EVOLVING NEURAL NETWORKS FOR CROSS-ADAPTIVE AUDIO EFFECTS

Iver Jordal, Øyvind Brandtsegg and Gunnar Tufte

Norwegian University of Science and Technology
iver56@hotmail.com, oyvind.brandtsegg@ntnu.no, gunnart@idi.ntnu.no

ABSTRACT

In cross-adaptive audio effects, effect parameters are dynamically informed by features of sounds other than the sound that is processed by the effect. Cross-adaptive audio effects can be applied in a wide range of research fields, including live music performance and audio mastering. Toward a toolkit for signal interaction we present a system that can exploit dynamic audio parameters of signal sources to control effect parameters, and thereby dynamically process audio. The vast number of possible combinations of parameters makes empirical experimentation tedious and unfeasible for live performance. Artificial Intelligence (AI) methods, herein Genetic Algorithms (GAs) and Artificial Neural Networks (ANNs), are exploited to find parameters for useful signal interactions in cross-adaptive audio effects. An experimental approach is taken to combine GAs and ANNs to control the audio effect parameters of one sound (input) by extracting audio features from another audio source (target) as to process the input to sound as close to the target as possible. Such results are shown to be feasible by using evolved ANNs.

1. INTRODUCTION

The problem of extracting audio features for control of effect parameters is here defined to two problem domains; the extraction and selection of audio/signal features and the mapping of such features to control parameters for audio effects. That is a selection of features from the source audio stream, mapping process to control the effects that can manipulate the target audio stream toward a signal that include sought audio properties. The system presented is part of the development of a toolkit for experimentation with signal interaction [1]

To handle the mapping of features to effect parameters an evolved ANN is used. The chosen neural network is based on NeuroEvolution of Augmenting Topologies (NEAT) [2]. The architecture of the ANN in a NEAT approach allow evolution, e.g. a Genetic Algorithm [3], to define weights and topology of the network. Further, the training of the network is based on performance, i.e. fitness, instead of supervised learning, e.g. backpropagation [4].

The set of audio features for extraction is predefined, i.e. the evolved network exploits favorable features within the available feature set. The audio effect is also predefined.

To explore the possibility of exploiting AI methods toward cross-adaptive audio effects, a system for conducting and evaluating signal interaction experiments has been implemented. As a test case the system is set to make one sound similar to another by applying audio effects controlled by extracted audio features.

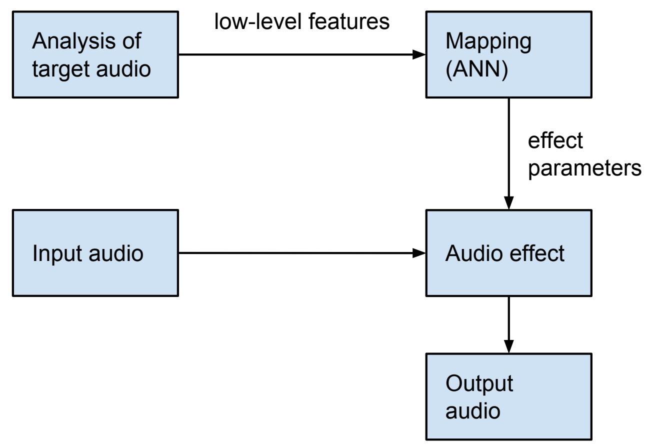

Figure 1: Cross-adaptive audio effect process with two audio streams: input audio and target audio

Figure 1 roughly illustrates the system setup. The low-level features are extracted audio features from the target sound. The features are mapped by the evolved ANN to effect parameters that are used to manipulate the input audio. The output audio is the result of effects applied to the input audio.

The system described produces large amounts of data in various forms, including audio features, effect parameters and output sounds. To handle the data for evaluation, an interactive visualization tool was made to make it easier to evaluate results and understand what the system is doing. The system and the visualization tool are open source and available on GitHub[1].

2. EXPERIMENTS AND RESULTS

The presented experiment's target goal was to make white noise sound like a drum loop with snare drum and bass drum. The selected and applied audio effect was distortion and resonant low-pass filter. The audio features used were spectral centroid and the first two Mel-Frequency Cepstral Coefficients (MFCC). Audio features were calculated for each

[1] https://github.com/iver56/cross-adaptive-audio

frame of 512 samples. The set of features in one frame is the feature vector for that frame. The fitness function used in the experiment was

$$1/(1+e) \tag{1}$$

where e is the average euclidean distance between feature vectors of the target sound and the corresponding feature vectors of the output sound. This means that fitness values are between 0 and 1. The population size was 20, the mutation rate was 0.25 and the crossover rate was 0.75. The experiment was run 20 times, with different Pseudo-Random Number Generator seed for each run. The fitness values were aggregated and are shown in Figure 2. Some of the sounds produced have been published in the project's blog[2].

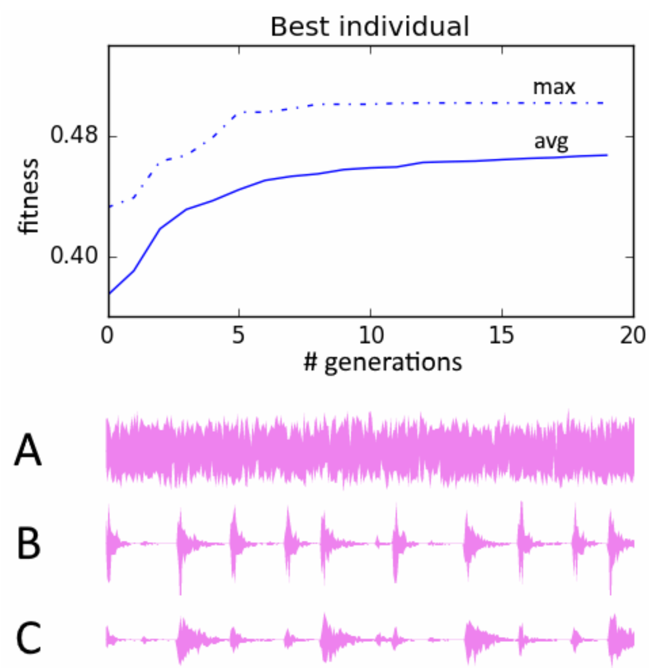

Figure 2: This plot shows the fitness (from expression 1) of the best individual in each generation. Below are waveforms of A) the input sound (white noise), B) the target sound (drum loop) and C) an output sound

3. FUTURE WORK

Future work may address the following:

- Conduct experiments with other audio effects. Let a genetic algorithm decide which audio effects to apply.
- Develop methods for dealing with long and complex sounds, such as music with many instruments.
- Make the system work on live audio streams, with pre-trained neural networks.
- Explore possible applications, such as mixing/mastering and novel sound effects.
- Experiment with other audio features. Use machine learning techniques to create high-level features.
- Implement the system on a Field-programmable gate array (FPGA) or other parallel computing environments for the sake of decreasing computational time. This may make it possible to train useful neural networks in seconds, making the system more flexible in live performances

4. CONCLUSION

Output sounds from the system demonstrate that it is possible to make white noise sound like a drum loop by applying a cross-adaptive audio effect. It also proves that NEAT can train a neural network to work as a musically interesting mapping from a set of audio features to a set of audio effect parameters. A comprehensive toolkit has been developed. It includes an interactive visualization tool that makes it easier to evaluate results and understand the neuroevolution process.

5. REFERENCES

[1] Ø. Brandtsegg, "A toolkit for experimentation with signal interaction," in *Proceedings of the 18th International Conference on Digital Audio Effects (DAFx-15)*, 2015, pp. 42–48.

[2] K. O. Stanley and R. Miikkulainen, "Evolving neural networks through augmenting topologies," *Evol. Comput.*, vol. 10, no. 2, pp. 99–127, Jun. 2002. http://dx.doi.org/10.1162/106365602320169811

[3] D. Goldberg, *GENETIC ALGORITHMS in search optimization & machine learning*. Addison Wesley, 1989.

[4] P. J. Werbos, *Applications of advances in nonlinear sensitivity analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1982, pp. 762–770. ISBN 978-3-540-39459-4. http://dx.doi.org/10.1007/BFb0006203

[2]http://crossadaptive.hf.ntnu.no/index.php/2016/06/27/evolving-neural-networks-for-cross-adaptive-audio-effects/