Intelligence Algorithm for the Speech Analysis in Security & Detection Applications

Swati Mangesh Khandare[1]

[1] Research & Training Department

Abstract. Speech analysis creates wide application in security & forensic detection and it is source of accurate speech detection. This intelligence algorithm based on speech analysis will analyses accurate vocal pitch across the full spectrum of speech & sound signals.

Intelligence algorithm will easily extract the speech or sound signals for particular duration from previous recordings of a person, bird or animal's sound after that it will analyses the voice signals and calculate the different specifications of the voice signals like pitch, volume & stereo.

Intelligence algorithm will improve the quality of voice signals using different methodologies like autocorrelation, histogram spread & cumulative sum.

Keywords: Intelligence, Algorithm, Speech Signal, Quality of Voice (QoV).

1 Introduction

The speech or sound signals are broadly classified into Continuous Time Signal and Discrete Time Signal. The speech signal gives oscillations that propagate as an acoustic wave, generally through air as a medium. There are different ways to record sound waves. The human have the hearing voice frequencies from 20 Hz to 20 kHz are called as audio frequencies. The range of audio frequencies of different animals and birds varies from 10 Hz to 150 kHz. The table 1 shows the range of frequencies of different birds & animals.

Noise is the unwanted signals added to the speech signals. This is unwanted extra added component in sound. Mostly natural sound frequencies are complex in nature with different range of frequency components. The acoustic is the combination of kind of sound frequencies. Generally sound signals generated by human are called soundscape is the part of acoustic signals.

The speech signals generated by human like sound waves are generated by different sources like vibrating diaphragm of the stereo speaker. Sound can propagates through a medium as longitudinal waves and also transverse wave in solids. There are medium through which sound travels are different but at the reception sound can be represented by two physical quantities pressure and time. This is simple representation of sound waves which is in form of sinusoidal waves of different frequencies.

Table 1. Frequency range of different birds, animals.

Name	Range of frequencies
Rabbit	96Hz-49kHz
Horse	55Hz-33.5kHz
Cow	23Hz-35kHz
Owl	200Hz-12kHz
Sheep	125Hz-42.5kHz
Dog	64Hz-44kHz
Elephant	17hz-10.5kHz

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 07 | July - 2021 ISSN: 2582-3930

2 Measures of Analysis for Speech or Voice Frequencies

The speech signal is clear in terms of intensity range so can be recorded, giving broad-spectrum and obtained for analysis purpose. To analyze the speech signals of human, bird or animals over time the temporal envelope and temporal fine structure as perceptually relevant analysis is best method. There are aspects like loudness, pitch and timbre perception and spatial hearing will help in speech analysis and leads to detection of voice of human, bird or animal & the measurements can be utilized in security applications based on speech or sound.

The main task is to record the sound generated by different means of communication or by different medium. To analyze the signals need to input the signals like we can analyses different signals are airborne microphone signal, contact microphone signal, filtered signals, LPC residue signal, speech material.

There are measures of analysis of different sound signals generated by many sources but, frequency and loudness are the major characteristics of sound can be utilized for the analysis.

3 Artificial Intelligence Algorithm for Analysis of Speech or Voice Frequencies

In this algorithm the input is sound signal transducer by microphone and further saved as recorded sound file. In this algorithm the recorded sound file converted to wave file for further processing.

To analysis of different sound waves the sampling rate should be proper to achieve the Nyquist Rate. If we want to convert continuous sound into digital or discrete form then digitization is process where first stage is sampling after acquisition of sound signal. These signals are sampled at particular sampling rate in Hz or KHz it has to be twice of input signal. The different sound samples can be collected at different sampling rate, here in this algorithm sampling rate is 18000 Hz and sampling scale maximum amplitude is one.

The following figures 1, 2 & 3 showing input signals with centre frequency, right channel and left channel frequency in both the channels negative peak and positive peak of amplitude with respect to different time limit. The signals are sinusoidal continuous signals.

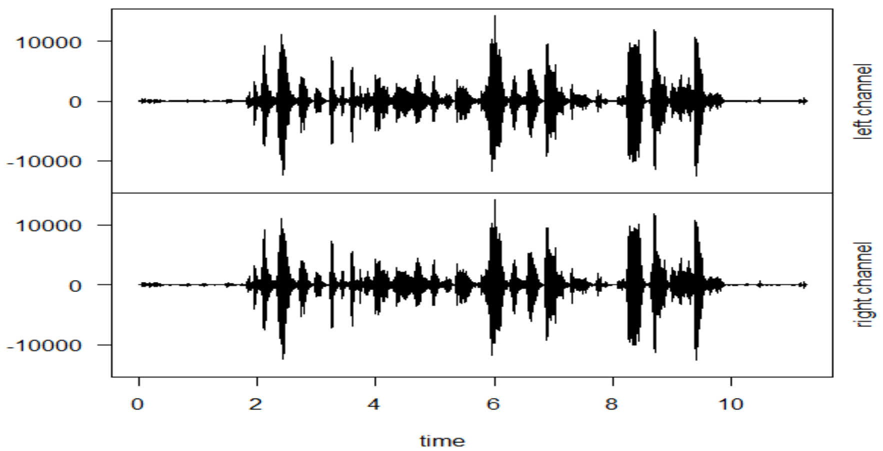

Fig. 1. A figure shows input signal stereo plot for ten seconds.

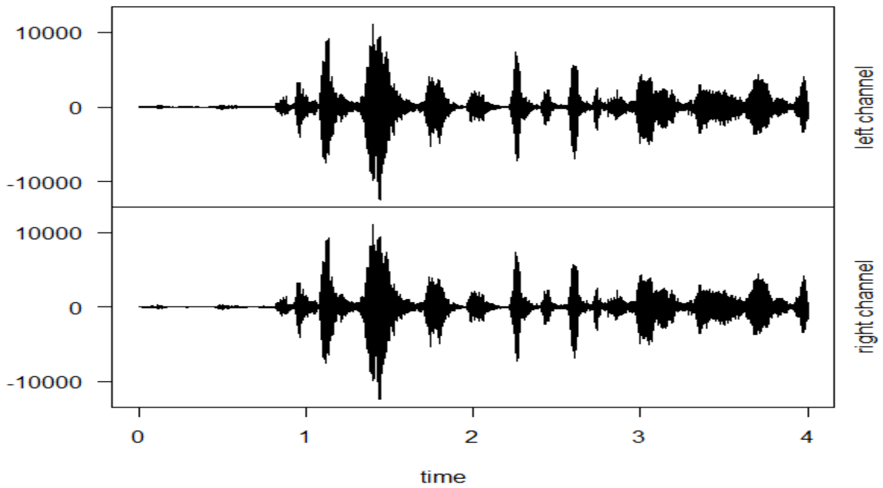

Fig. 2. A figure shows input signal stereo plot for four seconds.

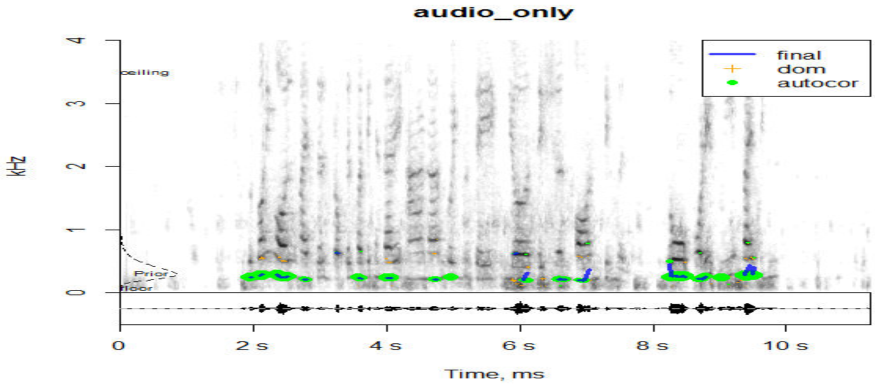

Fig. 3. A figure shows sampled signal at sampling rate of 18000 for time ten seconds.

The Fourier Transform converts time domain signal into a sum of finite series of sine or cosine functions. The Fast Fourier Transform is tool of transform the signal quickly into sum of finite series basically it will convert time domain signal into frequency domain. Here the sampled signals are in discrete form so here Discrete Fourier Transform had specified frequency & amplitude at maximum at one scale.

The true value from synthesized signal is useful signal. Before applying DFT estimation of true pitch value is normalized value. After applying DFT the signal will be represented for Short Term Fourier Transform.

The whole signal no need to use for analysis as it will increase bandwidth and requires more time for analysis. The information can be extract from small time duration signal. The solution is DFT on successive sections along the signal. A window is then convolved along the signal and a DFT is computed at each sampled signal. The whole transformation is part of analysis of sound signal. Short Term Fourier Transform followed by averaging data frames then transformed & average transformed signal displayed in figure 4 below.

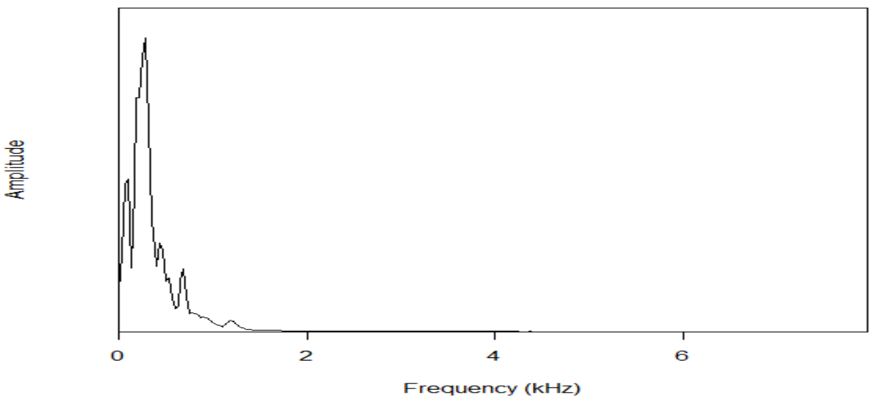

Fig. 4. A figure shows average data frame with frequencies and their strength

4 Quality of Voice (QoV)

In pitch Perturbation pathological voices tend to show unusually large cycle-to-cycle fluctuations in the fundamental period. The phenomenon of cycle-to-cycle fluctuations in the fundamental period is referred to variously as pitch perturbation, fundamental frequency perturbation, or vocal jitter. Percent jitter is defined as mean jitter divided by the mean period, multiplied by 100.

Amplitude perturbation, or vocal shimmer, is defined as cycle-to-cycle fluctuation in the amplitudes of adjacent pitch pulses. The simplest is mean shimmer, which is simply the average absolute difference in amplitude between adjacent pitch pulses. Methods based running averages are also in widespread use.

Due, in part, to a variety of technical problems in measuring perturbation from natural speech signals, there is no consistent evidence for a straightforward relationship between perturbation values measured from natural speech and perceptual dimensions such as roughness, hoarseness, or overall severity of dysphonic prompting some investigators to question the utility of perturbation measures for characterizing voice quality[1].

In the speech processing here I have analyze self-audio record data base for the processing and to analysis the quality of voice for fundamental period. The fluctuation in voice is clearly shown with the phenomenon in sinusoidal cycles. The voice is containing the pauses which referred as vocal shimmer. The voice contains jitter which is less in percentage due to good recording sound. Due to technical problems in speech the amplitude of voice is low.

The voice quality is depending on less percentage of perturbation present in sound. There is very less possibilities of quality of voice present in nature.

5 Autocorrelation for Voice Determination

After transform spectrum to Power Density Function, need to process each frame after storing. The fundamental frequency in the sampled voice signal is Pitch. For finding the speech signal that is actual Pitch from the spectrum here the function of autocorrelation is deployed. The autocorrelation function computes the two speech signals at the highest peak of sinusoidal signal amplitude. The first signal is actual discrete sampled signal and second is same signal delayed by some time duration. The result of autocorrelation function is a measure of Pitch Amplitude. Here the difference will lead to detection of required speech samples.

The loudness analysis is measure of intensity of speech. Skew analysis is a measure of asymmetry of the probability distribution of real valued random variables in spectrum. Sweep is cumulative sum signal for some duration will give pure speech signal over that duration of spectrum.

6 Conclusion

Speech is important feature of human, in various applications like security and detection we need clear voice signal for detection. Intelligence algorithm has extracted, processed & analyzed by deploying autocorrelation function. Figure 6 shows the output of autocorrelation function. Autocorrelation function clear the noise and improved quality signal further analyzed using histogram spread spectrum.

The analysis part displayed using different plots, figure 7 shows polyphonic pitch, autocorrelation function, smoothness, cepstrum method, short term phase spectrum and measure of sub-harmonics.

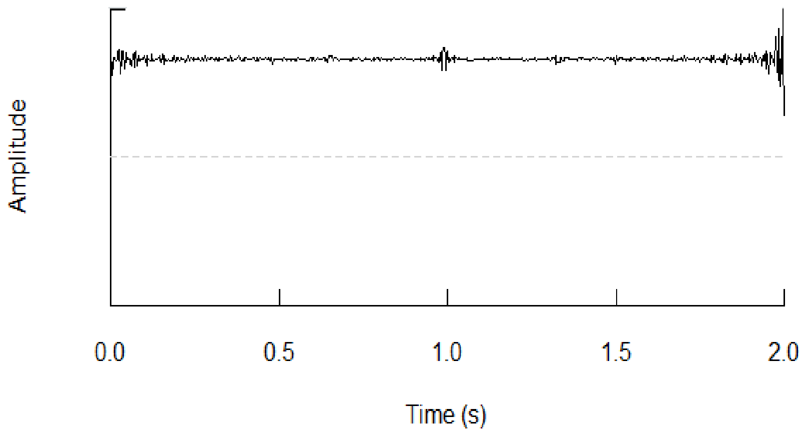

Fig. 6. A figure shows output of autocorrelation function

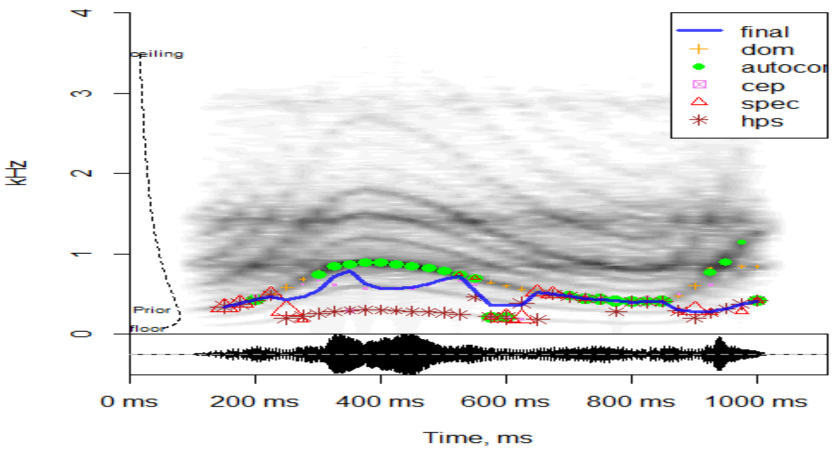

Fig. 7. A figure shows polyphonic pitch, autocorrelation function, smoothness, cepstrum method, short term phase spectrum and measure of sub-harmonics.

References

1. James M. Hillenbrand, Acoustic Analysis of Voice: A Tutorial https://www.researchgate.net/publication/270860066.
2. Jerome Sueur, Seewave, A very short introduction to sound analysis for those who like elephant trumpet calls or other wildlife sound Paris, France May 28, 2020.