ANALYZE THE PRINCIPALS OF
PITCH-SHIFTING

Xiao Han 420149996(SID)

Digital Audio Systems, DESC9115, Semester 1 2014

Master in Audio and Acoustics

Faculty of Architecture, Design and Planning

University of Sydney

Abstract

Nowadays, with the development of digital music, more and more digital technologies are used in electronic music. Digital signal processing is one of the most powerful technologies that will shape science and engineering in the twenty-first century [1].

Pitch shifting, as one of the most important processing, always plays significant role in the whole digital system. This paper introduces some basic knowledge about pitch-shifting like the development and working principles. After that, the paper introduces some different methods to produce the pitch-shifting effect. And then it will mainly introduce the time-stretching approach and explain the principals.

The main objective of this written review is to illustrate the principals of pitch shifting and different approaches to produce it. Finally, it stress on some problems we need to overcome during pitch shifting.

Introduction

What is pitch shifting?

The pitch of a sound corresponds to the set of frequencies the sound is made [2]. We can regard it as a choir. There are both men and women singing but the women will have a higher pitch than the men. What pitch shifting does is it takes the sound produced at a given pitch and changes its frequencies. For instance, by shifting up the pitch of a man who is singing, we could end up with a sound where it sounds like a woman is singing instead. Pitch shift relates to two aspects: pitch and timbre.

Pitch

Firstly, pitch means one note being produced at a particular level. It partly depends on the fundamental frequency of one signal. It is a kind of very subjective sense. Maybe influenced by a lot of music and harmonizer, therefore in humans’ conscious there is a kind of inborn ability to distinguish different pitch of a sound. Pitch has a very close relation with frequency, but it doesn’t mean frequency can decide the pitch. Usually, when you feel pitch is growing up, the frequency is also increasing at the same time. For example, when pitch goes up one octave, frequency will be improved twice. Especially, when it arrives the hearing sensitive range, this kind of change will
be more distinguished. But when the tone is too high or too low, this kind of distinguish ability will become more and more ambiguous. However, “pitch” is still a totally subjective notion. For example, when frequency exceeds 4000Hz, the change of pitch caused by frequency will become more and more blur. Therefore, timbre is a quite important part of pitch. The same frequency sound, if we set different timbre, it will also lead to the difference of pitch distinguish. In addition, the more complex the sound is, the less accuracy of pitch distinguish ability.

Secondly, pitch partly depends on the fundamental frequency of one signal. Also, if we change the ratio of the loudness of each band, pitch will be also changed. For example, if we increase the low frequency band, pitch will drop, while if we increase the high frequency band, pitch will rise. In addition, the change of pitch will also lead to the loudness change in one sound. For example, if we lower the pitch, it will generate louder sound, while if we improve the pitch, sound will be weaker than before. One signal can be derived where an excitation part is modified by a resonance part. In the case of the voice, the excitation is provided by the vocal chords, hence it is related to the frequencies of the spectrum, whereas the resonances correspond to the formants. When a singer transposes a tune, he has, to some extent, the possibility of modifying the pitch and the formants independently.

Thirdly, pitch is a kind of very subjective sense. Maybe influenced by a lot of music and harmonizer, therefore in humans’ conscious there is a kind of inborn ability to distinguish different pitch of a sound. Pitch has a very close relation with frequency, but it doesn’t mean frequency can decide the pitch. Usually, when you feel pitch is growing up, the frequency is also increasing at the same time. For example, when pitch goes up one octave, frequency will be improved twice. Especially, when it arrives the hearing sensitive range, this kind of change will be more distinguished. But when the tone is too high or too low, this kind of distinguish ability will become more and more ambiguous. However, “pitch” is still a totally subjective notion. For example, when frequency exceeds 4000Hz, the change of pitch caused by frequency will become more and more blur. Therefore, timbre is a quite important part of pitch. The same frequency sound, if we set different timbre, it will also lead to the difference of pitch distinguish. In addition, the more complex the sound is, the less accuracy of pitch distinguish ability.

**Timbre and tone quality**

Firstly, Timbre is the quality of a musical note that distinguishes different types of sound production. It is a more subjective sense and assessment. It is a very subjective description to different types of sound source. Timbre decided by a lot of different factor, including different types of resonances. Even though a very short sound, the first impression to listeners should be the timbre, namely, the property of the sound source. Timbre mainly depends on harmonizer and the organization of its spectrum. So if we can control the harmonizer and the organization of its spectrum, that means we can control the timbre.

Secondly, any natural sound can be divided into four stages: attack, decay, sustain and release. They made up the envelope of a sound. Therefore, sometimes two signals have same frequency, but they bring totally different feeling to listeners. This is because the envelope of each signal is different. If we put the attack part to sustain part, it sounds no changes.

Thirdly, Timbre and tone quality is a little different. Timbre is a kind of “color
expression "of music, it is the emotion expression of music, we can describe it by using “bright” and “dark” while tone quality is a wider and more subjective conception. It relates to the “quality” of sound and it is a physical expression of a sound.

Finally, the distinguish ability to timbre and tone quality are mostly based on the hearing experiences. Because most people will compare different sounds in subconsciousness and search the suitable words to describe it. The process of comparing is based on the hearing experiences in listeners’ memory. If a listener has been trained for long time, he will distinguish the different timbre and sound quality very quickly and accurately.

Development

Traditional way

Before digital audio signal appears, we just can control the pitch by adjusting the strings or other part’s size of the instruments. It is a kind of very traditional ways to generate pitch-shifting effect. For example, if we want to shift up the guitar’s pitch, we usually use a capo to lock the strings at a specific spot on the neck. This reduces the length of each string and thus changes its fundamental frequency, which results in a new tuning at a higher pitch. Each time the performer transposes the melody, he makes use of a different register of his instrument. By doing so, not only is the pitch of the sound modified, but also the timbre is affected.

Figure 1

![Diagram of traditional pitch-shifting method](attachment://image.png)

In order to shift the pitch down, the strings need to be slacked. The tension in the strings is reduced by turning the tuning keys (Figure 2).
Figure 2

Digital can also produce it

Nowadays, with the development of digital sound effect, we can get the same sound effect by using pitch-shifting function. Because it is “digital”, we don’t need consider about the timbre modification or whether the characteristic timbre of the instrument has to be maintained in each of its registers.

As I illustrate before, the timbre of a sound heavily depends on the organization of its spectrum. A signal model can be derived where an excitation part is modified by a resonance part. In the case of the voice, the excitation is provided by the vocal chords, hence it is related to the frequencies of the spectrum, whereas the resonances correspond to the formants. When a singer transposes a tune, he has, to some extent, the possibility of modifying the pitch and the formants independently. In a careful signal-processing implementation of this effect, each of these two aspects should be considered. If only the spectrum of the excitation is stretched or contracted, a pitch transposition up or down, with a constant timbre, is achieved. If only the resonances are stretched or contracted, then the pitch remains the same, but the timbre is varied. Harmonic singing relies on this effect. If both excitation and resonance are deliberately and independently altered, then we enter the domain of effects that can be perceived as unnatural, but that might have a vast musical potential.

The digital implementations in the form of the harmonizer might allow for a better quality, but there are still severe limitations. For transpositions of the order of a semitone, almost no objectionable alteration of the sounds can be heard. As the transposition ratio grows larger, in the practical range of plus or minus two octaves, the timbre of the output sound obtains a character that is specific to the harmonizer. This modification can be heard both in the frequency domain and in the time domain and is due to the modulation of the signal by the chopping window.

Approaches and principals to generate pitch shift

There are mainly three methods to produce pitch-shifting effects time stretching and resampling, delay-line modulation, PSOLA and formant preservation. In digital signal, the method of time stretching and resampling is the most common and the simplest way to shift the pitch, so this part mainly describes time stretching and resampling.
Time stretching and resampling

Time stretching is the process of changing the speed or duration of an audio signal. If we apply it in signal processing, it will change the pitch due to the change of the duration. The main task of time-stretching algorithms is to shorten or lengthen a sound file of M samples to a new particular length $M' = \alpha \cdot M$, where $\alpha$ is the scaling factor[3].

We can clearly see from the following figure 3, when $\alpha$ is decreased from 1 to 0.5, the duration of signal becomes shorter, and at the same time in the corresponding spectra, the notch between around 500Hz disappears. When $\alpha$ is increased from 1 to 2, the duration of signal becomes longer, at the same time in the corresponding spectra, the notch between around 500Hz becomes deeper. Therefore that means with the decrease of factor $\alpha$, the duration of signal becomes shorter and we can cut low frequency. This phenomenon is very common when we play the tape. If we play it in fast forward, the speed will be very fast but we can hear the pitch of the music becomes higher than before.

Figure 3

![Figure 3](image)

Resampling is the simplest way to change the duration or pitch of a digital audio clip is to resampling it. This is a mathematical operation that effectively rebuilds a continuous waveform from its samples and then samples that waveform again at a different rate. When we apply the time stretching and resampling in digital signal. There is different variables lead to a compression or expansion of the duration of a sound and to a pitch shift. This is accomplished by resampling in the time domain. Figure 4 has already illustrated the discrete-time signals and the corresponding spectra. The spectrum of the sound is compressed or expanded over the frequency axis. Figure 4 clearly illustrates the change of the signal after time stretching and resampling. We can see by changing $\alpha$, we can control the duration of the signal and the frequency ratio. With the increase of $\alpha$, we can effectively increase high frequency and cut low frequency.
Figure 4

Delay-line modulation
The following block diagram describes the principal of delay-line modulation. It a pitch shifter based on an overlap-add scheme with two time-varying delay lines is proposed. We can clearly see that the signal is divided into two parts a and b. Stream b is delayed, and then both a and b are added the cross-fade function.

Figure 5

Pitch shifting by PSOLA and formant preservation
PSOLA means Pitch Synchronous Overlap and Add, it is a digital signal processing technique used for speech processing and more specifically speech synthesis. It can be used to modify the pitch and duration of a speech signal. At the core is a PSOLA based pitch shifter, which is controlled by the output from a pitch extractor from another thread [4]. Manipulation of fundamental frequency is achieved by changing the time
intervals between pitch markers [5]. The modification of duration is achieved by either repeating or omitting speech segments. We can clearly see from the following figure that to increase the pitch, we need cut the time intervals between pitch markers while in order to decrease the pitch we need to increase the time intervals.

![Graphs showing increase and decrease of pitch](image)

**Discussion**

Although we can use digital signal process to shift the pitch of the music, however there are still many problems we need to conquer in processing. Firstly, some signal before processing has its original tune, so if we want to change to tune totally, we need flat the pitch firstly and then shift the pitch. Secondly, when we shift the pitch, especially shift the signal to very high frequency, some of the sound is distorted. That means some original signal is damaged because of pitch shift. So we need be careful about it. When a distortion occurred fist of all we need to find where is the problem and measure the signal level at as many points as possible. Then shift the pitch within the feasible range.

**Conclusion**

In conclusion, pitch plays an important role in sound, and it relates to fundamental frequency and harmonizer. Pitch shifting is a very common way to change the pitch of a sound. Generally speaking there are three methods to produce it: time stretching and resampling, delay-line modulation, PSOLA and formant preservation. Each approach has its own features. However, there are still many problems need to solve in future, for example how to flat the pitch and overcome distortion.
Reference