IS DISENTANGLEMENT ENOUGH? ON LATENT REPRESENTATIONS FOR CONTROLLABLE MUSIC GENERATION (SUPPLEMENTARY MATERIAL)

Ashis Pati
Center for Music Technology
Georgia Institute of Technology, USA
ashis.pati@gatech.edu

Alexander Lerch
Center for Music Technology
Georgia Institute of Technology, USA
alexander.lerch@gatech.edu

1. MODEL ARCHITECTURES & TRAINING DETAILS

The VAE model architectures used to conduct the experiments in this paper are based on dMelodies-RNN and dMelodies-CNN architectures proposed in our prior work [1].

For both models, a 32-dimensional latent space is chosen. The 9 attributes from the dMelodies dataset are regularized along the first 9 dimensions of the latent space.

2. ADDITIONAL RESULTS

2.1 Attribute Disentanglement

Figure 1 shows the disentanglement performance for all learning methods for the two different model architectures. Overall, dMelodies-RNN has a lower variance in performance.

2.2 Reconstruction Accuracy

Figure 2 shows the reconstruction accuracy for all learning methods for the two different model architectures. While supervised methods have better reconstruction performance, there is not much difference between the two model architectures.

2.3 Independent Control during Generation

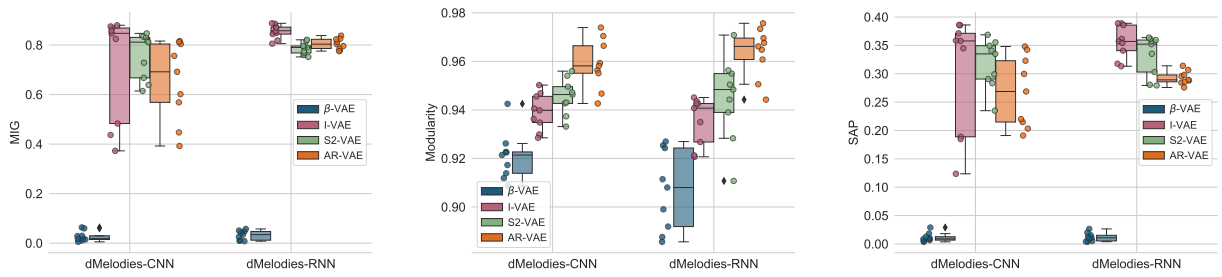
Figure 3 shows the performance of different learning methods in independently controlling the attributes for both model architectures.

2.4 Latent Space Visualization

Figures 4, 5, 6 show plots for latent space visualization experiments for rest of the attributes for the I-VAE, S2-VAE, and AR-VAE methods respectively

3. REFERENCES

[1] A. Pati, S. Gururani, and A. Lerch, "dMelodies: A Music Dataset for Disentanglement Learning," in *Proc. of 21st International Society for Music Information Retrieval Conference (ISMIR)*, Montréal, Canada, 2020.



(a) *MIG* (b) *Modularity* (c) *SAP Score*

Figure 1: Overall disentanglement performance (higher is better) of different supervised methods on dMelodies. Individual points denote results for different hyperparameter and random seed combinations. Results for β-VAE are also shown for comparison.

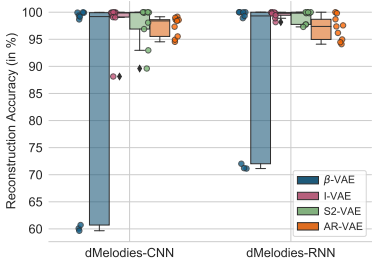
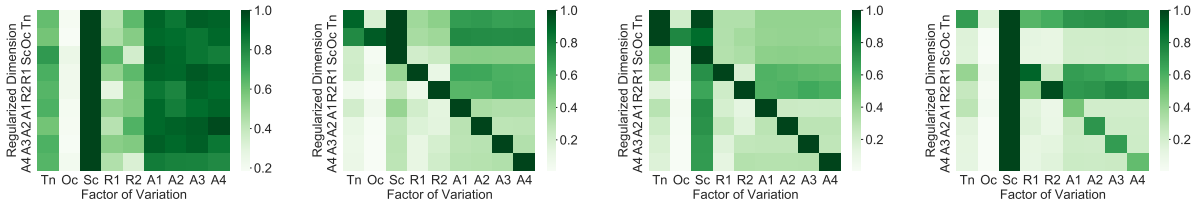
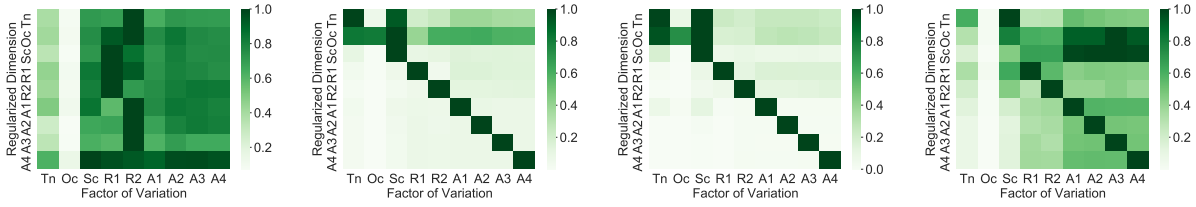


Figure 2: Overall reconstruction accuracies (higher is better) of the supervised methods on dMelodies. Individual points denote results for different hyperparameter and random seed combinations. Results for β-VAE are also shown for comparison.

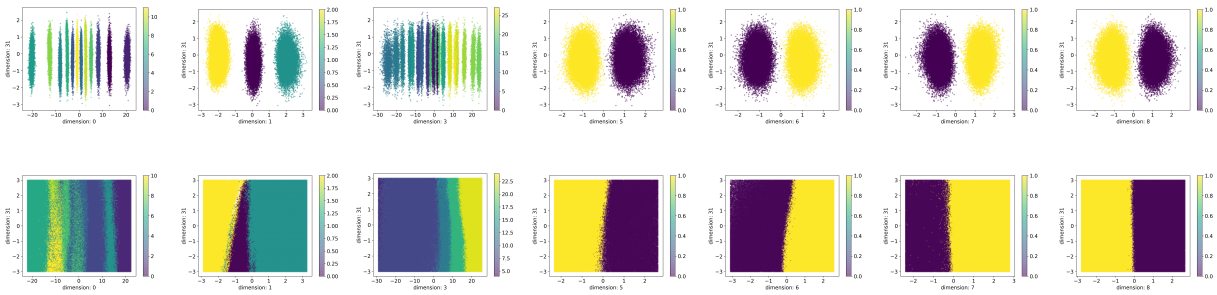


(a) β-VAE (dMelodies-CNN) (b) I-VAE (dMelodies-CNN) (c) S2-VAE (dMelodies-CNN) (d) AR-VAE (dMelodies-CNN)



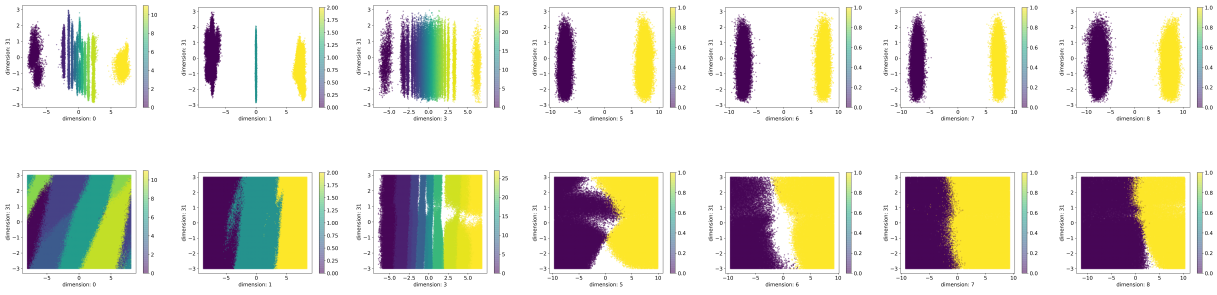
(a) β-VAE (dMelodies-RNN) (b) I-VAE (dMelodies-RNN) (c) S2-VAE (dMelodies-RNN) (d) AR-VAE (dMelodies-RNN)

Figure 3: Attribute-change matrices corresponding to different methods. Top row shows results for dMelodies-CNN, bottom row shows results for dMelodies-RNN. The columns show the normalized net change in a particular attribute value as one traverses the regularized dimension for the attribute corresponding to the rows. Tn: Tonic, Oc: Octave, Sc: Scale, R1 and R2: rhythm for bars 1 and 2 respectively, A1-A4: arpeggiation direction for the four chords.



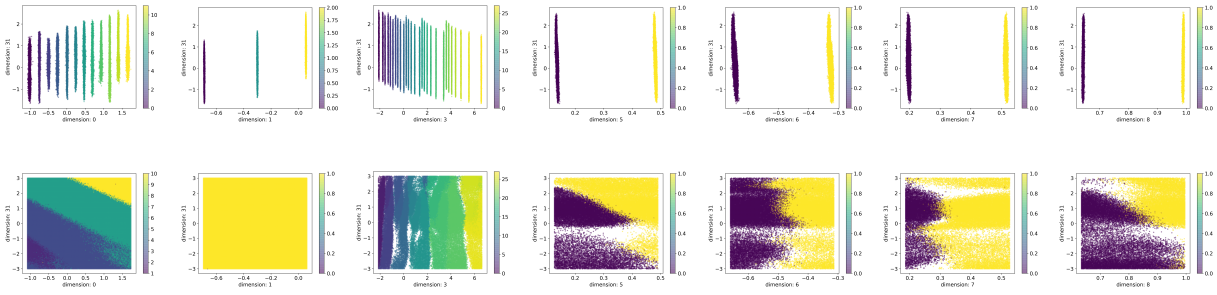
(a) (b) (c) (d) (e) (f) (g)

Figure 4: Data distribution (top row) and surface plots (bottom row) for I-VAE using dMelodies-RNN for different attributes. *a) Tonic, b) Octave, c) Rhythm Bar 1, d) Arp Chord 1, e) Arp Chord 2, f) Arp Chord 3, g) Arp Chord 4.* The x-axis corresponds to the regularized dimension for the attribute and the y-axis corresponds to a non-regularized dimension. Empty regions in the bottom row surface plots denote undefined or out-of-distribution attribute values.



(a) (b) (c) (d) (e) (f) (g)

Figure 5: Data distribution (top row) and surface plots (bottom row) for S2-VAE using dMelodies-RNN for different attributes. *a) Tonic, b) Octave, c) Rhythm Bar 1, d) Arp Chord 1, e) Arp Chord 2, f) Arp Chord 3, g) Arp Chord 4.* The x-axis corresponds to the regularized dimension for the attribute and the y-axis corresponds to a non-regularized dimension. Empty regions in the bottom row surface plots denote undefined or out-of-distribution attribute values.



(a) (b) (c) (d) (e) (f) (g)

Figure 6: Data distribution (top row) and surface plots (bottom row) for AR-VAE using dMelodies-RNN for different attributes. *a) Tonic, b) Octave, c) Rhythm Bar 1, d) Arp Chord 1, e) Arp Chord 2, f) Arp Chord 3, g) Arp Chord 4.* The x-axis corresponds to the regularized dimension for the attribute and the y-axis corresponds to a non-regularized dimension. Empty regions in the bottom row surface plots denote undefined or out-of-distribution attribute values.