Music dictation is a very popular way to play music without using musical scores. However, this skill depends on one’s experience and sense. Moreover, the effects units such as distortion, chorus, flanger and so on control the tones of guitar sounds used in recent music. This makes it difficult to take down in musical notation for not only human also software. In this paper, we focus on the effect of distortion that is one of the most famous guitar effect units. Our motivation is to improve music dictation regardless of one’s skills and experience in the sense of not the pitch but the quality of the sound. Distortion units involve a nonlinear transform, which is constructed by the following two steps: amplification and clipping. It is difficult to analyze distortion sounds by Fourier method, since the distortion unit process has clipping; the nonlinear transform. We have defined the methods to extract the feature of distortion sound with the wavelet. That is, we defined the index of distortion sound feature as the correlation between the sound signal and the Haar wavelet. Since the wavelet function has the strong localization, we can extract the feature of distortion sound. Therefore, we propose the describing method based on the index of distortion written above and the center of gravity of the spectra, which corresponds to the tone of the sound. The subjective experiments were conducted to validate the similarity between the proposed indexing method and the human perception. The result of the experiment showed that the sounds with same feature value and same center of gravity of the spectrum are recognized as the similar sounds. This means the proposed method could be used as the index of the distortion sounds which human and machine could commonly recognize.

Keywords: distortion sound, Fourier transform, wavelet transform,

1. Introduction

Music dictation is a very popular practice among musicians to reproduce music they have heard without referring to a musical score. This requires them to master the identification and differentiation of various musical elements by hearing. Thus, music dictation relies highly on the musician’s experience and music sense. We will interested in dictation of guitar music involving distortion sounds. Distortion sound is one of the most famous electric guitar effects. Hereinafter, we assume the signal \( f(t) \) satisfies that \( \max_t |f(t)| \leq 1 \).
A distortion effect is a nonlinear transform that involves two steps, amplification and clipping. An amplification is a transformation from an original signal \( f(t) \) to \( \tilde{f}(t) = Cf(t) \) for some constant \( C > 1 \). Clipping means cutting the signal off. It is a nonlinear transformation from \( \tilde{f}(t) \) to \( f(t) = \max\{-1, \min\{1, f(t)\}\} \) (see Fig. 1). Clipping produces non-differentiable points that make analyzing with Fourier method difficult because it can yield high frequency (see Fig. 2). This means that we cannot measure the level of distortion sound when we see only the frequency of the distortion sound.

### 2. Previous Method

In previous work, there are several methods used for analyzing distortion sound. One of the values of the describing distortions level is Total Harmonic Distortion (\( D_{\text{THD}} \)) (see [2] [7]) which is defined by

\[
D_{\text{THD}} = \frac{\sqrt{H_2^2 + H_3^2 + \cdots + H_N^2}}{\sqrt{H_1^2 + H_2^2 + H_3^2 + \cdots + H_N^2}}
\]  

or

\[
D_{\text{THD}} = \frac{\sqrt{H_2^2 + H_3^2 + \cdots + H_N^2}}{H_1}
\]

where \( H_N \) denotes Harmonic response of \( N \)-th harmonic and \( H_1 \) does Fundamental response. Nevertheless, \( D_{\text{THD}} \) does not consider the original sound harmonic. If we try to measure the level of distortion sound with total harmonic distortion, we have to recognize the original sound frequency. Therefore, total harmonic distortion is not effective in this situation (see [8]).

Our motivation is to aid in music dictation regardless of one’s skill or experience. We can get the pitch of the sound from using autocorrelation of itself. However, it is difficult to know the distortion

Figure 1: the structure of distortion sounds

Figure 2: the spectrograms of distortion sounds
level of the sound from just frequency information. As mentioned, previous method including Fourier analysis is not an appropriate method. Therefore, we shall propose to use wavelet analysis to extract the feature of distortion sounds.

3. Proposal Methods

3.1 Wavelet analysis

The wavelet transform is one of the time-frequency representations (see [3]). The word "wavelet" means the "small wave" and wavelet is the function which has the localization that trigonometric function does not have. That is, wavelet analysis enables us to obtain the local information of the signal more than the Fourier analysis does. The first wavelet is the Haar wavelet in the early 20th century, when the concept of the wavelet did not exist at that time. There are a lot of wavelets and we can choose them in accordance with the signal which we want to analyze.

![Image of Haar wavelet]

Figure 3: Haar wavelet

Wavelet can analyze with dilation and a shift translation of the wavelet function. We call $\psi$ a wavelet if it satisfies the admissible condition

$$C_\psi = \int_{-\infty}^{\infty} \frac{\hat{\psi}(\omega)^2}{|\omega|} d\omega < \infty,$$

where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(t)$:

$$\hat{\psi}(\omega) = \int_{-\infty}^{\infty} \psi(t) \exp(-it\omega) dt. \quad (4)$$

For $a, b \in \mathbb{R}, a \neq 0$, and wavelet $\psi$, we define the continuous wavelet transform (CWT) of signal $f(t)$ by

$$W[f](a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t) \overline{\psi} \left( \frac{t - b}{a} \right) dt \quad (5)$$

where $\overline{\psi(t)}$ denotes the complex conjugate of $\psi(t)$. In (5), the parameters $a$ and $b$ represents dilation and shift translation respectively. If $\psi(t)$ is the real valued function, wavelet transform is same as taking correlation between $f(t)$ and $\psi(t)$. The inverse continuous wavelet transform is also defined by

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W[f](a, b) \psi \left( \frac{t - b}{a} \right) \frac{da db}{a^2}. \quad (6)$$

3.2 Distortion Unit

Distortion pedal has mainly three kinds of knobs to change the quality of the guitar sound continuously. One is the level knob. It just changes the "volume" (or level) of the sound. Second is the "gain"
(or distortion) knob. This can alter the level of distortion. Third knob "tone" can change the frequency band that musician wants to emphasize. The level knob does not affect the quality of sound. That is, we have to take the gain and the tone into consideration to feature extraction of distortion sound. However, since distortion sound includes the irreversible transform, Fourier method is not proper to grasp the feature of distortion sound. Therefore, we will propose the method to extract the feature of distortion sound to approximate sound quality.

### 3.3 Feature of Distortion Sound

The Haar wavelet has much strong localization (see Fig. [3]). The Haar wavelet is defined as

$$\psi(t) = \begin{cases} 1 & (0 < t \leq \frac{1}{2}), \\ -1 & (\frac{1}{2} < t \leq 1), \\ 0 & \text{(otherwise)}, \end{cases}$$  \hspace{1cm} (7)

We will make the best use of Haar wavelet’s character, strong localization in two ways.

First, we focus on the amplifying of the distortion units. We define the feature of distortion sound $f(t)$ with Haar wavelet as follows:

$$E_1(a) = \max_b \left| \frac{1}{a} \int_{-\infty}^{\infty} f(t) \psi \left( \frac{t - b}{a} \right) dt \right|,$$  \hspace{1cm} (8)

and we recognize the large value of (8) represents the high level distortion sound. The absolute value of integration in (8) becomes higher when the translation of the Haar wavelet overlaps with the amplified part of distortion signal, and lower when the translation of the Haar wavelet overlaps with the clipped part of distortion signal. That is, the quantity (8) bring out where the gradient is higher. The parameter $a$ can be changed in accordance with the pitch of the sound since the wavelength depends on it. To be dimensionless of feature quantity, the coefficient of $E(a)$ is changed from CWT. Feature value $E_1$ can measure the gradient of distortion sound $f(t)$ even $f(t)$ is discrete signal. Moreover, the value $E_1$ is independent of the signal state, steady or unsteady, since it takes maximum of the translation parameter $b$. We have already shown this effectiveness to extract feature of distortion sound (see [8]).

Second, we focus on the clipping part of distortion sound $f(t)$. The high level distortion sound has a lot of clipping part. Therefore, we define the feature of distortion sound based on the amount of clipping as follows:

$$E_2(a) = \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} \left| f(t) \right|^{1/4} \psi_H \left( \frac{t - b}{a} \right) dt \right| db \hspace{1cm} (9)$$

and we recognize the small value of (9) represents the high level distortion sound. Since the mean of the Haar wavelet is zero, that is,

$$\int_{-\infty}^{\infty} \psi(t) dt = 0,$$  \hspace{1cm} (10)

the value of (9) vanishes when the translated Haar wavelet overlaps the clipped part of distortion sound $f(t)$. Accordingly, clipped part of $f(t)$ makes the value of (9) smaller. The value $E_2$ depends on the signal state. We should take care the state and if the level of the signal decreases rapidly, we have to pick up the part where the signal is state.

Moreover, to approximate the quality of sound, we should take the center of gravity of the spectra $E_0$ into the consideration:

$$E_0 = \frac{1}{S} \int_{-\infty}^{\infty} \xi \hat{f}(\xi) d\xi \hspace{1cm} (11)$$

where $S$ denotes the area of the $\hat{f}$. Although the center of gravity of spectra is not proper to estimate level of distortion, it is effective to approximate the quality of sound since which corresponds to the tone of the sound.
4. Effectiveness of Proposed Method

To verify the utility of proposed method is, we did some experiments.

4.1 The Result for Simple and Distortion Sound

First, we applied the above methods (8) and (9) for simple sound and distortion sounds. We prepare the seven sounds \( f_n(t) = \max\{-1, \min\{1, n \sin t\}\} \) \( (n = 1, 2, 3, 4, 5, 10, 100) \) (Fig. 4) and applied the (8) and (9). One wavelength of each \( f_n \) equal 80. The results are shown in from Fig. 5 to Fig. 8 where \( |\text{supp } \psi| \) denotes the length of the interval that \( \psi(x) \neq 0 \).

In each graph, the vertical axis represents the value of proposed formula and the horizontal axis numbers correspond to the values of n (6 corresponds to the case \( n = 10 \) and 7 does to the case \( n = 100 \)).

With the method (8), both graphs are growing. This means that higher level distortion sound takes the larger value of \( E_1 \) and we can extract the feature of distortion sound. We can say that the method (9) can also extract the feature of distortion sound since the graphs are decreasing with the method (9).
To compare our method with human perception, we conducted an experiment to fifteen people that ranging from eighteen years old to twenty eight years old. We prepared eight distortion simple sounds. For each kind of the sounds, the experiment involves the following two steps:

a. The subjects listen to the some sample sounds and identify among them the distortion sounds.

b. The subjects listen to two sounds and determine the one with higher level of distortion.

We can see the result of the experiment for simple guitar sounds with the formula (8) in Fig. 9 and with the formula (9) in Fig. 10. In each figure, the horizontal line represents the feature values $E(a)$ of the proposed method while the vertical lines represent the order of distortion levels recognized by human perception. The topmost is the highest level of distortion sound that human perception recognized and the bottommost is the lowest level of distortion sound that human perception recognized. That is, the large number in the vertical line means human perception judges sound to be high level distortion sound.

The results say that the method (8) has much correlation with human perception since the Fig. 9 grows with the value of $E_1(a)$. The method (9) also has the correlation if the distortion level of the guitar sound is high. That is, especially for high distortion, we can see the level of distortion sound as the feature value.

Figure 7: the result of the method (9) with $|\text{supp } \psi| = 5$

Figure 8: the result of the method (9) with $|\text{supp } \psi| = 10$

4.2 The Result of Subjective Experiment

Figure 9: the result of the method (8)

Figure 10: the result of the method (9)
4.3 The Results of Effectiveness of Spectra

Moreover, we did the experiment about the center of gravity of spectra of distortion sound. We prepared the two sounds $S_1$ and $S_2$ that have the same values of $E_1$ and $E_2$ and the different values of $E_0$. We also prepared seven sounds that have the same values of $E_1$ and $E_2$ and the different values of $E_0$. For each of the sounds, the experiment consists of the following two steps:

a’. The subjects listen to two sound $S_1$ and $S_2$ and recognize them.

b’. The subjects listen to other sounds and determine the one from two sound $S_1$ or $S_2$ that is similar to.

The result is that about 93 percent of the answers are ones that close to the center of gravity of the spectrum. This shows that human perception recognize that two distortion sounds are similar if they have nearly centers of gravity of the spectra. Therefore, the center of spectra is effective to approximate distortion sound.

5. Conclusion

Since the distortion unit has irreversible transform, Fourier transform is not quite enough to analyze it. So we defined the feature values of distortion sound $E_1$ and $E_2$ with the Haar wavelet. Since the Haar wavelet has the strong localization, proposed methods are correlated with human perception. Moreover, with the center of gravity of spectra of distortion, we can make similar distortion sound for the human perception.

REFERENCES


