Preliminary Study on Morphing of Chord Progression

Aiko Uemura[1] and Tetsuro Kitahara[1]

Nihon University
{uemura, kitahara}@chs.nihon-u.ac.jp

Abstract. Reharmonization is changing the chord progression of music to alter the impression of harmony. To generate a variety of reharmonizations, morphing between a small number of instances of manual reharmonization is expected to be a practical approach. In this paper, we propose a method of morphing between reharmonized chord progressions using a variable autoencoder (VAE). Experimental results show that our method generates new chord progressions based on interpolation between prepared chord progressions in the latent space obtained through the VAE.

Keywords: reharmonization, chord morphing, chord progression

1 Introduction

Reharmonization is one way to change the impression of harmony. It refers to the technique of replacing the original chord with another chord without changing the melody but altering the style of the music. The theory behind the progress and connection of chords is called functional harmony. Even though the functional harmony is similar, the genre of the music may change when non-harmonic tones, such as tension or voicing, are used. Conventionally, tension and voicing have often been used in jazz improvisation. Many videos of arranged music have been recently posted on the video posting site and SNS, such as playing in a specific genre or another artists style. There remains an ever-increasing interest and challenge to rearrange existing harmonies.

Studies [1, 2] to execute reharmonization automatically have been proposed in this background. Hirata et al. proposed pa-pi-pun [1] and it reharmonized jazz-like chord progression using a knowledge representation and reasoning mechanism called Deductive Object-Orientation. They made it possible to reflect the users intention based on the case example and in regards to harmonic context. Yoo et al. investigated a tension curve[2], which represents the degree of change in tension of the chord progression. It allows the user to reharmonize the original music by editing the curve.

On the other hand, there is another approach to achieve a wide variety of reharmonization by obtaining an intermediate chord progression from a small number of reharmonized chord progressions. A method of generating medial data from multiple sources is called morphing. There are studies conducted on melodies, chords, drum patterns and baselines [3, 5, 6]. Hirata et al. proposed morphing the melody partially from the time span tree obtained as an analysis result of the Generative Theory of Tonal Music (GTTM) [3]. Roberts and colleagues suggested melody, drum and baseline-based morphing using a variational autoencoder (VAE)[4] in morphing independent from the music's structure. Murata et al. also performed morphing of the melody and chords [6]. The latent space representing features in low dimensions can be obtained in the VAE, and these researchers utilized the function to perform continuous complementation.

We follow the idea of Roberts et al [5]. and Murata et al [6]. The reharmonized chord progression with different styles are mapped into the latent space, and we generate a new chord progression by interpolating and extrapolating in this space. By inserting the chord progression of two styles on the latent space, it is preferable that the result generated is an intermediate between the two. Through experiments, this research verifies how this phenomenon occurs. If morphing generates reharmonized data, it is expected to be applied to data augmentation technology [8] which extends training data.

This paper also reports on the applications that can be used in latent space. The user can generate and view a new chord progression concerning the original chord progression plotted on latent space and the reharmonized chord progression. Even users who can not understand tension or voicing by chord name or musical score can intuitively choose chord progressions from the sounds of the chord.

2 Chord Morphing using VAE

We derive a latent space representing the chord progression given to a melody and the chord progression data reharmonized manually to another genre by the generation model based on a VAE. The morphing is performed as shown in Fig.1 since the features are complemented continuously in the latent space.

A VAE is a neural network for acquiring characteristic expressing data and assumes a multivariate standard normal distribution for latent variables. Therefore, modeling that specific data is generated based on the abstract representation of the latent variable.

In this study, we assume the latent variable $\boldsymbol{z}$ which represents the character of chord progression. As usual, let $\boldsymbol{z}$ obey the multivariate standard Gaussian distribution of the mean vector $\mathbf{0}$, covariance matrix I. The song $\boldsymbol{x}$ is a chord progression sequence and the onset series of note pitch is represented in units of quarter notes. Here, we treat four octaves of C2 to B5, and the onset of the pitch at a specific time is represented by bits $0, 1$. The number of input dimensions is assumed to be 48×60, using data of 15 bars excluding the anacrusis of each song. MIDI data is converted so that the pitch onset interval becomes 1 when reading and 0 otherwise and the model is learned.

In learning of VAE, we use $\boldsymbol{x}$ as the song data and $\boldsymbol{z}$ as the latent variable corresponding to $\boldsymbol{x}$ to maximize the marginal likelihood $p_\theta(\boldsymbol{x})$ We calculate

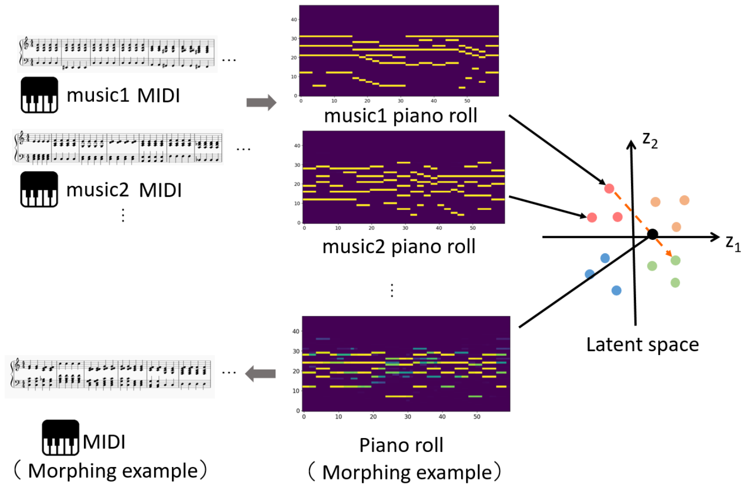

Fig. 1. Conceptual diagram of chord morphing.

the distribution of $p_\theta(\boldsymbol{x}|\boldsymbol{z})$, but this calculation is difficult. Therefore, we make a distribution of $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ which generates a latent variable distribution in the VAE. The log probability likelihood is given by

$$\log p_\theta(\boldsymbol{x}) = D_{KL}\left[q_\phi(\boldsymbol{z}|\boldsymbol{x})||p_\theta(\boldsymbol{z}|\boldsymbol{x})\right] + \mathcal{L}(\theta, \phi, \boldsymbol{x}). \tag{1}$$

where D_{KL} is the Kullback-Leibler (KL) divergence and $\mathcal{L}$ represents the variational limit. Hence, the variational limit is expressed by the following equation.

$$\begin{aligned} \mathcal{L}(\theta, \phi; \boldsymbol{x}) &= \log p_\theta(\boldsymbol{x}) - D_{KL}\left[q_\phi(\boldsymbol{z}|\boldsymbol{x})||p_\theta(\boldsymbol{z}|\boldsymbol{x})\right] \\ &= -D_{KL}[q_\phi(\boldsymbol{z}|\boldsymbol{x})||p(\boldsymbol{z})] \\ &\quad + \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}[\log p_\theta(\boldsymbol{x}|\boldsymbol{z})] \end{aligned} \tag{2}$$

Maximizing $\mathcal{L}(\theta, \phi; \boldsymbol{x})$ for θ and ϕ minimizes the KL divergence of $p_\theta(\boldsymbol{z}|\boldsymbol{x})$ and $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ from the Eq.2.

In the network structure, the encoder and decoder are two layers of Multi-Layer Perceptron (MLP), the optimization method is the Adaptive Moment Estimation (Adam), and the parameters follow the document [7]. We used tanh as the activation function like [4].

3 Experimental results

3.1 Conditions

We used MIDI files labeled manually based on the notation [9] of traditional music “The Water Is Wide” as learning data. The reharmonized songs include seven genres:, Jazz, Bossa Nova, R&B, Funk, Folk, J-Pop 1, and J-Pop 2. All keys

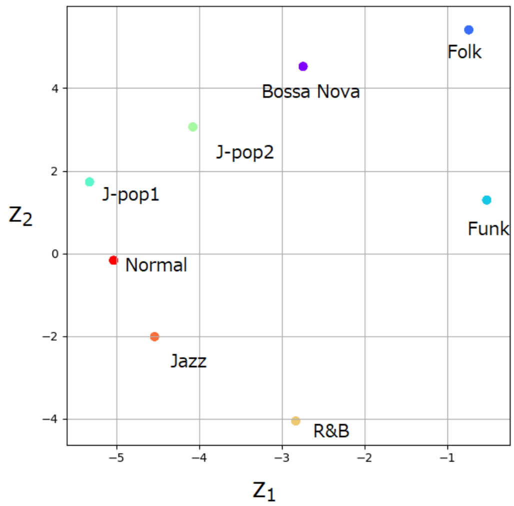

Fig. 2. Latent space of eight chord progresses.

are in C for the original song (Normal). We did not consider changes in dynamics in the songs and velocity was at 100. We split notes longer than quarter notes and omit notes shorter than quarter notes.

3.2 Results

Accuracy of reconstruction Fig. 2 shows a plot of the latent space of eight songs after learning. We move the coordinates in $-6 \leq z_1 \leq 0, -6 \leq z_2 \leq 6$ and analyze the generated chord progression since the plot of chord progression is distributed in $-5.03 \leq z_1 \leq -0.52, -5.03 \leq z_2 \leq 5.41$ in the latent space. Since the decoded piano roll results output was $[0, 1]$, we used a threshold value to perform processing and set the value to 0 below the threshold. The threshold value was experimentally set to 0.3. The velocity was set to an output multiplied by 100 to convert a piano roll to a MIDI signal.

Table 1 shows an example of chord analysis results. We analyzed the number of notes of 7th, ♭ 9th, 9th, ♭13th, and 13th corresponding to the lowest note (base note) of each chord for a song. We also investigated the average number of chord components.

It is considered that a reasonable number of chord compositions to listen to is three to six as a chord, because the learning data is a piano or guitar score. It tends to become an unclear sound if the average amount of chord compositions exceeds eight. As the number of constituent sounds of the chord increased, the more uncomfortable sound progressed throughout the whole. In the case of $z_1 = -0.50, z_2 = -2.00$, it seems that the progression is close to funk in latent space, but chord progression of unclear sound was generated in all the bars as is shown in Table 1. It shows that this plot was far from the data set

Table 1. An example of the average number of chord components and the number of notes of 7th, ♭ 9th, 9th, ♭13th, and 13th corresponding to the base note.

z_1	z_2	average of chord composition	tritone	7th	♭9th	9th	♭13th	13th
−0.50	−2.00	40.9	60	60	60	60	60	60
−1.5	4.00	6.3	0	42	5	46	24	1
−2.00	0.00	4.2	7	8	0	8	10	0
−3.00	0.50	3.7	0	9	0	8	19	0
−4.50	2.20	4.5	2	26	1	19	19	2
−4.50	2.25	7.9	9	41	9	48	30	9
−6.00	6.00	6.4	0	30	6	36	21	0
Bossa Nova		4.2	2	16	4	30	4	2
Folk		4.2	2	10	0	6	14	2
Funk		5.0	0	60	0	40	4	0
J-pop1		4.4	0	26	1	15	20	2
J-pop2		4.9	2	36	0	39	14	0
R&B		4.8	6	54	6	28	4	6
Jazz		4.8	2	46	8	30	10	2
Normal		4.0	0	0	0	0	0	0

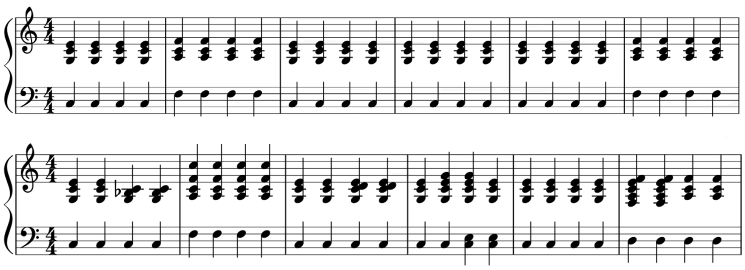

Fig. 3. The score at the first six bars (top: Chord progression before reharmonization, bottom: the output of $z_1 = -2.00, z_2 = 0.00$.)

expressed in the latent space, and the probability density became extremely low and could not be reconstructed.

Morphing (interpolation) As shown in Fig. 2, each data point is distance from the center, but the progression of $z_1 = -2.00, z_2 = 0.00$ with 7th, 9th, ♭13th is generated based on the original chord progression. Fig. 3 shows the first six bars of the original chord progression and the output of $z_1 = -2.00, z_2 = 0.00$. The output seems to include tension of 7th, 9th, ♭ 13th as shown in Table 1. This example shows that it would be a new sound based on the original chord progression using morphing.

Fig. 4 shows the score of J-pop 1 and J-pop 2. Regarding the chord change between the coordinates, the genre of chord progression tended to switch instead of changing step by step. The results of $z_1 = -4.50, z_2 = 2.00$, $z_1 = -4.50, z_2 =$

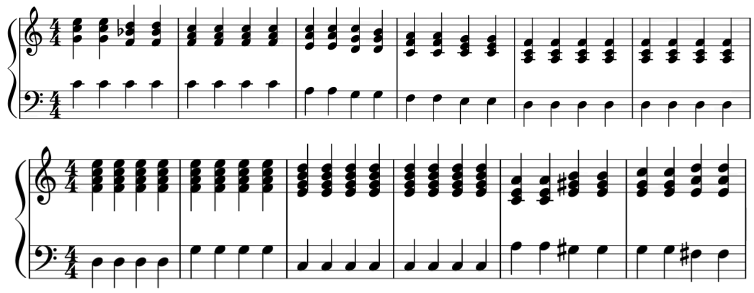

Fig. 4. The score at the first six bars (top: J-pop1bottom: J-pop2.)

Fig. 5. The score at the first six bars (top: $z_1 = -4.50, z_2 = 2.00$, bottom: $z_1 = -4.50, z_2 = 2.25$.)

2.25 are shown in Fig. 5. It is a small change and the latter increased z_2 by 0.25. The former is similar to J-pop 1 and the latter has many chord components but is close to J-pop 2. It is considered that this coordinate is located at the boundary of each distribution.

Fig. 6 shows an example of $z_1 = -3.00, z_2 = 0.50$. The average chord composition was 3.7 from Table 1, which was smaller than the other outputs. In fact, there are also 29 triads in the song. This is caused by it being far from the dataset represented in the latent space and the probability density has decreased.

Morphing (extrapolation) Fig. 7 shows the output resulting from extrapolation performed with $z_1 = -6.00, z_2 = 6.00$. Although the average of chord components was as large as 6.4, it was close to the J-pop 2. Compared to J-pop

Fig. 6. The score at the first six bars of $z_1 = -3.00, z_2 = 0.50$

Fig. 7. The score at the first six bars of $z_1 = -6.00, z_2 = 6.00$

Fig. 8. The application launched on the browser. The MIDI data is outputted based on the coordinates clicked by the user in the latent space and can be played.

2 in Table 1, the 7th and 9th decreased but the ♭ 9th and ♭ 13th increased. It seems that output is calculated according to the probability distribution by extrapolation in the latent space.

4 Implementation

Fig.8 shows the chord morphing application screen as it appears in the browser. This uses the model in the previous section. When the user clicks on the latent space, the chord progression is reconstructed based on the coordinates. The generated MIDI can be reproduced at the bottom of the screen, and the output can be confirmed interactively. The user interface uses JavaScript, Chart.js [10] for mapping latent space and MIDI.js [11] for MIDI playback. JavaScript passes the coordinates in the latent space clicked by the user, Python decodes, and MIDI creates, and JSONP data exchange is executed.

5 Conclusions

Our study focused on supplementation in the latent space of VAE and we analyzed chord progression generated by learning a small number of chord progressions before reharmonization and reharmonized chord progression. We also reported on applications that execute chord morphing by operating in latent space. The user can listen to the original chord progression and the reharmonized chord progression by tracing the plot in the latent space. Even if users cannot understand the chord names or music scores, they can select chord progressions from the way desired chord sounds.

We could not analyze the chord progression generated by VAE from the viewpoint of correct music structure. We should consider the generation and evaluation based on musical theory in the future as well. Since we used a small sample of data, the behavior of many chord progressions of the same genre arrangement in latent space is worth considering. In addition, the network model of the VAE needs to be discussed in future studies.

Acknowledgments. This work was supported by JSPS KAKENHI Grant Numbers 16K16180, 16H01744, 16KT0136, 17H00749.

References

1. Hirata, K., Aoyagi, T.: Pa-Pi-Pun. Creativity-support Tool for Generating Jazz Harmony, Information Processing Society of Japan. 42(3), pp.633–641 (2001)
2. Yoo, M-J.; Lee, I-K.: Musical Tension Curves and its Applications, In: International Computer Music Conference (ICMC). pp. 482486 (2006)
3. Hirata, K., Tojo, S., Hamanaka, M.: Melodic Morphing Algorithm in Formalism, Mathematics and Computation in Music. In: Third International Conference, MCM (2011)
4. Kingma, D. P., Welling, M.: Auto-encoding variational Bayes. In: International Conference on Learning Representations (2014)
5. Roberts, A.: Hierarchical Variation Autoencoders for Music, In: NIPS Workshop (2017)
6. Murata, S., Bando, Y., Itoyama, K., Yoshii, K.: Melody and chord morphing using Variational Auto Encoder. In: The 80th National Convention of IPSJ, 3N-08 (2018)
7. Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization. In: International Conference on Learning Representations, pp.1-13 (2015)
8. Tanner, M. A.; Wong, W. H.: The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association, 82 (398), pp. 528–540 (1987)
9. Sugiyama Y.: In the hands-on master (CD) with the music theory to help Genre code arrangement art composition and arrangement that are polished with Rihamonaizu. Ritto Music, Inc, Japan. (printed in Japan) (2011)
10. Chart.js,
`https://www.chartjs.org/`
11. MIDI.js,
`https://github.com/mudcube/MIDI.js`