Improved Perceptual Tempo Detection of Music

Bee Yong Chua and Guojun Lu
Gippsland School of Computing and Information Technology
Monash University, Churchill, Victoria 3842
Australia
Email: {bee.yong.chua, guojun.lu}@infotech.monash.edu.au

Abstract

Perceptual tempo refers to listener's tempo perception how fast the music goes, when he listens to a piece of music with fairly constant overall tempo. Existing work on automatically determining the tempo of a piece of music is usually not able to determine the Perceptual tempo. Therefore, our previous work on the Perceptual Tempo Estimator (PTE) improved on existing work and experimental results had shown its effectiveness in determining the perceptual tempo than the existing work. However there are music pieces that have overall uniform perceived tempo, but have some small parts (segments) with quite different tempo from other parts. The PTE may not work well for this type of music pieces if the energy level in these small segments is significantly high. In this paper, we propose an improved PTE (IPTE) to provide better tempo determination. Experimental results show that IPTE is more effective in determining the Perceptual tempo of the music signal with constant perceived tempo.

1. Introduction

To effectively and efficiently manage and use digital music database, we need a number of automatic music analysis techniques/tools. One technique needed is to automatically determine the tempo (beat rate) of music that corresponding to human perceived music speed (how fast or slow). We call this tempo *Perceptual tempo.* Thus, music perceived to be faster would have higher perceptual tempo than music perceived to be slower.

Determining perceptual tempo has many applications. For example, we can retrieve music based on the perceptual tempo. Further, perceptual tempo is one of the main factor that enable listener to identify and classify different emotion expression in music [5] and the empirical findings in music psychology research on how the tempo affects different emotion expressed in music are always in terms of perceptual tempo. Therefore, perceptual tempo can be used as one of the main features for classifying and retrieving music signal in a large music database based on different emotion expression in music. In addition, perceptual tempo can also aid in identifying and segmenting different emotion segments in video analysis.

Existing work on automatically determining the tempo of a piece of music mainly focus on estimating the *Score tempo* or the *'Foot-tapping' tempo*. Some of the works are [2,4,6,8-10]. However, both the Score and Foot-tapping tempo are usually not the same as the Perceptual tempo. *Score tempo* refers to the tempo reflected in the music score for musicians to follow as they played the music. Figure 1 shows one example where the two different music extracts have the same score tempo but different perceived tempo. Based on the estimated score tempo, the computer system will interpret that these two music extracts have the same tempo. But the actual perceived tempo of the bottom music extract is about three times slower than the top one. *Foot-tapping tempo* refers to the tempo listener sub-consciously tap along when he listens to a piece of music. It does not usually correspond to perceptual tempo because foot-tapping tempo is centred around average human normal heart beat rate of about 80 to 100 beats per minutes (bpm). Thus, for a piece of music with very fast-perceived tempo, foot-tapping tempo is usually half of the actual perceived tempo. Table 1 illustrates three pieces of music extract with different listener's tempo perception but having the same foot-tapping tempo. Based on these foot-tapping tempos, the computer system will interpret that these three music pieces have the same tempo. However, the actual perceived tempos are different.

Proceedings of the 11th International Multimedia Modelling Conference (MMM'05)
1550-5502/05 $20.00 © 2005 **IEEE**

Therefore, our previous work on the Perceptual Tempo Estimator (PTE) improved on existing work and experimental results had shown its effectiveness in determining the perceptual tempo than the existing work [1]. However, there are music pieces that have overall uniform perceived tempo, but have some small parts (segments) with quite different tempo from other parts. The PTE may not work well for this type of music pieces if the energy level in these small segments is significantly high. In this paper, we propose an improved PTE (IPTE) to provide better tempo determination.

The next section outlines the PTE algorithm and discusses the limitation of this algorithm. Following that, our proposed IPTE algorithm is detailed. And in the subsequent sections we describe the test music data used and the experimental results.

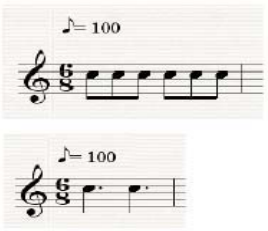

Figure 1: Illustration of two different music extracts with same score tempo

Listener's Tempo Perception	Foot-Tapping Tempo (bpm)	Perceptual Tempo (bpm)
Fast	80	160
Moderate	80	80
Slow	80	40

Table 1: Illustration of three music extracts with different perceived tempos but having the same foot-tapping tempo

2. PTE Algorithm

Figure 2 shows the block diagram of our previous proposed PTE algorithm. The followings are the outline of the algorithm, detail description of this algorithm is found in [1].

1. As in the common approach in estimating tempo, the input music signal is first divided into different frequency sub-bands. And in each sub-band, the signal is smoothed to produce amplitude envelops. In our algorithm, we had chosen six different frequency sub-bands (0-200Hz, 200-400Hz, 400-800Hz, 800-1600Hz, 1600-3200Hz and 3200-half of sampling frequency).
2. The periodicity of each of the un-normalized smoothed sub-band signal is detected using *Autocorrelation Function* (ACF). ACF compares one signal with the delayed signal of itself. The ACF plot is plotted with the degree of similarity between the signal and the delayed signal against time lag. Thus, the peaks in the ACF plot reflect the beat occurrence at different duration.
3. *Peak Finding algorithm* is then applied to find the tempo of each sub-band. Existing work estimates the tempo by finding the highest peak with time lag greater than zero. However, this highest peak usually does not reflect the Perceptual tempo. Our hypothesis of perceptual tempo is that it should correspond to the most salient event among various music events in a piece of music. Thus, based on this hypothesis, our peak finding algorithm determine each sub-band tempo by finding the peak with the shortest time lag, away from time lag zero, above 20% of the total energy. The threshold of 20% was determined empirically.
4. Since we are interested in how fast the listener perceived a piece of music, therefore the perceptual tempo of the music signal is determined by finding the highest tempo among the sub-band tempos.

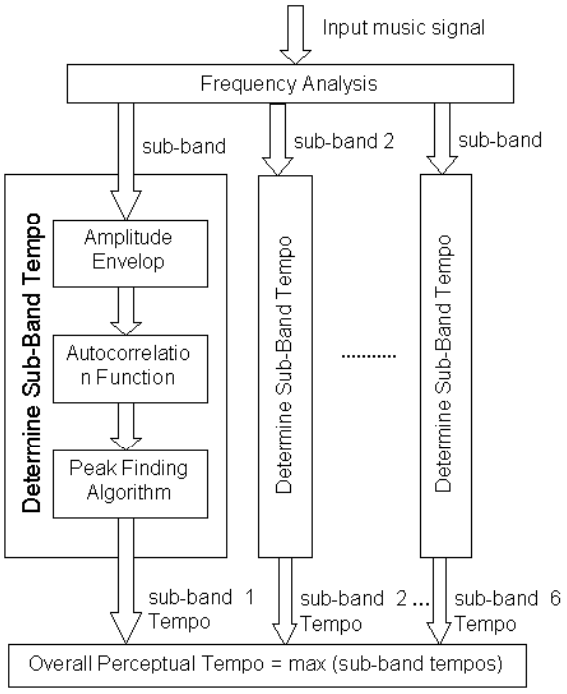

Figure 2: Block Diagram of PTE Algorithm

Experimental results and discussion detailed in [1] shown that this PTE algorithm is effective and more reliable than existing tempo estimation works in determining perceptual tempo of a piece of music.

However, there are some music pieces that have overall uniform perceived tempo, but have some small parts (segments) with quite different tempo from overall perceived tempo. When the energy level of these segments is high enough, our PTE algorithm may use the tempo of these segments as the overall perceived perceptual tempo, leading to wrong perceptual tempo estimation for this type of music.

3. IPTE Algorithm

To overcome the problem of PTE, we propose IPTE in this paper. In IPTE, a piece of music is divided into small fixed length segments. We then apply PTE algorithm on each time segment to estimate the tempo for each segment. The overall tempo of a piece of music is determined from these segment tempos by considering a number of factors based on the findings from music psychology studies. One of the finding is that in perceptual encoding of rhythmic patterns, listener attempts to find a regular beat pattern [7]. The other finding is that when a brain begins to sense a train of pulses, it continues to anticipate them even when individual pulses disappear into silence, or notes held long [4]. However, due to human short-term memory limitation, anticipation of these pulses can begin to fade unless these pulses are repeated within a time period [3].

Based on the above findings we determine the most regular beat pattern by finding the tempo that occurs most often among the segment tempos. Further, due to human anticipation ability as discussed above, we make the earlier segment tempos more likely to be the final overall perceptual tempo than the later segment tempos. Therefore, when we determine the final overall perceptual tempo from the segment tempos, we need to consider how similar the segment tempo is to the rest of the segment tempos and whether this segment belongs to the earlier time segment.

The overall IPTE algorithm is as follows:

1. Divide the music piece into fixed length (we used 10 seconds) segments.
2. Apply PTE algorithm to each segment to obtain the tempo for that segment.
3. Find the frequently occurred tempo among the segment tempos by determining the similarity weight in each segment.
4. Determine the likelihood weight for each segment tempo to make it more likely that the earlier segment tempo will be selected as overall perceptual tempo.
5. Determine the overall perceptual tempo of the music from the information obtained in steps 2, 3 and 4.

We describe steps 3, 4 and 5 next.

3.1. Finding the frequently occurred Tempo

In each of the fixed 10 seconds segments of the music signal, PTE algorithm is applied to estimate the tempo for each segment. The overall perceptual tempo of a music piece will be determined from one of these segment tempos. Thus, in determining the overall perceptual tempo, we need to consider how similar the segment tempo is to the rest of the segment tempos and whether this segment belongs to the earlier time segment. The followings will describe how we determine the frequently occurred tempo. And in the next section, we will describe how we determine the likelihood weight for each segment tempo to make it more likely that the earlier segment will be selected as overall perceptual tempo.

Each of the segment tempos estimated is in whole number ranging from a possible lowest of 20 bpm to a possible highest of 200 bpm. However, listener usually cannot detect slight difference in tempo. But rather, he perceives it as having the same tempo. Thus, in finding the frequently occurred tempo among the segment tempos, we need to regard those segment tempos with only slight difference as having the same perceived tempo. We do so by determining the similarity weight for each segment tempo. Segment tempos having similar tempo as other segments will have higher similarity weight than those segment tempos having different tempo as others. Thus, the segment tempo with higher similarity weight implies that the tempo estimated in this segment is occurring more frequently in other segments than those segment tempos having lower similarity weight.

The similarity weight for each segment is determined as follows:

1. We first check how similar the segment tempo is to each of the other segment tempos by calculating their tempo difference.
2. We then assign different weight based on the tempo difference. In order to give more weight to the segments having the same tempo, we set the weight to 1.2 if the tempo difference is zero, otherwise, the weight is inversely proportional to the tempo difference.

3. The similarity weight for the segment tempo is then the sum of all the weights determined in steps 2.

3.2. Determine the likelihood weight

As mentioned, in determining the overall perceptual tempo, we need to consider how similar the segment tempo is to the rest of the segment tempos and whether this segment belongs to the earlier time segment. The earlier section described how we determine the frequently occurred tempo. And in this section, we will describe how we determine the likelihood weight for each segment tempo to make it more likely that the earlier segment will be selected as overall perceptual tempo.

The segment tempos are obtained by applying PTE algorithm to each of the fixed 10 seconds segments of the music signal. Thus, the segment tempo for the first 10 seconds of the signal will be the earliest time segment tempo and the segment tempo for the last 10 seconds of the signal will be the latest time segment tempo. In order to give more weight to the earlier segment than the later segment, we determined the likelihood weight as follows:

1. We first calculate the likelihood weight for each segment as the inverse proportional to the segment number (with earlier segment having lower segment number than the later segment).
2. In our test data, the duration of the music extracts range from about 30 to 60 seconds. Thus, we will have some music signal having only 3 segment tempos in a piece of music and others having 4, 5 or the most 6 segment tempos in a piece of music. With the maximum numbers of segments varies in different music extracts, the third segment, for example, in a piece of music with maximum of 6 segments should have different weightage than the third segment in another piece of music extracts with only maximum of 3 segments. To achieve this, we normalized the likelihood weight for one segment by taking its likelihood weight calculated in step 1 and divide by the sum of the likelihood weights calculated in step 1 for each segments.

3.3. Determining the Overall Perceptual Tempo

The overall perceptual tempo of a piece of music signal is determined by:

1. Calculating the total weight for each segment tempos.
2. The final overall perceptual tempo is then determined by selecting the segment tempo with the highest total weight.

The total weight for each segment tempo is calculated by combining the similarity weight and the likelihood weight found in section 3.1 and section 3.2 respectively as follows:

$$totalWt(a) = [\alpha 1 * simWt(a)] + [\alpha 2 * segWt(a) * (N-1)]$$

where:

totalWt(a) = total weight at segment a
simWt(a) = similarity weight at segment a
segWt(a) = likelihood weight at segment a
N = total numbers of segments in a piece of music

In order to give more emphasise on the similarity weight than the likelihood weight, $\alpha 1$ and $\alpha 2$ are set to 0.8 and 0.2 respectively. These values are determined empirically.

Further, the reason for multiplying the likelihood weight by (N-1) is to give slightly more emphasize on the likelihood weight for those music extracts with higher maximum number of segments than those with lower.

4. Experimental Results

4.1. Test Data

Previously, we had used 25 music extracts to test our previous proposed PTE algorithm. In this paper, we have increased it to 50 music extracts. Each of these music extracts has overall fairly constant perceived tempo. In Table 2, the first 6 of the music extracts are converted to WAV format from MIDI and the rest are converted from CD recordings to WAV format. All the music extracts are in 16-bit mono audio with sampling frequency rate of 11025Hz. Duration of the music extracts is mostly the first one minutes of the music and a few exceptions with the first 30 seconds of the music.

The test data represents different perceived tempo, with some music pieces having very fast tempo, about 180 bpm, some with moderate tempo and some with very slow tempo, about 50 bpm. In addition, the test data are made up of a number of different musical genres, ranging from ballad, rock, waltz, tango, popular music with vocals, instrumental with and without drumming, classical without drumming, jazz and blues.

4.2. Results

The results comparing the Perceptual tempo estimated using the PTE and IPTE algorithm are shown in Table 2. Experimental results and discussion detailed in [1] had shown that the PTE is effective and more reliable than existing tempo estimation works in determining Perceptual tempo of a piece of music.

Column 2 reflects the Ground Truth or the listener's tempo perception. The ground truth is determined manually by measuring how fast each piece of test music is perceived.

The experimental results obtained are all in bpm and the values in Bold and Italics indicate the tempos that are not corresponding to Perceptual tempo, or the Ground Truth. Due to expressive timing in music, it is rare to find music pieces with fixed tempo throughout, although listener may perceived it to be fairly constant. But rather, for music with overall fairly constant perceived tempo, the tempo is usually vary slightly throughout the whole music piece. Therefore, if the result obtained is slightly different from the Ground truth, it is considered to be corresponding to the Ground truth.

Column 3 shows the results estimated using PTE algorithm and column 4 shows the results estimated using IPTE algorithm. On the whole, IPTE algorithm is more effective than PTE in determining Perceptual tempo.

5. Discussion

Column 2 of Table 2 shows results obtained using the PTE approach. Except for Blues 2, the rest of results in this column that are not corresponding to the ground truth (in Bold and Italics) are those music extracts having some small segments with quite different tempo from overall perceived tempo. Further, the energy level in these small segments is much higher than the rest of the segments. Thus, the PTE algorithm estimates the tempo based on these small segments. Blues 1 and 2 have many small segments with twice the tempo of the overall perceived tempo. However, PTE manages to estimate Blues 1 correctly but estimated Blues 2 as twice the perceived tempo. This is because in Blues 1, it happens that the energy belonging to the actual perceived tempo segments has slightly more energy than that with twice of the actual perceived tempo segments. However, in Blues 2, it was the other way round.

Column 3 of Table 2 shows results obtained using the IPTE approach. As shown, IPTE is able to estimate accurately those music extracts having some small segments with quite different tempo from overall perceived tempo. However, IPTE is not able to estimate the perceptual tempo of Blues 1 and 2. This is because in Blues 1 and 2, there are many small segments with different perceived tempo scattered throughout the whole music signal.

6. Conclusion

In this paper, we discussed the problem in our previous proposed PTE algorithm to determine the perceptual tempo. Further, we discussed about how we overcome the problem by proposing an improved PTE algorithm. In IPTE, a piece of music is divided into small fixed length segments. We then apply PTE algorithm on each time segment to estimate the tempo for each segment. The overall tempo of a piece of music is determined from these segment tempos by considering a number of factors based on the findings from music psychology studies.

We had shown from experimental results obtained that IPTE is more effective in determining the Perceptual tempo of the music signal with constant perceived tempo.

7. Future Work

As mentioned, PTE and IPTE approach can only estimate tempo with music signal with fairly constant perceived tempo, having the most, just some small parts with quite different perceived tempo. However, as in Blues 1 and 2, there are music pieces with many small parts with quite different perceived tempo scattered throughout the signal. Further, there are also music pieces with more than one perceived tempo in a piece of music. Thus, in one of our future work, we will try to determine the perceptual tempos for these types of music pieces. We are looking into the possibility of finding the segment boundary, where each segment will contain only one constant perceived tempo, rather than dividing the music piece into small fixed length segments.

Further, we will be using the perceptual tempo determined from the IPTE approach as a main feature in classifying different emotion expression in music.

8. Reference

[1] B.Y. Chua, G. Lu, Determination of Perceptual Tempo of Music, Computer Music Modeling and Retrieval (CMMR 2004), Esbjerg, Denmark, 2004.

[2] M. Goto, An Audio-based Real-time Beat Tracking System for Music With or Without Drum-sounds,

Journal of New Music Research, 2001, vol 30, p.159-171.

[3] D.A. Hodges, Handbook of music psychology, Institute For Music Research UTSA, 1999.

[4] R. Jourdain, Music, the brain, and ecstasy : how music captures our imagination, New York : W. Morrow, 1997.

[5] P.N. Juslin, J.A. Sloboda., Music and emotion : theory and research, New York : Oxford University Press , 2001.

[6] J. Laroche, Estimating tempo, swing and beat locations in audio recordings, in: IEEE Workshop on Application of Signal Processings to Audio and Acoustic (WASPAA), New Paltz, New York, USA, 2001, 135-138.

[7] D.J. Povel, The internal representation of simple temporal patterns, Journal of Experimental Psychology: Human Perception and Performance 7, 1981, p. 3-18.

[8] E. Scheirer, Music Listening System, Doctor of Philosophy, Massachusetts Institute of Technology, 2000.

[9] J. Seppanen, Computational Models of Musical Meter Recognition, Master of Science Thesis, Department of Information Technology, Tampere University of Technology, 2001.

[10] G. Tzanetakis, Manipulation, Analysis and Retrieval Systems For Audio Signals, Doctor of Philosophy, Princeton University, 2002.

Genre	Ground Truth	PTE	IPTE
Ballad 1	73	73	73
Ballad 2	70	69	69
Ballad 3	80	***40***	78
Rock 1	102	101	101
Waltz 1	91	***46***	91
Tango 1	121	122	121
Popular 1	55	52	65
Popular 2	170	163	164
Popular 3	69	69	68
Popular 4	64	64	63
Popular 5	140	134	133
Popular 6	132	133	132
Popular 7	71	69	72
Popular 8	150	151	148
Popular 9	72	88	89
Popular 10	80	92	70
Popular 11	78	***154***	78
Popular 12	150	152	151
Popular 13	80	80	80
Popular 14	173	172	174
Popular 15	63	31	68
Popular 16	102	102	102
Popular 17	117	***79***	119
Instrumental 1	170	170	170
Instrumental 2	150	153	153
Instrumental 3	175	180	181
Instrumental 4	70	71	75
Instrumental 5	110	108	108
Instrumental 6	155	156	157
Instrumental 7	53	50	57
Instrumental 8	121	119	121
Instrumental 9	180	183	184
Instrumental 10	90	90	90
Classical 1	56	57	59
Classical 2	60	59	60
Classical 3	55	32	23
Classical 4	64	66	62
Classical 5	83	87	89
Classical 6	131	128	131
Classical 7	176	***95***	176
Classical 8	175	179	183
Classical 9	103	97	100
Classical 10	130	133	133
Jazz 1	105	109	110
Jazz 2	121	121	121
Jazz 3	72	***155***	80
Jazz 4	100	136	104
Blues 1	67	65	***128***
Blues 2	60	***129***	***124***
Blues 3	101	156	99

Table 2: Experimental results. See section 4.2 for description.