Sound Visualization for Deaf Assistance Using Mobile Computing

Aiman A. Abu Samra[1] and Mahmoud S. Alhabbash[2]

[1]Associate Professor, Faculty of Computer Engineering, Islamic University of Gaza, Palestine.

[2]Master of Computer Engineering, Faculty of Computer Engineering, Islamic University of Gaza, Palestine.

Abstract:

This thesis presents a new approach to the visualization of sound for deaf assistance that simultaneously illustrates important dynamic sound properties and the recognized sound icons in an easy readable view. .In order to visualize general sounds efficiently ,the MFCC sound features was utilized to represent robust discriminant properties of the sound. The problem of visualizing MFCC vector that has 39 dimensions was simplified by visualizing one-dimensional value, which is the result of comparing one reference MFCC vector with the input MFCC vector only. New similarity measure for MFCC feature vectors comparison was proposed that outperforms existing local similarity measures due to their problem of one to one attribute value calculation that leaded to incorrect similarity decisions.

Classification of input sound was performed and attached to the visualizing system to make the system more usable for users. Each time frame of sound is put under K-NN classification algorithm to detect short sound events. In addition, every one second the input sound is buffered and forwarded to Dynamic Time Warping (DTW) classification algorithm which is designed for dynamic time series classification. Both classifiers works in the same time and deliver their classification results to the visualization model.

The application of the system was implemented using Java programming language to work on smartphones that run Android OS, so many considerations related to the complexity of algorithms is taken into account. The system was implemented to utilize the capabilities of the smartphones GPU to guarantee the smoothness and fastness of the rendering. The system design was built based on interviews with five deaf persons taking into account their preferred visualizing system. In addition to that, the same deaf persons tested the system and the evaluation of the system is carried out based on their interaction with the system. Our approach yields more accessible illustrations of sound and more suitable for casual and little expert users.

Keywords: Android, Mobile Computing, MFCC, sound Visualization.

1. INTRODUCTION

1.1 Sound awareness

People use sound mainly to gain awareness of the state of the world around them. For example, many everyday devices such as mobiles, doorbells. at street, one might hear the horn of cars and guess a passing car is becoming closer.

According to Palestinian Central Bureau of Statistics, more than 43617 people in Gaza and West Bank are deaf and 95% of them suffer from the illiteracy [1] as they need special equipment and learning criteria.

1.2. Assistive listening devices based on vision

Vision can help a hearing-impaired individual extract meaning (or assign meaning) to sound events, if the sound visualizing describes the sound properly for hearing impaired , e.g. the use of sign language can be extremely useful tool to those who cannot hear. The rapid development of video technology has inspired many researches for sound expression on visual displays. Our proposed system will pick the most suitable sound features, similarity measures, classifiers , renderingframe work to achieve the main goals of the system

1.3 Sound Features

Many different types of sound features have been proposed to describe sound coming from speech recognition community [2][3][4][5][6].

(a) Temporal shape features

(b) Temporal features

(c) Energy features

(d) Spectral shape features

(e) Harmonic features

(f) Perceptual features :

In this thesis we will focus on spectral shape features as it proved higher discrimination results than other features[7].

1.4 Similarity measures

There are many methods that can be used to compare and derive the differences between two vectors. They are grouped into main categories according to their functionality.

a) Local dissimilarity/distance measure, such as Euclidean [8], cosine[9].

b) Statistical similarity measures, such as Kullback Leibler distance [10], and the Hotelling T2-Statistic distance [11].

The local similarity measures are more suitable to our proposed system because it is hard to have full dataset for all environmental sounds, so statistical measures will be biased according to the dataset.

1.5 Methodology

Our method started by making interviews with deaf persons living in different environments by considering their profession, capabilities, and ages. Then we tried to pick the most discriminative sound features taking into consideration the computation power of the device where the system will be implemented. Finally, after combining the results of the whole previous steps we noticed that it is hard for the deaf to use our proposed system directly without continuous help, so we added recognition module to the system that classifies some prior known sounds to help the user.

2. RELATED WORKS

Audio visualization for hearing impaired and deaf has been proposed in [12][13][14]. In [12] the authors analyzed the techniques used by deaf people for sound awareness; they made interviews with deaf and based on the results, two sound displays have been presented. One is based on a spectrograph and the other is based on positional ripples. In the spectrograph scheme, height is mapped to pitch and color is mapped to intensity (red, yellow, green, blue, etc.). In the positional ripple prototype, the background displays an overhead map of the room. Sounds are depicted as rings, and the center of the rings denotes the position of the sound source in the room. As shown in Figure (1) the size of the rings represents the amplitude of the loudest pitch at a particular point in time. Each ring persists for three seconds before disappearing.

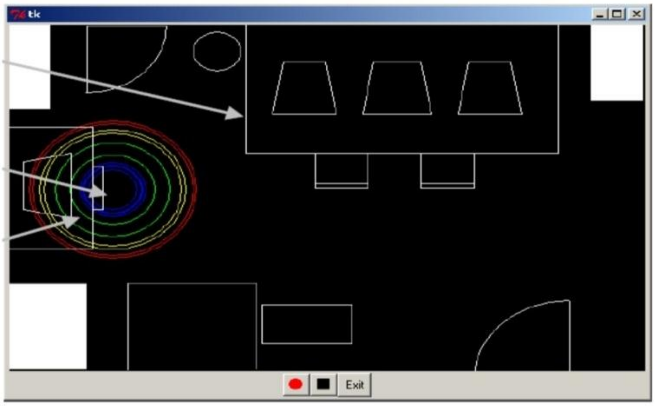

Figure 1: Speech Visualized by Positional Ripples[12].

This architecture however is impractical since it requires prior knowledge of the surrounding place (e.g. office); also it is expensive in terms of equipment setup (array of microphones placed at certain corners in the room) and is also not portable (bound to the workplace environment).

In [13], new models have been proposed, based on the proposed system in [12]. The authors proposed two models. The first model, based on single icon scheme, which displays recognized sounds as icons, located on the upper right corner of the user's computer monitor. It was used throughout the analysis and was shown to give good results.

According to the survey performed in [13], all participants liked it because it identified each sound event. The disadvantage of this method however is the actual need for prior knowledge of the type of sound to be detected which is very hard for a person who cannot hear well.

3. PROPOSED SYSTEM

3.1 Gathering design requirements using interviews

The properties of good sound visualization system must answer the following questions;

- What sounds are important to people who are deaf?
- What display size is preferred (e.g. mobile, PC monitor, or large wall screen)?
- What information about sounds is important (e.g. sound classes, location, or characteristics like volume and pitch)?
- How the person who is deaf can be aware with the visualizing system?

The initial data for the deaf participants was gathered by interviewing five of the deaf persons. The participants were chosen in different ages and jobs.

3.2 Sound input

The sound is sampled from the portable device microphone at 44100 samples per second, 16 bit per sample and mono. Further sound processing requires framing the sound to be processed in real time . A good solution was found using hamming window of size N=1024 samples with overlap of 50% at sampling rate of 44100 sample/second, which approximately corresponds to 23.2ms of sound input.

3.3 Feature extraction

The most well-known state of art-feature extraction methods are MFCC and LPC; by considering their popularity in sound recognition systems [15]. The widespread use of the MFCCs is due to its low computational complexity and better performance for most ASR systems [16]. MFCCs is used for speech data in most cases but it can be generalized for environment sound as in [16]. The characteristics of MFCCs that made it preferable for our

system is that it has lower computational complexity than many other algorithms ($n\log(n)$),], its discrimination rate , and for its simplicity in implementation .

Seven computational steps for generating MFCC vectors are summarized in Figure (2) and expressed as following;

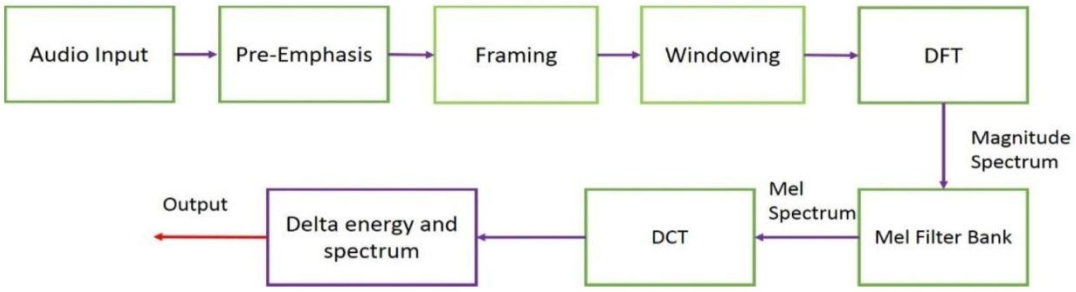

Figure 2: MFCC block diagram

3.4 The proposed similarity measure

In this subsection, we introduce efficient algorithm for measuring the similarity between two vectors of MFCC taking into account "shuffling property" and the different impact of the coefficients of the MFCC vector. Algorithm (1) handles the shuffling property of the MFCCs, as it differentiates between two MFCC vectors regarding to the places of the dominating

Purpose : to measure the distance between two vectors of MFCC
Input : MFCC vector A, MFCC vector B of length N
Output: distanc between A, B
Procedure:
1 Create two vectors Ai, Bi with the same length of A,B to store the indices of the elements in A,B
2 distance=0
3 Sort the elements of both A,B descending with corresponding indices in Ai, Bi
4 For i=0 to N-1
5 distance += wi|Ai(i)-Bi(i)|, wi is the corresponding weight of each attribute
6 Return (distance)

coefficients

Algorithm (1) : The proposed distance measure

We used Manhattan distance [17] as it completes the idea of the proposed distance measure of calculating the required steps of converting one MFCC vector into another. Finally, equation (1) summarizes the modified proposed distance.

$$P = M_D + 10^m * A1_D \quad (1)$$

Where M_D is the Manhattan distance, $A1_D$ is the distance calculated from alogrithm1, and m is the next power of then after the maximum value of M_D.

3.5 Visualization system framework

In this section, we present our proposed visualizing system based on our made interviews. Figure (3) illustrates the overall system separated in modules (Audio processing module, similarity measure module , classification module , and visualizing module).

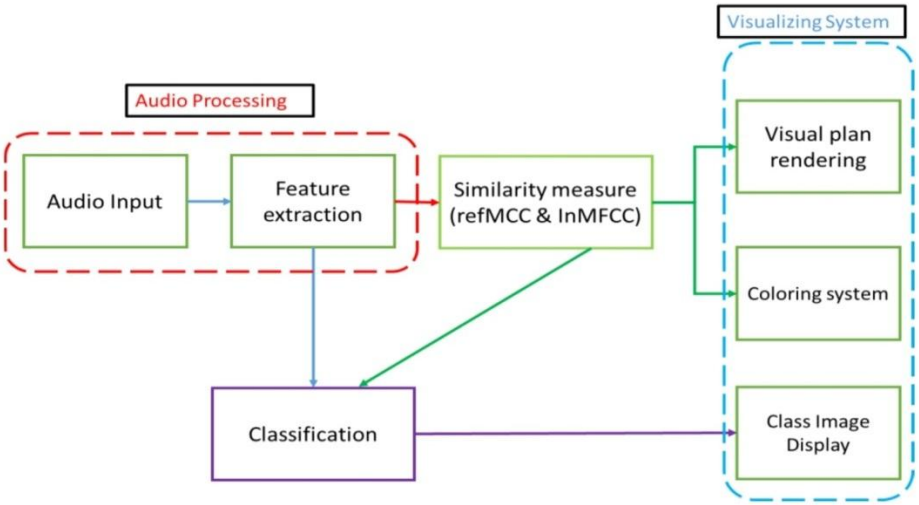

Figure 3: The overall system design

4. EXPERIMENTS AND RESULTS

The performed experiments were described and discussed. Different groups of sound features, similarity measures, and classifiers where tested and compared in order to choose the best of each group to build the proposed visualizing system.

4.1 Data set

We built a database of sound samples by collecting the preferred samples from well-known datasets [18] [19] [20]. The dataset contains 430 (80 door bells, 100 cars, 130 speech, 70 crowds, and 50 explosions) samples. All signals in the database have a 16 bit resolution and are sampled at 44100 Hz mono channel. In this way, all possible sound spectrum components can be introduced for experimentation purpose. This point is very important for environmental sounds, because some sounds show an important energy content in the highest frequencies, like glass breaks for example. The samples duration is fixed of four seconds but have different loudness levels. Each sound sample is assigned to exactly one of the five classes.

4.2. Algorithm choosing phase

The system that was used during this phase is MATLAB program version 7.8.0.347 (R2009a). We used a platform of Intel Core i5 with 4 GB RAM during the experiments.

The goal of this framework is to choose the most suitable sound features, similarity measure, and classifier for the proposed system to be implemented on smartphone. Thus, the challenges arise when considering the smart phone computation power and real time performance with complex algorithms.

4.3 Distance measures algorithms

The evaluated distance measures are considered local distance measure, so the evaluation criteria we used is the recognition rate of a classifier that uses local distance measure for classification. We used K-NN classifier for classifying every time frame in real time and

Dynamic Time Wrapping DTW for classifying input sounds with long duration to preserve the perceptual properties of the sound

Figure (4) shows the recognition rate for the K-NN classifier and DTW using the mentioned distances. We can notice the benefit of the proposed distance measure for increasing recognition rate for both classifiers more than any other distance measure.

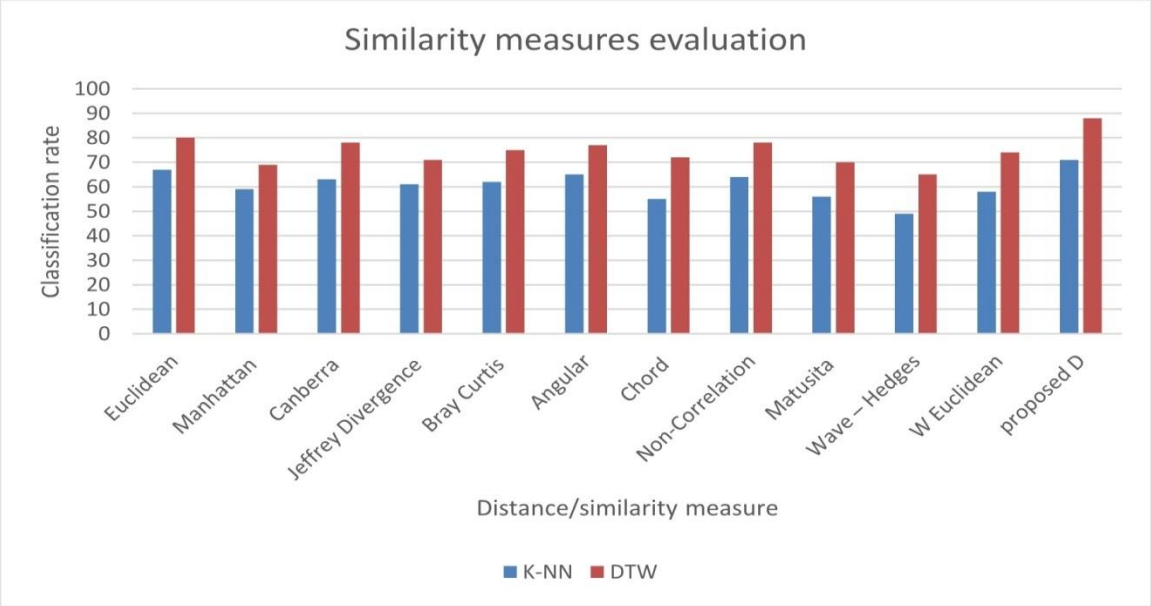

Figure 4: Similarity measures evaluation using k-NN and DTW

4.4 System implementation phase

The overall system was implemented on smart phone of model number Galaxy Ace 2 made by SAMSUNG co , which has a Dual-core 800 MHz , 4 GB storage, 768 MB RAM , and runs Android operating system.

Android is a free open source operating system for mobile devices, running on a Linux kernel, and owned by Google. Android provides various applications written in Java programming language. This operating system includes a set of core libraries [21] that provides most of the functionalities available in the core libraries of the Java programming language. In order to develop Android applications, developers use the Android System Development tool Kit

(SDK). It provides all the necessary tools to write, compile and run an Android application with or without a connected mobile device, as the emulator emulates an Android mobile phone. Once the latter is installed, it is easy and simple to use it with Eclipse IDE.

For fast video rendering of the visualized sound we used OpenGL ES 2 framework on Android [22], which uses the phone's GPU and provides simple API to call the native interfaces implemented inside.

4.5 Visualization

The visualization is drawn on a rectangular canvas with adaptive size to fit on any android device's display. It consists of two parts; the first part is the 3D colored visualization of the acquired sound while the second part series of images displaying the symbols of recognized sounds .The visualized sound flows from left to right as so as the additional icons that appear if the classifiers recognize any sound.

4.6 Speech

Figure (5) shows the visualizations of a number of different voiced Arabic vowels (ا , و , ي). Since vowels shows some constant representation of the sound during the voice, we can notice clearly the visualized sound even if it cannot recognized by the classifiers.

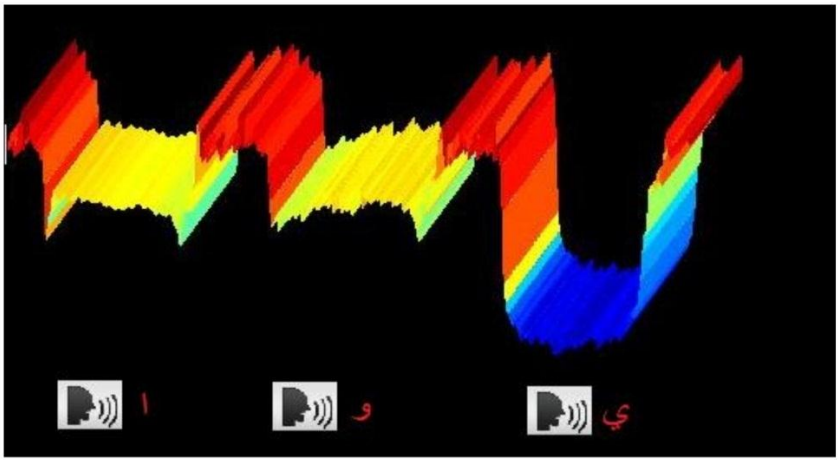

Figure 5: The visualized Arabic vowels

The reference sound used for the similarity measure and hence for the visualization system, is picked from the third vowel, so we can notice that the third vowel has the lowest height in 3D mesh.

The additional notices from Figure (5.4) is that the three different vowels show related colors (except the third, because we used the reference from one of its frames) because they belong to speech class. The third vowel shows blue color during the visualization for expressing our point of view only hence we use a reference sound represents silence in the real time application.

4.7 Door bells

Figure (6) shows the visualizing result for doorbell sound. The interesting thing about this visualized doorbell is that it displays icon of doorbell in yellow (warning color) above bird icon. This happened because in fact the doorbell is designed to output bird sounds. Since one of the classifiers, detect that this sound is likely to be a bird sound and the other for doorbell sound. The visualizing system displays the icon of both classes.

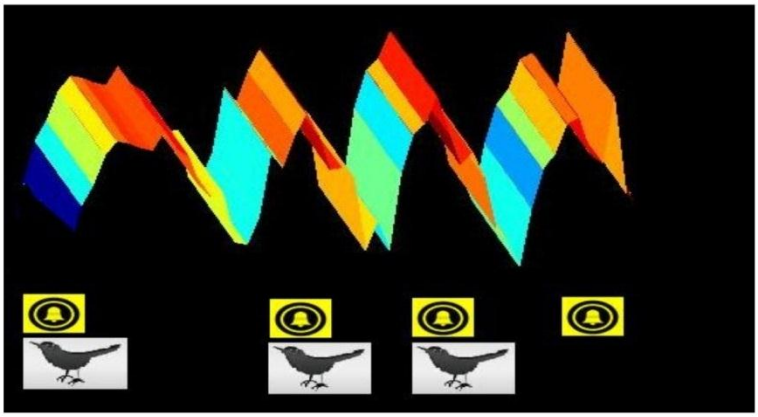

Figure 6: Visualized door bell tone

4.8 Explosions

This class represents the most severe case among all other classes. The mobile phone makes a vibration besides the visualization. The explosions includes gunshots, heavy falling mass and real known explosions.

Figure (7) shows the visualization of explosion sound. The visualizing system showed the explosion icon with vibration on the test smart phone.

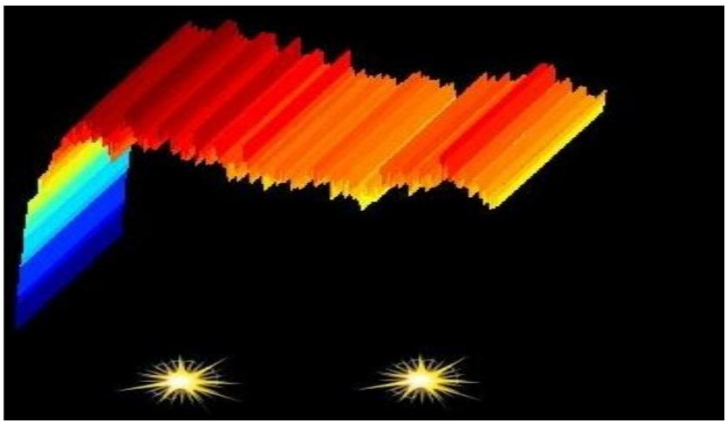

Figure 7 : Explosion sound visualization

4.9 Testing phase

For every sound, the answers of the trainees are collected with their response time for every answer is recorded for analysis. Due to the deaf inability to analyze sound from previous experience, the tests were made repeatedly and the results only picked in the last two sessions and only for correct answer rate of 90% and above.

Figure (10) shows the average duration of correct answers curve for the testing users. As we can note, the users at first find some difficulties for giving correct answers with the new sounds during session 1. In the next sessions , the users shows improvements in response time . The interesting notice about the final results is that the response time reached several few seconds this indicates that they can use the program in real time with little difficulties

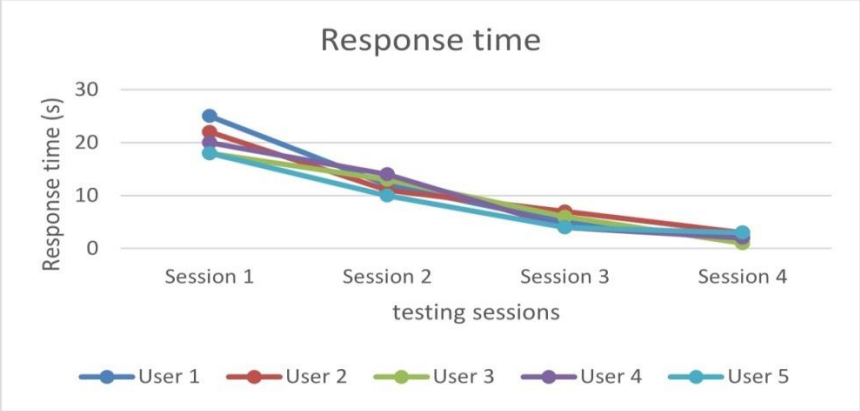

Figure 10: Testing sessions for users and their response time

5. CONCLUSION

We proposed a new visualization system to help deaf person to experience surrounding sounds. This system depends on vision sense of deaf to understand the sound visualization. Technically, the system depends on extracting robust sound features and comparing them with reference sound feature for using the comparison result for visualizing the sound in 3D curve with different colors. The building of the system involved in using feature extraction, similarity measures, classification, and rendering frameworks.

The sound feature that was used for representing sound is MFCC by evaluating many sound features and picking the highest recognition rate feature vector. Since, there is wide range of feature vectors proposed previously, our evaluation done on the most well-known features in open literature.

We formed sound database from other three databases to get different sound classes that fits the resulted application-working environment. We used our database for evaluating many sound features, similarity measures, and classification algorithms.

The visualization system renders the frames of sound as 3D dynamic mesh changing over time to give the user real time feeling with sound. The dynamic color and height of the visualized sound can be read easily by little experienced user

REFERENCES

[1] Palestinian Central Bureau of Statistics.(2012). [Online] Available at: <http://www.pcbs.gov.ps/default.aspx> [Accessed 10 June 2013].

[2] Foote, J. (1997). Content-base Retrieval of music and Audio Multimedia storage and Archiving Systems II, Multimedia Storage and Archiving Systems II. In Proc. of SPIE, vol. 3229, pp. 138-147.

[3] Scheire, E. and Saleny, M. (1997). Construction and evaluation of robust multifeature speech/music discriminator. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp. 1331 – 1334.

[4] Borwn, J. (1998). Music Instrument identification using autocorrelation coefficients. Proceedings International Symposium on Musical Acoustics ISMA 1998, Leavenworth, Washington, pp. 291-295.

[5] Matin, K and Kim, Y. (1998). Instrument Identification: a pattern recognition approach 136th Meet, USA

[6] Rabiner, L. and Jung, B. (1993). Fundamentals of speech recognition. Prentice-Hall,USA.

[7] GeoFroy, P. (2004). A large set of audio features for sound description (similarity and classification), Iracm project ,France.

[8] Foote, J. (1999). Visualizing Music and Audio Using Self-similarity. Conf. Multimedia, 7th ACM Int (Part 1), Orlando, pp. 77–80.

[9] Foote, J. (1999). Visualizing Music and Audio Using Self-similarity. Conf. Multimedia, 7th ACM Int (Part 1), Orlando, pp. 77–80.

[10] Siegler, M. Jain, U. Raj, B. and Stern, R. (1997). Automatic Segmentation, Classification and Clustering of Broadcast News Audio. DARPA Speech Recognition Workshop, USA, pp. 97-99.

[11] Zhou, B. and Hansen, J. (2000). Unsupervised Audio Stream Segmentation and Clustering via the Bayesian Information Criterion. Proc. ISCLP'00, China, vol.3, pp. 714-717.

[12] Wai-ling, F. Mankoff, J. James A. (2003). From Data to Display: the Design and Evaluation of a Peripheral Sound Display for the Deaf. In Proc. of CHI, p. 8.

[13] Matthews, T. Fong, J. and Mankoff, J. (2005). Visualizing Non-Speech Sounds for the Deaf. In Proc. of ACM SIGACCESS on Computers and Accessibility (ASSETS). Baltimore, pp. 52-59.

[14] Yeo, W. and Berger, J. (2006). A New Approach to Image Sonification, Sound Visualization: Sound Analysis And Synthesis. In Proc. of the International Computer Music Conference, New Orleans, pp.34-50.

[15] Temko, A. (2007). Acoustic Event Detection and Classification. PhD thesis, University Politecnica De Catalunya, Spain.

[16] Zhang, X. (2009). Audio Segmentation, Classification and Visualization. Ph.D. thesis, Auckland University of Technology, New Zealand.

[17] Deza, E. & Deza, M. Marie. 2009 . Encyclopedia of Distances. pp.94-236.

[18] DeWolf Sound effect Database .(2013).[Online] available at: < www.dewolfe.co.uk > [Accessed 5 January 2013].

[19] EFX Guns Library. (2013).[Online] available at: <www.efx-sound.com> [Accessed January 2013].

[20] Sound Spaces Environmental Sound Library.(2013).[Online] available at: <http://sounds.bl.uk/environment/soundscapes> [Accessed 5 June 2013].

[21] Saha, A. (2008). Developer's First Look at Android: Linux for You. In Proc. of Hot Mobile, pp. 48-50.

[22] Android Frame Wrok Samples. (2013). [Online] available at: <http://developer.android.com/resources/samples/ApiDemos/src/com/example/android/apis/graphics/index.html> [Accessed 13 June 2013].