2019 EE 6641 “Analysis and Synthesis of Audio Signals (ASAS)”

Final Project

Group 6

105061214 蘇淮安 105061132 吳柏澔 104061242 李宇凡 107591506 陳鵬安

**[Introduction]**

In this final project, we experimented on instrumental sound transformation, seeing if we can successfully transfer a piano sound to guitar and vice versa. This idea came from the motivation that in our daily lives, we may hear some beloved song performed in piano, for example, but we want to listen to the same song played in guitar or other more familiar instruments.

In comparison with another totally different method to approach this, which uses pitch detection or other music information retrieval (MIR) techniques to find out the exact score of a song then replaces it with other instrumental sound source, we decided to operate the transformation directly on signal. Prospectively, with this method, we can probably remain more musicality and performing style in the original audio during transformation, just like the same piece of music played with other instruments.

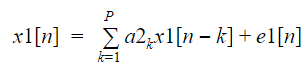
In this project, we mainly apply linear prediction (LP) model, which was used in previous lab. Although the model was used to process human's voice, we try to modify the model and see if it still work on instrumental sound. To reduce our experimental variables and make the work simple, we first deal with single notes, such as C4, D3, and a simple C major scale.

Though our project can now only deal with single pre-recorded note, we discuss some further application in the future. Firstly, we can use machine learning techniques to improve our model, for example, using neural network to train the LP coefficients, automatically finding their proper relation between different instruments. Besides, we can also combine audio input with our model, which reads a piece music in, analyzes it, and then re-synthesizes with other instruments components, approaching our goal indeed - the auto instrumental sound transformation.

**[Methods]**

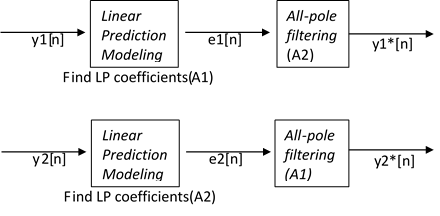
This project is implemented by the linear prediction model. In previous Lab 5 or “Introduction to digital speech processing” [1], the purposes are to analyze human voices and to observe the extracting parameters, such as linear prediction coefficients and excitations. Therefore, the aim of this project is to resynthesize signal from LP coefficients and excitations of two different instruments.

Before synthesizing the signals, recording samples need to be preprocessed. Preprocessing part is below the Fig1. In addition, Fig1 is the brief descriptions of process for this project. From records of the guitar and piano, we can obtain each LP coefficients and excitations by linear prediction. And then, if we want to transform a guitar sound to a piano sound, we take piano’s LP coefficients to filter guitar’s excitation. Besides, in the project, adjusting the frame length and orders is the main factor that affects the qualities of resynthesized signals.



*x1 is the signal of piano,a2 is guitar’s LP coefficients , e1 is piano’s excitatio*n

Transforming a piano sound to a guitar is also in the same way.



*Fig1*

Moreover, we gave an informal test to identify how the results work well. We give not only the resynthesized signals,guitar to piano or piano to guitar , but also the original clips to see whether subjects can distinguish difference.

Preprocessing:

There are a few things to notice before recording the guitar and the piano.

First we choose the upright piano and acoustic guitar. We use fingers play the guitar directly not the pick. In addition, due to the method we synthesize the signals, we try to make each note in the same length. That is, we want each segment of piano can correspond to each segment of guitar. However, the lengths of sound decay in two instruments are different. Lengths are based on how you play the instruments. Hence, it needs to be careful to keep the strength steady or synthesized signals will have some troubles.

**[Results]**

In the process of the linear prediction, there are some variables in this process required modification. The first issue needed to be dealt with is the frame length of the system. The frame length is set to 0.08 as the initial setting. The result turns out to be disappointing. The frame length at 0.08 causes the generated soundtrack to create constant knocking sound for literally every 0.08 second and is clearly not the desired output to be glad with. The frame length is then set higher and higher in order to get the best outcome. Frame lengths at 0.16, 0.32, 1.5, 2.5, and 3.5 are attempted. In the end, the frame length setting stops at 2.5 and receives the best possible sound generation. Any value lower or higher than this particular length is considered worse. As a whole, the result with 2.5 in frame length is considered the best that can be found. That is why we need to preprocess our records in advance.

In addition to frame length, order is also a factor to the sound. A desired soundtrack is referred to a clear, pitch-less knocking sound when listening to the excitation of the sound. For example, if the original sound track is a pitch scale of C3 to C4 (low Dol to high Dol in C3 major), the excitation of it should be 8 separated pitch-less knocking sound if the order is set properly. At first, the order is set to 50 and pitch is still clearly heard in the excitation, which is the opposite of the result required. The order is then set higher and higher for better results. After setting 50, 100, 150, 200, 250, 300, and 500, it turns out order at 250 has the best outcome among them. With the information gathered above, it is safe to say that 250 in order are the closest to a perfect excitation.

During the presentation in class, sound quality testing and statistics is made by interacting with the students and the teacher. The method behind this is mentioned in the methods section. *Form 1* shows the results to the statistics. P to G means converting piano soundtrack to the guitar sound track while G to P is the other way around. The total number of guessed people is different, which is not considered or discussed in the statistics.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Test Data | 1 | 2 | 3 | 4 | 5 | 6 |
| Conversion | P to G | P to G | G to P | P to G | G to P | G to P |
| Like guitar | 11 | 9 | 1 | 10 | 0 | 12 |
| Like piano | 10 | 8 | 17 | 9 | 18 | 10 |
| accuracy | 52.38% | 52.94% | 94.44% | 52.63% | 100% | 45.45% |

*Form 1*.

According to the statistics, the P to G conversion has an accuracy of 52.38%, 52.94%, and 52.63%. All three of them have similar accuracies close to fifty two percent. The G to P conversion, on the other hand, has an accuracy of 94.44%, 100%, and 45.45%. Despite the last data, G to P conversion has an overall decent performance with high accuracy. This leads to a conclusion established that the G to P conversion has a better work effectiveness than the P to G conversion.

**[Discussions]**

Our group think that converting guitar soundtrack to the piano soundtrack is more successful because of the habit of hearing these instrument sounds. Generally, we usually hear the higher pitch sound from the piano and the guitar sound is lower pitch in most musical occasions. Therefore, we also can see the result part to realize .

Otherwise, we also discuss about the limitation of this final project. First of all, these audios are recorded by ourselves. However, since we did not use high-level recorder, there are some noise in these soundtrack. This problem also leads the sound track conversion to have non-ideal sound and influence the final results.

On the other hand, we also think that these sounds of instruments are generated non-linearly. Nevertheless, in this final project, we only use the linear prediction model to convert soundtrack. Therefore, maybe there are some limitations we cannot simulate the non-linear part by the linear method.

Apart from the above reasons, using the machine learning method may improve our results. But a problem we may encounter is that the training datasets are too few. Therefore, in the part of distinguishing instrumental sounds may have more errors.

Finally, our group has a good idea as this application. We can record the different pitches of piano and guitar in advance and save these LP coefficients. Next someone can play the piano or guitar and use MATLAB to record the sound at the scene. Through a series of calculations, we can have the LP coefficients to find the minimum difference with recorded sound track. Therefore, we may distinguish which instrument being played and synthesis the other instrument sound to display. This way may become an instrumental converter and interact with the people at the scene.

**[Reference]**

[1] LR Rabiner and RW Schafer, “Introduction to digital speech processing,” in *Foundations and Trends in Signal Processing*, Vol. 1, No.1-2, pp. 1-194, 2007.