Concatenative Resynthesis for Extracting Bass Parts from Songs

Dzmitry Kasinets and Dr. Michael I Mandel, Department of Computer and Information Science, Brooklyn College of the City University of New York

Speech Lab @ BC-CUNY

Introduction

What is Concatenative Resynthesis?
• Concatenative Resynthesis is a new approach to source separation that is able to utilize a rich dictionary of speech signals. (Mandel et al. 2014).

What does Concatenative Resynthesis do?
• It is used to remove noise from speech recordings.
• This approach uses a deep neural network (DNN) to predict whether a noisy chunk of audio contains a given clean chunk. When used with a dictionary of clean speech 'units', this system is able to resynthesize clean speech from noisy observations (Mandel et al. 2014).

Why does Concatenative Resynthesis return high quality audio?
• The dictionary only consists of clean speech, and contains no noise. Therefore, the noisy observations are replaced with clean speech. As a result, the result has a very high audio quality with levels of noise close to those of original speech, i.e., absent. (Mandel et al. 2014).

Most importantly, the system can also be used for other tasks.

Objectives

• The goal of the current study is to use the Concatenative Resynthesis system for the purpose of extracting bass parts from songs.

In other words, we give Concatenative Resynthesis system a song as an input and receive the song's bass part as output.

Methods

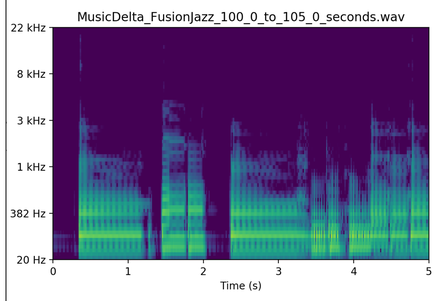
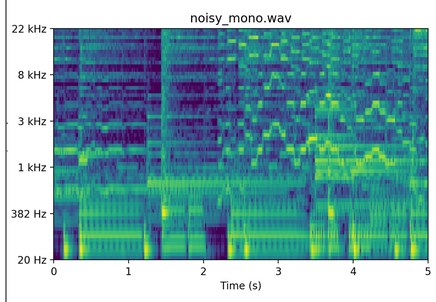
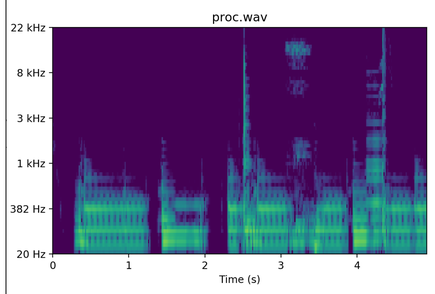
What materials do we use?
• For this task, we need a dataset of multitrack songs of many genres to train the model and evaluate the accuracy of the results.
• MedleyDB is a dataset of annotated, royalty-free multitrack recordings primarily developed to support a research on melody extraction. The dataset contains 122 songs, 108 of which include melody annotations (Bittner et al. 2014).

What is our Methodology?
• Each song has a stereo mix, and both dry and processed multitrack stems. This means that each song is provided as the final product, along with all of the recordings of individual instruments that were combined to make it.
• This lets us choose a test song with good bass and train the network on more songs with bass to resynthesize the test song's bass part.

Our experiment
• To create high quality resynthesized bass part, we tried new Gammatone Filterbank feature, and later created a custom one to better characterize bass at low frequencies.
• It calculates a spectrogram-like time frequency magnitude array based on an FFT-based approximation to gammatone subband filters.

• In addition, we used together different training datasets and configurations to select the combination that resynthesizes "the best sounding" bass part.

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1618061. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

	The bass guitar part of a song.
	All of the instruments in a song.
	The bass guitar part recreated by Concatenative Resynthesis system.

Results

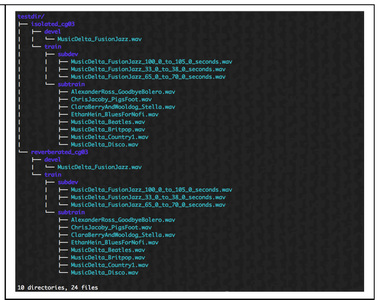
How do we analyse the results?
• We evaluate the results by comparing them to the actual bass recordings. We use Rhythm, Pitch, Noise Suppression and Quality as main criteria.

Rhythm – how well the system preserves the rhythm of the bass.
Pitch – how well it reproduces pitch, i.e., correct notes of the bass.
Noise Suppression – how well it suppress the noise, i.e., other instruments.
Overall quality – how well it sounds in general.

How well does it perform?
• Preliminary listening tests suggest that it is able to resynthesize bass with high audio quality, rhythm and pitch accuracy if it is trained on parts of itself in addition to different songs.
• It can resynthesize bass with good audio quality, rhythm and pitch accuracy if trained only on different songs.

Result using only other songs
• The system resynthesizes good quality bass part for songs when we choose "the best sounding" configuration, and train the system on a "nice" dataset that can have a few songs recorded by the same music group.

• The system produces similar good results for other songs when using the same configuration and training dataset.



File Configuration: 8 different songs

Result when adding parts of itself
• The system performs really well when it is trained on parts of the target song in addition to different songs.

• The system resynthesizes the high quality bass part when we create similar configurations and training datasets for other songs.

File Configuration: 8 songs + parts of itself

Conclusion

• Concatenative Resynthesis system can be used for extracting bass parts from songs. It is able to resynthesize bass with high audio quality, rhythm and pitch accuracy if it is trained on parts of itself in addition to different songs.

• It can resynthesize bass with good audio quality, rhythm and pitch accuracy if trained only on different songs.

Recommendations?
• Surveying multiple people can help better evaluate the Concatenative Resynthesis system, especially in comparison to other similar systems

Discussion
• This project is valuable because it has the potential to create much higher quality separations than current approaches, which could enable new musical applications of this promising source separation technique that has so far been applied only to speech.

References

Bittner, Rachel, Justin Salamon, Mike Tierney, Matthias Mauch, Chris Cannam and Juan Bello. "MedleyDB: A Multitrack Dataset for Annotation-Intensive MIR Research." Proceedings of the International Society for Music Information Retrieval (ISMIR), Taipei, Taiwan, 2014

Mandel, Michael, Young Suk Cho and Yuxuan Wang, "Learning a Concatenative Resynthesis System for Noise Suppression," 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Atlanta, GA, 2014, pp. 582-586.

Heeris, Jason. (2013). Gammatone Filterbank Toolkit.