International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 3 | ISSN : 2456-3307

Audio Signal Processing: A Review of Audio Signal Classification Features

Mittal C. Darji
Information Technology Department, G H Patel College of Engineering & Technology, V.V.Nagar, Anand, Gujarat, India

ABSTRACT

Digital audio processing applications are getting popularity with the time. Audio data compression, summarization, speech recognition, speaker identification, speech and music separation, music genre classifications, singer identification, gender detection and many more are there. Feature selection is the curial part of these applications. Various audio signal features are reviewed here especially considering classification as the purpose.

Keywords: Audio Features, Physical features, Perceptual features, Zero Crossing Rate, Short Time Energy, Spectral Centroid, Flux, Fundamental Frequency, Loudness, Pitch

I. INTRODUCTION

Human hearing system provides a great sense of environment with respect to location and variations of sounds of various objects – living or non-living. Our system is capable to process any complex mixture of sound at same time. The same system development using machines requires high level of efforts on processing the audio signals which actually is a complex task to do. Automation of tasks related to audio such as audio summarization, classification of speech and music, identification of genre, gender or speaker are highly required considering the rapidly growing archives of digital speech and music on internet. In this paper, basic features of signal processing are reviewed, mainly from the point of view of audio classification.

Audio possesses various features, which have specific behavior with respect to the category of audio. Such behavior can be used to differentiate various classes of audio. Picking up the features is a key of proper classification. Therefore most of the classification systems emphasis on feature selection and extraction.

II. CHARACTERISTICS OF AUDIO SIGNAL

Humans can hear sounds in to the frequency range of 20 Hz to 20 kHz based on the pressure applied on the eardrum. One more factor of sound is intensity, which is measured in dB – decibels. The lowest sound that we can here is of 0 dB and most of us will feel pain at 120 dB. This audible range we can classify into various categories. At the top level, root class can be considered as sound which we can broadly classify into pleasant and unpleasant sound. But these categories would not be enough to describe all types of sound. Therefore, we may classify using natural English nouns like: speech, music, noise, natural sound, artificial sound. These categories are having their own sub partitions, which make our hierarchy multi-dimensional. Figure 1 displays the same.

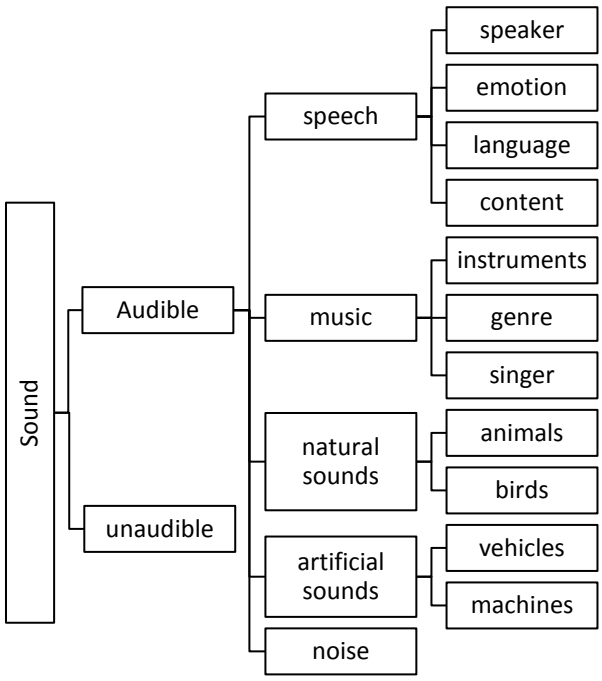

Figure 1. Sound Classification

Audio signal varies with the time. They can be represented either in time domain or in frequency domain. Each of our sound sensation is related with one or more spectral or temporal property of sound. Hence, features of both, time domain and frequency domain are jointly required for representation of sound.

III. AUDIO FEATURES

Heart of classification is to find features, which can provide a good discrimination among various classes. Years of research on audio signal features has given 2 broad categories of features: physical features and perceptual features. Features can also be classified as static and dynamic. Static features are the measurement of feature of signal at a particular time. Long time variations in static features can provide dynamic features. In this paper, physical and perceptual types of features are focused on.

A. Physical Features

Physical features are such low-level features, which can directly be measured from frequency or time representation of audio signal. They are physical as they can be computed directly from the amplitude values or other spectral values of signal.

1) Zero Crossing Rate

It is the rate of sign-changes along a signal. It is defined as the number of times zero crossed within a frame. [6]

ZCR is an important feature for many classifications like separating signal into voiced and unvoiced, gender classification etc. it provides useful information at very cost. It can be calculated as,

$$ZCR = \frac{1}{T-1}\sum_{t=1}^{T-1} F(St * St - 1 < 0) \quad (1)$$

Where, S is a signal of length T

$$F(A) = \begin{cases} 1, & \text{if } A=\text{true} \end{cases}$$

2) Short-Time Energy

It represents the total power spectrum of the frame [5]. Rather than its actual value at a time, its variations over the time can provide information that is more useful. It provides the representation of variations of amplitude over the time. It can be calculated as,

$$STE = \frac{1}{T-1}\sum_{t=1}^{T-1}(|St| * |St|) \quad (2)$$

3) Spectral Centroid

It indicates where most of energy of the signal is. It measures the brightness of the signal. Due to its effectiveness of describing spectral shape, centroid has been used in classification of audio. It can be given as,

$$C = \frac{\sum_{k=1}^{N/2} f[k]|X_r[k]|}{\sum_{k=1}^{N/2} |X_r[k]|} \quad (3)$$

Where, N is the number of FFT points, $X_r[k]$ is STFT of frame r and f[K] is the frequency of bean k.

4) Spectral Roll-off

This is another measure of spectral shape of signal. It was first used as feature to separate voiced and unvoiced signal. It is calculated as,
R = f[K]

If K is the largest bean having,

$$\sum_{k=1}^{K} |X_r[k]| \leq 0.85 \sum_{k=1}^{N/2} |X_r[k]| \quad (4)$$

5) Spectral Flux

Spectral flux is also called as delta spectrum magnitude. It measures rate of change in spectral shape. It can be calculated as the frame-to-frame magnitude spectral difference. High value of flux indicated sudden change in magnitude.

$$F=\sum_{k=1}^{N/2}(|X_r[k]| - |X_{r-1}[k]|)^2 \quad (5)$$

6) Fundamental Frequency

It is measured by time domain periodicity of signal. In human voice or musical instruments, it is non-trivial to measure the fundamental frequency due to the variations in waveform periods and due to other stronger harmonics. Autocorrelation function is used to find the periodicity of signal.

7) Mel-Frequency Cepstral Coefficient (MFCC)

It is most used in speech analysis and music signal processing. It is a compact representation of spectrum of an audio signal that takes into account the non-linear human perception of speech. It is generally being calculated as:

- Take the Fourier transform of a signal window.
- Map the power of the spectrum obtained above into the Mel scale, using triangular overlapping windows.
- Take the log of the powers at each of Mel frequencies.
- Take the discrete cosine transform of the list of Mel log powers, as if it were a signal.
- The MFCCs are the amplitudes of the resulting spectrum.

B. Perceptual Features

Human recognizes sound based on perceptual attributes of sound. Psychoacoustic models have also been proposed that measure the perceptual features of sound for classification. Most of the perceptions of sound are measured as loudness, pitch and timbre. Timber is mostly used distinguish between the sounds having same loudness and pitch.

1) Loudness

Loudness indicates the sensation of strength of signal. Loudness can be approximately related to intensity as follows.

$$L = kI^{\alpha} \quad (6)$$

Value of α is proven to be 0.23 in case of noise. Loudness also depends on frequency. Considering that, some of the models are defined there to calculate loudness.

2) Pitch

Pitch is a perceptual property of sounds that allows their ordering on a frequency-related scale, or more commonly, pitch is the quality that makes it possible to judge sounds as "higher" and "lower" in the sense associated with musical melodies. [14] Pitch is defined as the fundamental frequency of the excitation source.

IV. CLASSIFICATION

The general process of classification is simple. It is either a supervised leaning or unsupervised learning. In case of supervised leaning, the labeled sample data is already given to train the system and then unlabeled data is given to this trained system for labeling based on the features. While in unsupervised learning, training set is not labeled. System itself observes the samples for their features and creates classes. This process is called clustering. In both the ways, selection of features plays an important role.

V. VI. FUTURE WORK

Classification of audio signal into various categories as shown in figure 1 can be done by selecting suitable audio features.

VI. REFERENCES

[1]. Moore BCJ (2003) An Introduction to the Psychology of Hearing. Academic Press, San Diego.

[2]. McKinney M F, Breebaart J (2003) Features for Audio and Music Classifcation. Proc of the Intl Symp on Music Information Retrieval (ISMIR)

[3]. Tzanetakis G, Cook P (2002) Musical Genre Classifcation of Audio Signals. IEEE Trans on Speech and Audio Processing 10(5):293-302.

[4]. Burred J J, Lerch A (2004) Hierarchical Automatic Audio Signal Classification. J Audio Engineering Society 52(7/8):724-739

[5]. M.C. Darji, Dr. N.M. Patel, Z.H. Shah, "A Review on audio features based extraction of songs from movies", International Journal of Advance Engineering and Research Development (IJAERD) e-ISSN: 2348 – 4470, print-ISSN: 2348-6406, 2015.

[6]. M. Casey, "General sound classification and similarity in mpeg-7," Organized Sound, vol. 6:2, 2002.

[7]. Zhang T, Kuo C C J (2001) Audio Content Analysis for Online AudioVisual Data Segmentation and Classifiation. IEEE Trans on Speech and Audio Processing 9(4):441-457.

[8]. M.C. Darji, Dr. N.M. Patel, Z.H. Shah, "Extraction of Video Songs from Movies using

Audio Features", IEEE, Advanced Computing and Communication (ISACC), Print ISBN: 978-1-4673-6707-3, 2015.
[9]. Wold E, Blum T, Keisler D, Wheaton J (1996) Content-based Classi¯cation, Search and Retrieval of Audio. IEEE Multimedia 3(3):27-36
[10]. Meribeth Bunch. Dynamics of the Singing Voice. Springer-Verlag, New York, 1982.
[11]. Chris Chafe and David Jaffe. Source separation and note identification in polyphonic music. In icassp, pages 1289–1292. IEEE, 1996.
[12]. Information on Tempo available: http://en.wikipedia.org/wiki/Tempo
[13]. Information on Zero-Crossing Rates available: http://en.wikipedia.org/wiki/Zero-crossing_rate