FULL-AUTOMATIC DJ MIXING SYSTEM WITH OPTIMAL TEMPO ADJUSTMENT BASED ON MEASUREMENT FUNCTION OF USER DISCOMFORT

Hiromi Ishizaki
KDDI R&D Laboratories Inc.
ishizaki@kddilabs.jp

Keiichiro Hoashi
KDDI R&D Laboratories Inc.
hoashi@kddilabs.jp

Yasuhiro Takishima
KDDI R&D Laboratories Inc.
takisima@kddilabs.jp

ABSTRACT

This paper proposes an automatic DJ mixing method that can automate the processes of real world DJs and describes a prototype for a fully automatic DJ mix-like playing system. Our goal is to achieve a fully automatic DJ mixing system that can preserve overall user comfort level during DJ mixing.

In this paper, we assume that the difference between the original and adjusted songs is the main cause of user discomfort in the mixed song. In order to preserve user comfort, we define the measurement function of user discomfort based on the results of a subjective experiment. Furthermore, this paper proposes a unique tempo adjustment technique called “optimal tempo adjustment”, which is robust for any combination of tempi of songs to be mixed. In the subjective experiment, the proposed method obtained higher averages of user ratings on three evaluation items compared to the conventional method. These results indicate that our system is able to preserve user comfort.

1. INTRODUCTION

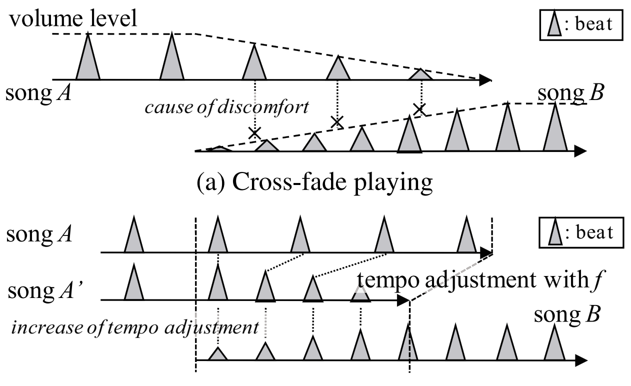
Due to the development of various audio compression methods, many online music distribution services have provided the opportunity for users to listen to songs from huge music collections. Furthermore, the increasing popularity of portable music players has enabled users to carry around thousands of songs. However, the variety of methods for the common user to enjoy listening to the songs in their collection is basically limited to “shuffle” play, which simply plays songs in the collection (and/or playlists) in random order. In order to extract a set of songs that match user preferences from large-scaled music collections , there are many useful techniques such as [1–3]. These techniques can provide users a set of songs as playlists, from which users select and play songs. In order to provide users new experience, it is important to play the songs in an entertaining way. For instance, Basu proposed a method which can blend two songs smoothly to create different aspects of the songs [4].

In the real world, DJs (disk jockey), *i.e.*, people who select and play music in clubs and discos, are able to maintain the excitement of the audience by continuously playing songs with the utilization of various DJ techniques: selections of songs, beat adjustment, *etc.*. One fundamental DJ technique is to gradually switch from one song to the other, while adjusting the beats of the songs. This technique enables the DJ to switch songs smoothly without disturbing the listener. A similar method should be effective in providing an entertaining music experience for common music listeners. However, such music playing requires skilled techniques and/or specialized equipment, which are both difficult for casual users to utilize.

In this research, we propose an automatic DJ mixing method that can automate real world DJ processes and describe a prototype for a fully automatic system. The objective of this research is to develop an automatic music playing system that can play a variety of different songs consecutively in an entertaining way without causing the users any discomfort. Specifically, we define the measurement function of user discomfort based on the results of a subjective experiment. Furthermore, we propose an optimal tempo adjustment technique that is robust for any combinations of the tempi of songs to be mixed.

2. CONVENTIONAL PLAYING METHOD

As mentioned in the previous section, DJs effectively utilize the cross-fade playing (*CFP*) technique to maintain the entertain level of the music they play. Naive *CFP*, *i.e.*, cross-fading two songs without any tempo/beat adjustment, is a simple and effective approach in avoiding silence between songs, and can be easily implemented in any music playing application. This method is effective in avoiding silence between songs, which may be distracting to listeners who prefer that the music play continuously. However, especially in situations where the tempi of the two songs to be cross-faded are significantly different (Figure 1-(a)), naive *CFP* may result in a negative listening experience, since the beats of the two songs occur asynchronously. Therefore, it is necessary for DJs to conduct *CFP* while adjusting the tempo and beat of one song to the other. The adjustment of tempo can be done by simple signal expansion (in cases where the song is to be played slower than the original) or contraction [5].



(a) Cross-fade playing

(b) naive DJ mixing

Figure 1. Conceptual illustrations of cross-fade playing and naive DJ mixing.

3. PROBLEMS

However, there are two problems in realizing such DJ techniques automatically.

One problem is the degradation in the acoustic quality of music, which may occur in the tempo adjustment process, especially in conditions where the tempi of the two target songs are significantly different (Figure 1-(b)). Such quality degradation may cause discomfort for listeners. Furthermore, the double or half tempo error is common for any existing automatic tempo extraction algorithm, as mentioned in [9]. Although a highly accurate tempo and beat extraction method is obviously essential for the implementation of a fully automatic DJ mix playing system, it is unrealistic to expect any system to achieve 100% accurate beat extraction. If the fully automatic DJ mix playing system adjusts the tempo based on tempo extraction results with double/half errors, the resulting factors of tempo adjustment will be two times the actual requirement. It is obvious that such excessive tempo adjustment is a cause of acoustic quality degradation, and ultimately, discomfort for music listeners. Furthermore, in the cases of adjustment of the song/songs that result in double/half tempo errors, strong beats and weak beats are adjusted to each other, which causes user discomfort.

The other problem is that there is no previous work on the effective measure of tempo adjustment to preserve the comfort level of users. It is not clear that users feel discomfort with regard to the degree of tempo adjustment or the manner in which the tempo was adjusted for the songs to be mixed. Actually, it is essential to define some kind of measure in order to achieve the fully automatic DJ mixing system. Additionally, it is important to investigate the threshold and the applicable range of tempo adjustment for songs to be mixed in order to achieve a comfortable DJ mixing system.

4. DEFINITION OF MEASUREMENT FUNCTION

In this section, we conducted a subjective experiment to define the measurement function of user discomfort. The objective of this experiment is to define the measurement function of user discomfort to determine the level of user discomfort given the tempo adjustment ratio.

In this experiment, we assume that the difference between the original and adjusted songs is the main cause of user discomfort. We investigate the correlation between user discomfort and tempo adjustment factors with actual tempo adjusted songs using time-scaling algorithms. Details of this experiment are presented as follows.

4.1 Experimental method

The methodology of this experiment, namely, details on the method of generating the sample audio and the subjective measure, are explained. In this experiment, we generate the actual songs for which the tempo will change. Subjects listen to these songs and input the time when they feel discomfort.

The experimental data set consists of 18 popular songs selected from the RWC music database [11]. For each of the selected songs, tempo changes are applied to the song excerpts. The adjusted tempo is obtained by multiplication of the original tempo of song and the factor of tempo adjustment f, $f > 1$ means the speedup factor and $f < 1$ means the slowdown factor. The speedup and slowdown factors for tempo changes are set from 1.00 to 2.00 and 1.00 to 0.30, respectively. For each experiment, the song is played in its original tempo for the first 15 seconds. After this initial period, the tempo of the song is repetitively increased (in the case of speedup) or decreased (in the case of slowdown) by a scale of 0.05, for every three seconds, until the tempo change factor reaches its maximal/minimal value. This range is decided empirically enough to investigate the correlation.

In the tempo adjustment, we have changed the time scale of the songs, while maintaining the original pitch. As tools of tempo adjustment, we use the two time-scaling algorithms: the audio processing library *SoundTouch Library* [1] and the *SOLA* [10] time-scaling algorithm. *SoundTouch* is a high quality means to change tempo, *SOLA* is a low quality means. A total of 72 excerpts are generated for this experiment (44.1 kHz, 16-bit, WAV).

In this experiment, the 96 subjects are divided into two groups. Each group listens to half of the excerpts (36 excerpts per group). In the listening task, the subject is to submit the time when they feel discomfort to the tempo change of the song. The submission results are accumulated to analyze the effects of tempo change factors.

4.2 Results

Table 1 shows the averages of tempo adjustment factors that subjects feel discomfort to the song associated with each time-scaling algorithm. In this table, there are differences between speedup and slowdown factors where the subjects feel discomfort. These results show that the subjects are more sensitive to effect of slowdown as opposed to speedup. Furthermore, the averages of tempo adjustment factors for *SoundTouch* and *SOLA* are approximately

[1] SoundTouch Library: http://www.surina.net/soundtouch/

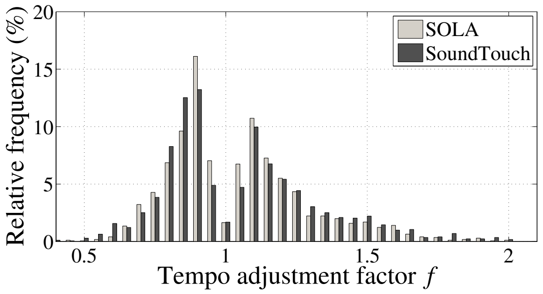


Figure 2. Histogram of user discomfort to factors of tempo adjustment.

Table 1. Averages of tempo adjustment factors

method	speedup	slowdown
SoundTouch	1.227	0.852
SOLA	1.226	0.852

equal to each other. These results indicate that user discomfort depends on tempo adjustment factors rather than the method.

Figure 2 shows the histogram of user discomfort and factors of tempo adjustment with each time-scaling algorithm. In this figure, the factors at the peaks of each algorithm are 1.10 (speedup) and 0.90 (slowdown). The percentages of subjects that feel discomfort inside these factors of each algorithm are 15.42% (*SOLA*) and 11.31% (*SoundTouch*). In the area near the original tempo, there are differences between the algorithms. *SoundTouch* is better able to preserve the comfort level of subjects under the condition in which the factor satisfies $0.90 < f < 1.10$ than *SOLA*.

4.3 Definition from the result

In order to define the measurement function based on the results in the previous section, we assume that the difference between the original and adjusted songs is the main cause of user discomfort. On the basis of this assumption and previous results, we define the level of discomfort (L_{dc}) expressed by the following equation:

$$L_{dc}(f) = \begin{cases} a(f-1) & f > 1 \\ 0 & f = 1 \\ b(1/f - 1) & f < 1 \end{cases} \tag{1}$$

In Eq.(1), parameters a and b are to be weighted because the level of user discomfort is different between the adjustment from the speedup factor and from the slowdown factor as described in the previous section. Hence we extract the weighted parameters a and b as $a = 0.765$ and $b = 1.000$, these are extracted to make the score computed by speedup and slowdown factors equal when the factors are given as those written in Table 1. These weighted parameters are assumed to be effective in preserving the users' level of comfort in the song-to-song (*StS*) transition of DJ mixing. For example, Eq.(1) is able to decide which factor is appropriate (speedup or slowdown) in the DJ mixing. Additionally, we extract the stricter and average applicable ranges from the factors at the peaks (mentioned in Section 4.2) and the averages shown in Table 1. Specifically, we extract the $0.90 < f < 1.10$ as the stricter applicable range and $0.852 < f < 1.227$ as the average applicable range.

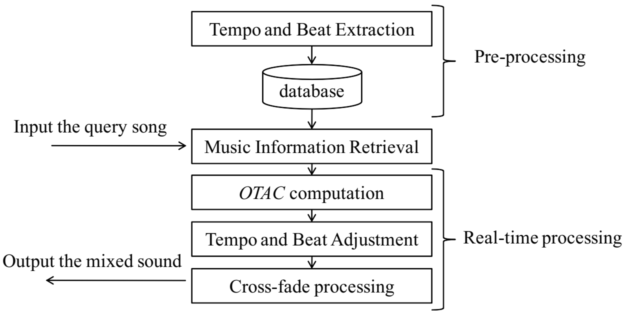


Figure 3. An overview of prototype of fully automatic DJ mixing system.

5. SYSTEM

In this section, we describe the prototype of the fully automatic DJ mixing system, which can solve the problems of tempo/beat adjustment, described in Section 3. By applying the score of the measurement function, which is computed based on the tempi of the target songs, our system is designed to be able to preserve the overall level of user comfort during the transition between songs.

Fig. 3 shows the overview of the prototype for the fully automatic DJ mixing system. This system mainly consists of five processes: tempo and beat extraction, music information retrieval (*MIR*), optimal tempo adjustment coefficients computation, tempo and beat adjustment, and cross-fade playing. In this system, we propose a unique tempo and beat adjustment method, which is able to deal with double or half tempo errors in the tempo and beat extraction technique: optimal tempo adjustment is able to compute the optimal factors of tempo adjustment to minimize the amount of tempo adjustment by dealing with tempo octave relationships. Details of the main processes of the system are described as follows.

5.1 Tempo and beat extraction

In this section, we describe the method of automating the DJ processes: tempo and beat extraction. As concerns the tempo and beat extraction process, there are many research efforts in tempo and beat extraction techniques, such as [6–8]. Although these techniques have the common problem of double/half error, there are practical mean to extract the tempo and beat automatically. Such methods can be useful to automate the tempo and beat extraction in DJ mixing processes. In our proposal, we apply *BeatRoot*[2] as the method of extracting the beat in the pre-process to the database.

[2] http://www.elec.qmul.ac.uk/people/simond/beatroot/

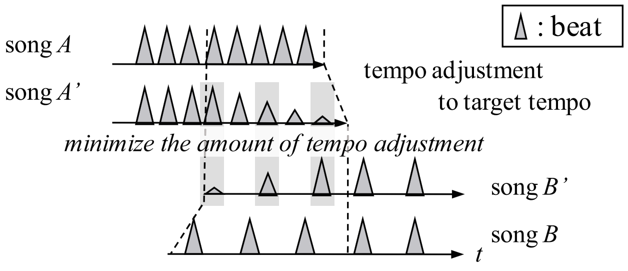


Figure 4. Conceptual image of dual tempo adjustment

5.2 Music information retrieval

In this section, we describe the method of automating the DJ processes of selecting the songs to be mixed. As mentioned in Section 1, there are many research efforts in music information retrieval/recommendation. Although these are not specialized to DJ mix playing, these have achieved highly accurate retrieval/recommendations. Hence these are practical ways of substituting and selecting the song manually. In this system, we apply the content-based MIR technique [2], which can retrieve songs from the database by means of content-based similarity to the users' query.

5.3 Proposed DJ mixing

5.3.1 Optimal tempo adjustment coefficient computation

In order to automatically generate a smooth *StS* transition, we propose a unique tempo adjustment technique. Our proposal computes the optimal tempo adjustment coefficients, hereafter described as *OTAC*, which expresses the factors of tempo adjustment for the songs to be consecutively played, thus is capable of automatically generating smooth *StS* transitions for any given combination of songs. Namely, two *OTAC*s are computed and optimized for each song in the combination. As previously mentioned, the naive tempo adjustment approach may result in user discomfort, especially under conditions where the tempo of song *A* (T_A) and song *B* (T_B) are significantly different, which causes the tempo adjustment factor to be extremely high.

In order to solve this problem, the proposed method considers the individual position of beats in the two songs to compute the *OTAC*s, which will hereafter be denoted as f_{opt}. Figure 4 shows the conceptual image of proposed DJ mixing. We focus on the position of beats in the two songs, and it is clear that the beats of the two songs can match the smaller factors of tempo adjustment compared to naive DJ mixing. The proposed method computes *OTAC*s by utilizing the double/half characteristics to reduce the score for user discomfort.

The following describes the computational procedure for *OTAC*s, which expresses the factors of optimal tempo adjustment of the two target songs. In this procedure, we reduce the amount of tempo adjustment and user discomfort in a *StS* transition by dual tempo adjustment, for example, song *A* with a 5% speedup factor and song *B* with a 5% slowdown factor, instead of song *A* with a 10% speedup factor and song *B* untouched. In the following explanation, song *A* is defined as the target song to compute *OTAC*s.

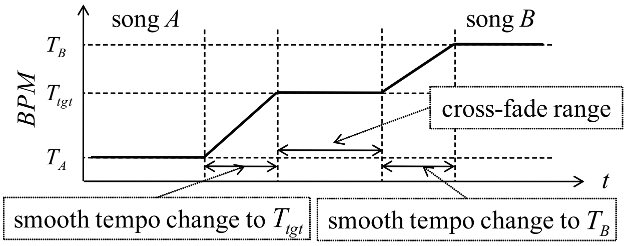


Figure 5. Shifts of tempi of target songs in *StS* transition.

First, a candidate set of adjusted T_A is computed using the following Equation:

$$T'_A = 2^C \times T_A \quad (2)$$

where $C = \{-2, -1, 0, 1, 2\}$. From the set of T'_A, we select the result which is closest to T_B. This is equivalent to determining $C_{opt} = \text{argmin}(|T'_A - T_B|)$.

Next, parameter b_{opt} is computed with the following Equation:

$$b_{opt} = 2^{C_{opt}} \times T_A \quad (3)$$

In Eq.(3), multiple values of b_{opt} can be computed in certain combinations of T_A and T_B. In such cases, the value b_{opt}, which results in a smaller $|C_{opt}|$, is selected. For example, given tempo combination as $(T_A, T_B) = (50, 75)$, possible solutions of Eq.(3) are $b_{opt} = 50, 100$. In this case, $b_{opt} = 50$ is selected as the final parameter.

The target tempo T_{tgt}, which the adjustment of the tempi of songs *A* and *B* will match, is computed with the following equation:

$$T_{tgt} = \frac{(a-b)T_{low} + \sqrt{(a-b)^2 T_{low}^2 + 4abT_{high}T_{low}}}{2a} \quad (4)$$

where T_{high} denotes the tempo of the song with a higher tempo, and T_{low} denotes the lower in b_{opt} and T_B. T_{tgt} is designed to divided the score based on Eq.(1) equally between the two songs, *i.e.*, T_{tgt} is computed in order to satisfy that the L_{dc} of speedup and slowdown is equal. Figure 5 shows the shifts in the tempi of target songs in the transition, which is the case where the tempo of song *A* is lower than song *B*. These shifts are optimized for reducing the score of user discomfort based on Eq.(1).

Finally, the *OTAC*s f_{optA}, f_{optB} are computed based on b_{opt}.

$$f_{optA} = \frac{T_{tgt}}{b_{opt}}, \quad f_{optB} = \frac{T_{tgt}}{T_B} \quad (5)$$

The proposed method is capable of computing the factors of optimal tempo adjustment for any combination of two songs. For instance, where the tempi of songs *A* and *B* are 60 and 120 BPM, the result of the computed *OTAC*s is $f_{optA} = f_{optB} = 1$, which is equal to the ideal rate for preserving the overall acoustic quality of the DJ mix result. It is also notable that the proposed method is capable of applying the DJ mix regardless of the existence of double/half tempo estimation errors, since the effect of such errors is disregarded during the *OTAC* computational procedure.

5.3.2 Beat adjustment and cross-fade playing

Next, we explain the procedure to generate the *StS* transition of the mixed sound. This procedure is necessary to reduce the discomfortness of the mixed sound, which assume to occur when the strong beats of a song are adjusted to the weak beat of the other song during the cross-fade range. In this procedure, we utilize the power of the beats in the cross-fade sections, to avoid the mismatching of strong and weak beats in the two songs to be mixed.

In order to generate the *StS* transition that matches the strong beats precisely, our method computes the score for the cross-correlation of the beats of target songs within the range of the cross-fade. When the powers of beats within the range of the cross-fade of songs *A*, *B* are described as Pow_A and Pow_B. The following describes the power of n-th beat as $Pow_A(n)$ and $Pow_B(n)$. The score between the songs *A*, *B* is described as Equation (6):

$$score(\tau) = \frac{\sum_{k=1}^{\tau}(Pow_A(N_A - k + 1)Pow_B(k))}{\tau} \quad (6)$$

where τ denotes the number of beats within the range of the cross-fade and N_A denotes the number of beats of song *A* as the former song in the mixed sound. Specifically, the beats of song *A* are matched to the beats of song *B* when $\tau_{max} = \mathrm{argmax}_\tau(score(\tau))$ is satisfied. Pows are computed by the power located near the beat ($\pm 50ms$). The powers of the spectrogram are computed by the FFT of the audio signal low-pass filtered (20th order FIR, cutoff freq. 1500Hz). Finally, cross-fade is applied to the overlapped range based on the highest score computed by τ_{max}.

6. EXPERIMENT

In this section, we will describe the experiment to subjectively evaluate our system and the proposed DJ mixing method. The objective of this experiment is to evaluate the effectiveness of the proposal.

In order to conduct this evaluation, two sets of DJ mixed sounds are generated; one by naive DJ mixing, and the other by the proposed method. The experiment is evaluated in a subjective manner. Namely, subjects of the experiment are to listen to the mixed sounds and provide preference ratings for each sample. Details of the experiment are described as follows.

6.1 Data

Experimental data consist of 1434 songs, which are collected from *Jamendo* [3] , a web site which distributes music licensed by Creative Commons. The source audio used for the experiments is extracted from the songs in the data collection. The length of each source is 30 seconds including the chorus. Note that, for all source audio, meta-information, such as the position of each beat, and tempo (BPM) are applied by *BeatRoot*.

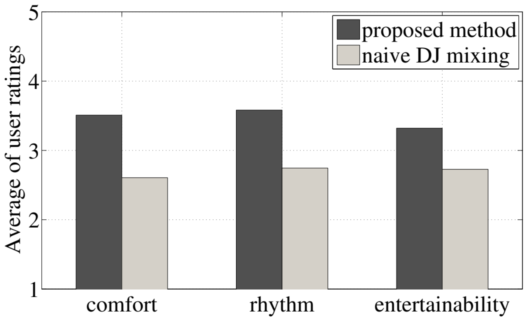


Figure 6. Average of user ratings in proposal and naive DJ mixing.

6.2 Experimental method

6.2.1 DJ mixed sound generation

The mixed sound files are generated by applying one of the previously described methods using five selected source audio extracted by MIR system [2] as the target songs. In total, six mixed sounds are generated by naive DJ mixing and the proposal, respectively. For the methods that utilize tempo adjustment, we have added interval periods to gradually change the tempi from/to the original to/from the target tempo, as shown in Fig.5. This interval period, which is fixed as 5 seconds for all mixed songs, is inserted in order to avoid abrupt changes in tempo, which is obviously uncomfortable. The period in which *CFP* is conducted begins immediately after the 5 second interval. For tempo adjustment, we use the *SoundTouch Library*.

6.2.2 Subjects and evaluation measures

A total of 27 subjects participated in the experiment. Each subject listened to all of the generated DJ mixed sounds and were asked to provide subjective ratings in five ranks for all sounds. In total, 165 ratings were collected on naive DJ mixing and the proposed method, respectively. Evaluation measures consist of the following three items: "*comfort*": the level of listener comfort during *StS* transition (1: discomfort – 5: comfort), "*rhythm*": the smoothness of the rhythm through the sound (1: bad – 5: good), and "*entertainability*": the overall preference rating (1: bad – 5: good).

6.3 Results

Average of user ratings in proposed method and naive DJ mixing are shown in Figure 6. It is clear from this figure that the proposed method was given a higher rating for all evaluation items compared to the conventional method, proving the overall effectiveness of the proposed method. According to the result of paired t-test, there are statistically-significant differences ($p < 0.001$).

[3] http://www.jamendo.com/

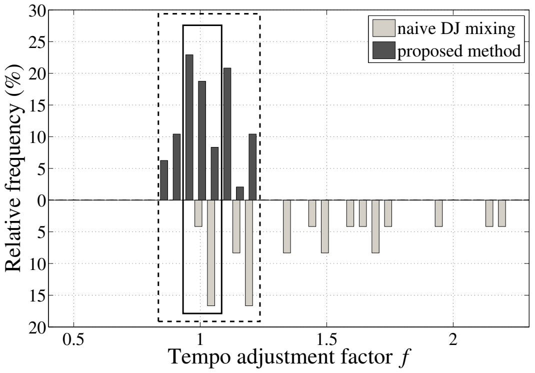


Figure 7. Histograms of the relative frequency of factors in *StS* transitions of proposal and naive DJ mixing.

Figure 7 shows the histograms of the relative frequency of factors in each of the *StS* transitions in each mixed sound. In this figure, stricter and average applicable ranges described in Section 4.2 are plotted as solid and dashed lines.

It is clear from this figure that the proposed method can keep factors near the original tempo compared to naive DJ mixing in a transition. The proposed method is able to deal with the difference in tempi between the former and latter songs. Furthermore, it is notable that the proposed method can almost satisfy the stricter applicable range and perfectly satisfy the average applicable range. Specifically, the percentage of factors inside the stricter range of the proposed method is 50.00% and inside the average range is 100.00%.

For further analysis, we investigated the averages of user ratings for each mixed sound. There were some cases that although L_{dc} of the proposed method were lower than naive DJ mixing, the score of user ratings was lower than naive DJ mixing. These cases tended to be adjusted strong beats and weak beats. In this case, user ratings of the proposed method about the evaluation item *RH* is lower than that of naive DJ mixing, which is able to adjusted appropriately. Furthermore, the correlation between *CF* and *RH* has a strong positive-correlation to each other. These results indicate that appropriate beat adjustment is one of the important factors. Generation of a smooth *StS* transition in the aspect of *RH* is essential to achieving a high quality DJ mixing method.

7. CONCLUSIONS

In this paper, we proposed an automatic DJ mixing method with optimal tempo adjustment with a function to measure user discomfort, described a prototype for a fully automatic DJ mixing system. The measurement function is defined by a subjective experiment, and our proposed method is designed to optimize the score of the function. In order to generate a smooth song-to-song transition, this paper proposes an optimal tempo adjustment based on the computation of optimal tempo adjustment coefficient. Furthermore, the proposed DJ mixing method is designed to preserve user comfort. The proposed DJ mixing is capable of generating a smooth song-to-song transition for any given combination of songs that includes double or half tempo errors. The advantages of the proposed method were proved by comparing the subjective evaluations of the samples generated by the proposed and conventional methods.

However, it is also obvious that tempo is just one of many elements in music that affect user preferences. For example, some combinations of source songs were unacceptable to subjects in the experiments, regardless of the DJ mixing method implemented to generate the sample audio. Therefore, we plan to further pursue research to develop a way to effectively apply the measurement function and a fully automatic music playing method, including the extraction and utilization of features other than tempo and beat position.

8. REFERENCES

[1] S. Pauws and B. Eggen: "PATS: Realization and user evaluation of an automatic playlist generator," *Proc. ISMIR 2002*, pp. 222-230, 2002.

[2] K. Hoashi, *et al.* : "Personalization of User Profiles For Content-based Music Retrieval Based on Relevance Feedback," *Proc. ACM Multimedia 2003*, pp.110-119, 2003.

[3] K. Yoshii, *et al.* : " Improving Efficiency and Scalability of Model-based Music Recommender System Based on Incremental Training," *Proc. of ISMIR* , pp.89-94, Vienna, Sep. 2007.

[4] S. Basu: "Mixing with Mozart," *Proc. of ICMC 2004*

[5] A. Inoue, *et al.* : "Playback and Distribution Methods for Digital Audio Players" *IPSJ SIG Notes 2006(9)* pp.133-138 (*in japanese*)

[6] M. Alonso, *et al.* : " Tempo and beat estimation of musical signals," *Proc. ISMIR 2004*, pp.158-163, 2004.

[7] S. Dixon: "Automatic extraction of tempo and beat from expressive performances," *J.New Music Res.*, Vol.30, No.1, pp.39-58, 2001.

[8] E. Scheirer: "Tempo and beat analysis of acoustic musical signals," *J. Acoust. Soc. Amer.*, Vol.103, No.1, pp.588-601, 1998.

[9] F. Gouyon, *et al.* : "An experimental comparison of audio tempo induction algorithms," *IEEE Trans. Audio, Speech, and Lang. Process.*, IEEE Transactions on. Sept., pp.1832-1844, 2006.

[10] S. Roucos and A. M. Wilgus: "High quality time-scale modification for speech," *IEEE ICASSP*, pp.493-496, 1985.

[11] M. Goto, *et al.* : "RWC Music Database: Popular, Classical, and Jazz Music Databases," *Proc. of ISMIR 2002*, pp.287-288, October 2002.