Special Issue Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, www.ijtrd.com

Review on Sound Signal Processing: Domain, Features, Noise, Classification and Applications

Simna Rassak[1], Nachamai M[2] and Krishna Murthy A[3]
Department of Computer Science, Christ University, Bangalore, India

***Abstract*:** Sound is an important part of our communication. For a human being, sound is something that they hear and understand but cannot be seen or touched. Sound can be defined as the regularly generated vibration from any object when it comes in contact with other object or air which travels through various media. The sound spectrum for human and animals are different. By processing these sound signals we can extract a lot of useful information. The main objective of this paper is to give a brief idea about the sound signal processing. It explains about the various domains of the sound signals and features. The paper also gives a brief explanation about noises and their types. The paper profiles drum sound signal classification and brief on applications of audio signal processing.

Keywords—*Sound, Signal Processing, Noise, Sound features, Drum Sound Classification*

I. INTRODUCTION

Sound is the vibration of objects, which create energy. These energies can transfer from one air particle to another and it is known as sound wave. Every sound wave has a frequency and different features, which can help us to identify what kind of sound it is. But there are many sounds which cannot be understood by Human being such as ultrasound, infrasound etc. The sound wave in the air can be decomposed into useful information. Human brain processes the sounds in order to understand them. Signal processing help us to extract the time-series data for analysis and development. Signal processing is mainly based on various domains, some of the most commonly known domains are time and frequency. Every signal has many features which can be used to identify and extract information from them. In signal processing there are sounds which is known as noise.

Sound signal processing helps not only to understand the sound but help us to create application which can be used in many fields like defense, business, education, communication, astronomy, science, etc.

This paper is organized as follows. Section II gives a brief idea about various domains of the sound signal processing and most commonly used features of the sounds in section III. The types of noise and color of noise is explained in Section IV. Section V discuss about drum sound signal classification followed by a brief on the application in section VI. Finally, conclusion is given in the section VII.

II. DOMAINS OF SOUND SIGNAL PROCESSING

Sound signal processing has various domains. Domains are nothing but the purview of the sounds which help to understand and extract information from them.

A. *Time Domain*

Time domain is an analysis which shows the change of any signals over time [1]. The time domain can be used to extract the amplitude and peak of a signal [2]. The below figure (Fig.1) shows the sound representation in time domain.

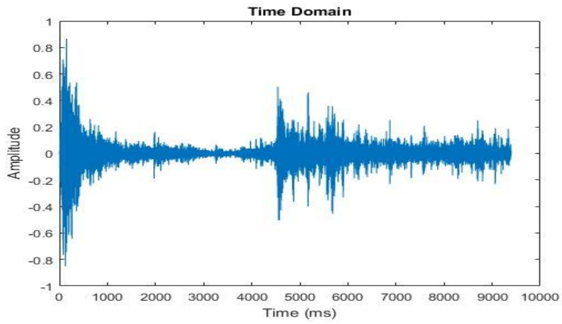

Figure 1: Sound Signal in Time-Domain

B. *Frequency Domain*

In a general definition the frequency domain is a representation of number of frequencies present in a signal [1]. It shows where the sound signal present in a frequency band [3]. It gives information of phase shift. In digital signal processing most of the time the signals are processed by converting them from time to frequency domain and vice versa, this process is known as transform. One of the most commonly used transform is Fourier series [3]. The below figure (Fig.2) shows the sound representation in frequency domain. The figure (Fig.3) give a comparative representation of time and frequency domain

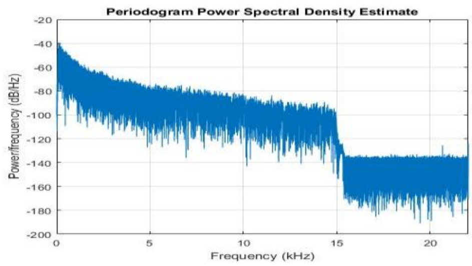

Figure 2: Sound Signal in Frequency Domain

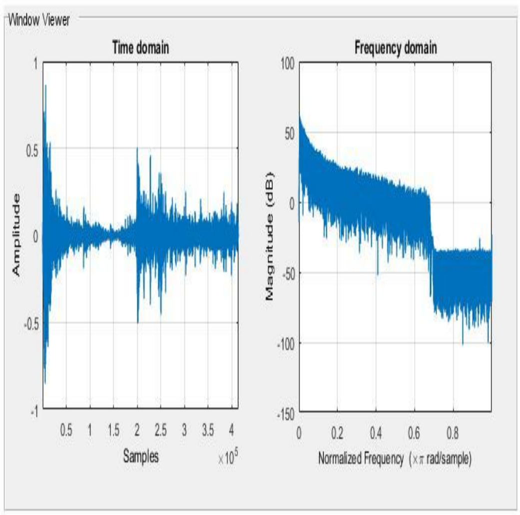

Figure 3: Sound Signal Comparison on Time and Frequency Domains

C. Time-Frequency Domain

The most commonly known domains in signal processing are time domain and frequency domain. To extract more information from the signal a combination of time domain and frequency domain was used, which is known as time-frequency domain [1]. Short Time Fourier Transformation is a time-frequency domain algorithm for analyzing the signals. Spectrogram give the signal representation in time-frequency domain [4]. The below figure (Fig.4) shows the sound representation in Time-frequency domain.

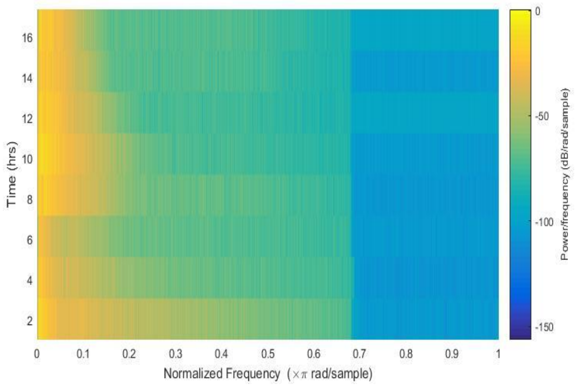

Figure 4: Time-Frequecy Domain Spectrum of Sound Signal

The frequency variation can be seen using a spectrogram. The increased and decreased frequency over time can be seen from the energy representation of the signals. The time-frequency domain is most useful in analyzing the frequency variation over time. In a signal there can be frequency which has low frequency and present only for a short time period. It is identified that the noise signals are distributed over time-frequency domain and it is very efficient domain to compute the signal to noise ration from this domain [3]. Below figure (Fig.5) shows the different frequency present in the signal.

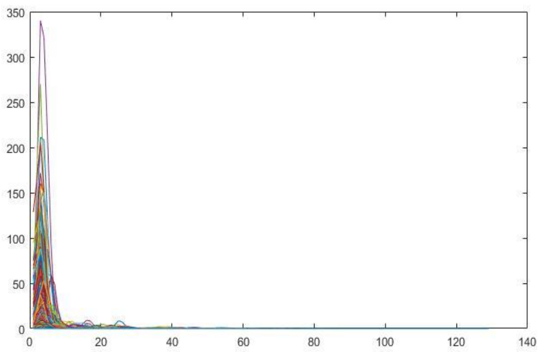

Figure 5: Frequency variations Representation of Time-Frequency Domain

D. Cepstrum Domain

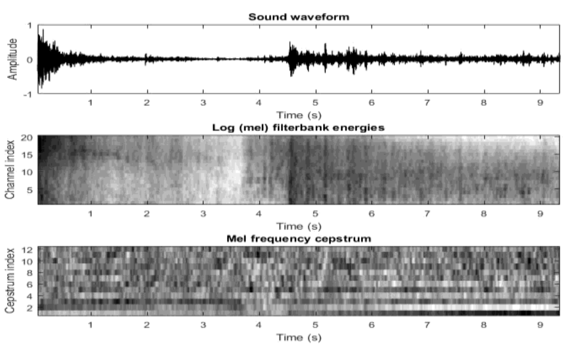

Figure 6: Cepstrum of MFCC

Cepstrum is nothing but the inverse of a Fourier transform of a log spectrum. Cepstrum is originated from spectrum and quefrency is its independent variable. Quefrency is nothing but a measure of time. Cepstrum gives the periodicity of the spectrum [1]. The below figure (Fig.6) shows the sound-MFCC representation in cepstrum domain.

III. FEATURES OF SOUND SIGNALS

Sound wave has four major characteristics and many features. The four characteristics of sound waves are frequency, wave shape, amplitude and phases and some of the most commonly used features in digital signal processing are pitch, amplitude, timber, energy, rhythm, loudness etc. [5].

A. Pitch

The pitch of a sound wave can be defined as the sensations of a frequency. It can also be defined as the rate of vibration of object. A high pitch is a sensation of high frequency wave and a low pitch is a sensation of low frequency wave [5].

There are many algorithms which can be used to extract the pitch. They are either in time-domain or in frequency domain. In time-domain pitch can be extracted using peaks, the Matlab function used to find peak is find peaks (data) [6]. Most of the time sound signal consists of other harmonic sounds along with the sounds that we process to extract the peak, due to the other signals the chances of getting wrong results are high. In the frequency domain using the filters the unwanted signals can be filtered and the pitch can be extracted. Most commonly a low pass filter is used to filter out the sounds.

B. Amplitude

Amplitude is the degree of the atmosphere pressure change due to a sound wave. The amplitude can be expressed as the peak value in time-domain. The time envelope is used to measure the maximum absolute amplitude [5].

C. Timbre

Timbre is the tone or harmonic content of a sound. It can also call the tone quality. Timbre is the characteristic which is used to distinguish different sounds [5]. Some of the timbre features which are commonly found in sound signal processing are

1) Mel- Frequency Cepstral Coefficients (MFCC)

Mel-Frequency Cepstrum (MFC) is a representation of power spectrum of sound signals. The coefficient of MFC is MFCC. Where mel is the measurement unit of pitch. This is one of the most commonly used feature of sound signal [7].

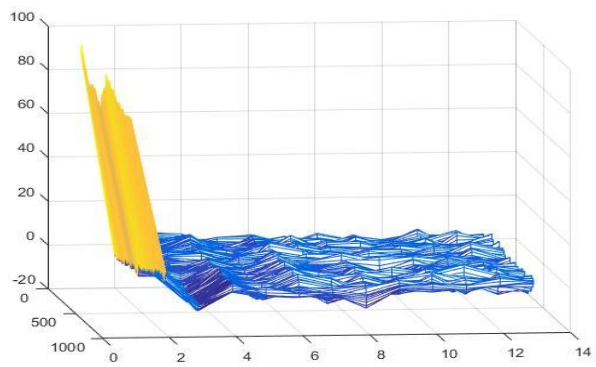

Figure 7: MFCC Mesh

The method of extracting MFCC involve various steps like transforming the signal using Fourier Transform and windowing using Hamming function. Then it maps the power into mel scale using triangular overlapping window. In the next step it takes the logs of the power and Discrete Cosine Transform. Finally, it extracts the amplitude of the resulted spectrum as MFCC [8]. The below figure (Fig.7) shows the sound feature MFCC Mesh.

2) Spectral Flux

Spectral Flux is a measurement of the speed of the signal power spectrum change [7]. It compares the frames to extract the change [5].

3) Spectral Centroid

Spectral Centroid is the computation of the center of the mass or gravity of a spectrum. It uses Fourier Transform to compute [7].

4) Zero Crossing

Zero crossing is the estimation of sign change in the sound signal [9]. It gives the fundamental frequency of a sound signal.

D. Energy

The energy of a sound signal can be defined only based on the context of the usage. One of the general view point of energy is the loudness of the sound. Most commonly the energy of each frame in a sound signal can be extracted using the Root Mean Square [5].

E. Rhythm

This is a regular pattern of cyclic occurrence of frequency over the time. A beat histogram can be used to represent the rhythm. The beat histogram can be used to extract the peaks of rhythm. It gives a single valued feature [5].

F. MPEG-7

MPEG-7 is used to standardize the features for the audio signal classification [10]. It is used for the multimedia content description. It is related to time. In sound signal processing it is used to tag the sound with time to extract the event. MPEG-7 audio descriptors are of two types. They are low level and high level [11].

1) Low Level descriptors

The audio features which can be used for any audio classification and audio applications. They are either scalar type like power or vector like spectrum. The main low level features are basic signal features, basic spectral features, signal parameter features, timbral temporal features, timbral spectral features, MPEG-7 silence descriptor features etc. Some of the examples of features are waveform envelope, audio spectrum envelope, audio spectrum centroid, audio spectrum spread, audio spectrum flatness, pitch tracking, audio harmonicity, audio spectrum projection etc. [11].

2) High Level descriptors

They are more specific towards the application to be used. Some of the features are audio signature, melody, music instrument timbre, sound signal recognition and indexing, spoken content description etc. [11].

IV. Noise in Sound

Noise is nothing but sound itself. In digital signal processing the noise is considered as the unwanted sound, which need to remove from signal in order to get the expected output [12]. In general noise is a subjective view of sound, which is useful or not useful based on the scenario. There are many types of noise. Few of them are Continuous, Intermittent, Impulsive and Low frequency [13]. The noise can also be represented as colors and most common colored noises are white, pink, blue and brown or red.

A. Types of Noise

1) Continuous

The noise that is continuously present on the spectrum of the signal, for example the sound of a machinery which produce a sound without any interruption. This kind of sound can be due to the processing of machine, its heat or from a ventilation system [13].

2) Intermittent

The noise that increases and decreases over time, for example the sound of a train passing in the background, an electronic machine working on cycles etc. The duration of noise over time can be calculated [13].

3) Impulsive

These noises are fast or sudden burst, for example explosion, piling, drilling etc. We can calculate the peak in these kind of noises [13].

4) Low Frequency

This noise can be a background sound from a far away power plant, or a heavy vehicle or even horns of vehicles etc. In general it is a background humming sound. This noise is hard to process [13].

B. Color of the Noise

The colors of a noise signals are basically come from the spectrum. In spectrum different noises are represented in different colors. It is considered to be one of the most important characteristics of a power spectrum. Each and every colors of noise represent different properties [12].

1) White Noise

This noise has same power in any intervals of frequencies [12]. Examples are fan sound, muffled sound of television, wind blowing, rain, waterfall etc. The name came from the white light in the power spectrum [14]. The below figure (Fig.8) shows the White Noise and figure (Fig.9) represent the comparative representation of white noise in time and frequency domains.

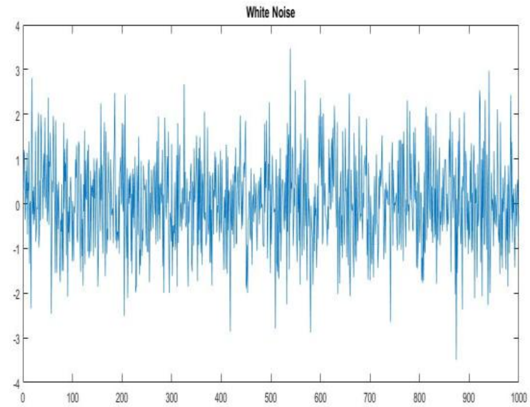

Figure 8: White Noise Representation

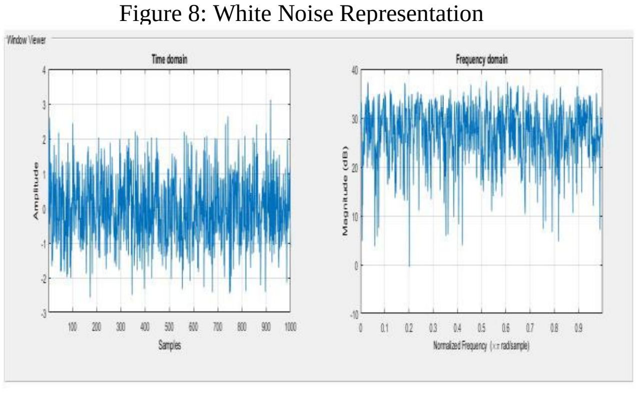

Figure 9: White Noise representation in Time and Frequency Domains

2) Pink Noise

This noise is inversely proportional to the frequency of a sound signal [12]. The name came from the power spectrum where this noise appears as pink in the visible light. The electronic noise such as flicker noise is a pink noise. The pink noises are even present in Biological and Physical systems. For example, heart beat and electromagnetic radiation of some of the space objects [15]. The below figure (Fig.10) shows the Pink Noise and figure (Fig.11) represent the comparative representation of pink noise in time and frequency domains.

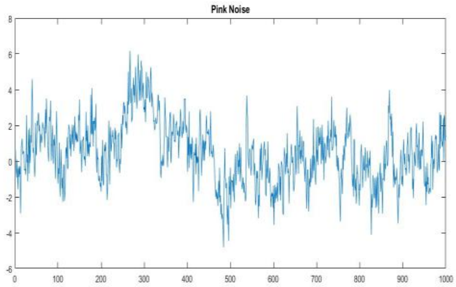

Figure 10: Pink Noise Representation

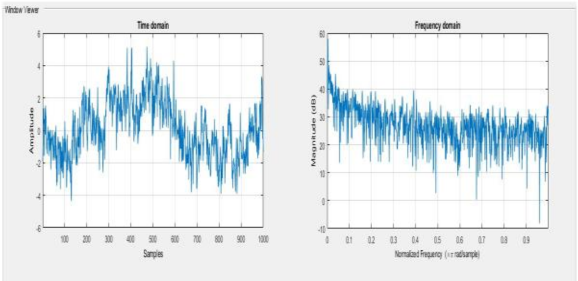

Figure 11: Pink Noise Representation in Time and Frequency Domain

3) Blue Noise

The noise which has 3dB increased power density over the increasing frequency is called blue noise [12]. It is commonly added with sound signal in order to smooth (reduce the distortion of low signals) the audio signals [16]. Some of the blue noises are buzzing, whistling in the ear, kind of sound. Another example for this noise is the Cherenkov radiation [17]. The below figure (Fig.12) shows the Blue Noise and figure (Fig.13) represent the comparative representation of blue noise in time and frequency domains.

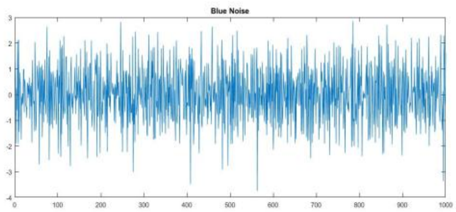

Figure 12: Blue Noise Representation

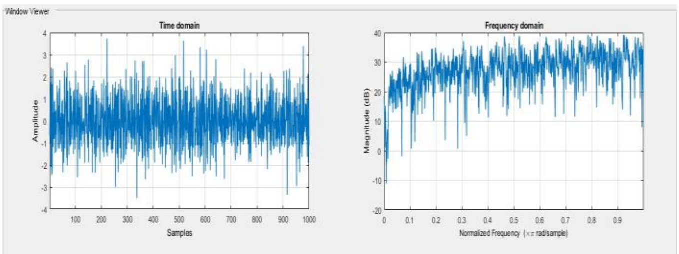

Figure 13: Blue Noise Representation in Time and Frequency Domain

4) Red or Brown Noise

The noise which has 6dB in the power density over the increasing frequency is called brown noise or red noise [12]. Unlike other noises its decreases the power when the frequency increases. This is a noise produced from Brownian motion so this is also known as Brownian noise [18]. This is the integral of a white noise. This noise sound like a low roar. The below figure (Fig.14) shows the brown Noise and figure (Fig.15) represent the comparative representation of brown noise in time and frequency domains.

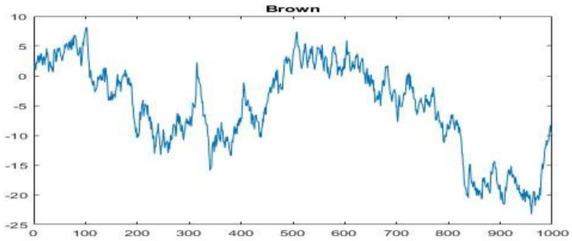

Figure 14: Bown or Red Noise Representation

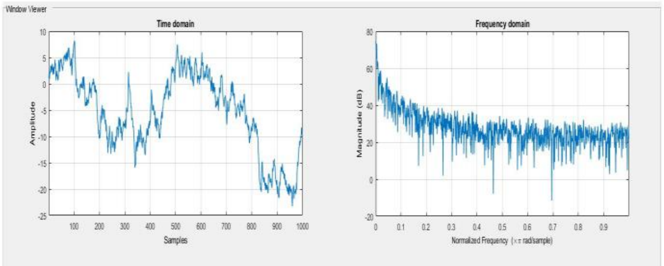

Figure 15: Brown or Red Noise Representation in Time and Frequency Domains

V. CLASSIFICATION

The sound signal processing is not complete without discussing about the classification method used. Classification is a process of grouping the same set of data under desired categories which can be used for identification and analysis. In this paper we discuss about previous work on drum sound classification from 2001 to 2015 and a music classification of drum and guitar sound using k-NN (k-Nearest Neighbor).

1) Previous work

TABLE 1: DETAILS ABOUT PREVIOUS WORK ON DRUM SOUND SIGNAL-PROCESSING

Sl. No	Paper Title	Research Methodology	Result
		Methodology	*Accuracy*
1.	"Automatic classification of drum sounds : a comparison of feature selection methods and classification techniques [19] "	Tree based, statistics and instant based.	97-99%
2.	"Drum Transcription in the presence of pitched instruments using Prior Subspace Analysis [20]"	Prior Subspace Analysis	82.8%
3.	"Prior Subspace Analysis for Drum Transcription [21]"	Prior Subspace Analysis	82.8%
4.	"Retrieval of Percussion Gestures using Timbre Classification Techniques [22]"	Artificial Neural Network	90%

Sl. No	Paper Title	Research Methodology	Result
		Methodology	*Accuracy*
5.	"A simulated annealing optimization of audio features for drum classification [23]	Simulated Annealing and Support Vector Machine classification	95%
6.	"Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine [24]"	Support Vector Machine	93%
7.	"Adamast: A drum sound recognizer based on adaptation and matching of spectrogram templates [25]"	Template Matching	72.8%
8.	"Drum sound recognition for polyphonic audio signals by adaptation and matching of spectrogram templates with harmonic structure suppression [26] "	Template Adaptation and Harmonic Structure Suppression	83%
9.	"Audio Classification from Time-Frequency Texture [27] "	Time-Frequency based classification algorithm is proposed	85%
10.	"Automatic Drum Sound Description for Real-World Music [28] "	Template Adaptation and Template matching	90%
11.	"Musical Notes Identification using Digital Signal Processing [29] "	Method which uses Instant finding and zero padding	-

The above table (TABLE.I) give a brief on previous work done on drum sound classification

2) Classification Drum Sound Signals using k-NN

The drum sound is a heterogeneous sound, but the sound can be classified based on the types of drum used. In the current work the sounds can be classified as snare and kick. A set of 40 drum sound data [30] was used to show the classification of sound signals. In the dataset, each drum audio had mixed sounds such as snare, kick, hit-hat, toms and cymbals and each of these had open, close, high, medium and low levels [31] so the data was considered as heterogeneous. A basic classification experiment was performed. The figure (Fig.16) shows the classification steps.

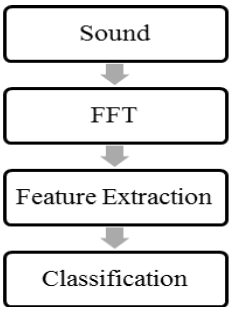

Figure 16: Classification Flow

First step in the classification is a sound data transform from time domain to frequency domain, which was done. After that the features such as centroid, bandwidth, skew, zero crossing, kurtosis [32] are extracted from the signal data. Then the k-Nearest Neighbor method was used for classifying. It used Euclidean distance to get the nearest neighbors. We used 70% data for train and 30% for the test. The function classified the sounds as snare and kick with an accuracy of 78.6.0%. The figure (Fig.17) shows the confusion matrix for the classifier. 100% of the kick sound and 57.1% snare sound was classified correctly. 42.9% of snare value was incorrectly classified as kick.

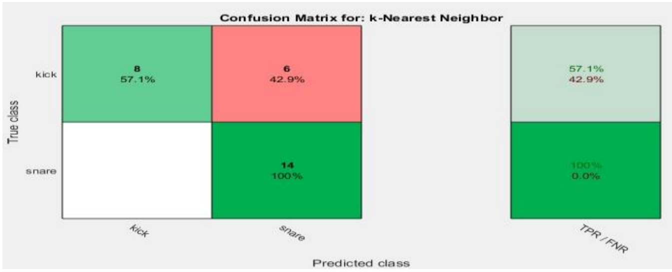

Figure 17: Confusion Matrix for k-NN

The figure (Fig.18) represent the k-Nearest Neighbors in the scatter plot. It shows the drum sound data, [which has both snare and kick sound data values], features and nearest neighbor values. It uses chisquare distance to search k- Nearest Neighbors in a heterogneous dataset. The cross-'x' represent the sample, 'square' represent the features which are used to classify and the 'circle' represent the nearest neighbors.

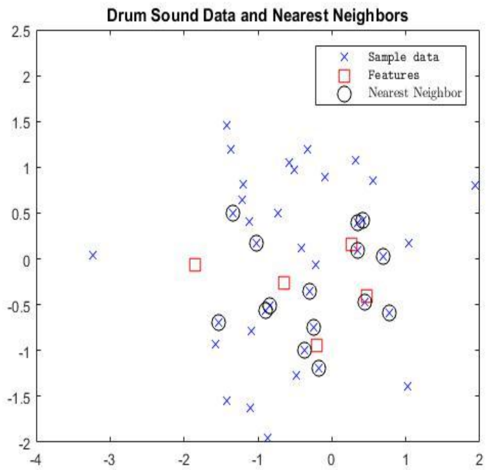

Figure 18: k-NN Classification

VI. Applications

The major application of audio signal processing is for signal transmission. It is an important part of both context aware systems and Human Computer Interaction such as voice recognition systems, audio event search, speech translators, voice navigation, voice assistants etc. Other applications of audio signal processing are hearing aids, audio water marking etc.

Conclusion

The main aim of this paper was to give a brief idea about sound signal processing. There are several areas in this field which has very high scope for research. This paper provides a comprehensive report on domains of sound signal processing, various sound features and noise. There are many different features and noise that are present but this paper discusses only about selected features and noise which has highest priority in the audio/sound signal processing world. Classification of sound signal is a vast research area. This paper gave a brief on the previous research works on the drum sound classification and explained k-NN based drum sound classification. Finally, it gives a brief idea about the applications of audio signal processing.

REFERENCES

[1] Veggeburg. Kurt, " Advanced Signal Processing Algorithm for Sound and Vibrations" [online]. Available: https://www.sem.org/.../Veggeburg_Advanced%20Signal%20Processing

[2] "Time Domain," [online], Available: https://en.wikipedia.org/wiki/Time_domain

[3] "Frequency Domain," [online], Available: https://en.wikipedia.org/wiki/Frequency_domain

[4] Chu, S.; Narayanan, S. and Kuo, C.-C.J, "Environmental Sound Recognition With Time Frequency Audio Features," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, pp. 1142-1158, 2009.

[5] Subramanian, H, " Audio signal classification," *IIT Bombay M. Tech Credit Seminar Report*, pp. 1-17, 2004.

[6] Matlab functions, [online], Available: http://in.mathworks.com/help/signal

[7] Herrera, Perfecto; Dehamel, Amaury and Gouyon, Fabien, " Automatic Labelling of Unpitched Percussion Sounds," *Aes 114Th Convention*, 2003.

[8] "Mel-frequency cepstrum," [online], Available: https://en.wikipedia.org/wiki/Mel-frequency_cepstrum

[9] Zhang, Tong and Jay Kuo, C. C., "Audio content analysis for online audiovisual data segmentation and classification," *IEEE Transactions on Speech and Audio Processing*, vol. 9, pp. 441-457, 2001.

[10] Shih-Fu Chang ; Sikora, T. and Purl, A., "Overview of the MPEG-7 standard," *IEEE Trans. on Circuits and Systems for Video Technology*, pp. 688-695, 2001.

[11] Li, Beinan, " MPEG-7 Audio Overview," [online], Available: www.img.lx.it.pt/~fp/cav/Additional.../MPEG7_Audio_overview_1.pdf

[12] Vaseghi, Saeed. V, "Advanced Digital Signal Processing and Noise Reduction", Ed.4, John Wiley & Sons, pp. 35-50, 2008.

[13] " What are the 4 Types of Noise?," [online], Available: http://www.cirrusresearch.co.uk/blog/2015/01/4-different-types-noise.

[14] Wu, Zhaohua; Huang, Norden E and Zhaohua, By W, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proceding of The Royal Society London*, vol. 1221, pp. 1597-1611, 2004.

[15] Keele, D. B. "The Design and Use of a Sample Pseudo Random Pink-Noise Generator," *Journal of The Audio Engineering Society*, vol. 21, pp. 33- 41, 1973.

[16] Ulichney, Robert a., "Dithering with Blue Noise," *Proceedings of the IEEE*, vol. 76, pp. 56-79, 1988.

[17] "Blue Noise," [online], Available: https://en.wikipedia.org/wiki/Colors_of_noise#Blue_noise

[18] "Brownian Noise," [online], Available: https://en.wikipedia.org/wiki/Brownian_noise

[19] Herrera, Perfecto; Yeterian, Alexandre and Gouyon, Fabien, "Automatic classification of drum sounds : a comparison of feature selection methods and classification techniques," *Journal of Music and Artificial Intelligence*, pp. 69-80, 2002

[20] FitzGerald, D ; Lawlor, B and Coyle, E, "Drum transcription in the presence of pitched instruments using prior subspace analysis," *Issc*, 2003

[21] FitzGerald, Derry ; Lawlor, Robert and Coyle, Eugene, "Prior Subspace Analysis for Drum Transcription," *114th Convention of Audio Engineering Society*, vol. 1, pp. 5808-5813, 2003.

[22] Tindale, Adam ;Kapur, Ajay ;Tzanetakis, George and Fujinaga, Ichiro, "Retrieval of percussion gestures using timbre classification techniques," *Proceedings of the International Conference on Music Information Retrieval*, pp. 541-545, 2004

[23] Degroeve, Sven and Tanghe, Koen"A simulated annealing optimization of audio features for drum classification," *Proceedings of the 6th International Conference on Music Information Retrieval*, pp. 482-487, 2005

[24] Helén, M and Virtanen, T, "Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine," *Proc. EUSIPCO*, vol.1, 2005

[25] Yoshii, K; Goto, M and Okuno, Hg, "Adamast: A drum sound recognizer based on adaptation and matching of spectrogram templates," *1st Annual Music Information Retrieval Evaluation eXchange (MIREX)*, 2005

[26] Yoshii, Kazuyoshi; Goto, Masataka and Okuno, Hiroshi G, "Drum sound recognition for polyphonic audio signals by adaptation and matching of spectrogram templates with harmonic structure suppression," *IEEE Transactions on Audio, Speech and Language Processing*, vol.15, pp. 333-345, 2007

[27] Yu, Guoshen and Slotine, Jean-Jacques, "Audio Classification from Time-Frequency Texture," [0nloine], Available: http://arxiv.org/abs/0809.4501, 2008.

[28] Yoshii, Kazuyoshi; Masataka Goto, and Hiroshi G. Okuno. "Automatic Drum Sound Description for Real-World Music Using Template Adaptation and Matching Methods." *ISMI*, pp. 184-191, 2004

[29] Patel, Jay K and Gopi, E.S, "Musical Notes Identification using Digital Signal Processing," *Procedia Computer Science*, vol.57, pp.876-884, 2015

[30] "Freesound," [online], Available: https://freesound.org

[31] Herrera, Perfecto ; Yeterian, Alexandre and Gouyon, Fabien, "Automatic classification of drum sounds : a comparison of feature selection methods and classification techniques," *Music and Artificial Intelligence*, pp. 69-80, 2002

[32] "CCRMA Workshop on Music Information Retrieval at Standford University," [online], Available: https://github.com/stevetjoa/ccrma/blob/master/README.md