DiscStitch

Towards audio-to-audio alignment with robustness to playback speed variabilities

Joren Six - IPEM, Ghent University

Introduction

This work presents an **audio-to-audio alignment algorithm** with different tradeoffs than traditional DTW-based approaches.
DiscStitch is designed to align and mix digitized audio originating from **acetate discs**. These discs are recorded at variable speed which requires the audio alignment algorithm to be robust aginst audio speed modifications.
The ideas behind DiscStitch are presented together with two implementations and a preliminary evaluation.

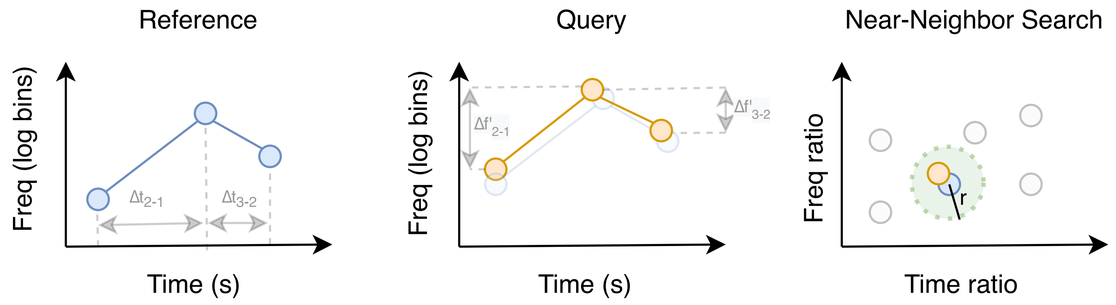


Fig: Spectral peaks are extracted from reference audio and combined in lists of three peaks (blue). A slightly sped up query (orange) has a slightly higher frequency and a slightly faster succession of peaks. The ratios between the time deltas $\Delta t_{2-1} / \Delta t_{2-1} = \Delta t'_{2-1} / \Delta t'_{2-1}$ and frequency deltas $\Delta f_{2-1} / \Delta f_{2-1} = \Delta f'_{2-1} / \Delta f'_{2-1}$ stay nearly identical for query and reference. If each list of three peaks is mapped to a 2D plane a near-neighbour search with a small circular radius r yields spectral peak lists originating from duplicate audio.

Why DiscStitch?

The archives of the national broadcasting company of Belgium contains about ten thousand digitized acetate disc sides. This type of medium was popular between about 1930-1960 and had a short recording duration. Long recordings were done using overlapping discs.

To make the audio - spread over several discs - fit for reuse the 'album' of discs needs to be identified, the audio needs to be aligned and mixed. DiscStitch is designed to facilitate tis reuse..

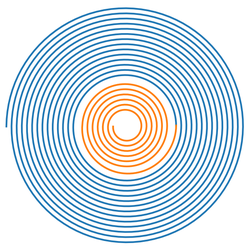


Fig: Acetate or laquer disc with a physicial marker identifying audio overlap.

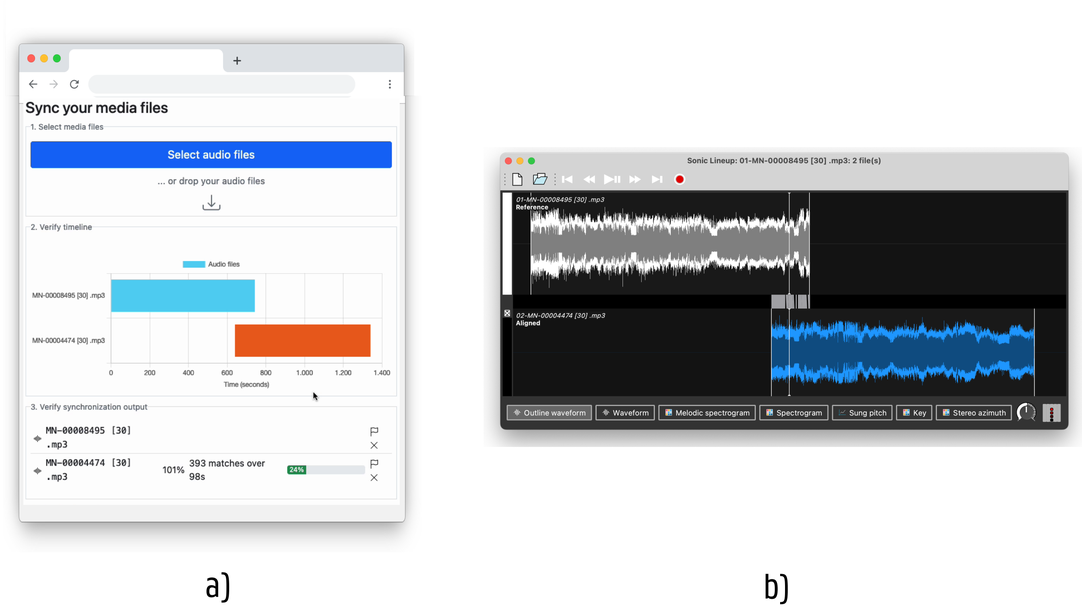
Fig: Overlapping discs need to be identified from a large set of disks. The overlapping parts (in orange) need to be precisely aligned. A mixing step reassembles the original recorded event. Identification, alignment and mixing is problematic due to speed differences and speed variabilities in overlapping discs.

Audio-To-Audio alignment

The DiscStitch audio-to-audio alignment algorithm works by extracting peaks in a spectral representation with a max filter. Each peak has a time and frequency component: (t,f). These lists of peaks reduces the information drastically. Now the audio alignment problem is reduced to aligning these lists of peaks.

Alignment of peaks is difficult due to variable speed differences. To cope with this we bundle neighbouring peaks in sets of three and calculate ratios as in the figure above right. Next the ratio's are mapped to a 2D pane the with the time-ratio in the horizontal and frequency-ratio in the vertical axis. Points close to each other in this pane have nearly the same ratios and might originate from spectral triplets with the same shape which may mean that they originate from similar audio.

Implementations



a) b)

Fig: Two implementations of DiscStitch. a) is a browser based implementation in Javascript/WASM b) is a Java based implementation. Sonic Lineup is used to visualize audio-to-audio alignment results

Preliminary evaluation & Conclusion

The Java version of DiscStitch audio-to-audio alignment algorithm is evaluated by chopping up a long audio recording and stitching it back together. Either a part is speedchanged or not. Afterwards the remixed audio duration is checked.

For the straight remix the time difference is small (mean=0.1ms, stdev=1.2ms, N=80). The speed changed version time differences are larger (mean=1.3ms, stdev=237.2ms, N=80)

Discstitch offers a promising way to do audio-to-audio alignment.

Further Reading

[1] Six, J., & Leman, M. (2014). Panako: a scalable acoustic fingerprinting system handling time-scale and pitch modification. ISMIR-2014)
[2] Wang, A. L. C., & Culbert, D. (2009). U.S. Patent No. 7,627,477.
[3] Velasco, G. A., Holighaus, N., Dörfler, M., & Grill, T. (2011). Constructing an invertible constant-Q transform with non-stationary Gabor frames. Proceedings of DAFX11,.

UNIVERSITEIT GENT

More Info

joren.six@ugent.be

https://github.com/JorenSix/SyncSink.wasm

https://github.com/JorenSix/DiscStitch