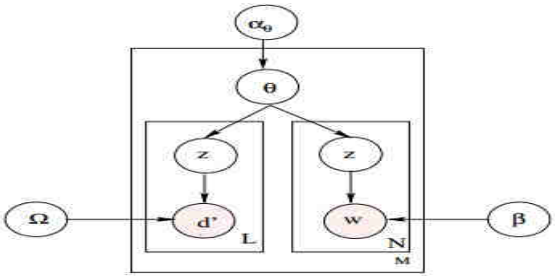

Figure:2 Representation of LDA model

LDA attempts to model the difference between the classes of data. DLDA is similar to the LDA , but with a diagonal covariance matrix estimate.

2. Quadratic Discriminant Resolution

Quadratic discriminant resolution (QDA) is closely related to linear discriminant resolution (LDA), where the measurements are normally distributed. In QDA, there is no assumption that the covariance of each of the classes is identical. Diagonal QDA is similar to QDA, but with a diagonal covariance matrix estimate**.**

3. Mahalanobis Discriminant Resolution

This classifier uses the Mahalanobis distances with covariance estimates. It is the distance measure based on correlations between variables by which different patterns can be identified and analyzed. It evaluates the similarity of an unknown sample set to a known one.

4. K-Nearest Neighbor (KNN)

The K-Nearest Neighbor algorithm is the method for classifying objects based on closest training examples in the characteristics space. K -NN is a majority vote method. Classification using an instance-based classifier is capable of locating the nearest neighbor in instance space and labeling the unknown instance with the same class label as that of the known neighbor.

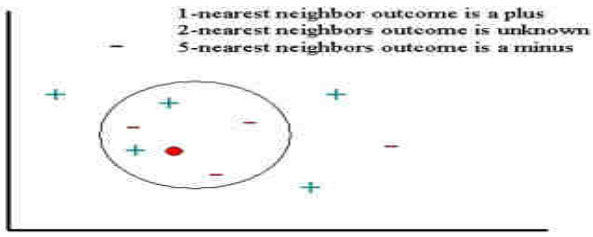

Figure:3 KNN

This approach is referred to as a nearest neighbor classifier. There arises a possibility for the occurrence of noise errors. Larger 'k' values help reduce the effects of noisy points within the training data set. The nearest neighbor rule is calculationally intensive. The training phase of the algorithm consists only of storing the characteristics vectors and class labels of the training samples. In the classification phase, 'k' is a user-defined constant, and a test point is classified by assigning the label which is most common among the 'k' training samples nearest to that test point.

4.1 Algorithm:

a. Consider the test vector to be 'X' and the training data set 'G'.
b. Find the 'k' closest observations in the training data.
c. Predict the class of the test sample that is most common among those 'k' neighbours.
d. Euclidean distance is used as the distance metric.
e. The neighbours are taken from a set of objects for which the correct classification is known.

III. RESULTS

In this research, the wave files in the Berlin Database which are sampled at 16 kHz are read. The Denominator AR coefficients, Gain G coefficients, Reflection REF coefficients of an Pole model are computed for few orders. For these values, the eight statistics are determined. The statistics is considered for training and testing of data. 302 files are allotted for training and 233 files for testing purposes. The wave files in the database consist of seven different agitations. For the conjecture of the maximum success rate, the statistics of the orders 8,9,10,11,12 are further classified by the help of twenty Five classifiers which are LDA, DLDA, QDA, DQDA , MDA and KNN for k= 1 to 20.

Table 1: Authenticate rate

Agitations	Max authentication rate(%)	Classifier	Coefficient	Orders
Anger(A)	76.471	KNN 19,20	11,12, 13,14, 15	8,9,10,11,12
Boredom (B)	70.270	MDA	9	12
Discussed(D)	78.5710	LDA	19,20, 21,22	9,10,11, 12
Fear(F)	59.2590	DQDA	16,18	10,11
Happy(H)	64.8650	DLDA	3	12
Sad(S)	75	DLDA	14,15, 16,17	9,10,11, 12
Neutral(N)	60	MDA	12,13, 14,15	8,9,10,11,12

Table I gives the authentication rate for the combined statistics of all the coefficients for the orders 8, 9,10,11,12. From the twenty five classifiers used, the agitation Disgust acquires the maximum of 78.571 % authentication rate by the LDA classification methods. Anger is authenticated next by the success rate of 75% with the help of DLDA technique.

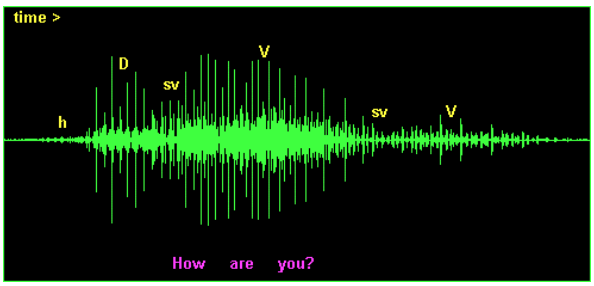

Figure 4: Human voice signal

Table 2: Comparison of individual statistics

Agitations	Max Authentication Rate (%)	Classifiers	Coefficients	Statistics	Orders
A	86.275	KNN 19,20	AR	MAXIMUM	8
B	85.1350	QDA	GAIN	STANDARD DEVIATION	9,10,12
D	87.5000	QDA	ALL	KURTOSIS	12
F	87.50	DQDA	REF, GAIN +REF	KURTOSIS	9,10,12
H	90.5410	QDA	GAIN	KURTOSIS	9,10,12
S	100	MDA	GAIN	SKEWNESS	9,10,1112
N	67.61	DQDA	AR	MAXIMUM	10

The resolution of the ALL, AR, GAIN, REF, AR+GAIN, GAIN+REF coefficients can be made for the individual statistics of mean, median, maximum, minimum, quartile, standard deviation, skewness, kurtosis.

Table 2 gives the overall comparison for the seven different agitations with the individual statistics for different orders. The agitation Sad reaches it's max success rate 100% with the GAIN coefficients for the Skewness Statistics for the orders 9,10,11,12 under the classification of MDA. The agitation Disgust attains it's max success rate 87.500% with ALL coefficients for the Kurtosis Statistics for the order12 classified by QDA.

The classifiers used for improved performance were LDA, DLDA, QDA, DQDA, MDA, and KNN – k of higher values. It can be seen that the authentication rate for the negative agitations are higher than that of the other agitations.

Figure 5: The resulting digitized signal is then analyzed in both time and frequency domains.

Also the result shows that the individual statistics of the characteristics shows a higher authentication rate when compared with the combined statistics for all the coefficients.

IV.CONCLUSION

Voice Agitation authentication based on a voice signal is one of the recent research topics in Affective Computing. The ability of the denominator, gain and reflector coefficients of the Pole model play the most important role in this case. The individual statistics for the different agitations were determined along with the combined statistics for the coefficients for the orders 8, 9,10,11,12. The result shows that individual statistics attained higher authentication rate for the negative agitations than the combined statistics of the coefficients. The negative agitations Sad, Anger, Boredom, Disgust and Fear have achieved a higher authentication rate than the other agitations. The agitation Sad has the maximum success rate of 100% for the GAIN coefficients with the individual statistics SKEWNESS under the classification of MDA classifier. The authentication rate can be further improved for higher orders.

REFERENCES

[1] S.D.Apte, "Innovative Wavelet Based Voice Model Using Optimal Mother Wavelet Generated from Pitch Synchronous LPC trajectory" Int J Voice Technol, Springer Science+Business Media, Jan 2009.

[2] Chun Chen, Jia Liu, Jiajun Bu, Mingyu You ," Voice Agitation Authentication Using An Enhanced Co-training Algorithm" , ICME Proceedings ,2007, pp 999- 1002

[3] Chul Min Lee and Shrikanth S. Narayanan ,"Toward Detecting Agitations in Spoken Dialogs", IEEE Transactions on Voice and Audio Processing, Vol. 13, No. 2, pp 293-303, March 2005 .

[4] Dmitri Bitouk , RaginiVerma , Ani Nenkova , "Class-level Spectral Characteristics for Agitation Authentication" Science Direct, Voice Communication 52 , 2010, pp 613–625.

[5] Fatema N Julia, Khan M Iftekharuddin,"Detection of Agitational Expressions in Voice", IEEE, 2006, pp307-312.

[6] Hao Tang, Stephen M. Chu, Mark Hasegawa-Johnson, Thomas Huang, "Agitation Authentication from Voice via Boosted Gaussian Mixture Models", ICME Proceedings, 2009, pp 294-297.

[7] John Makhoul, "Spectral Linear Prediction: Properties and Applications", IEEE Transactions on Acoustics, Voice and Signal Processing, pp 283-296, June 1975 .

[8] K.B.Khanchandani and Moiz A. Hussain , "Agitation Authentication Using MLP AND GFFNN " Journal of Scientific & Industrial Research, Vol 68, pp 367 – 371, May 2009.

[9] Lawrence R.Rabiner and Ronald W.Schafer. "DIGITAL SIGNAL PROCESSING OF VOICE SIGNALS" pp 396-406.

[10] Lee C.M , S. Narayanan. R. Pieraccini, "Authentication of Negative Agitations from the Voice Signal ", IEEE, 2002, pp 240- 243.

[11] LiqinFuXia Mao Lijiang Chen "Speaker Independent Agitation Authentication Using HMMS FUSION System with relative characteristics", First International Conference on Intelligent Networks and Intelligent Systems, IEEE, 2008, pp 608-611.

[12] Mohammad H. Sedaaghi, Constantine Kotropoulos , and Dimitrios Ververidis "Using AdaptivE Genetic Algorithms To Improve Voice Agitation Authentication", IEEE, 2007 , pp 461-464.