AMPLAB-freesound Audio Mosaicing

Dougal Shakespeare
dougalian.shakespeare01@estudiant.upf.edu

March 2020

1 Introduction

In this report, we present the results of audio mosaicing experiments using the FreeSound API [1] to query source files used for reconstruction. Drawing from the early works Robert Silvers' Photomosaics [3], in which thousands of (often) semantically related photographs are assembled to form a whole, audio mosaicing is the creative process of recreating target audio from a range of chunks (frames) of source audio clips [2]. In the music domain, the concept of semantic similarity translastes to the union of audio features extracted from target and source audio clips. Commonly, target audio clips are first analysed to extract a range of audio features that can be used to subsequently retrieve source chunks which are similar to the target, thus maintaining some coherence and similarity to the original audio in the reconstructed version. In audio mosaicing, the possibilities of creative endeavors are endless such that one has almost infinite choice of source and target audios and subsequently features to extract. Such choices made are highly dependent on the overall end state the user wishes to achieve, one of replication or abstraction. In this report, we present two applications of audio mosaicing such that we: (1) Explore rhythmic qualities to recreate the infamous amen break-beat. (2)Explore tonal melodic reconstruction with a piano extract with the audio clips from JP-X Roland synthesiser drawing influence in final sound construction from the works of artists such as Aphex Twin and Boards of Canada. The report is structured as follows: Section 2, presents the methodology deployed for rhythmic and melodic target reconstruction. Section 3. Presents results and conclusion alongside a discussion of the success of each approach applied. Code repositiores are made openly available to replicate all results detailed [1].

2 Methodology

All features extracted from source and target audio clips are performed using the Python interface for Essentia, an open source library and tools for audio

[1] https://github.com/dshakes90/AMPLAB-Freesound-Audio-Mosaicing

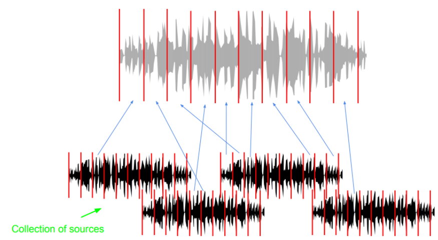


Figure 1: An example of target audio recreation using source chunks

analysis, description and synthesis [2].

2.1 Target and Source Audio Selection

In this work, we recreate two target clips, a chopped amen break-beat [3] commonly heard in the music of drum and bass and jungle and a short melodic piano extract [4]. Both samples are retrieved from FreeSound. With respect to the amen break, this was selected with the idea of trying to recreate a deconstructed 90s rave tune. To elaborate, the idea was to replicate some of sounds which one might hear outside a rave/club whilst a DJ is playing a set, (i.e. strong low sub bass, occasional hi frequency vibrations from the sound system vibrating near by objects ect). I believe such an idea to be interesting as the environment (synthesised in this case) is not just reacting to the DJ / soundsystem (i.e. the target audio) in our fictional club, but instead going further to react and emulate sounds created. To try to achieve this idea I select source audio clips including the keywords, 'bass', 'sub', '303', 'clap'. Additional environmental sounds were also experimented with (such as smashing glass, people smoking / talking ect.) however this was not implemented in our final version released.

With respect to the melodic piano clip here, the goal was to recreate some of the droney, slightly dissonant sounds commonly heard in the works of artists such as Aphex Twin and Boards of Canada. In an abundance of Aphex Twin compositions (such as in the infamous 'Vordhosbn'), detuned pianos are commonly utilised to create eerie, confusing soundscapes for the listener. To achieve this effect, audio clips from the Roland JX-8P synthesizer as shown in Figure 2 are retrieved from Freesound as source clips. The idea for this creative endeavor was built upon the observation that not enough sounds would be retrieved to fully recreate all pitches the piano melody was based upon and thus, creative directions should be explored which utilise the potential for incoherence with the target whilst still maintaining relative pitch contours.

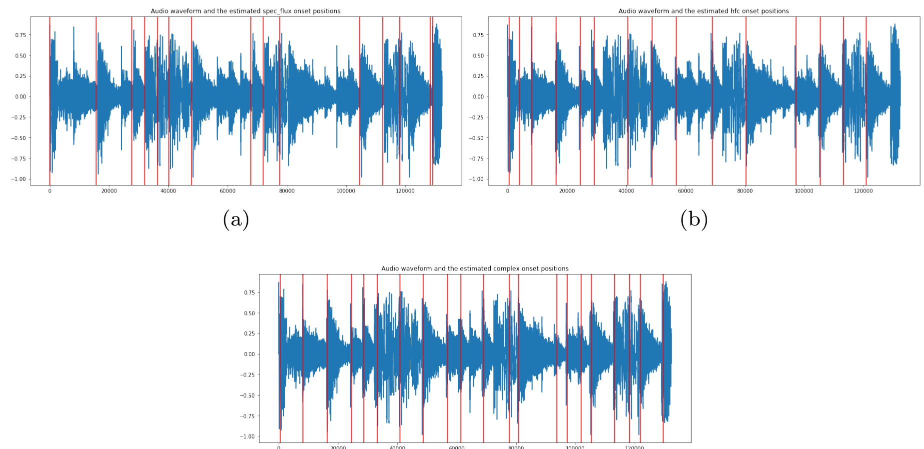
[2] https://essentia.upf.edu/index.html
[3] https://freesound.org/people/csum/sounds/194817/
[4] https://freesound.org/people/tkky/sounds/486472/

Figure 2: The infamous Roland JX-8P used in our piano target recreation

2.2 Temporal features - Onset Detection

Preliminary experiments performed initially focused on splitting target and audio clips based on beat positions using the Essentia function, BeatTrackerDegara. Whilst the results were shown to be reliable for both target clips, this approach was not included in our final implementation. Code for beat tracking is left in the Python notebook submitted allongside this report for the sake of completeness.

For the amen break beat sample considered, we observe that a substantial amount of audio events exist in between downward beats, (for example complex snare rolls, glitches ect). Hence, we observe that splitting the target audio clip by only beats before performing audio feature extraction will possibly result in a substantial amount of features missed in the target audio. Instead, we deploy a different methodology which splits the target audio clips using onset detection functions from Essentia, (to capture subsequent events commonly occurring between beats). To achieve this, we first compare 3 onset functions from Essentia: (1) SuperFluxExtractor (2) Complex based approach (3) HFC based approach. Plots for the first 3 seconds of onsets assigned to the amen break are displayed in Figure 2. From these results, we observe 3c the complex functionality to best capture onsets with the Amen break and thus, we use this algorithm. For the amen break, both the target and source audio are segmented using this methodology. For the piano sample, we observe that the complex onset detec-



(a) (b) (c)

Figure 3: Comparison of Essentia Onset Detection functions with the Amen Break

tion function identifies many onsets (when in reality onsets are sparse). Hence, in this context, we use the hfc onset detection algorithm to segment audio and target clips as we observe it gives better results. Full plots for this can be found in the Python notebooks provided alongside this report.

2.3 Melodic features - Pitch detection

For the piano sample retrieved from Freesound, we observe that the sample is tonal and contains much melodic content. Thus, to simply recreate the rhythmic qualities of the target audio through onset segmentation is not adequate to recreate a coherent version of the target audio which holds some resemblance to the original. To achieve this, we use the PitchMelodia and PitchContourSegmentation function from Essentia to retrieve a final MIDI note symbolising the pitch of the chunk of audio. This approach was not applied when recreating the amen-break as little tonal content is available for this sample.

2.4 Other Audio Features Extracted

Additional feature extractions applied include MFCCs and attack time to compute both textural and temporal aspects of the signals.

The zero crossing rate of audio to analyse the noisiness of the signal was also computed alongside the spectral centroid in the time domain however these were not included in our final implementation. After much experimentation and trial and error, it was deemed that MFCC's paired with attack time (also pitch information in the case of piano reconstruction) resulted in the most coherent target audio reconstructions.

3 Results and Conclusions

Recreation of the Target audio for the amen break-beat was successful. Rhythmic quality was maintained by applying segmentation via onset detection, improving the quality of results in comparison to the baseline code provided in the Python notebooks. With respect to the creative endeavor of this project, I also believe this to be a success. The reconstructed audio (although somewhat abstractly) resembles a club like context, resembling some of the acoustic properties commonly observed (such as low sub bass, objects rattling to produce hi end frequencies). In future works, we suggest for composers to continue to explore the idea of modeling sound through environment using field recordings as opposed to elements of the original sound performed in this work.

With regards to the recreation of the piano segment, I believe this to be also be a success but to less of a degree. Whilst the atmospheric, discordant sound commonly heard in the music of Boards of Canada is achieved to some extent, the audio mosaic downfall is the randomness of some of the pitches of the source samples played. We suggest for future works to include pitch information only if an average pitch confidence level is achieved for both source and target audio. In addition, we also suggest future works to apply filtering to the Pandas dataframe which stores audio chunk features which have been extracted to impose further selection criteria which could result in a more coherent target audio reconstruction.

References

[1] Frederic Font, Gerard Roma, and Xavier Serra. Freesound technical demo. pages 411–412, 10 2013.

[2] Adam T. Lindsay. Journal of New Music Research: Introduction. *Journal of New Music Research*, 35(1):1–2, 2006.

[3] Robert S. Silvers. Photomosaics : Putting Pictures in their Place by. 1996.