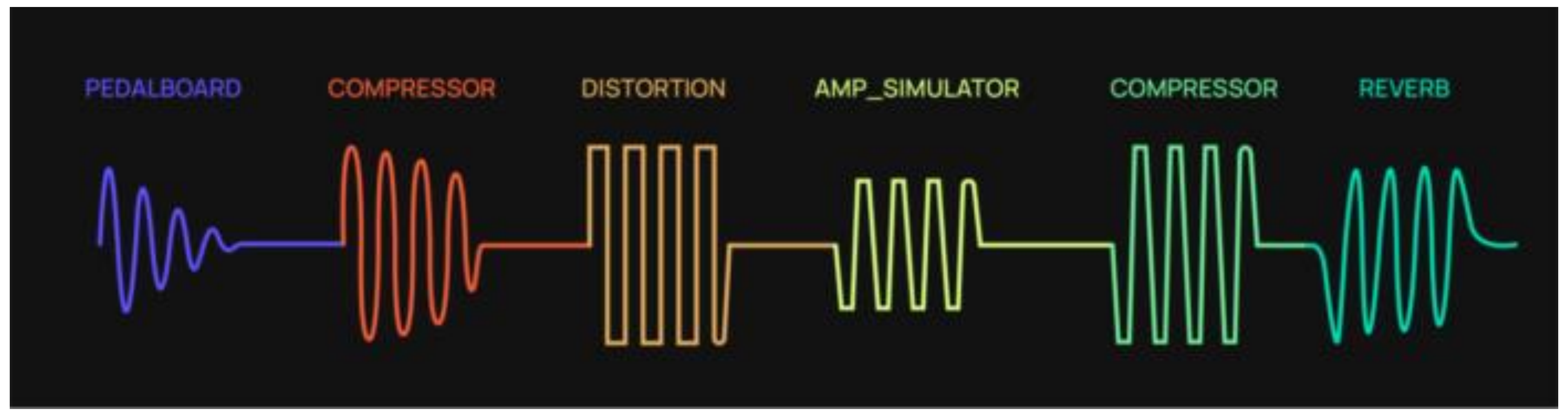
Deep Learning for Style Transfer and Experimentation with Audio Effects and Music Creation

Ada Tur

McGill University, Montreal, QC, Canada

Abstract

- Recent advancements in deep learning have potential to transform process of writing and creating music
- Models that have potential to capture and analyze higher-level representations of music and audio can change neural DSP
- Set of Music+AI methods for audio generation, modelling and transferring of timbres/effects, applying effects, including research into experimental audio effects, and production of audio samples using style transfers

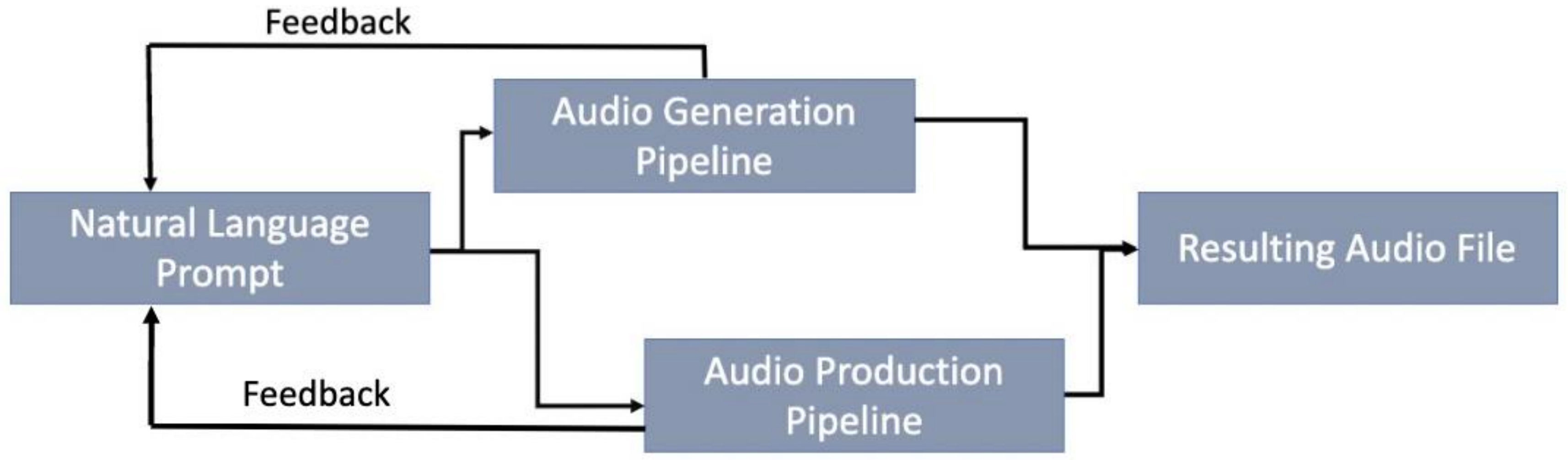


Source: Spotify Engineering

- Natural language prompt to finalized audio result pipeline
- Writing and producing music requires money, time, and knowledge
- All-encompassing framework for music processing would make process much more accessible and simple

Proposal

- Natural language prompt → vectorized representation
- Either **edit** or **generate** audio
- **If edit:**
  - Encode audio to vectorized representation (EnCodec)
  - Apply/edit learned effects/timbres
    - Transformer-based encoder OR temporal convolutional network
  - Apply style transfer
- **If generate:**
  - Generation using vector quantization and auto-regressive transformer-based decoder (MusicGen)



Evaluation + Methods

- How do we define "good" music?
  - Fréchet Audio Distance
    - Low score implies generated audio is plausible
  - Kullback-Leiber Divergence
  - CLAP Score
    - Audio-text alignment
  - Audio effect alignment classifier
- Human Evaluation: Set of participants receive audio sample prior to and after model alteration
  - Use set of criteria to describe final product (”Excellent”, ”Better”, ”Terrible”, etc)
  - Musicians experienced with using audio software can utilize framework and return feedback on performances

References

- Copet, J.; Kreuk, F.; Gat, I.; Remez, T.; Kant, D.; Synnaeve, G.; Adi, Y.; and Defossez, A. 2023. Simple and Controllable Music Generation. arXiv preprint arXiv:2306.05284.
- Dhariwal, P.; Jun, H.; Payne, C.; Kim, J. W.; Radford, A.; and Sutskever, I. 2020. Jukebox: A generative model for music. arXiv preprint arXiv:2005.00341.
- Paissan, F.; Wang, Z.; Ravanelli, M.; Smaragdis, P.; and Subakan, C. 2023. Audio Editing with Non-Rigid Text Prompts. arXiv preprint arXiv:2310.12858.
- Steinmetz, C. J.; Bryan, N. J.; and Reiss, J. D. 2022. Style transfer of audio effects with differentiable signal processing. arXiv preprint arXiv:2207.08759.
- Steinmetz, C. J.; and Reiss, J. D. 2021. Steerable discover of neural audio effects. arXiv:2112.02926.