End of Funding Report: DiffSynth Project

Project Overview

The DiffSynth project, supported by a $5000 Agile Seed Fund from CIRMMT, aimed to explore the creative potential of generative models of artificial intelligence, particularly the real-time audio generation capabilities of diffusion models. The team, led by me, Dominic Thibault, investigated the usability of diffusion models in the generation of real-time audio, specifically in the context of audio synthesis.

Team Contributions

Co-researcher Sean Wood, professor of Electrical and Computer Engineering at Université Sherbrooke, was consistently involved in the project. Along with his master's student, Adam Mezaber, he contributed to the adaptation of the DiffWave[1] model to make it real-time. Caroline Traube was an important collaborator, particularly in developing the concept of "timbre space" and its exploration as a means of musical expression. Her advice greatly aided the project's development. Jean-Philippe Jullin and Sarah Lecompte-Bergeron, students from my lab, played crucial roles as research assistants. Jean-Philippe updated the codebase, tested the models, and optimized the algorithm's learning process. Sarah developed the original datasets that are still used for testing the models' learning and inference processes.

Technical Development

We based our prototype on the DiffWave library, which implements an AI method based on diffusion models to generate spoken words from text. We adapted the algorithm to make it real-time (low latency) and interactive. The resulting codebase (in Python) can train and generate audio with interesting results. The algorithm trains on 3 to 15 minutes of audio, and within two hours of training, the resulting model can generate audio in real-time based on parameters sent via OSC. The resulting audio shares timbre resemblances with the original audio but can be shaped in radically new ways.

We aim to publish our results in the next round of international conference : ICMC, NIME, AIMC and DAFX. Equally importantly, we aim to create music that employs the trainable models in the coming year, for example at an Akousma concert that will take place at SAT in January 2025.

---

[1] Zhifeng Kong et al., “DiffWave: A Versatile Diffusion Model for Audio Synthesis” (arXiv, March 30, 2021), https://doi.org/10.48550/arXiv.2009.09761.

Creative and Future Directions

The project has produced a significant amount of knowledge on AI audio generation. We are gathering information on the creative possibilities rendered by such models. We believe there are many musical opportunities in interacting with the models created in our project. Consequently, we plan to apply for a CRSH Savoir grant in Fall 2024 to study the creative opportunities and ethical challenges of creating music with deep-learning models such as the one we've created.

Our long-term objective is to make DiffSynth a musical instrument. We see a strong possibility for embedding the algorithm in portable micro-controllers, leading to the creation of a Eurorack module. The implementation on a small factor computer will open the door to innovative musical applications.

Budget

- Research assistant contracts: $4500
- Purchase of equipment (NVIDIA Jetson Nano dev kit): $500

The Agile Seed Fund provided the necessary resources to put this research-creation project in motion, and we are excited about the future possibilities it holds for AI-generated music.