EQ2341 Pattern Recognition and Machine Learning

Assignment 2: Song Recognition

Hanqi Yang hanqi@kth.se

Qian Zhou qianzho@kth.se

April 29, 2022

1 Introduction

The goal of this project is to design a pattern recognition system capable of distinguishing between a few brief snippets of whistled or hummed melodies. In this particular assignment, we will design a feature extractor suitable for this melody recognition task, which will involve some knowledge about music signals, Fourier analysis, music theory, and human pitch and loudness perception.

2 Feature Extractor Design

In this part, we will design and implement a set of features for melody recognition, based on signal pitch, correlation, and intensity series from *GetMusicFeatures*. The features have the following properties:

1. They should allow distinguishing between different melodies, i.e., sequences of notes where the ratios between note frequencies may differ.

2. They should allow distinguishing between note sequences with the same pitch track, but where note or pause durations differ.

3. They should be insensitive to transposition (all notes in a melody pitched up or down by the same number of semitones).

4. In quiet segments, the pitch track is unreliable and may be influenced by background noises in your recordings. This should not affect the features too much, or how they perceive the relative pitches of two notes separated by a pause.

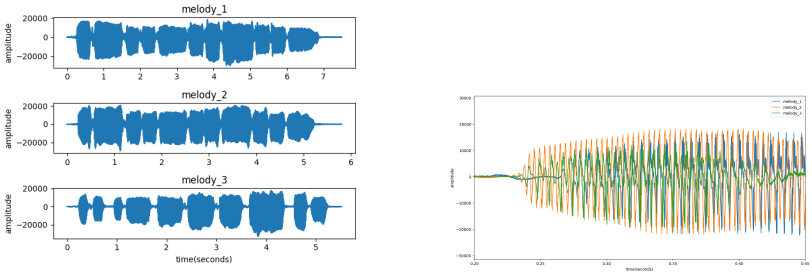
5. They should not be particularly sensitive if the same melody is played at a different volume.

6. They should not be overly sensitive to brief episodes where the estimated pitch jumps an octave (the frequency suddenly doubles or halves).

First we plot the the sound waves of the three melodies, which is shown in Figure 1(a). It can be seen that the period and frequency of the wave changes depending on its tone. There is a high level of similarity between the first two melodies, while the third one is entirely different from the first two melodies. Then we zoom in on a range of 500 samples in Figure 1(b), we can find that the signal in this section has a "wavy" appearance.

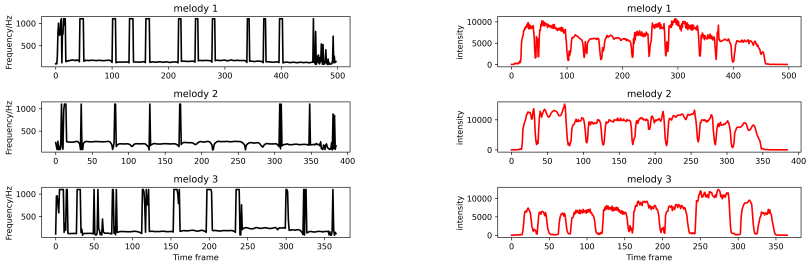
Then, to help develop the features, we make a plot of the three pitch profiles of these recordings together, and another plot with the three corresponding intensity series, which are shown in Figure 2(a) and Figure 2(b) respectively.

From the picture of pitch frequency, we can't tell which two pictures are from the same melody, but from the picture of intensity, we can roughly guess



(a) All samples (b) 500-1000 samples

Figure 1: Audio Signal for melodies



(a) pitch (b) intensity

Figure 2: Pitch and intensity of melodies

that the first two of these are from the same melody, while the third one is from a different melody. For more robust note detection, we plot the estimated correlation coefficient between adjacent pitch periods of the three melodies.

From Figure 3, it can be found that the correlation coefficient is lower in the non-harmonic (noisy or silent) regions. By setting a threshold and removing the components below the threshold, we can cancel part of non-harmonic components. Here, we choose the threshold as 0.75.

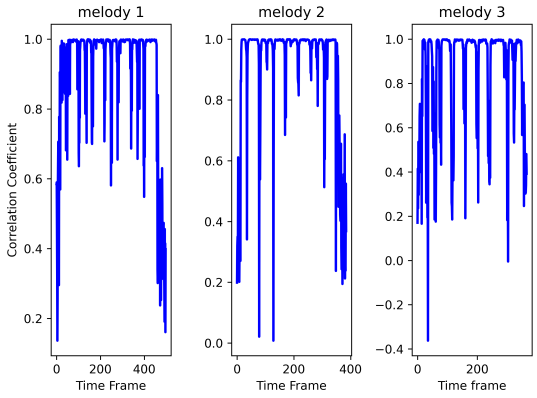


Figure 3: Correlation coefficient of melodies

We plan to use GaussMixD output distributions for our HMMs, so we add Gaussian noise to the non-harmonic components. The mean of the Gaussian noise is chosen at the pitch frequency of the harmonic components which is closest to the mean of the whole pitch frequencies, which helps reduce the number

of Gaussian distributions in the Gaussian Mixture Model.

The octave is further subdivided into twelve smaller steps, known as semitones. A melody is usually formed by a series of semitones with different duration. Therefore, we subdivide the pitch using the transformation relationship:

$$s = 12log_2(\frac{f}{f_0}) \tag{1}$$

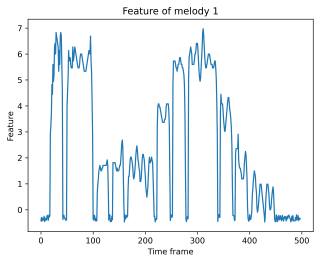
where f is the pitch frequency and f_0 denotes the threshold, which is the minimal pitch frequency.

3 Performance Check

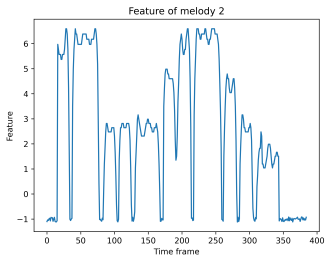
3.1 Features of Melodies

We apply the feature extractor in three melodies. The features are shown in Figure 4. It can be seen that the features of melody 1 and 2 are reasonably similar, and we can say they are from the same source, while the feature of melody 3 is quite different from 1 and 2 and we can say it is from another source. It is clear that this feature extractor can allow distinguishing between different melodies.

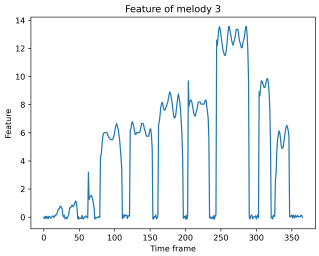
From the different characteristic between the features of melody 1 and 2, the feature extractor can also distinguish between note sequences with the same pitch track, but where note or pause durations differ. The property 1 and 2 of the feature extractor are illustrated in this part.



(a) melody 1



(b) melody 2



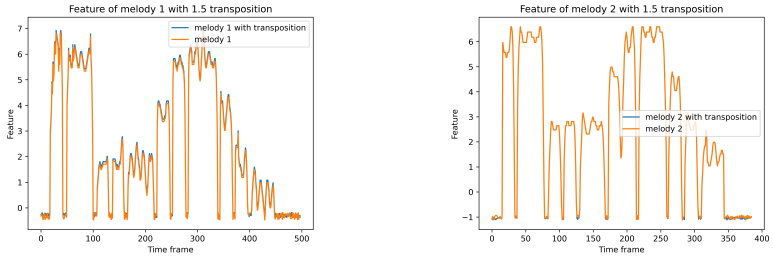
(c) melody 3

Figure 4: Extracted features of melodies

3.2 Transposition Independent

To verify that the feature extractor is transposition independent, we mulitply the pitch track of melody 1 and 2 by 1.5 and the result is shown in Figure 5.

Comparing the features before and after the transposition, it can be sees that the feature of both melodies stay unchanged. We can say that the feature extractor is transposition independent to some extent.

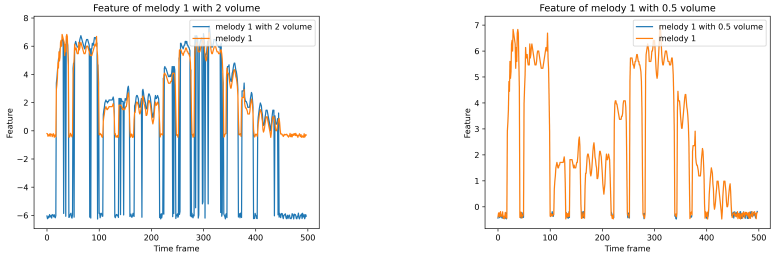


(a) melody 1 (b) melody 2

Figure 5: Features of 1.5 pitch of melodies

3.3 Volume Sensitivity

To verify the property 5, we double and halve the volume of melody 1 and the features are shown in Figure 6. In the feature with low volume, the feature matches the original feature perfectly. In the feature with high volume, however, the low value of the feature is significantly reduced, although the upper contour is similar to the original feature. We have to say the feature extractor is sensitive to the volume amplification.



(a) high volume (b) low volume

Figure 6: Features of melody 1 with volume changed

3.4 Brief Episode Sensitivity

We insert 10 jump points into pitches of melody 1 randomly, 5 of which are doubled and 5 halved, and then observe the result of the feature, which is shown in Figure 7. The result shows that the feature with 10 jump pitch stays consistent, which means the feature extractor is not very sensitive to brief episodes where the estimated pitch jumps an octave, i.e. frequency suddenly doubles or halves.

4 Discussion

After designing and testing the feature extractor, we would like to discuss some limitations of this system in this section.

First of all, the feature of melodies will be affected by the volume amplification. The high values are lower compared to the original values, but the upper

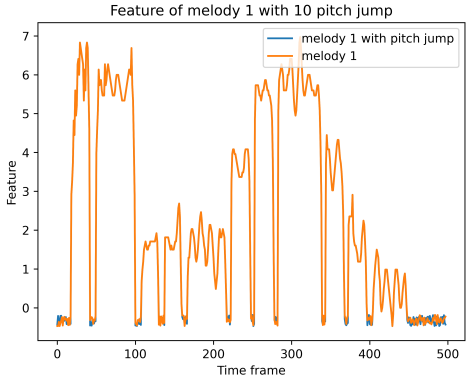


Figure 7: Feature of melody 1 with 10 pitch jump

contours basicly match, while the low values drop from around 0 to around -6 from the original values. And many lower values also increase abruptly. In our opinion, the amplification of volume changes the correlation coefficient of the melody, the feature extractor, however, uses a fixed threshold 0.75 to choose the pitch, which may cause this result.

Almost the same problem, the threshold 0.75 basically only applies to the case of these three melodies, and once the number of melodies increases, it is necessary to change it to a dynamic threshold, which will be able to fit into more situations.

Secondly, the feature extractor is only based on the pitch, correlation coefficient and intensity of each frame, which means many other properties of melody, such as tempo, timbre, are not taken into consideration. If we change the instrument of the melody, or we speed up the tempo, the feature may become different.

5 Conclusion

In this article, we design a feature extractor with many given properties and test it. After using the given functions, we obtain the necessary parameters of melodies. We then set threshold to choose the needed pitch, add noise to the pauses and convert the pitch into semitones. The feature of melody is extracted by such a process. We test the six properties of feature extractor and find that it can distinguish between different melodies and note sequences with the same pitch track, be insensitive to transposition and the noise in the pauses. It is also not sensitive to volume reduction and brief episodes like pitch jump, but it will be affected by the volume amplification. We finally discuss some limitations of this feature extractor. Overall, this feature extractor basicly meets the requirements of the project.