Evaluation Plan

Kieran Grant - 2357351G

How do we know our encoder works, i.e. are embeddings meaningful?

- Show latent embeddings for different classes of DAFX under UMAP projection, both seen/unseen DAFX.
  - *Are they well separated?*
  - *Are similar effects close to one another?*
- Show domain colouring/parameter interpolation for seen/unseen DAFX.
  - *Do similar parameter configurations get mapped to similar places in latent space?*
  - *Are some changes in parameters better represented in the latent space (e.g. frequency changes, dynamic changes, temporal changes)?*

Does the style matching work?

- For general style transfer similarity, use both seen (VTCK) and unseen (DAPS) datasets. Use the same methodology as in validation step to calculate metrics (MRSTFT, SCE, MSD, RMS) for seen and unseen audio effects.
  - *Maybe*: use MusDB18 dataset to see if trained model can style match for musical audio as well?
  - Example results table for a single dataset given below (**DAFX unseen in encoder training*):

Effect	MRSTFT	SCE	MSD	RMS
Overdrive	0.0	0.0	0.0	0.0
Delay	0.0	0.0	0.0	0.0
RingMod	0.0	0.0	0.0	0.0
Dynamics	0.0	0.0	0.0	0.0
Combo*	0.0	0.0	0.0	0.0
Ambience*	0.0	0.0	0.0	0.0
Flanger*	0.0	0.0	0.0	0.0

- Compare our model to DeepFx-ST [1] for style transfer.
  - Only effect that both models style match for is *compression*.
  - Generate input/target pairs using mda MultiBand (effect unseen by both models).
  - Style match with both models (ours using mda Dynamics) and store predictions.
  - Compare metrics similar to the above table.
- MUSHRA Listening test (***not sure about time contraints***):
  - Listeners are given clean and effected audio, along with 3 different style match attempts:
    - The true effected audio (hidden reference)
    - A random parameter setting (anchor)
    - The model predicted style match
    - *Maybe DeepFx-ST for compression in addition to the above 3?*
  - Listeners should rate how well each of the samples matches the reference given (scale from 1-10). The hidden reference and anchors then act as upper and lower bounds for user ratings.
  - Only do this for a couple of effects (maybe 1 which the encoder was trained on and 1 which is unseen by the encoder?) - 5-10 different examples per effect.
  - Should potentially aim for ~10 participants (they do not necessarily need to be experts).

Does the low-dimensional latent space for parameter adjustment work?

- Plot interpolation across 2D latent space and respective changes to parameter curves.

  - e.g. going from (0,0) to (1,1) in the latent space.
  - *Do the parameters change smoothly, or are they all over the place?*
- Use a digital ear (e.g. AudioCommons) to see if audio features change with the above interpolation.
  - *Does brightness, depth, etc. increase/decrease in an obvious way?*
- Perform both of the above steps with a couple different DAFX.
- User (expert) evaluation:
  - **Test 1:** plot points A and B on 2D latent space and let them listen to the generated audio. Ask them to think about a point C midway between them. Let the user then listen to C.
    - *How well does it match their expectation? (1 = not at all, 10 = matches exactly).*
    - Two different DAFX, 5-10 examples each.
  - **Test 2:** Give the user time with latent space (can ask me to point to different points in latent space and play audio).
    - Then, start at a point in the latent space and give the user a prompt of a change to make to the audio (e.g. *"make it brighter"*, where should the new point be placed?)
      - *How well does it match their expectation? (1 = not at all, 10 = matches exactly).*
    - Same DAFX as in Test 1, make semantic descriptors relevant to effect used (e.g. *"make grittier"* for Overdrive, *"make repetitions further apart"* for Delay)
  - End with general evaluation of how intuitive they found using the latent space to control parameter setting.

References

[1] Steinmetz, Christian J., Nicholas J. Bryan, and Joshua D. Reiss. "Style transfer of audio effects with differentiable signal processing." *arXiv preprint arXiv:2207.08759 (2022).*