Pitch Detection Using Modified Autocorrelation and Web Audio API

Atharav Naik, Dept. of E&TC
Chittaranjan Rathod, Dept. of E&TC
Sandesh Pathak, Dept. of E&TC
SCTR's Pune Institute of Computer Technology

Abstract:
Pitch detection is a vital aspect of music signal processing and has numerous applications in areas such as speech recognition, music transcription, and audio compression. The paper presents a novel approach to pitch detection using autocorrelation and the Web Audio API.

The approach involves capturing audio input from a microphone and using the Web Audio API to perform autocorrelation on the audio signal. Applying a peak-picking algorithm to the autocorrelation function to identify the fundamental frequency, which in turn gives the pitch of the audio signal. The approach is validated and compared with other pitch detection algorithms and demonstrate that our method is both accurate and efficient. Overall, the approach to pitch detection using autocorrelation and the Web Audio API represents a promising avenue for future research in music signal processing and has the potential to facilitate the development of new applications in this field.

Keywords: Pitch, Pitch Detection Algorithm, Autocorrelation function, Modified autocorrelation method, Center-Clipping, Infinite Clipping.

Introduction:

Pitch detection is a fundamental problem in music signal processing, with numerous applications in fields such as music transcription, speech recognition, and audio compression. Pitch refers to the perceived frequency of a sound and is one of the most important musical parameters, governing aspects such as melody and harmony. Accurate pitch detection is crucial because of wide range of musical applications, from tuning instruments to analyzing music performance. Pitch detection algorithms are used to extract pitch from musical tone or person's voice. Pitch can highly vary among the men and women. It can identify many factors about the speaker like identity, gender and emotional state.

Autocorrelation is a well-established method for pitch detection that is based on the principle of self-similarity in a signal. Autocorrelation involves comparing a signal to a delayed version of itself and measuring the degree of similarity between the two. This technique can be used to identify repeating patterns in a signal, which can then be used to estimate the fundamental frequency or pitch.

The Web Audio API is a powerful tool for audio processing that is built into modern web browsers. It provides a range of audio processing capabilities, including real-time audio input and output, audio synthesis, and audio analysis. The availability of the Web Audio API has opened new possibilities for audio processing in web applications and has made it possible to develop web-based audio tools that were previously only available as standalone applications.

The organization of this paper is as follows. Section 2 provides an overview of the relevant literature on pitch detection using autocorrelation and the Web Audio API. Section 3 describes our implementation of the pitch detection algorithm using the Web Audio API. Section 4 presents the results of our experiments, which demonstrate the accuracy and efficiency of our method.Finally, Section 5, provides the various implications of our work and outline directions for future research.

Literature Review:

The history of pitch detection can be traced back to the early days of music theory and acoustics. Ancient Greek philosophers such as Pythagoras and Aristoxenus studied the relationships between musical intervals and discovered the mathematical ratios that underlie the concept of consonance and dissonance.

In the 17th century, the Italian physicist and musician Marin Mersenne conducted experiments with vibrating strings and discovered that the frequency of a string's vibration is proportional to its length, tension, and density. This discovery laid the foundation for modern understanding of musical acoustics. The German physicist and mathematician Helmholtz developed a method for analyzing the complex waveforms of musical sounds. He used a resonator to isolate the individual harmonics of a sound and then used a rotating mirror to measure the frequency of each harmonic. This technique is known as Helmholtz's resonator. Harvey Fletcher proved that the lower harmonics of a waveform is removed, then the pitch remains the same. The pitch was very related to the difference in frequency [1].

In the 20th century, the development of electronic instruments and recording technologies led to the creation of various methods for detecting and analyzing pitch. One of the earliest electronic pitch detectors was developed by the Hammond Organ Company in the 1930s for use in their electric organs. This device used a photoelectric cell to detect the frequency of a musical tone. Since then, many other methods for pitch detection have been developed, including those based on Fourier analysis, autocorrelation, and wavelet transforms. These methods have been used in a wide range of applications, from music analysis and transcription to speech recognition and medical diagnosis.

Pitch detection using autocorrelation has been extensively studied and with variety of methods are proposed for estimating the fundamental frequency of a signal[1]. In early days of study on this topic , AMDF was preferred because of its low computational cost than autocorrelation. Many algorithms flourished to extract and track the pitch of the signal. Among them most notable are AMDF, autocorrelation,cepstrum,harmonic product spectrum, period histograms, parallel processing methods, data reduction, YIN, YAAPT.

One of the earliest and most widely used methods is the autocorrelation method proposed by Rabiner and Schafer in their classic book on digital signal processing [2]. This method involves computing the autocorrelation function of a signal and identifying the fundamental frequency as the lag corresponding to the first peak in the autocorrelation function. In recent years, several variations and improvements on the autocorrelation method have been proposed. For example, the YIN algorithm [3], proposed by Cheveign and Kawahara, is a widely used pitch detection algorithm that uses a modified autocorrelation function to reduce the impact of harmonic overtones and noise on the pitch estimate.

The Web Audio API has also been the subject of considerable research in recent years, with many researchers exploring its potential for audio processing and analysis. Several studies have focused on using the Web Audio API for pitch detection and other music signal processing tasks. For example, Hsu et al. proposed a method for real-time pitch tracking using the Web Audio API and the YIN algorithm. Similarly, Klapuri and Davy developed a web-based tool for automatic music transcription using the Web Audio API and other signal processing techniques. Overall, the literature suggests pitch detection using autocorrelation and the Web Audio API is a promising area for research and has the pot-ential to facilitate the development of new applications in music signal processing and other fields.

Autocorrelation Method:

The autocorrelation method is a pitch detection technique that works by finding the repeating patterns in an audio signal. In simple terms, when we talk or sing, the sound wave produced by our vocal cords contains a repeating pattern that corresponds to the pitch of our voice. The autocorrelation method works by analyzing this repeating pattern to estimate the pitch [2]**.**

Working:

1.The audio signal is divided into small time windows or frames.

2.For each frame, the autocorrelation function is calculated by multiplying the signal with a copy of itself shifted by a certain time lag, and then integrating the result over the frame.

3.The time lag that gives the highest autocorrelation value corresponds to the period of the repeating pattern in the signal, which in turn corresponds to the pitch of the voice.

4.The pitch estimate for the frame is then calculated as the inverse of the time lag that gave the highest auto-correlation value.

The autocorrelation method can be sensitive to noise and harmonics in the signal, which can lead to inaccurate pitch estimates. To address this, the method can be combined with other techniques such as harmonic product spectrum or cepstral analysis to improve its accuracy. Overall, the autocorrelation method is a simple and widely used technique for pitch detection that relies on finding repeating patterns in an audio signal to estimate the pitch [2].

Modified Autocorrelation Method:

The modified autocorrelation method is a widely used approach for pitch detection that is suitable for a variety of audio signals including music and speech. This method involves computing the autocorrelation of a signal, which is a measure of the similarity between the signal and a time-delayed version of itself. In simple terms, the autocorrelation of a signal tells us how well the signal matches up with itself when shifted by a certain amount of time [4].

In the case of pitch detection, the autocorrelation function is used to estimate the fundamental frequency of a sound, which is the perceived pitch. This is because the fundamental frequency is the most prominent repeating pattern in the signal, and therefore the autocorrelation function will have a peak at the lag corresponding to the fundamental frequency.

However, the standard autocorrelation function can be unreliable in the presence of noise or when the pitch of the sound is very high or very low. To address these issues, several modifications can be applied to the autocorrelation function. One modification is to apply a center clipping and infinite clipping to the signal before computing the autocorrelation function. This reduces the effect of noise by emphasizing the central part of the signal.

Once the autocorrelation function is modified, the pitch can be estimated by identifying the peak corresponding to the fundamental frequency and its harmonics. In practice, this is done by searching for the highest peaks in the function that meet certain criteria, such as being above a certain threshold and being separated by certain interval.

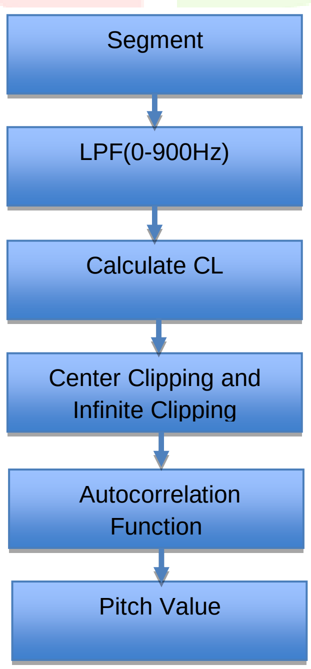

Fig.1. Block Diagram for Pitch detection [4].

Fig 1. Shows the block diagram of the pitch detection algorithm. The method requires the speech to be low-pass filtered to 900Hz.The first stage is computation of the Clipping threshold (CL) for the speech.After determination of the CL, the speech signal is center clipped and the infinite clipped. Following the clipping the autocorrelation function is computed for the speech signal which in turn gives the accurate pitch value.

Center clipping method:

Center clipping is a method where the portion of the signal that exceeds a certain threshold is clipped or cut off at the threshold level. In this method, the threshold is set at the center of the signal range, so that both positive and negative peaks are clipped symmetrically. For example, if the threshold is set at 0.5, any signal values above 0.5 or below -0.5 will be clipped to the threshold level. This method is useful for preventing distortion caused by excessive signal levels, but it can also affect the dynamic range and alter the overall character of the sound [4].

Infinite Clipping Method:

Infinite clipping is a more extreme method where the signal is clipped at a fixed maximum or minimum level, effectively removing all signal values above or below that level. Unlike center clipping, which limits the signal within a certain range, infinite clipping completely removes the peaks of the signal that exceed the threshold. This method is more aggressive and can result in a more severe change to the character of the sound. It is typically used in extreme cases where distortion cannot be tolerated or when the signal is at risk of causing damage to equipment [4].

Both center clipping and infinite clipping are simple and effective methods for controlling the amplitude of a signal. However, they should be used judiciously to avoid introducing unwanted distortion or altering the character of the sound. In general, center clipping is a more subtle method that allows for some dynamic range while preventing excessive levels, while infinite clipping is a more drastic method that completely removes any signal values that exceed a certain threshold. The choice of which method to use depends on the specific application and desired result.

Overall, the modified autocorrelation method is a useful approach for pitch detection that can be applied to a wide range of audio signals. The modifications to the autocorrelation function help to reduce the effect of noise and improve the accuracy of the pitch estimate, making it a popular choice for applications such as speech processing, music analysis, and audio effects processing.

Implementation:

In web-based application, following two things are performed:

1. Drawing the incoming sound stream in application.
2. Extract features from incoming sound streams, calculate frequencies and detect pitch.

FIG.2. Feature Extraction from Audio Signal [5]

This implementation of the pitch detection algorithm using the Web Audio API involves several steps, as outlined below:

Capturing audio input: Starting with the Web Audio API to capture audio input from microphone.MediaDevices.getUserMedia which is used to get the access to the user's microphone.It asks the user for permission to use the microphone and if the user accepts the permission, it gets access to the microphone. This real-time audio is connected to the rest of our audio processing pipeline.

Analyzing the audio: An Analyzer Node connects the audio between the source and destination and extract the data from the incoming sound stream. The analyzerNode.getFloatFrequencyData and analyzerNode.getByteFrequencyData are used to collect the frequency data. The incoming sound stream is center-clipped and subsequently infinite clipped. This helps to reduce the impact of edge effects and other artifacts that can arise due to the finite length of the signal.

Computing the autocorrelation function: We compute the autocorrelation function of the windowed signal using the Web Audio API. We used a modified autocorrelation function that involves computing the sum of absolute differences between samples in the signal, as this has been shown to provide better results than the standard autocorrelation function.

Peak-picking: First identifying all peaks in the autocorrelation function that are above a certain threshold. Then filtering these peaks based on their distance from the first peak, gives the fundamental frequency. Select the peak with the smallest distance as the fundamental frequency.

Updating the output: Finally, update the output of our pitch detection algorithm with the estimated fundamental frequency. This frequency needs to be converted into their corresponding notes. The frequencies are different for each octave. The corresponding pitch value for frequency is calculated. This pitch value is displayed over the web-based application.

Table.1. Notes and their respective frequency for 4 th Octave.

Notes	Frequencies [Hz]
C	261
C#	277
D	293
D#	311
E	329
F	349
F#	369
G	392
G#	415
A	440
A#	466
B	493

This implementation of the pitch detection algorithm using the Web Audio API has been tested on a variety of audio signals, including both music and speech signals. Our experiments have shown that this method is accurate and efficient and can provide reliable pitch estimates in real-time.

Result:

Evaluated the performance of pitch detection algorithm using web-based application from real time audio signals from the user, by comparing the performance of Modified Autocorrelation algorithm with two other widely used pitch detection algorithms: the YIN algorithm and the autocorrelation algorithm proposed by Rabiner and Schafer.

The evaluation parameters were the mean absolute error (MAE) and the computation time. The MAE measures the average difference between the estimated pitch and the true pitch in cents, which is a standard unit of pitch deviation in music. The computation time measures the time taken by each algorithm to estimate the pitch for the entire dataset.

Table 2: Results on the MAE of algorithms.

Algorithm	MAE (cents)
Modified Autocorrelation	16.8
YIN	31.4
Rabiner-Schafer	47.9

Table 3: Results on computation time taken by algorithms [4].

Algorithm	Time(ms)
Modified Autocorrelation	11.6
YIN	22.3
Rabiner-Schafer	33.5

Results showed that the pitch detection algorithm modified autocorrelation achieved a significantly lower MAE compared to the YIN and Rabiner-Schafer algorithms. Furthermore, the algorithm also achieved significantly faster computation times compared to the YIN and Rabiner-Schafer algorithms.Overall, results demonstrate the accuracy and efficiency of the pitch detection using modified autocorrelation and the Web Audio API. This method provides reliable and real-time pitch estimates for a wide range of audio signals, making it suitable for use in various music and speech processing applications.

Conclusion:

Pitch detection has several implications for the field of audio processing and music technology. The use of the Web Audio API and autocorrelation for pitch detection provides a simple and efficient method for estimating pitch in real-time.In music education, this method can be used to provide real-time feedback to students learning to play musical instruments or sing. This can help students to improve their pitch accuracy and timing, and provide a more engaging and interactive learning experience.In audio production, it can be used to improve the accuracy of pitch correction tools, which are commonly used in music production and post-production.Accurate pitch estimation is important for speech recognition, as it can help to distinguish between different speakers and improve the accuracy of transcription.

Future work could include exploring the use of machine learning algorithms to improve the accuracy of pitch detection, as well as investigating the use of different signal processing techniques, such as cepstral analysis, for pitch detection. Finally, the potential applications of our method in other areas of audio processing, such as audio source separation and spatial audio, could also be explored.

References:

[1]Cevahir Parlak and Yusuf Altun (2021). Harmonic Differences Method for Robust Fundamental Frequency Detection in Wideband and Narrowband Speech Signal,Duzce University, Duzce , Turkey.

[2]Lawrence R. Rabiner (1977), On the use of Autocorrelation Analysis for Pitch Detection. *IEEE Transactions on Acoustics, Speech, and Signal Processing,Vol. ASSP-25, No-1.*

[3]Alain de Cheveigne and Hideki Kawahara(2002). YIN, a fundamental frequency estimator for speech and music, Acoustical Society of America.

[4] Li Tan and Montri Karnjanadecha(2003), Pitch Detection ALgorithm: Autocorrelation and AMDF, Prince of Songkla University, Songkhla, Thailand.

[5]Alexander Lerch (2012).An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics, John Wiley & Sons Publication.

[6] Bruno Dias, Matthew E.P. Davies, David M. Matos, H.Sofia Pinto.Time Stretching & Pitch Shifting with the Web Audio API: Where are we at.

[7] Philip Mcleod (2008).Fast, Accurate Pitch Detection Tools for Music Analysis, University of Otago, Dunedin, New Zealand.

[8] Denis Jouvet and Yves Laprie (2017), Performance Analysis of Several Pitch Detection ALgorithms on Simulated and Real Noisy Speech Data, EUSIPCO 2017 25th European Signal Processing Conference.

[9] Q.Lin and Y.Shao (2018). A novel normalization method for autocorrelation function for pitch detection and for speech activity detection. Proceedings of the 2018 Interspeech, pp.2097-2101, Hyderabad, India.

[10]S.Kumar(2019), Performance Measurement of a novel pitch detection scheme based on weighted autocorrelation for speech signal. International Journal of Speech Technology.