

# A STRING ALIGNMENT METHOD TO FIND TIME CUE POINTS TO MIX MULTI-TRACK AUDIO FILES

Rui Guo

Nusic  
rui@nusic.com.cn

Adam Place

Nusic  
adam@nusic.com.cn

Joseph Horton

Nusic  
jo@nusic.com.cn

## ABSTRACT

Multi-track Audio-encoding formats such as mogg and poly-wav allow multiple track sections to exist in one file. This separation of music can reduce the complexity of the algorithms necessary to analyse the music in order to prepare it for mixing. In this paper, a method is proposed to determine the time cue points in a song to mix different parts of songs. The chord progression of each song is represented by a string with a corpus from 1 to 7, and that string is divided into several parts based on similarity. Then the string global alignment method from bioinformatics is applied on the chord string in each part to match the chord string from another song in the proper section. This method has value for DJing, music production; especially remixing, and music games.

## 1. INTRODUCTION

The aim of mixing is to make seamless transitions from one section of music to the next, so that the listener can experience continuous music without interruption. These transitions can be gradual fades between songs over a period of time or abrupt transitions, as long as the flow of the music is not interrupted this is acceptable from the perspective of the listener.

In this paper a method is proposed to find suitable time cue points for transitions based on the bass chord progression. It adjusts the key of the songs to be mixed and also the tempo with information pulled from the drumbeats. It makes use of a multitrack audio format and a string alignment method is applied to find the time cue points to synchronise songs for mixing.

## 2. PREVIOUS WORK

A number of methods have been proposed to mix songs. Ishizaki et al. [1] adjusted tempo to mix songs at any point. Hirai et al. [2] proposed a DJ system based on tempo and chroma features. When a song is not mixed in a multi-track format, there is a variety of sound in addition to the melodic content, and the chroma may contain too



© Rui Guo, Adam Place, Joseph Horton. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).  
**Attribution:** Rui Guo, Adam Place, Joseph Horton. "A string alignment method to find time cue points to mix music based on multi-track audio files", Extended Abstracts for the Late-Breaking Demo Session of the 18th International Society for Music Information Retrieval Conference, Suzhou, China, 2017.

much information to identify the musical characteristics accurately. To our knowledge, no existing system makes use of a multi-track audio encoding format and incorporates bass chord progression, music structure analysis, tempo and pitch together to build a song mixing system.

## 3. METHOD

### 3.1 Bass note extraction

Our method is based on chord progression similarity. In the STEM format [3], the bass track is separate from other tracks (i.e. Vocal, Melody, Drums) and bass progressions are generally monophonic, which makes it easier to identify the note which is playing. The beat time is estimated from the drum track by functions in the librosa [4] library, and the bass note at each beat time is extracted. Constant-Q transform is applied to get the chromagram of sound at the onset of the beat (samples in 0.3s), and the note with maximum energy in that chromagram is found. In music, the bass note often follows the song's chord progression, e.g. B note following C note would usually represent the G major chord, not B chord. For that reason, the bass alphabet from 1 to 12 is reduced to only 7, representing C to B. The mapping between the redundant note and target note is C# -> A, F# -> D, D# -> B, G# -> E, A# -> G, as the redundant notes are contained in the target note chord. To ease the comparison of bass notes, the bass notes of each song are also adjusted to the key of that song. If a song is in key of D major, the note sequence will be subtracted by 2 to change that sequence to the same representation as if the song is in key of C major. Pitch class distribution is used to find the key of each song. The idea above is depicted in Figure 1.

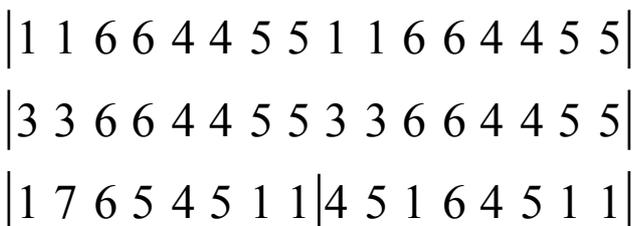
Original bass:	D	D	D#	B	G	G#	A	A
Alphabet reduction:	D	D	B	B	G	E	A	A
Key change:	C	C	A	A	F	D	G	G
Numeric notation:	1	1	6	6	4	2	5	5

**Figure 1.** Bass note extraction process. The bass notes map to only 7 notes from C to B, and the song key is used to adjust the numeric notation to easily compare with songs in a different key.

### 3.2 Bass sequence separation

If two bass sequences are compared by the sequence global alignment algorithm [5] directly, it is not likely to find the ideal alignment. In the previous work Bello [6], string global alignment is used to find cover songs. Be-

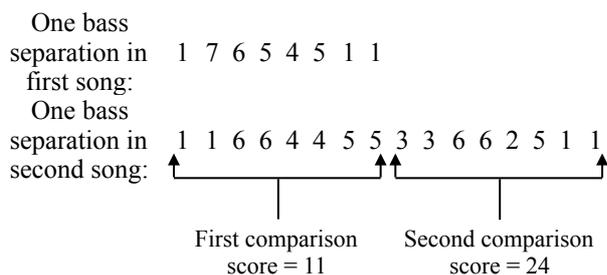
cause the chord progressions between cover songs are nearly the same as the original song, that method is feasible. However, in this setting multiple songs with different chord progressions will be compared together. According to our tests, it is more likely sequences will align incorrectly if the global alignment algorithm is used on two whole bass strings. One possible reason for a false score higher than the true meaningful alignment is greater or lesser number of notes may be inserted or deleted in the string alignment. To alleviate this problem, we have segmented the strings into several sections based on similarity between each adjacent section. The idea is in Figure 2.



**Figure 2.** Bass sequence separation result. A sliding window is used to compare bass strings in adjacent windows. If the string is not the same in the adjacent window, the bass string will be separated.

### 3.3 Bass alignment

With the help of the string segmentation, we can compare separate segments of bass between two different songs. The string global alignment is applied and a large penalty score is used to eliminate the gap in the alignment because a gap in chord will cause dissonance of music. A score matrix is designed to incorporate musical knowledge into the string alignment process. If two strings are compared with length 8 and 16 respectively, the shorter string will be compared with first 8 characters in the string of length 16, then the last 8 characters. As the bass should be compared phrase by phrase, and because the use of string separation and elimination of gaps in string comparison represents a more meaningful musical comparison.



**Figure 3.** Bass alignment process. One bass sequence in the first song is compared with a bass separation from the second song. The higher alignment score of the second comparison indicates a better alignment.

After the best alignment place is found, the time point at that position is traced back, and possible pitch and tempo changes are implemented to make the transition seamless. For each song pair in the song library, the pro-

cedure above is applied to find musically appealing cue points. The bass alignment idea is depicted in Figure 3.

## 4. RESULT

Our method is applied to a number of existing songs manipulated and remixed into a multitrack format with clear bass progressions. On average the beat time and onset time has 0.02s difference. The bass sequence separation works, but if the phrase structure is irregular (rests for one more beat in some bars), the method will need some further refinement. The bass alignment matches the best alignment based on the alignment score, and refinements can also take the song structure knowledge (verse, chorus etc) into consideration. In the track *You Exist in My Song* from Qu Wanting and song *Fairytale* from Michael Wong, the chorus part can be mixed seamlessly. Similar results were found with other songs trialled.

## 5. CONCLUSION

In this paper, we devise a string alignment method to find the time cue points to mix different songs. Based on the bass track in isolation, chord progression is found according to the beat time and the corpus can be reduced. From these simple tests we can see that by separating out and analysing different elements independently we are able to gain a level of clarity which is impossible with a single channel format. In the future we can exploit the value of multi-track audio by devising methods to analyse and integrate information from each separate track.

## 6. REFERENCES

- [1] H. Ishizaki, K. Hoashi, and Y. Takishima, "Full-Automatic DJ Mixing System with Optimal Tempo Adjustment based on Measurement Function of User Discomfort.," *Proceedings of the International Symposium on Music Information Retrieval*, pp. 135–140, 2009
- [2] T. Hirai, H. Doi, and S. Morishima, "MusicMixer: computer-aided DJ system based on an automatic song mixing," *Proceedings of the 12th International Conference on Advances in Computer Entertainment Technology*, p. 41, 2015
- [3] "STEMS," [Online]. Available: <http://www.stems-music.com/>. [Accessed: 16-Oct-2017].
- [4] B. McFee *et al.*, "librosa: Audio and music signal analysis in python," *Proceedings of the 14th python in science conference*, pp. 18–25, 2015.
- [5] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *Journal of molecular biology*, pp. 443–453, 1970.
- [6] J. P. Bello, "Audio-Based Cover Song Retrieval Using Approximate Chord Sequences: Testing Shifts, Gaps, Swaps and Beats.," *Proceedings of the International Symposium on Music Information Retrieval*, pp. 239–244, 2007