REAL-TIME PITCH TRAINING SYSTEM FOR VIOLIN LEARNERS

Jian-Heng Wang  Siang-An Wang  Wen-Chieh Chen  Ken-Ning Chang  Herng-Yow Chen
Department of Computer Science and Information Engineering, National Chi Nan University, Taiwan
E-mail: {s96321521,s99321516,s99321504,klim,hychen}@ncnu.edu.tw

ABSTRACT—This paper specifically targets violin learners who are working on their pitch accuracy. We employ a pitch tracking algorithm to extract the pitch played. Through volume thresholding and region detection, only parts of frames are processed. So our system can provide real-time feedback to show violin learners whether they played the right pitch. The system also provides the major scales and arpeggio scores as teaching materials, and violin learners can choose different tempos to practice, depending on their level. The user-friendly system interface allows violin learners to easily perceive the pitch differential between the pitch of the target note and the pitch played, allowing users to precisely adjust their playing. The statistical feedback records progress and analyzes error patterns, enabling violin teachers to evaluate student progress precisely, and correct common error patterns effectively.

Keyword: violin, pitch detection, feedback, HPS

I. INTRODUCTION

Shinichi Suzuki was the inventor of the educational philosophy known as the Suzuki method of music education [1]. Considered an influential and controversial pedagogue, he often spoke of the ability of all children to learn musical instruments well, in the right environment.

Suzuki emphasized what he referred to as a “mother-tongue” method [2]. He believed that by listening to lots of musical input, learners could train their pitch perception. Pitches are compared as “higher” and “lower” in the sense that allows the construction of melodies. In other words, it is a subjective sensation in which a listener can assign tones to relative positions on a musical scale based primarily on the frequency of vibration [3]. However, his methods are not easily accepted by all beginning music students. Teacher Jennifer Wade notes that many children in western culture, particularly in the Northern Virginia region, come to violin at a slightly older age (five years old and up), often coming from families where both parents work [4]. Takako Nishizaki notes that the Suzuki method’s focuses on listening instead of understanding made children not capable of understanding music score [5]. Therefore, we designed a system which through providing the pitch training and real-time feedback on the score augments learners’ practice sessions and promote their learning performance.

This paper provides an overview of a real-time pitch training system for violin learners to play the right pitch through a specified scale and arpeggio with a specified tempo. A pitch detection algorithm is employed to extract a pitch as the player creates it. Visual feedback is presented in real time, showing the degree of difference between the correct pitch and the one performed, enhancing overall learning progress [6]. The system also provides assistance to music teachers, creating a means by which they can analyze statistical feedback and evaluate error patterns to better monitor the learning processes their students.

There are only a few published works on real-time pitch training. One system, the Piano Tutor [7], provides computer-based instruction for novice pianists. Using score-following technology, the Piano Tutor claimed to help students identify errors made during a performance. Another program, Family Ensemble [8], proposed a collaborative musical edutainment system to allow a parent and his or her child to enjoy ensembles together. Using a score-tracking algorithm, Family Ensemble could cope with the particular errors commonly made by beginners. Interactive Digital Violin Tutor [9, 10, 11] combined audio, video, 2D and 3D animation to provide feedback. These animations were driven directly by a score or by a note table transcribed from a sequence played by a teacher. There are also some research regarding violin transcription in instrument tutoring [12], visual analysis of fingering [13], audio and video fusion [14], and violin music detection [15]. Those research concludes that computer-assisted feedback systems positively affect music learning.

The above mentioned methods of violin audio transcription require a very quiet environment to prevent environmental noise. Our system does not have this restriction and instead provides a new method to overcome this obstacle. We also designed a new interface where learners can adjust the pitch played using real-time feedback. It is our belief that by using our system, learners can improve their progress in violin learning.

II. SYSTEM ARCHITECTURE

The system architecture is as shown in Figure 1. It provides two modes to train pitch. Under mode 1, learners have to follow the specified tempo. Mode 2 is open and free. Only when they practiced correctly can they then switch to the next note. The audio signal from the violin is captured via a microphone, which is then sent to a pitch detection algorithm to extract different pitches. The pitches will be compared with the score provided by the performance evaluator. To give visual feedback to the learner, an evaluation result is displayed directly on the score in real time. The system also counts any errors...
played, and gives statistical feedback to help the teacher to improve the learner’s skills.

“Bee” as an example. The energy gap of two notes, which is between the release state of the note and the attack state of its next note, can be located.

An obvious characteristic of a bowed string instrument like a violin is overtone. Figure 3 depicts waveforms of piano and violin. No obvious energy decrease (decay state) or sustain state is shown in the waveform of a violin. Additionally, its envelope varies according to the player’s musical skills. The figure also shows that the energy gap between two successive notes is clear enough to separate them.

Figure 2. Attack, Decay, Sustain, and Release in the segmentation of violin audio

Figure 3. Piano and violin waveforms

IV. REAL – TIME PITCH DETECTION

There are many different ways to detect audio signals [16, 17]. However, for violin, note segmentation is particularly difficult [16]. Recently, several papers focused on violin audio signals have been published [18, 19, 20]. In order to be both robust and fast, our system sets an appropriate threshold to filter out environmental noise and determine the precise note duration. The following sections will describe some acoustic features of the audio signal analysis and the methods used to extract and derive these features. There are three goals for our pitch detection:

- **Accuracy**: To be useful, the performance of our pitch detection should be highly accurate; otherwise, our system cannot provide effective feedback.

- **Robustness**: Our system is designed for violin learners to practice the violin at home. We cannot assume that the environment is quiet and the recording equipment is professional. Therefore, our system should be able to handle the audio signal captured by a non-professional microphone, even in a noisy environment.

- **Speed**: Since our system provides visual feedback for the violin learner in real time, the processing delay should be as short as possible.

To achieve these goals, the envelope of the violin audio signal was analyzed and it was found highly varied. Determining the special envelope pattern for the violin can improve the accuracy of our algorithm. Our system is designed to be used in a non-studio environment, so an appropriate volume threshold is used to overcome any environmental noise issues. The volume slope and time duration can be used to help detect the valid region in the audio signal more effectively in real time. Finally, a

III. Violin Sound Background

The envelope of the audio signal waveform has a four-state structure, i.e., attack, decay, sustain, and release state (ADSR).

The ADSR parameter is often used to control the start and end of a note in synthesizers. Figure 2 shows the waveform of two notes extracted from the nursery rhyme
harmonic product spectrum is employed to extract the pitch in a valid region.

A. Environment Noise Eliminating
To reduce environmental noise, we need to consider the distance between the environmental noise (global minimum) and the main violin sound. However, for segmenting a valid region, we focus on the difference between the local maximum and local minimum.

Our research uses two methods to calculate the audio signal energy (volume). Method 1 uses the sum of absolute samples within each frame to find the violin sound differential between a local maximum and a local minimum.

\[
volume = \sum_{i=1}^{n} |S_i| \quad (1)
\]

Method 2 uses 10 times the 10-based logarithm of the sum of square of the sample to find the differential between the global maximum and the global minimum.

\[
volume = 10 \times \log(\sum_{i=1}^{n} S_i^2) \quad (2)
\]

1) Preliminary Thresholding
In general, pure violin sound and the environmental noise are produced when the violin is being played. To minimize environmental noise, we employ the volume threshold as the demarcation line to differentiate them (Figure 5).

![Figure 5. Filtering environmental noise via the volume threshold](image)

However, several factors create variation in environmental noise, such as recording equipment and environment. We need to ensure that the volume threshold is set properly. The volume threshold may require adjustment according to the environmental noise volume. Our system will calculate the average volume of environmental noise during the first five seconds of operation, and the volume threshold will be set to twice the recorded average.

2) Further Thresholding
To realize real-time processing, we adjust the volume threshold to reduce the execution cost. We attempt to process the strongest energy region, i.e., the top of the envelope of the note. In Figure 6, the shape of the volume is quite similar to the amplitude of the audio signal. Therefore, the volume threshold to extract the region can be set as:

\[
threshold = p \times 50\% \quad (3)
\]

where \( p \) is the first peak of volume. The volume between peaks and valleys will decrease about 60%. The volume threshold should be ensured that the information in the audio signal will not be cut too much to extract the pitch. Therefore, we set the volume threshold to be half that of the first peak.

![Figure 6. Reducing the execution cost by volume threshold adjustment](image)

In Figure 6, we promote the volume threshold value from the original to a new position which is located in the half of the first peak. Using this method, we can filter out environmental noise and reduce the execution cost by the newly-positioned volume threshold.

B. Valid Region Detection
To provide feedback for the learner, our algorithm has to be able to separate adjacent notes with the same pitch value. Depending on the skills of the musician, silent regions and obvious energy decreases may not exist in violin performances. Even if there are regions of obvious energy decrease, the levels are quite different. That explains why our first approach, zero-crossing rate, failed and why we proposed valid region detection. The method, employing volume slope and time duration, can not only separate different notes properly but also reduce execution time by extracting the highest volume region.

1) Volume Slope
By observing the envelope of the audio signal, we find there is a peak in each note. A slope that changes from positive to negative can be used to mark a note. Thus, individual notes can be separated. In Figure 7, the peak regions between where the volume increases and decreases are clear. The region where the volume decreases must be
sustained for a short period for our system to evaluate if the volume is truly decreasing, or if it is decreasing as a result signal disturbance. To be effective, it is necessary for our system to scan about 70% of a note region to extract a pitch.

![Figure 7. The volume slope](image)

2) **Time Duration**

Our system employs a tempo parameter (Mode 1) to train the learner to play at preset time positions. In Figure 8, the front side line of a note is the tempo position \( t, 2t, 3t \), and we can assume that the beginning of the note must be close to the tempo position.

![Figure 8. Valid region detection](image)

To constrain the valid region extracted by our algorithm, a duration \( S \) is defined. In Figure 8, there are four valid regions located in the tempo position \( t, 2t, 3t, 4t \) through duration \( S \). This method, which scans about 50% of a note region to extract a pitch, is more effective than the previous method.

**C. Harmonic Product Spectrum**

In general, there are two important issues relating to pitch detection: overtone and playing skills. The former issue exists because it is not easy to detect the overtone of a string instrument. The latter issue is special to playing skills that can generate an audio signal which consists of complex frequencies. Therefore, it is necessary for us to solve these two problems.

Our first choice was autocorrelation function. But the experimental results indicated that the signal cannot avoid interference with a strong overtone, creating an error when strong overtones occur simultaneously with the overtones from a previous frame. Therefore, we employ the harmonic product spectrum which detects the pitch in the frequency domain. The harmonic product spectrum can discriminate the fundamental frequency using overtones, and it then multiplies downsamplings to emphasize the fundamental frequency. The harmonic product spectrum works effectively, as it ignores the overtone influence over fundamental frequency (correct pitch).

**D. Violin Sound Pitch Detection**

We first set the threshold then apply valid region detection. Only parts of frames are extracted. For each of them, we apply the harmonic product spectrum to find the correct pitch. Following these steps, we can find pitches correctly and efficiently. As shown in Figure 9, a combination of the valid region with the volume threshold is more effective than the original volume threshold. The valid region contains the most important part of the audio signal because it covers the area of strongest energy (the attack state) and fewer overtones (the sustain state and release state). The volume threshold must be smaller than the volume of the tempo positions, or the volume threshold will sample only a part of the valid region.

![Figure 9. Combination of valid region detection and the volume threshold](image)

**V. SYSTEM IMPLEMENTATION**

**A. Tuner**

Before learners begin practicing violin, they need to correctly tune the strings. Mastering the precise pitch is a fundamental skill for music students. Therefore, we have implemented a tuner (see Figure 10(a)), to help the learner tune their strings correctly. The interface of the upper bar chart and the lower bar chart will show whether the pitch is high or low as the violin is being tuned, greatly simplifying the tuning process.

**B. Score and Tempo Selection**

After learners are able to play the accurate pitch in the scale scores, they may want to challenge themselves with some more difficult training methods. Tempo is an important element of the music composition. Tempo is an important element of the music composition. It can express deep emotion and affect mood. For violin beginners, how to play at a fixed tempo and how to play faster are important. Thus, we have implemented a timer and use a tempo cursor to help users follow the tempo. Figure 10(b) shows the score and tempo selection region.
Training Mode

There are two training modes designed for different learning targets. Mode 1 is designed to allow the learner to practice playing with different tempos and shows real-time feedback (see Figure 11). The system will collect and analyze errors, and then provide statistical feedback. By using statistical feedback, learners can understand what kind of mistakes they often make.

Under Mode 2, the learners have to continue playing until they played the previous notes correctly. When the learner plays a note incorrectly, the system will provide immediate feedback to signal the learner how adjust the pitch. Finally, errors are counted and shown in the statistical feedback.

C. Feedback

It is important for the violin beginners to get visual feedback intuitively. For violin beginners, especially children along their parents, providing visual feedback is effective. The visualization is easily understandable by people who have little music knowledge [9]. Thus, we use upward and downward triangles to tell learners if the pitch is high or low. Figure 10(c) shows real-time visual feedback directly on the score. This method can help learners self-correct and achieve accurate pitch. To accomplish this we first need to calculate the pitch differential (pd) between the target pitch and acoustic pitch:

\[ pd = |pitch_{target} - pitch_{play}| \]  

The level of the pitch differential is shown in Table 1. The differential is represented by triangles. To tolerate the interference within the audio signal, a pitch differential which is lower than 0.3 of a semitone will be considered correct. If the pitch differential is higher than a semitone (cf. Level 3 in Table 1), it is considered a serious mistake and three upward or downward triangles will be displayed.

Our system makes use of suggestive symbols to help learners achieve the target pitch. These two symbols provide feedback similar to what a teacher would provide for learners to revise their mistakes. This real-time visual feedback is intended to maintain the learner’s motivation when practicing violin alone.

VI. CONCLUSION

We have proposed and implemented a real-time pitch training system that can help violin learners to practice violin independently and effectively. Our system proposes an accurate, robust and fast pitch detection algorithm to analyze the learner’s play in real time.
To motivate the learner, our system provides a visual feedback to help correct mistakes. This may encourage the learner practice more effectively. We also provide statistical feedback for teachers, allowing them to monitor learner progress efficiently.

VII. REFERENCES


