Techniques for Enhancing Monophonic Audio to Classify Instruments in Polyphonic Music

1st Dony Armstrong Dsouza, 2nd Dr. Veena Devi Shastrimath V

[1]Senior Assistant Professor/Research Scholar, [1]Electronics & Communication Engineering Department,
[1]MITE, Moodabidri, INDIA
Visvesvaraya Technological University, Belagavi-590018
[1]dony.armstrong@gmail.com
[2]Professor, [2]Electronics & Communication Engineering Department
[2]NMAMIT, Nitte, INDIA
Visvesvaraya Technological University, Belagavi-590018
[2]veenadevi@nitte.edu.in

Abstract:- The field of music has been a subject of great fascination since the inception of human civilization, characterized by continuous advancements and developments. The availability of music teachers has been a challenge for anyone seeking to learn to play music, particularly due to the diverse range of percussion and string instruments that have been invented over time. Consequently, a significant challenge faced by many artists pertains to acquiring the necessary materials for learning, such as understanding the background or comprehending the specific role of a single instrument within an audio clip. Therefore, locating a suitable music track for practice is a considerable challenge, leading many musicians to encounter difficulties in this regard. Numerous attempts have been made by researchers and practitioners to extract the signal of a musical instrument from an audio track as a means of resolving this issue. However, the identification of a single instrument, specifically the violin, is being considered. This study considers various sets of instrument data. There exist numerous possibilities within this expansive domain, and a multitude of methodologies have been developed and explored by various academics. Certain individuals have utilized statistical models for computation, while others have experimented with other filter models. The features to be extracted for the identification of the musical instrument, specifically the violin, are determined by analyzing the spectrogram. Nevertheless, there exist several prospects in this domain that can be explored through various methodologies for future implementation.

***Keywords* -** MFCC, GFCC, tonal, Multitonal, Chords, Spectrogram.

I. INTRODUCTION

The prevailing method of playing the violin involves the utilization of a bow, which is drawn across the instrument's four or five strings. These strings are typically tuned to the notes G3, D4, A4, and E5, with the intervals between them conforming to perfect fifths [1]. On certain occasions, the violin can be played through the technique of strumming and plucking the strings with the fingers, known as pizzicato. Additionally, in limited and uncommon situations, the wooden portion of the bow may be utilized to strike the strings in order to achieve specific desired effects. The violin is often regarded as a prominent and indispensable instrument in several genres of music across different cultures worldwide. The primary instrument in western classical music, as well as a key component in various ensembles ranging from chamber music to symphonies, exhibits the most dominance. Moreover, violins assume a prominent position among other folk music genres, including jazz, country, and bluegrass. The endeavor of acquiring proficiency in playing the violin is frequently accompanied by difficulties that often discourage individuals from pursuing violin instruction. The violin is extensively utilized in various musical contexts, including symphonies and melodies, so establishing its widespread usage. Moreover, numerous endeavors were made to isolate the violin melody from the tracks, with the aim of focusing solely on the individual instrument during the process of learning. Certain electric violinists utilize solid-body instruments equipped with piezoelectric pickups.The Violin occupies a significant position within the realm of musical instruments due to its extensive array of dynamic features. The sound produced by a violin can exhibit significant variation depending on the specific bowing techniques employed [2]. When the violin is played in conjunction with other instruments, it exhibits a remarkable ability to harmonize seamlessly with the music and align with the specific genre in which it is being performed [7]. The violinist encounters the task of discerning the appropriate finger placement on the fingerboard of the instrument. Therefore, for a novice and inexperienced learner, the task of discerning the sound of a violin from an audio recording proves to be a challenging endeavor. Therefore, the task of discerning and isolating the auditory characteristics of a violin from a pre-recorded audio track presents significant challenges [6]. The violin is an instrument that produces a range of sounds, starting with a low G3 with a frequency component of 296 Hz, and to its greatest practical limit at E7 with a frequency of 2637 Hz [3]. The violin has a remarkable degree of versatility and use as a musical instrument, owing to its expansive tonal range, dynamic capabilities, and its capacity to generate a diverse array of textures, timbres, and articulations. Therefore, the process of associating the sound of a violin with its distinctive characteristics yields a wide range of unique and varied sounds for categorization.

II. Related Works

The characteristics of the violin mostly pertain to its architectural design [4] and the manner in which it is crafted. The physical composition of the violin also significantly contributes to its distinctive sound. The violin encompasses several elements of bowing techniques and pertinent details, primarily taking into account the wood composition and hull specifications [5]. The instrument exhibits a wide spectrum of sounds [6]. Each instrument possesses unique qualities in relation to its device properties. Timbre and instrument are crucial factors in delineating the features of an instrument for purposes of identification.

Duration is the primary characteristic that may be readily differentiated from timbre. It exemplifies the rationale underlying the definition of timbre, despite the absence of a universally accepted description for the perception of duration. A perceptible alteration in auditory perception occurs when a singer sustains a musical tone for an extended duration compared to a briefer duration. It is important to note, however, that this alteration does not correspond to a modification in the quality of the sound, commonly referred to as timbre. Once the duration of a sound exceeds the temporal frame used to generate the magnitude spectrum, the magnitude spectrum is not affected by the duration. Typically, the duration of sustained notes in music tends to be longer [7], [8].

Wang and Qi Bao have conducted research on the recognition of individual instruments of a similar type, a task that poses significant challenges and has received limited attention in the literature. The violins were examined in order to be individually detected and identified using the utilization of a source filter model. Once the spectral components have been plotted, they are subsequently classified into tonal and non-tonal content [16]. This phenomenon is frequently observed in the context of the instrument's timbre. The spectrum is described using gamma tone, which is content similar to MFCC and GFCC. The researchers utilized a Gaussian mixture model for the purpose of parameterization [7]. Various music signal processing techniques utilize ADSR curves to simulate musical instruments, whereas the application of filters is prevalent in numerous cases [9], [10]. Gaussian mixture models are widely regarded in the literature [11]. However, it is important to note that the complexity of the system is higher in this context. Additionally, the Support Vector Machine method requires a significant number of computational calculations. The researchers utilized a Gaussian mixture model for the purpose of parameterization [7]. Various music signal processing techniques utilize ADSR curves to simulate musical instruments, while filters are commonly employed in numerous cases [9], [10]. Gaussian mixture models are widely regarded in the literature [11]. However, it is important to note that the complexity of the system is higher in this context. Additionally, the Support Vector Machine method requires a significant number of computational calculations.

III. METHODOLOGY

The suggested system primarily focuses on the acquisition of the sample violin tone and its subsequent processing using signal processing techniques. The fundamental unit of the system is depicted in Figure 1.

The Acquisition of the Violin Sound Signal is done by considering the violin being in this context is a conventional 4/4 violin, and the arrangement for collecting data on the A4 note is depicted in Figure 2.

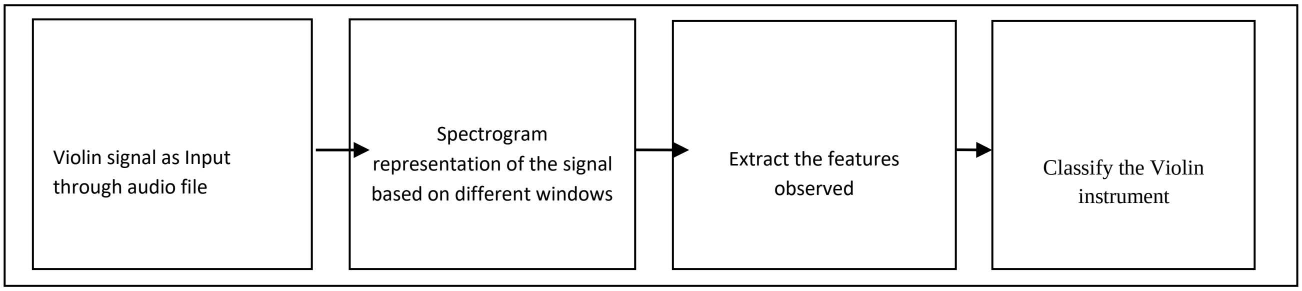
Violin signal as Input through audio file
Spectrogram representation of the signal based on different windows
Extract the features observed
Classify the Violin instrument

Figure 1. Basic block diagram of the System

The Acquisition of the Violin Sound Signal is done by considering the violin being in this context is a conventional 4/4 violin, and the arrangement for collecting data on the A4 note is depicted in Figure 2

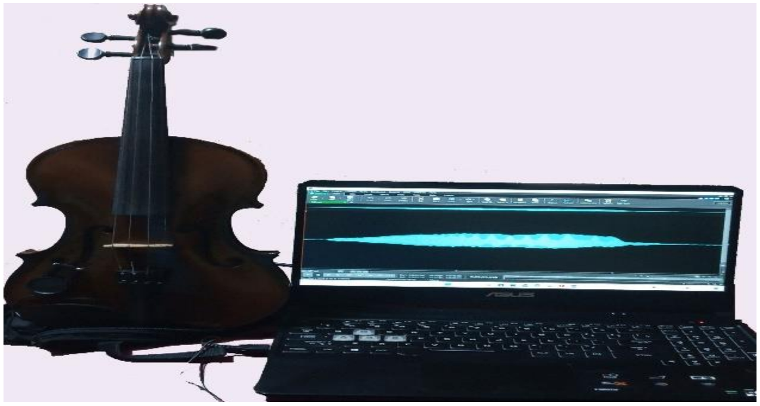

Figure 2. Setup for Data Acquisition

The musical note denoted as A in notation for a single beat is depicted in Figure 3.

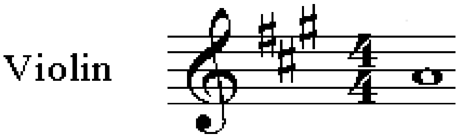

Figure 3. Musical Notation representation of Note A4

The experimental configuration involves the utilization of a piezoelectric pickup to connect the violin, as opposed to the conventional coil wrapped pickups often employed for guitar pickups [9]. The guitar pickups, on the other hand, exhibit a narrower frequency response that selectively captures the vibrations generated by metal strings. The piezoelectric pickup utilized in violins exhibits an expanded dynamic frequency range, hence enabling the capture of ambient sounds. The process of isolating these sounds and restricting acceptance just to the violin melody presents certain challenges. The sound produced by the violin is captured at a sampling rate of 44.1KHz and subsequently stored as data using the wavread command in MATLAB, which employs a sampling rate of 8KHz. The A4 note on the violin is commonly recognized as the standard frequency for tuning musical instruments using a 440Hz signal, which is widely accepted across various contexts.

The violin tone is performed in a frequently used time signature of 4/4, utilizing the note A, and subsequently captured in a recording. The captured audio is subsequently imported into the MATLAB environment in the WAV file format [10]. The reference pitch for the violin is often designated as A4 with a frequency of 440Hz. This pitch is produced by playing an open string on the instrument. It is important to note that the sound produced by the open string has additional dynamics. Additionally, it is important to acknowledge that the open string may be subject to various constraints, such as the introduction of additional disturbances and harmonics during the recording process. The technique employed involves applying upward pressure to generate the waveform of the violin sound sample. The sound signal produced by the Violin can be captured using the wavrecord function or any other recording tools often used for general purposes. The recording is conducted utilizing a wave recorder, namely an audio grabber, for the intended purpose. The technique employed involves applying upward pressure to generate the waveform of the violin sample. The auditory composition's musical notation is depicted in Figure 4. The musical composition comprises polyphonic notes that coalesce to create the harmonic structure, namely in the shape of E and E5th triad chords. The subsequent recording of a guitar being played follows the same recording technique and solely consists of the rhythmic chord progression in the key of A.

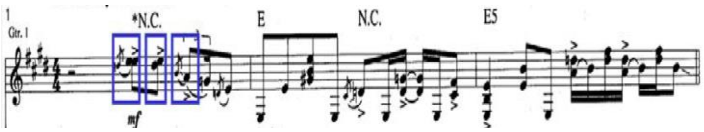

Figure 4. Musical Notation representation of the audio file.

The instrument employed in this context is an electric guitar equipped with magnetic pickups. The acquired signals are a consequence of being sampled at a frequency of 44,100Hz, denoted as the sampling frequency Fs. Figure 5 depicts the flow chart of the proposed system.

The processing of the signal. The resultant signal acquired through the process of sampling is represented as a vector or matrix, which contains the monophonic audio signal

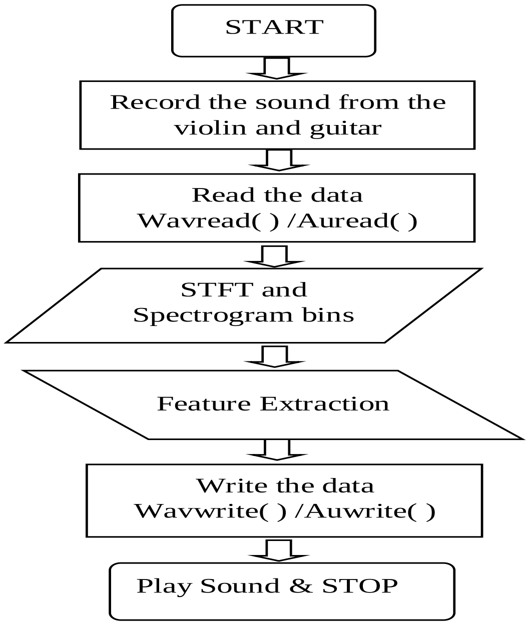
START
Record the sound from the violin and guitar
Read the data Wavread( ) /Auread( )
STFT and Spectrogram bins
Feature Extraction
Write the data Wavwrite( ) /Auwrite( )
Play Sound & STOP

Figure 5. Flowchart depicting the methodology of the proposed system

. The signal is subjected to sampling at a frequency denoted as Fs, and subsequent processing is performed using the MATLAB software. In order to extract spectrograms and gain a more comprehensive understanding of the features, the frequency spectrum is analyzed by taking the absolute Fast Fourier Transform (FFT) of the recorded input violin audio sample. The user has provided a numerical reference, indicating the presence of a citation or source. The spectrogram is utilized to visualize the frequency components of the signal in the frequency domain. The identification of components in the spectral domain involves considering the single note produced by a violin and the polyphonic nature of a guitar chord. The spectrogram is comprised of the temporal and magnitude representations of the frequency components. The signal undergoes processing using equations that take into account the Short-Time Fourier Transform (STFT) with window sizes of 256 and, in certain instances, 512 [13]. The spectrogram approach is commonly employed in numerous research and has consistently produced favorable outcomes when applied to log frequency plots. Moreover, it has been predominantly utilized for the purpose of identification. The identical methodology is employed to achieve the differentiation of various musical instruments utilized.

The timbre of the instrument is a significant feature that should be considered when extracting audio.As per the established definition of timbre, it denotes the distinctive attribute of auditory perception that enables a listener to differentiate between two sounds that are presented in a comparable manner, possessing identical volume and pitch [6].The concept is closely associated with the capacity to discern the origin of an auditory stimulus. For example, despite sharing similarities in terms of loudness and pitch qualities, the sounds produced by the flute and violin can nevertheless be readily distinguished from each other. Furthermore, in the context of polyphonic music, it is common for individuals to perceptually categorize the distinct sounds based on their sources by utilizing timbre information[14].Timbre is a complex concept characterized by numerous underlying auditory characteristics. There exist five primary acoustic components that exert an influence on timbre: 1) The differentiation between the tonal and noise-like attributes of a sound; 2) the spectral envelope; 3) the temporal envelope; 4) the fluctuations in the fundamental frequency and spectral envelope; and 5) the moment at which the sound deviates noticeably from the sustained vibration [15].To calculate the Fourier Transform for specific segments of a lengthy signal sequence, such as in the case of generating the Short Time Fourier Transform, it is necessary to extract brief duration samples. Typically, the generation of a short duration sequence of a lengthy signal involves the utilization of a window function, denoted as w[n], with a significantly narrow bandwidth. This window function is either shifted across the signal or multiplied by the lengthier signal to produce the desired outcome. The often employed window function for obtaining this response is typically a short-duration window, such as the Rectangular Window. This approach is employed to extract the necessary brief sequence without making any further modifications to the signal. Furthermore, it is advisable to utilize tapered ends in order to enhance the accuracy and improvised depiction within the realm of frequency. A commonly desired and employed window, such as the Hamming window, is recommended for this purpose. When considering a continuous frequency variable, the Short-Time Fourier Transform (STFT) can be characterized as.

$$X[n,w] = \sum_{m=-\infty}^{\infty} w[n-m]x[m]e^{-jwm} \text{------ (1)}$$

By use a window of finite length, we can derive the equation as follows.

$$X[n,k] = \sum_{m=n-(N_w-1)}^{n} w[n-m]x[m]e^{-j\omega_k m}$$

$$= \sum_{m=n-(N_w-1)}^{n} w[n-m]x[m]e^{-j2\pi mk/N}$$

Through careful observation, it becomes evident that the variables of frequency and time are integral components of a function with two variables: Time and Frequency. This function, denoted as X [n, k], is characterized by its discrete nature. By assuming that the signal is band limited, we can effectively analyze its frequency distribution and selectively isolate the frequencies that are not needed.

The STFT formula generated by considering the Band pass filters is not a systematic representation. Therefore, in order to obtain an equivalent and mathematically accurate equation, a slight modification is made to the formula. The block diagram is seen in Figure 6. The input sequence, denoted as x[n], is subjected to the Short-Time Fourier Transform (STFT) by applying a window function and a tie function in exponential form. The resulting output is represented as x[n,k].

The STFT formula generated by considering the Band pass filters is not a systematic representation. Therefore, in order to obtain an equal and mathematically accurate equation, a little adjustment is made to the expression. The block diagram is seen in Figure 6. The input sequence, denoted as x[n], is subjected to a window function and an additional tie function, both expressed in exponential form. This process yields the output of the Short-Time Fourier Transform (STFT), represented as x[n,k].

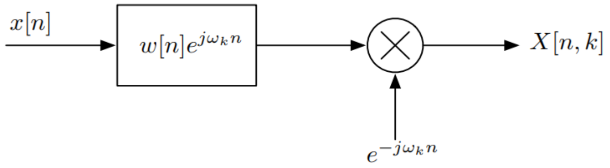

Figure 6. Block representation of the expression

The given phrase yields a bandpass implementation of a filter that is initially designed as a low pass filter window.

$$X[n,k]=\sum_{m=n-(N_w-1)}^{n} w[n-m]x[m]e^{-j\omega_k m}$$

This results in the equation to be,

$$=\left(\sum_{m=n-(N_w-1)}^{n}\left(w[n-m]e^{j\omega_k(n-m)}\right)x[m]\right)e^{-j\omega_k n}$$

IV. Results and Conclusion

The input signal under consideration is the monophonic sound produced by a violin playing a single note. In this particular instance, the note d5 is selected with the intention of observing its spectral characteristics. The time-domain representation of the violin signal is depicted in Figure 7. The violin string 1a5 is regarded as significant, and its associated dynamics are depicted in relation to time and amplitude. The frequency of the given signal is 440Hz. What is the internationally accepted standard for tuning musical instruments and serves as the reference frequency.

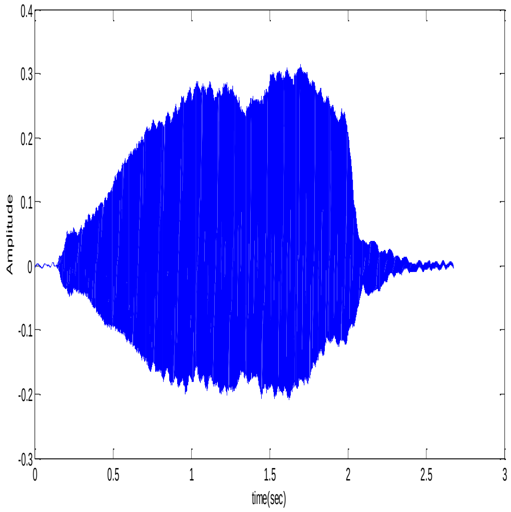

Figure 7. Sample Violin note played for d5

Figure 8 displays the audio recording of the chords played on an electric guitar. The initial peak corresponds to the auditory signal that occurs prior to the act of strumming the strings. The graph displays the strumming pattern of the guitar for several concurrently played notes, which corresponds to the chord patterns.

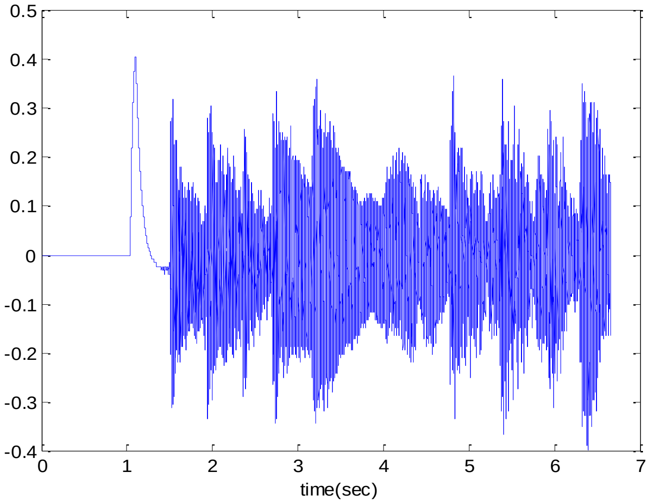

Figure 8. Audio clip played on electric guitar

The Figure 9 displays a mixed signal. The two audio recordings of the violin and guitar are observed, each exhibiting their distinct qualities. The guitar patterns and violin tune are readily observable. The audio of the violin and guitar is combined in a ratio of 60%.

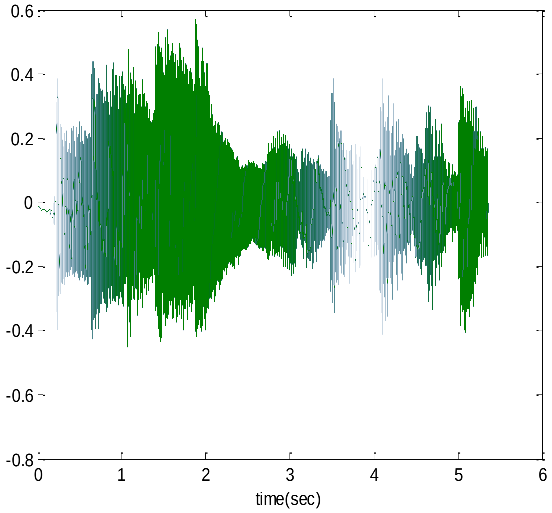

Figure 9. Mixed signals in time domain

The spectrogram of the Violin d5 sample is depicted in Figure 10. The auditory output produced by a violin.

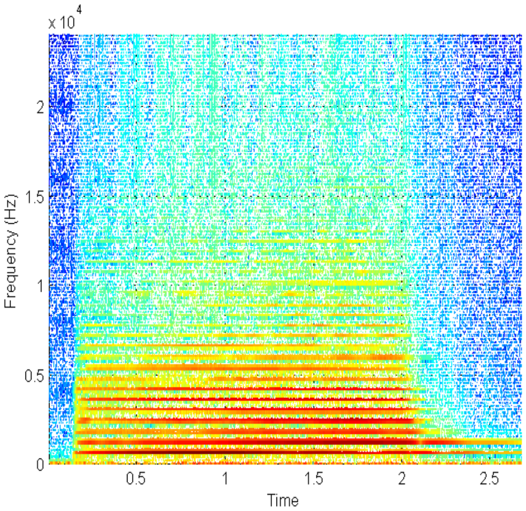

Figure 10. Spectrogram of violin signal played for note d5

The spectrogram of the audio sample is depicted in Figure 11. The spectrogram depicted in Figure 11 can be interpreted as follows: initially, there is an absence of signal, resulting in a mostly blue spectrum extending up to a duration of one second. The spectrum encompasses the harmonics produced by the guitar chords and is played at the higher regions of the guitar's fretboard. Consequently, the signal manifests as a collection of frequency traces arranged vertically, representing various ratios of frequencies that combine to produce chords.

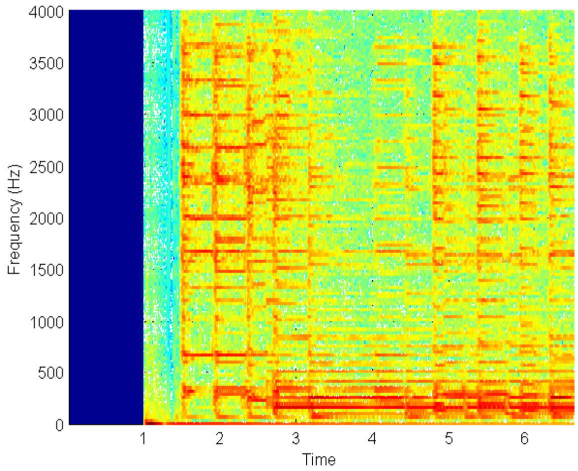

Figure 11. Spectrogram of the Audio sample.

The results obtained through mixing of the signals are shown in Figure 12.

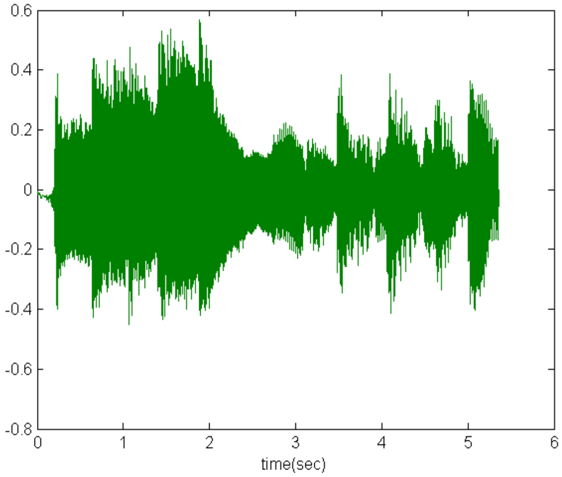

Figure 12. Results of mixing the two signals in time domain.

The spectrogram depicted in Figure 13 displays the outcomes derived from the mixing of the signals.Figure 13 displays the spectrogram outcomes derived from the analysis of the mixed signal. By analyzing the waveform and generating a spectrogram within a certain time frame, it becomes evident that distinct notes and the primary strings of the violin (E, A, D, G) can be discerned as they coincide in a musical excerpt. Through the examination of the acquired waveform, it is possible to discern the distinct tonal characteristics of the instrument. By making minute adjustments, estimation can be derived from the spectrogram.

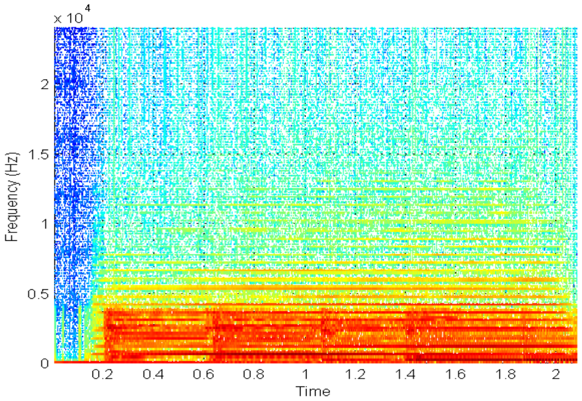

Figure 13. The spectrogram of the two mixed signals.

In the context of orchestration, when multiple instruments possess similar tonal qualities and exhibit overlapping frequency and power bands of comparable magnitudes, alternative methods of observation are necessary to identify frequency-power formants and, in certain instances, patterns within the signals. Due to the non-periodic and unpredictable nature of these signals, their identification becomes challenging. Nevertheless, the Short-Time Fourier Transform (STFT) technique serves as the fundamental framework for conducting study related to the majority of source recognition and identification endeavors. However, this research examines the spectrogram and analyzes the plotted values to identify the differentiating characteristics of the instruments. These qualities are then isolated to create a distinct track. Future research endeavors involve the incorporation of more instruments and the exploration of mixture models such as Gaussian distributions. This exploration may encompass the examination of Gabor transforms and chirping, with the aim of understanding and accounting for the detection and perception skills of the human auditory system.

The spectrogram of the instrument signal was analyzed to identify and extract the necessary frequency components. The process involved studying the identifiable characteristics of the signal and performing instrument identification and separation.

TABLE 1. Identification of Instruments

Sample No	**Comparison with different instruments**		
	Instrument Considered	***Second instrument considered***	***Recovery percentage***
1	Violin	Guitar(polyphonic)	60%
2	Violin	Flute	95%
3	Violin	Viola	30%
4	Mouth organ	Violin	60%

Table 1 displays the experimental findings pertaining to various musical instruments. This was done while considering the computational complexities imposed by the given constraints. By taking into account the limited time frame available for identifying the quantity of frequencies and their fluctuations in relation to the formants, it is possible to attain a certain degree of separation.

V. REFERENCES

[1] M. Mannone, “Introduction To Gestural Similarity In Music. An Application Of Category Theory To The Orchestra,” Journal of Mathematics and Music, vol. 12, no. 2, pp. 63–87, May 2018, doi: 10.1080/17459737.2018.1450902.

[2] S. Giraldo et al., “Automatic Assessment Of Tone Quality In Violin Music Performance,” Front Psychol, vol. 10, no. MAR, 2019,doi: 10.3389/fpsyg.2019.00334.

[3] V. Ronchetti, “Violin Timbre Analysis with Mel-Frequency Cepstral Coefficients.” University of Rochester.

[4] A. Livshin, “Automatic Musical Instrument Recognition And Related Topics.” .fr/tel- 00810688.

[5] J. Charles, D. Fitzgerald, and E. Coyle, “Violin Sound Quality Detection.” At arrow.tudublin.ie/ argcon

[6] G. Agostini, M. Longari, and E. Pollastri, “Musical Instrument Timbres Classification With Spectral Features,” 2003.

[7] Q. Wang and C. Bao, “Individual Violin Recognition Method Combining Tonal And Nontonal Features,” Electronics (Switzerland), vol. 9, no. 6, pp. 1–18, Jun. 2020, doi:10.3390 /electronics9060950..

[8] Y. Wu, Q. Wang, and R. Liu, “Music Instrument Classification Using Nontonal “ MFCC,” 2017.

[9] D. Armstrong D’Souza and V. Veena Devi Shastrimath, “Modelling Of Audio Effects For Vocal And Music Synthesis In Real Time,” in 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Mar. 2019,pp.1–4.doi:10.1109/ICCMC.2019.8819852.

[10] D. Dony Armstrong, S. v. Veena Devi, and V. N. Ganesh, “Pedal Effects Modeling For Stringed Instruments By Employing Schemes Of DSP In Real Time For Vocals And Music,” 2020, pp. 97–113. doi: 10.1007/978-981-15-0626-0_9.

[11] Sumit Kumar Banchhor, Arif Khan “Musical Instrument Recognition Using Spectrogram And Autocorrelation” International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2 Issue-1, March 2012”.

[12]Y.-N. Hung and Y.-H. Yang, “Frame-Level Instrument Recognition By Timbre And Pitch,” Jun.2018,[Online].Available:http://arxiv.org/abs/1806.09587

[13] Z. Otčenášek and J. Štěpánek, “Directional Timbre Spaces Of Violin Sounds.” Available at http://marcdata.hamu.cz/vyzkum/dokumenty/Lit77.pdf

[14] J. A. Charles, D. Fitzgerald, and E. Coyle, “Violin Sound Quality Detection,” in IET Conference Publications, 2008, no.539 CP,pp. 282–287. doi: 10.1049/cp:20080676.

[15] Y. Qi, J. W. Paisley and L. Carin, "Music Analysis Using Hidden Markov Mixture Models," in IEEE Transactions on Signal Processing, vol. 55, no. 11, pp. 5209-5224, Nov. 2007, doi:10.1109/TSP .2007.898782.

[16] Douglas A. Reynolds, Thomas F. Quatieri, And Robert B. Dunn." Speaker Verification Using Adapted Gaussian Mixture Models” Digital Signal Processing 10, 19–41 (2000) doi:10.1006/dspr.1999. 0361,