Sleepy Style Music through Variational Autoencoder

Wenwen Ouyang[1], Ruoyu Gao[1], Pujana Paliyawan[2], Ruck Thawonmas[3], Tomohiro Harada[3]
[1]Undergraduate School of Information Science and Engineering
[2]Research Organization of Science and Technology
[3]College of Information Science and Engineering
Ritsumeikan University, Japan
ruck@is.ritsumei.ac.jp

***Abstract*—This paper proposes a method that use a Variational Autoencoder (VAE) neural network to change the style of audiences' favorite music to sleepy style one, by which our purpose of making them relaxed is achieved. Not only music can influence our emotion, but also sleepiness. Some people like to listen to sleepy music to aid soothing sleep, and responding to this demand, there are currently many online channels streaming sleepy music all day. Nevertheless, the variety of music being streamed is limited. This study aims at applying style transfer techniques on music data to generate sleepy-style music from existing non-sleepy-style music. We believe our proposed technique will not only solve the above limitation but also bring music composers new ideas and inspirations.**

***Index Terms*—music style, style transfer, midi, Variational Autoencoder, MIDI-VAE**

I. Introduction

Music has a significant influence on our emotion condition. Listening to music is also one of the most popular strategies people of all ages choose to aid their sleeping [1] [2]. A survey on urban people in Finland [3] revealed that music was the second most important factor in promoting soothing sleep. Over past decades, systematic use of music as a treatment in various medical conditions has been a popular subject for academic studies. Due to this demand, one can see many live streaming channels of various platforms playing sleepy music all day.

To provide more choices of music for aiding sleeping and to demonstrate creative adaptation of existing music in the way that inspires music composers, we demonstrate the use of a deep learning algorithm in turning a style of any piece of music into a sleepy style. In this study, we use a state-of-art music generation model to generate sleepy music and compare generated music with original music in user evaluation. The evaluation is based on Stanford Sleep Standard [4], a metric for evaluating the degree of sleepiness.

II. Related Work

Style and domain transfer using neural networks has become exciting machine learning showcases. Most prior work focused on the image domain. For example, Selim et al. [5] demonstrated that neural networks can be applied to render photographs in a painting style of a certain artist.

We investigate whether neural networks can be used to change any kind of music style to a sleepy style. From literature review, we found three deep learning models, C-RNN-GAN [6], CycleGan for midi [7], and MIDI-VAE [8], that have a potential for this task. After a pilot study, the MIDI-VAE model was selected to generate sleepy music for our experiment.

III. Model Architecture

MIDI-VAE is a neural network model based on Variational Autoencoder [9] that is capable of handling polyphonic music with multiple instrument tracks, as well as modeling the dynamics of music by incorporating note durations and velocities [8]. It can perform style transfer on symbolic music by automatically changing pitches, dynamics, and instruments of a music piece.

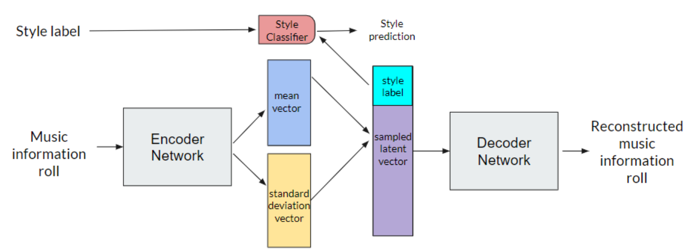

Fig. 1. The brief structure of MIDI-VAE

The architecture of MIDI-VAE is shown in Fig. 1. The model operates on a symbolic music representation that is extracted from MIDI files. This representation consists of note pitches, velocity and instrument rolls and hence models the most important information contained in MIDI files. MIDI-VAE uses separate encoder/decoder pairs that share latent space. Recurrent neural networks are used to implement the encoder and the decoder. VAEs are known to give representations with disentangled factors [10]. Due to isotropic Gaussian priors on the latent variables. Modeling those variables as Gaussians allows each dimension in the representation to push themselves as far as possible from the others. This model encodes a given music style to the top k dimensions (k denotes the number of styles. In this paper, k is two.) of the latent space and a style classifier is attached to a part of the latent space. Having such disentangled latent space will be helpful to make sure the encoder learns a compact latent style label that we can then use to perform style transfer.

IV. EXPERIMENT

Among the music domain, we chose jazz music as the original music style to be transformed into the sleepy style because we consider that jazz has a variety of subgenre, and thus it is a challenge to transfer jazz songs into the sleepy style. Two hundred and fifty pieces of jazz music and one hundred and fifty pieces of sleepy music were selected as our dataset. We randomly assigned 90 percent of music to a training set and 10 percent to a testing set.

We trained the MIDI-VAE model for 1000 epochs. Several combinations of hyperparameter values were tested. However, we finally found that default values of hyperparameters, recommended by the original paper, performed best.

We then randomly selected three pieces of generated music and their respective jazz music. Three pieces of sleepy music created by human composers were selected as baselines for evaluation via an online survey. As a result, there is a total of nine combinations[1] in the questionnaire where each combination consists of one piece of generated music, one piece of its original jazz music, and one piece of human-composed sleepy music. The survey randomly selected one of the nine combinations for a participant and let she or he rank the sleepy degree based on the Stanford Sleep Standard(SSS) [4]. The seven degrees of SSS is shown as Table I.

TABLE I
STANFORD SLEEPINESS SCALE

Degree of Sleepiness	**Scale Rating**
Feeling active, vital, alert, or wide awake	1
Functioning at high levels, but not at peak; able to concentrate	2
Awake, but relaxed; responsive but not fully alert	3
Somewhat foggy, let down	4
Foggy; losing interest in remaining awake; slowed down	5
Sleepy, woozy, fighting sleep; prefer to lie down	6
No longer fighting sleep, sleep onset soon; having dream-like thoughts	7

V. RESULTS

From 23 participants' results (Fig. 2), we ran a Friedman test [11] to evaluate the differences among the three music clusters. The p-value is found less than 0.01 indicating that there is a staticsally significant difference between jazz, sleepy and generated music at a confidence interval of 99%.

Apart from the Friedman test, we also ran a Wilcoxon Signed-Rank test [12] to evaluate the difference between any combination of two of the three conditions. Hence, there are totally three groups to compare. After the tests, the p-value is less than 0.01 for the jazz and generated music groups, less than 0.01 for the jazz and sleepy music groups, 0.68 for the generated and sleepy music groups, which means the style of

[1] http://www.ice.ci.ritsumei.ac.jp/%7eruck/sleepy-music-segah2019.htm

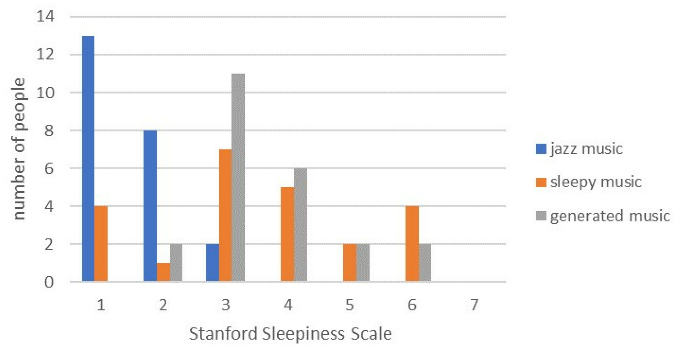

Fig. 2. Histogram of the results

our generated music is more similar to sleepy music than jazz music.

VI. CONCLUSIONS

We used MIDI-VAE to generate sleepy music from jazz music. The generated music was compared to their original and sleepy music created by human composers. Stanford Sleep Standard [4] was used as a metric to measure the degree of sleepiness. Our results showed that generated music sounded sleepier than the original jazz one. Beyond our expectation, the sleepy degree even reached the level of the carefully selected sleepy dataset in use.

However, the music generated by the current deep learning model still has some small flaws in rhythm. For our future work, we will design deep learning techniques that increase the harmony of generated music.

REFERENCES

[1] H. Lai, M. Good, "Music improves sleep quality in older adults," Journal of advanced nursing, 2005 Feb, vol. 49, no. 3, pp. 234-244.
[2] L. Harmat, J. Takcs, R. Bodizs, "Music improves sleep quality in students," Journal of advanced nursing, 2008 May, vol. 62, no. 3, pp. 327-335.
[3] G. De Niet, B. Tiemens, B. Lendemeijer, G. Hutschemaekers, "Musicassisted relaxation to improve sleep quality: metaanalysis," Journal of advanced nursing, 2009 Jul, vol. 65, no. 7, pp. 1356-1364.
[4] E. Hoddes, "The development and use of the stanford sleepiness scale (SSS)," Psychophysiology, 1972, vol. 9, p. 150.
[5] A. Selim, M. Elgharib, L. Doyle, "Painting style transfer for head portraits using convolutional neural networks," ACM Transactions on Graphics (ToG), 2016 Jul 11, vol. 35, no. 129, pp. 1-18.
[6] O. Mogren, "C-RNN-GAN: Continuous recurrent neural networks with adversarial training," 2016 Nov 29, pp. 1-6, arXiv preprint:1611.09904.
[7] G. Brunner, Y. Wang, R. Wattenhofer, S. Zhao, "Symbolic music genre transfer with cyclegan," In2018 IEEE 30th international conference on Tools with Artificial Intelligence (ICTAI), 2018 Nov 5, pp. 786-793, IEEE.
[8] G. Brunner, A. Konrad, Y. Wang, R. Wattenhofer, "MIDI-VAE: Modeling dynamics and instrumentation of music with applications to style transfer," 2018 Sep 20, pp. 1-8, arXiv preprint:1809.07600.
[9] D.P. Kingma, M. Welling, "Auto-encoding variational bayes," 2013, arXiv preprint:1312.6114.
[10] Higgins I, Matthey L, Glorot X, et al. Early visual concept learning with unsupervised deep learning[J]. arXiv preprint arXiv:1606.05579, 2016.
[11] D.W. Wayne, "Friedman two-way analysis of variance by ranks," Applied Nonparametric Statistics (2nd ed.), Boston: PWS-Kent, 1990, pp. 262274, ISBN 978-0-534-91976-4.
[12] R.F. Woolson, "Wilcoxon signedrank test," Wiley encyclopedia of clinical trials, 2007, pp. 1-3.