Journal of Computers Vol. 32 No. 2, 2021, pp. 222-232
doi:10.3966/199115992021043202019

Design of a Pitch Detection and Intonation Correction System Based on LabVIEW

Bing-Xia Shen*, Kai-Yi Wang, Jin-He Zhou

School of Information and Communication Engineering, Beijing Information Science & Technology University, Beijing, 100101, China
shenbingxia@bistu.edu.cn

Received 2 January 2021; Revised 2 February 2021; Accepted 31 March 2021

Abstract. Amidst the global popularity of music programs, the music attainment of audiences has kept enhancing, thus setting higher bars on intonation for music performers. For this reason, correct intonation is essential for both instrumentalists and singers. Despite various kinds of tuners and tuning software on the market today, software and system capable of both tone tuning and correction are still in short supply. Therefore, a system that corrects intonation of musical instruments and people's voice is designed and implemented herein. The system delivers pitch detection, intonation judgment and correction for simple notes. By collecting audio via NI myDAQ and analyzing the frequency spectrum of the audio through LabVIEW on PC, it realizes pitch detection and judges intonation by comparing the audio with the standard tone, and utilizes the pitch shifting algorithm to correct intonation. Test results show that the system can accurately detect pitch and judge intonation, with a desirable correction effect. The system described herein provides references for tuning of simple musical instruments and guidance for training of singers.

Keywords: Audio collection, pitch detection, intonation correction, myDAQ, LabVIEW

1 Introduction

In recent years, a wide range of music programs have been introduced amidst the musical culture boom, thus enhancing the music appreciation and attainment of audiences, who have increasingly higher requirements on music performers. This makes intonation an important criterion for the audience to make a judgment in music appreciation. Correct intonation is essential for both instrumentalists and singers. During a performance, singers' voice may deviate from the standard pitch due to breath, emotion or external interference. Similarly, musical instruments may be out of tune because of materials, temperature and other factors. To ensure the broadcast effect, it is very common for music programs to post-process live recordings, wherein intonation correction is a key step to correct the deviation and error made by performers. There are many kinds of musical instrument tuners on the market [1-2], including professional tuning software that is commercially available, all of which help more or less guarantee the intonation of music performers.

A well-developed tuning system that delivers pitch detection and intonation judgment and correction is designed and described herein. The system, which is equipped with an external sound pickup, collects sound of people or musical instruments through myDAQ and transfers the sound to a computer, and then the sound programmed by LabVIEW is subject to spectral analysis to obtain the pitch of the sounding body. This is how the system performs pitch detection. Besides, FFT (Fast Fourier Transformation) is carried out to obtain the sound frequency, which is converted into the corresponding cent based on the frequency-cent relationship in twelve-tone equal temperament. By comparing the deviation of corresponding cent from the cent of standard musical instrument tuning, the system can judge the intonation of the sounding body. Thanks to pitch detection and intonation judgment, people can monitor

* Corresponding Author

their performance level in real time from the visualized front panel when they are practicing or performing, thus correcting the intonation of themselves or musical instruments. At last, the system uses SOLA-FS (Synchronized Overlap-Add Fixed Synthesis) to correct the wrong simple notes, thus achieving intonation correction. Compared with the existing uni-functional tuners and pitch shifting software on the market, the present system can detect pitch, judge intonation and correct the simple note with intonation error. It is truly a full-featured intonation correction system.

2 Theoretical Basis

2.1 Basis of Music Theory

In the scenario of music, pitch reflects the fundamental frequency of a sound psychologically perceived by humans. Pitch is determined by the vibration frequency of a sounding body, wherein it raises as the vibration frequency of the sounding body increases, and vice versa. In the study of the field, music is graded into seven scale steps that are circularly represented as CDEFGAB according to its pitch level. The pitch range from any one of the scale steps to the seven other adjacent scale steps is what we call “an octave”. In practice (taking piano as an example), the octave is equally divided into 12 levels, each of which represents a semitone. The vibration frequency of two adjacent semitones is in proportion with a common ratio of. If the interval value of an octave is set to 1200 cents, in twelve-tone equal temperament, a semitone and a whole tone respectively represent 100 cents and 200 cents. According to the principles of twelve-tone equal temperament and standard tone, through increasing all the frequency components by times or decreasing them by times, we can raise or lower the pitch by one semitone.

2.2 Pitch Detection and Intonation Judgment

Sound is a wave, of which frequency and amplitude are important attributes. Sound can be broken down into the superposition of sine waves of different amplitudes and frequencies by the Fourier transform. The sound source is recorded by an external microphone, with its sound waveform being transmitted to the computer through the audio acquisition system, and then is subject to Fourier transform and corresponding data analysis through software programming on computer. After FFT, the amplitude peak is extracted as a fundamental frequency, which is converted into cents and corresponding temperaments and compared with preset standard values. In this way, the pitch can be detected and the intonation can be judged based on the cent differences between the recorded sound and the preset values.

2.3 Pitch Shifting Algorithm

SOLA-FS is a time-domain pitch shifting algorithm [3-4]. It adds or subtracts sampling points by increasing or decreasing data points of the sound signal at an equal interval, to realize pitch shifting and intonation correction. Both interpolation and decimation are applied. When the interpolation multiple is greater than the decimation multiple, the original data will be elongated to achieve falling tone; When the interpolation multiple is smaller than the decimation multiple, the original data will be shortened to achieve rising tone. In addition, time scale modification is also required to ensure a constant speed of sound.

3 Design Scheme

3.1 Pitch Detection Scheme

The sound is recorded by a microphone and processed with NI myDAQ for data acquisition. The acquired data is programmed by LabVIEW for FFT after being transferred to a computer. The original sound signal is converted into a frequency domain, and then its fundamental frequency is extracted and converted into the corresponding cents. A sub VI is written, capable of comparing the entered cent value with the cent value of the preset standard tone, wherein the pitch name corresponding to the cent can be identified within a certain range, thus realizing pitch detection. See Fig. 1. for the pitch detection process.

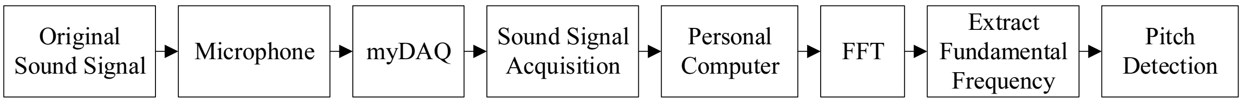

Fig. 1. Diagram of pitch detection process

3.2 Intonation Judgment Scheme

The cent value extracted from the pitch detection subprogram is used as the input of the intonation judgment subroutine. The value is rounded up to the nearest whole number to identify the standard tone that the entered sound deviates from. Next, the difference value between the cent of entered sound and that of the corresponding standard tone is calculated, to judge whether the sound pitch is lower or higher than the standard pitch, and the difference therebetween is the deviation of the sound from the standard tone. See Fig. 2. for the intonation judgment logic.

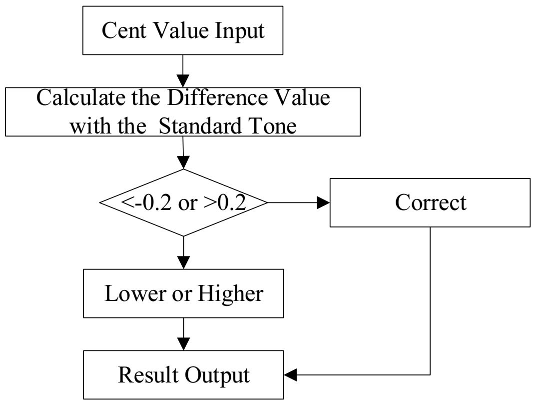

Fig. 2. Logic diagram of intonation judgment

3.3 Tone Correction Design Scheme

This paper adopts SOLA-FS to make tone correction. This pitch shifting algorithm includes two key steps, namely, sampling rate conversion and time scale modification [4-5]. The specific process is shown in Fig. 3. Interpolation and decimation are applied synchronously to make the multiple of pitch shifting more flexible. After pitch shifting, the sound signal is subject to time scale modification, to ensure a constant speed of the original sound signal. During data decimation, the frequency spectrum is multiplied based on the original value, thus may result in spectrum aliasing. In theory, a proper low-pass filter shall be added before the sequence decimation, but thanks to the linear interpolation of the resampling function in LabVIEW, the present system does not need a low-pass filter to implement.

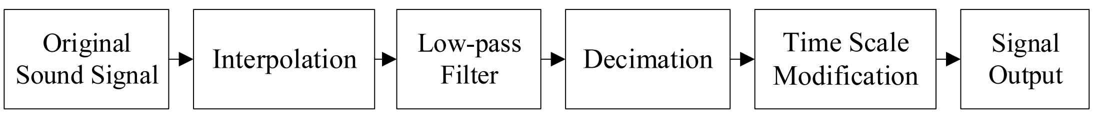

Fig. 3. Diagram of SOLA-FS pitch shifting process

4 System Implementation

4.1 Implementation of Pitch Detection and Intonation Judgment

4.1.1 Pitch Frequency Extraction

After connecting myDAQ to a computer, a DAQ Assistant Express VI is enabled on the program panel, with sampling rate set to 16000Hz and the number of sample to be read set to 200, for continuous sampling. All the pitch detection, pitch frequency extraction and intonation judgment programs are packed into the While circulation, wherein the main program is connected with a flat sequence structure.

Before sound acquisition, the sampling rate is set to 16000Hz, the quantity of queue data is set to 32000, and the number of one sampling is set to 200. After connecting the DAQ Assistant Express VI and setting the parameters, the acquired sound data is exported, and the sound signal can be converted into array type via dynamic data conversion Express VI, thus obtaining a group of one-dimensional array. By inverting the one-dimensional array, performing FFT, and rounding for maximum index, the pitch frequency of the input sound can be obtained. See Fig. 4. for pitch frequency implementation process

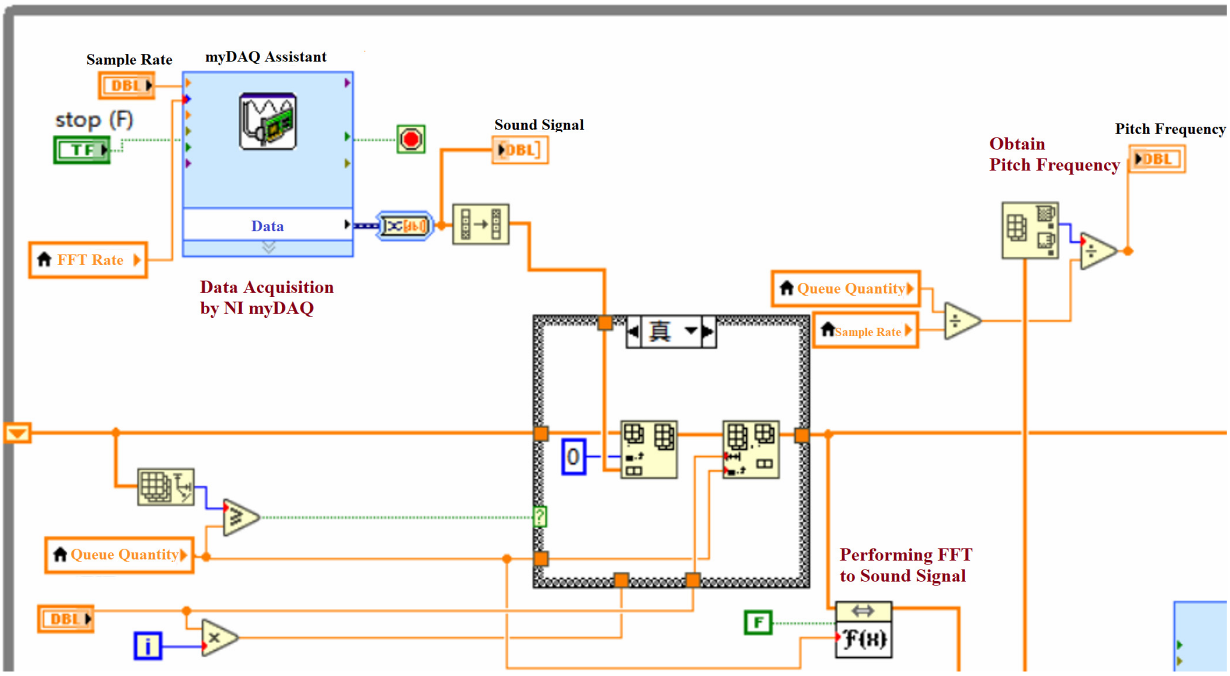

Fig. 4. Block diagram of pitch frequency extraction process

4.1.2 Conversion of Pitch Frequency into Temperament

When performing the pitch detection, the system needs to compare the collected sound data with the standard tone, thus detecting the pitch of the collected sound. Due to the non-linear frequency differences between standard tones, detecting by frequency is theoretically feasible, but not an accurate method. To minimize the error, the frequency needs to be converted into a temperament that is corresponding to the cent.

According to the calculation relationship among temperament, cent and frequency, the temperament frequency value sequence is a geometric progression with standard item of = 440Hz and common ratio of. The frequency value sequence corresponding to the cent of each pitch name section is a geometric progression with the first item of the pitch name frequency and common ratio of [6]. According to the general term formula of the geometric progression, the temperament and the frequency are represented as x and y respectively:

$$y = 440 \times 2^{\left(\frac{x-49}{12}\right)}. \tag{1}$$

Equation (1) can be simplified to:

$$y = 27.5 \times 2^{\left(\frac{x-1}{12}\right)}. \tag{2}$$

According to Equation (2), the frequency can be converted into temperament by LabVIEW, with the implementation process as shown in Fig. 5. The cent value of the temperament corresponding to the pitch frequency is obtained by multiplying the temperament by 100. That is, there is a 100-cent difference between the two adjacent temperaments, and after quantification, the quantity value of 1 cent is 0.01.

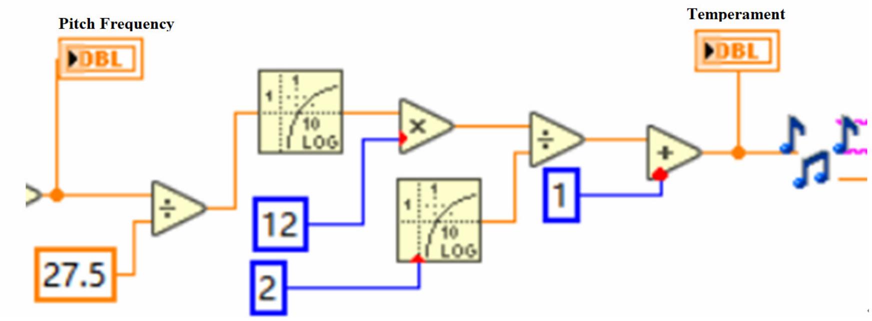

Fig. 5. Block diagram of pitch frequency-temperament conversion process

4.1.3 Pitch Corresponding to Temperament

A sub Vi is written, in which a temperament value entered is rounded to a whole number. It should be noted that LabVIEW has no function to implement rounding, so this is realized by adding 0.5 to the temperament value entered and then rounding down. After rounding, each temperament obtained corresponds to a standard tone. For example, the frequency of standard tone A4 is 440Hz. When converted into a temperament through Equation (2), it becomes 49, corresponding to the pitch name of A4.

4.1.4 Implementation of Intonation Judgment

Based on the tuning requirements of standard musical instruments, the deviation percentage between cent and pitch is approximately linear, and the average ratio coefficient is around 17, namely, 1% deviation corresponds to about 17 cents. From this, the intonation is judged as "approximately correct" when the deviation is less than 17 cents. For ease of calculation, the deviation is set to 20 cents. The conditional construct is used to judge the intonation, with the intonation and cent difference displaying on the front panel by string controls. To sum up, display of pitch and intonation judgment is implemented as shown in Fig. 6.

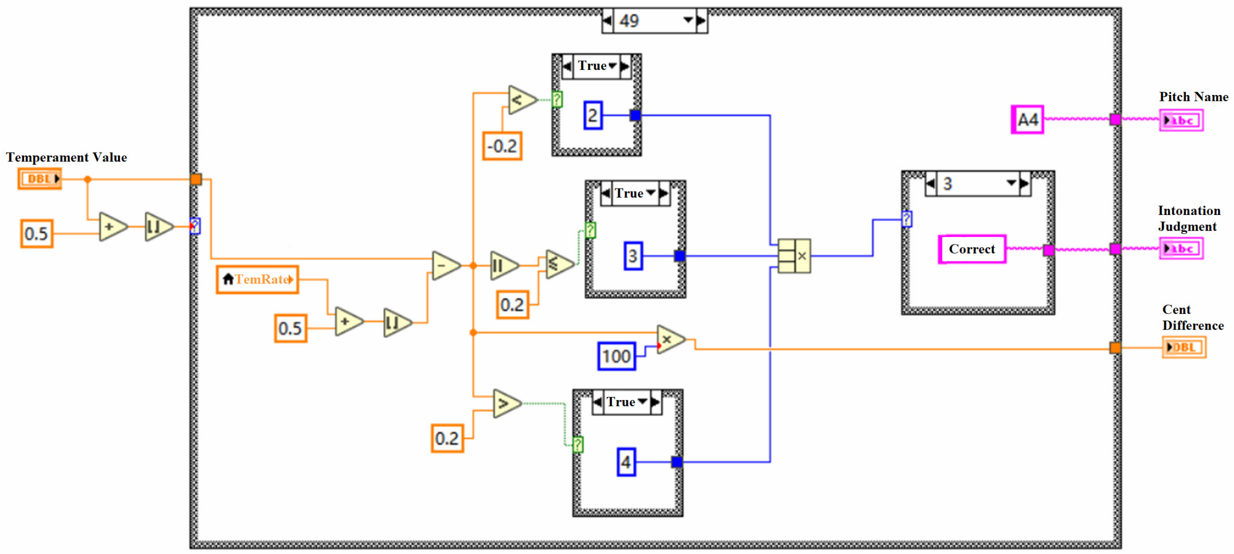

Fig. 6. Block diagram of pitch and intonation judgment display

4.2 Implementation of Intonation Correction

In this article, SOLA-FS is employed to make intonation correction. It includes two key steps, namely, sampling rate conversion and time scale modification. Time Scale Modification (TSM), namely time-scaling without pitch-scaling, can change the speed of sound playback while ensuring a constant tone and pitch [7]. Time scale modification includes framing of original data, neighborhood adding and frame selection and integration.

4.2.1 Framing

First, framing of original sound data is required in time scale modification. *W* represents data length, *L* represents the length of each frame, *S* represents analysis frame shift and *K* represents frame overlap, wherein the frame length is preferably not less than twice the frame period; frame overlap $K= L-S$; framing number is represented as *n*; data length is n times the length minus (n-1) times the frame overlap; thus obtaining an equation:

$$nL-(n-1)K=W\ . \tag{3}$$

From Equation (3), the framing number can be calculated as $n=(W-K)/(L-K)$, where $L-K$ is the analysis frame shift *S*. The frame length L=400 and the analysis frame shifting S=200 herein. The process is implemented as shown in Fig. 7.

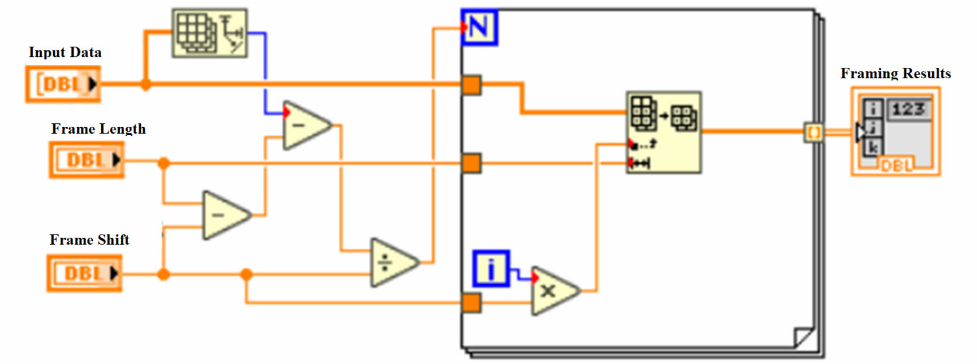

Fig. 7. Block diagram of framing process

4.2.2 Adding of Neighborhood

After framing, each frame needs to be added with neighborhood. Adding of neighborhood is required in SOLA-FS to solve the phase discontinuity of fundamental waves in each adjacent domain. The neighborhood value added before and after the first frame is 0. The neighborhood values added before and after the last frame are respectively the data at the end of the previous frame and 0. As for frames other than the first and the last frames, the neighborhood values added before and after the frame are respectively the data at the end of the previous frame and the data at the beginning of the next frame. All the neighborhoods are close to half a frame period. The process is implemented as shown in Fig. 8.

4.2.3 Calculation of Synthesis Frame Shift based on Cent Difference

It is necessary to calculate the synthesis frame shift before frame selection and integration. According to Equation (3), assuming that the original data length, frame length, and analysis frame shift are known, the number of framing can be calculated. By changing the data length to ensure the constant framing number and frame length, we can obtain the synthesis frame shift. According to the synthesis frame shift calculated, the ratio of the analysis frame shift to the synthesis frame shift can be calculated. The ratio is a pitch shifting coefficient, also known as pitch shifting factor. The cent difference corresponding to the pitch shifting coefficient can be calculated based on the twelve-tone equal temperament. In reverse, the pitch shifting coefficient can be calculated based on a certain cent difference, thus obtaining synthesis frame shift. The process is implemented as shown in Fig. 9.

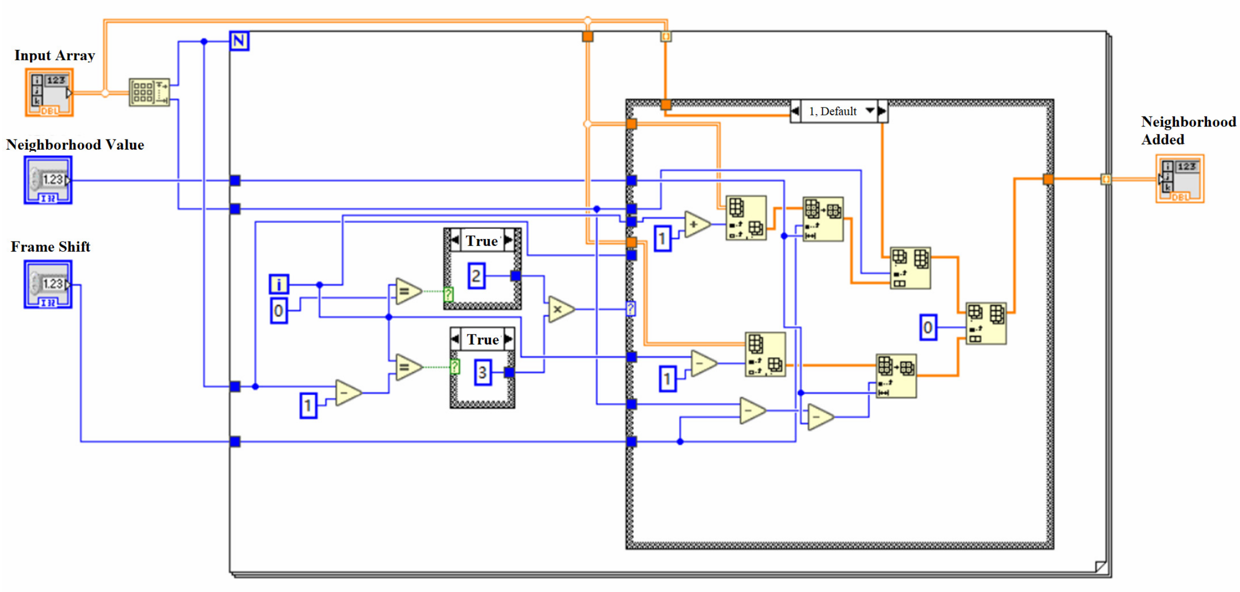

Fig. 8. Block diagram for adding of neighborhood

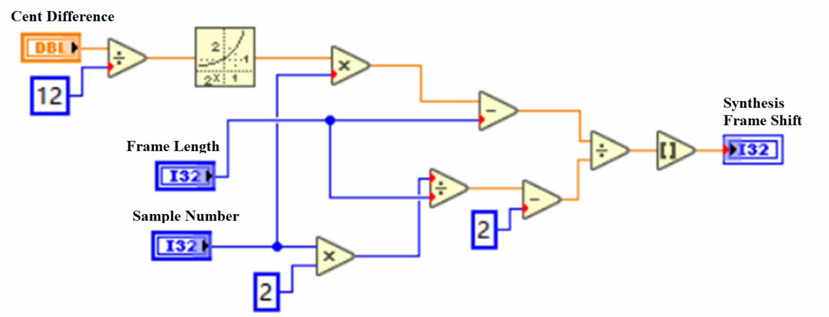

Fig. 9. Block diagram for calculation of synthesis frame shift based on cent difference

4.2.4 Process of Frame Selection

After the framing and adding of neighborhood for each frame, frame selection is required. The principle is to select frames with good continuity. As the cent difference changes, the synthesis frame shift changes accordingly, during which the previous frame shifts from back to front, and the next frame shifts from front to back. Next, the difference between the corresponding values of the previous frame and the next frame is calculated and summed. The frame under the synthesis frame shift corresponding to the minimal sum meets the requirement. See Fig. 10. for the process.

4.2.5 Process of Frame Integration

After frame selection, frame integration, the last step for time scale modification, is carried out, in which the position of each frame shit is also determined by the synthesis frame shift. Through frame integration, the previous two-dimensional array is integrated into a two-dimensional array before being output. See Fig. 11. for the process.

After frame integration, the entire time scale modification is finished. After that, the sampling rate needs to change, which can be realized by alignment and resampling Express VI in LabVIEW. The pitch shifting process is as shown in Fig. 12.

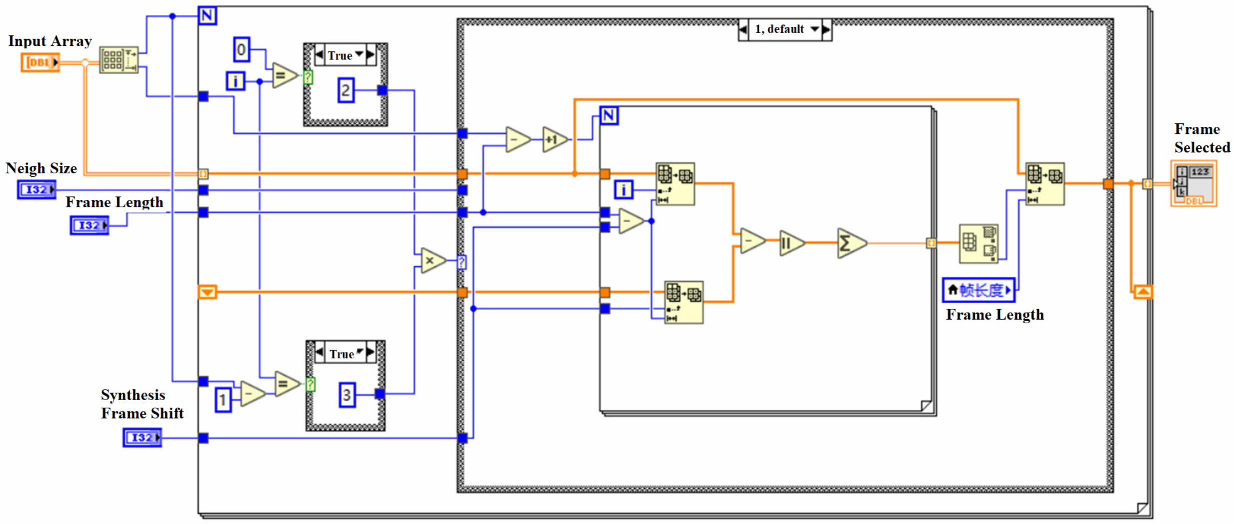

Fig. 10. Block diagram of frame selection

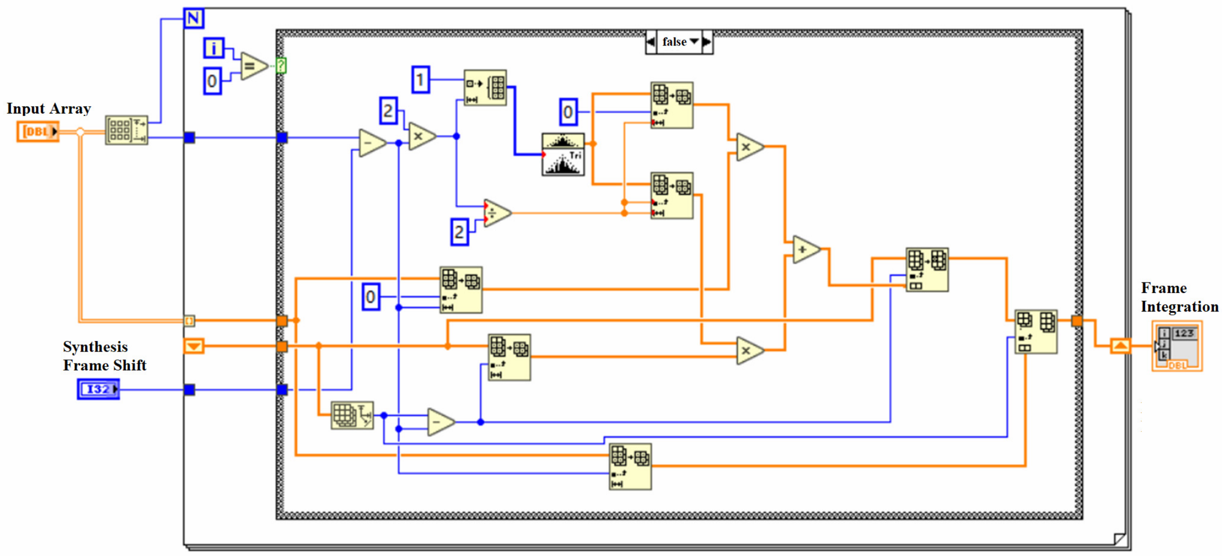

Fig. 11. Block diagram of frame integration process

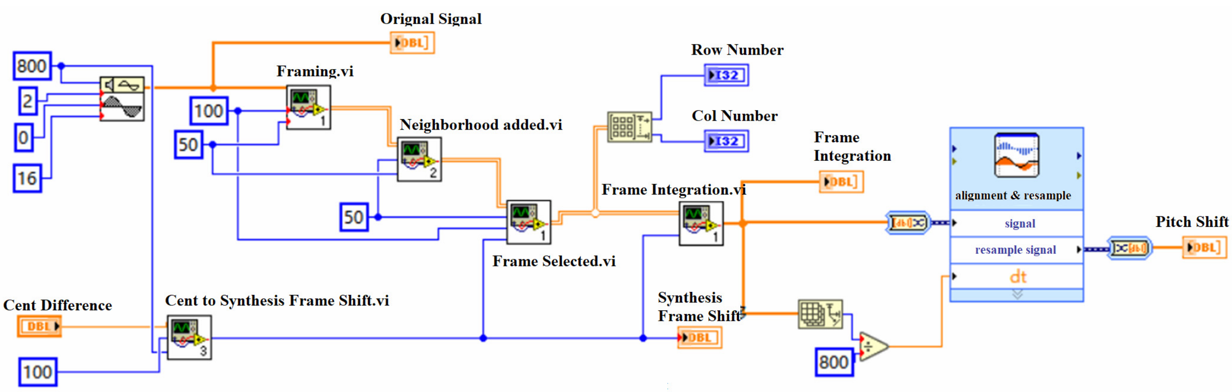

Fig. 12. Pitch shifting block diagram

5 Test Results

5.1 Results of Pitch Detection and Intonation Judgment

The vocal singing of A3 is tested to see whether it is pitched accurately or high. The judgment of pitch and intonation is observed, with test results as shown in Fig. 13 and Fig. 14. According to Fig. 13, the pitch frequency extracted from the first vocal singing of A3 is 218.5Hz, corresponding to a temperament of 36.8816, with a deviation from the standard A3 temperament of -11.8443. Since the deviation is less than 20, it is judged as correct-pitched. According to Fig. 14, the pitch frequency extracted from the second vocal singing of A3 is 224Hz, corresponding to a temperament of 37.3119, with a deviation from the standard A3 temperament of 31.1943. Since the deviation is greater than 20, it is judged as high-pitched.

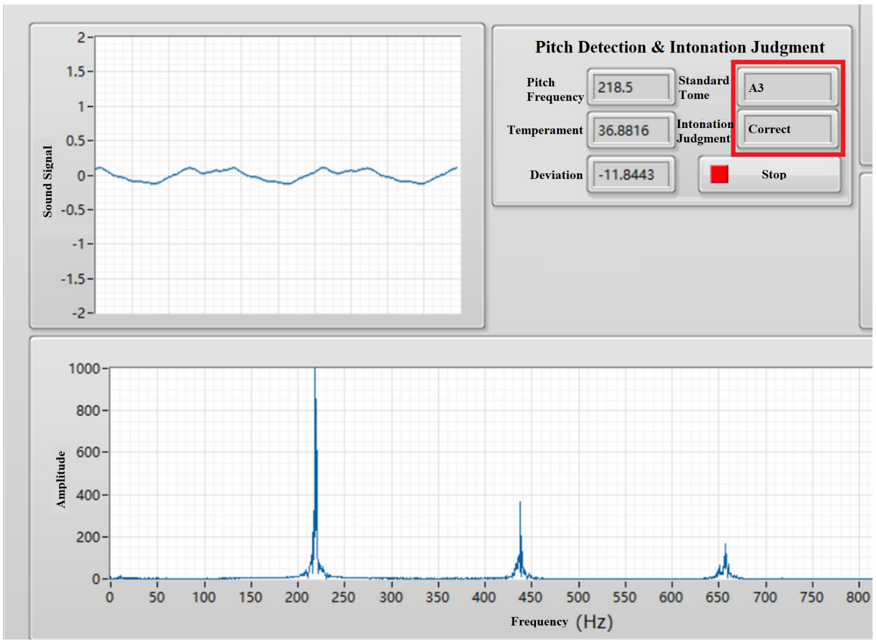

Fig. 13. Chart for correct-pitched test

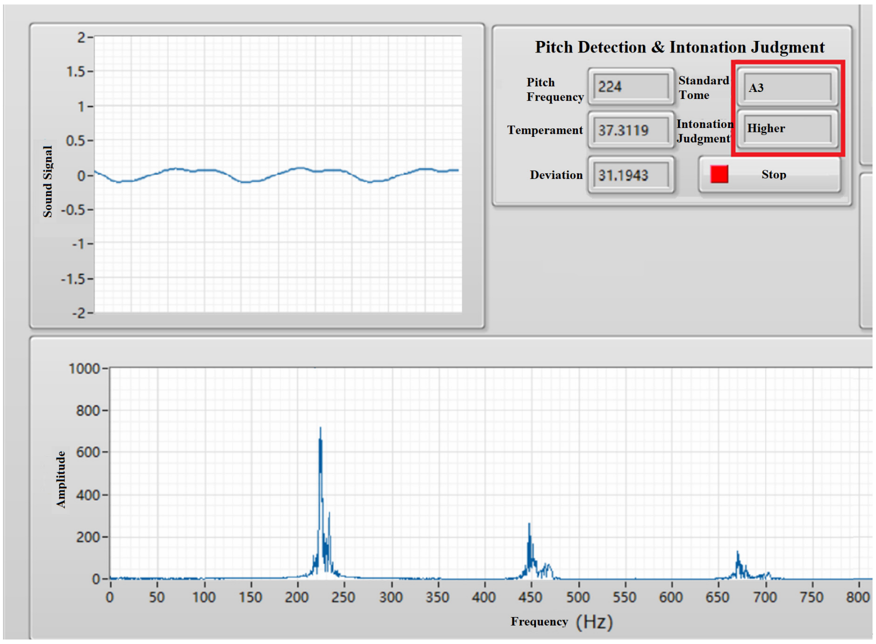

Fig. 14. Chart for high-pitched test

5.2 Result of Intonation Correction

Singing D3 in a low-pitched manner on purpose, and then performing correction. As shown in Fig. 15, the pitch frequency detected during singing is 145, the cent difference between the temperament and the D3 standard tone is -21.74, indicating that the intonation is lower than the standard. After corrected by SOLA-FS, the pitch detection results show the pitch frequency is 145.5, with deviation from the standard tone of -15.78, which can be judged as accurate intonation and achieve the correction effect. Comparison of frequency spectrum before and after correction is as shown in Fig. 16.

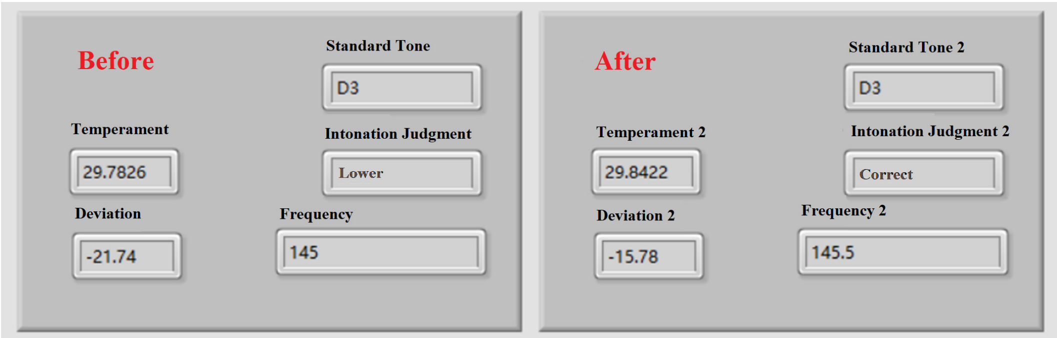

Fig. 15. Comparision of results before and after correction

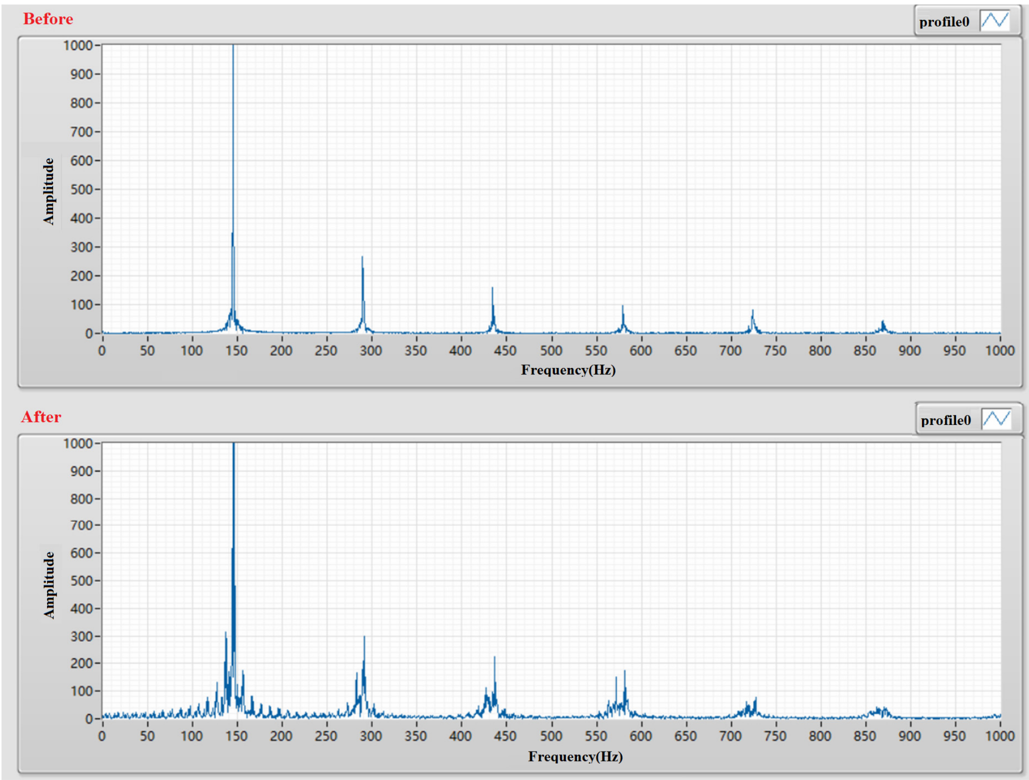

Fig. 16. Comparision of frequency spectrum before and after correction

6 Conclusion

The present intonation detection and correction system is based on NI myDAQ for audio acquisition and is implemented by SOLA-FS and LabVIEW. Test data and waveforms observed in several tests show that it delivers accurate pitch detection. Theoretically, the system can detect the pitch range corresponding to 88 keys on the piano. Such a wide detection range can basically cover the sound frequency of general musical instruments and people. Besides, it can accurately judge the intonation in real time by comparing the sound with the standard tone. Tests show that the system can correctly judge the intonation and basically meet the tuning requirements of simple musical instruments and singing practice.

As for the intonation correction, by comparing various data and waveforms before and after correction, it can be seen that within a certain range of tonal difference, the system can correct inaccurate tone to a certain extent. Such corrections, however, have certain limits. Taking D3 as an example, the tone can only be corrected when the deviation is not greater than 50 cents, otherwise, it will be judged as another tone, thus changing the original one. Therefore, the intonation correction remains to be improved. It can only correct within a certain range of tonal difference. For example, C3 cannot be corrected to D3. When playing a sound, a non-professional person can hardly differentiate the intonation before and after correction, but the data comparison indicates the tone has been corrected.

There is still room for improvement in future research due to these limitations. Compared with the existing uni-functional tuners and pitch shifting software on the market, the system described herein can detect pitch, judge intonation and correct the simple note with intonation error. It is truly a full-featured intonation correction system. The system provides references for tuning of simple musical instruments and guidance for training of singers. It helps music performers better control the intonation during the performance. It is hoped that this type of system can make people like music and feel its beauty.

References

[1] C. Cheng, Design of a pitch quantization and pitch correction system for real-time music effects signal processing, Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific. Hollywood, CA, USA: IEEE Conferences (2012) 1-6.

[2] S. Jilt, P.A. Manoj Kumar, A.M. Hema, An analysis of the high resolution property of group delay function with applications to audio signal processing, Speech Communication (2016) 81.

[3] W. Michael, I.P. Weiss, Ability to process musical pitch is unrelated to the memory advantage for vocal music, Brain and Cognition (129)(2019) 35-39.

[4] L. Wu, Study on the Technology of Pitch-Scale Modification, China Shenyang Ligong University (2016).

[5] C. Gussenhoven, Correction: Analysis of Intonation: the Case of MAE_ToBI, Laboratory Phonology (1)(2016) 7.

[6] H. Mizuno, M. Abe, Voice conversion algorithm based on piecewise linear conversion rules of formant frequency and spectrum tilt, Speech Communication (2)(1995) 16.

[7] J. Zhou, Y. Chen, W. Tan, Y. Gao, Parameter set Choices of SOLA Algorithm for Time Scale Modification of Speech, Microelectronics & Computer (04)(2007) 54-58+62.