IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

K-POP Composition with Generative AI

Yoonhee Jang

Saint Johnsbury Academy Jeju, 10, Global edu-ro 304beon-gil, Daejeong-eup, Seogwipo-si, Jeju Special Self-Governing Province, 63644, Republic of Korea; jangyoonhee0809@gmail.com

ABSTRACT: This research aimed to create a K-POP music composition model with Long Short-Term Memory and analyze K-POP datafication by integrating technology and other fields. K-POP was converted with a graph by Librosa, and it differed from people listening to music with their ears and looking at the data. It shows more relationships between pitch, time, and frequency. The LSTM model trains three data sets: all K-POP genre songs, BLACKPINK's songs, and BTS's songs. The output has a repeated melody when a model trains songs on all K-POP genre songs. It was a repetition of high notes. Conversely, when it focuses on learning one specific group, the output of the model has a wide range of notes. The model trained on a specific group's song generates a song that is closer to K-POP.

KEYWORDS: Robotics and Intelligent Machines, Machine Learning, K-POP Composition, LSTM, Music Composition.

Introduction

Artificial Intelligence, one of the most ideal technologies ever invented, has changed human lives. It brought convenience to individuals and the world. AI is inseparable from people's daily lives through Siri by iPhone[1] and Bixby by Samsung.[2] YouTube, an online video platform, uses AI for its video recommendation algorithm and distinguishes malicious comments to ban.[3] Instagram used AI to analyze and collect data about user preferences. They analyzed user preferences and exposed feeds related to user preferences on the user's screen.[4] AI is no longer separate from humans. For example, generative AI penetrates our lives.

The generative AI resolves diverse tasks. For instance, an AI drawing generator, Midjourney, creates an image or drawing depending on the user's prompt in a few minutes.[5] Large Language model, ChatGPT, is helpful in writing a novel or letter and speech outline. The sentences made by ChatGPT are very natural. It has proficiency in generating sentences.[6] Alexa, made by Amazon, is an AI personal secretary for people. It serves to enhance human comfort. It answers the questions humans ask.[7] As generative AI is in the spotlight, it keeps developing. AI generative models in the music field have also advanced. AI voice cover videos are especially favored on social media. On YouTube, a song cover video using AI got millions of views. People are using AI voice covers to make their favorite song cover videos. Through this example, I noticed the possibility of using AI to compose music. Thus, I attempt to create a K-POP music composition model.

K-POP is a category of Hallyu (Korean wave). After 2015, K-POP was moving forward to the global market. Then, in 2023, K-POP was fashionable worldwide. Its fame is centered around teenagers and those in their twenties. It won the Billboard Award, which is globally recognized and well-regarded. K-POP stands for Korean Pop. In a broader context, it means all Korean songs, but from a narrow perspective, it represents the end of 20th-century music's dance, hip-hop, R&B, and electronic music. K-POP shows breakneck development. Originally, K-POP was renowned in Asia. However, due to the commercialization of the internet, other countries can access K-POP through social media. This has resulted in K-POP becoming famous globally and promoting Korea. K-Pop has increased the number of tourists in Korea. Because of this, K-POP is also called South Korea's greatest export. Before the existence of K-POP, there was Hallyu (Korean Wave). There has been an increase in international interest in South Korean popular culture, especially music, film, fashion, and food. Hallyu has started getting attention at the end of the 20th century. Beginning with the export of Korean drama, exporting K-Drama expanded to songs and became famous. In the 21st century, a phenomenon favored Korean cultures, such as kimchi, electronics, and food. As K-POP became famous, as a new context of the Hallyu, K-POP was established.

The most prominent feature of K-POP is the ocular effect. It is considered a factor of K-POP fame. Usually, other singers and bands didn't sing a song while dancing. Conversely, almost all K-POP idols dance while they sing a song. Most K-POP songs are dance-pop. Not only these, but K-POP also has attractive melodies and lyrics. K-POP idols work in bands and both girls' and boys' groups. This paper divided K-POP into two categories: Group A is a boys' group, and Group B is a girls' group.[25]

This proves that K-POP has the value of research about the incredible speed of spreading its influence worldwide. To study K-POP, it requires a unique tool, MIDI. I used a MIDI file to construct a training dataset. Similar to how other songs are composed, and similarly to other songs, K-POP is also composed with verse and hook. It has repeated parts. Since K-POP's time series data is repetitive, I used LSTM to create a model. This research shows the positive side of AI and various possibilities for improvement and development. It is not only

© 2025 Terra Science and Education

DOI: 10.36838/v7i7.14

used in computer science; it can also connect with other fields. Thus, the whole society can improve together at the same time. I wish other education fields were developed through generative AI, not only art and writing. The development of technology has brought a lot of comfort to citizens. After the invention of generative AI, such as Midjourney, the quality of the student's project work has rapidly increased. For example, by developing drawing books for children, using my methods, students could submit high-quality work with the help of AI.

Methods

K-POP Data:

To analyze the characteristics of K-POP more clearly, I chose songs and K-POP groups that are publicly well-known. The data was collected in 2016 when K-POP started to attain sudden acclaim. The criteria for data (K-POP idol group) are the groups that at least won the music TV show and have a music video with over 100 million views on YouTube. I collected data from Twice, BLACKPINK, BTS, and NewJeans.

Table 1: K-POP MIDI File Data. It shows how many MIDI files were collected from the K-POP IDOL group. All the songs are after 2015. Some of the K-POP IDOL groups in the list debuted around 2021.

Group	Artist	Number of Songs
A	BTS, Tomorrow by Together, Stray Kids, NCT, EXO, ENHYPEN, Seventeen	79
B	BLACKPINK, Twice, NewJeans, IVE, NMIXX, Asepa, (G)-IDLE	75

I made up seven groups per group A and B. Group A had 79 songs, and Group B had 75 songs. In total, it consists of 154 songs. I collected at least five songs per group (Table 1).

MIDI file:

MIDI stands for Musical Instrument Digital Interface. It is a digital music sheet for playing digital instruments. It shows when and what pitch has to be played. Moreover, using MIDI files, requires special tools, such as PrettyMIdi and Py Fluidsynth.[28]

The song should be in a MIDI file format to model and learn the music. The MIDI file data used in this research paper are from the Musescore website. The MIDI file data is piano MIDI data for the consistency of the collected data. Musescore's website shares a music sheet or MIDI file with everything needed to write a song.

Table 2: "Still With You" MIDI note. A new note is played in less than one second.

Pitch	Start	end	Step	duration
73	0.685714	1.1010714	0.000000	0.325000
71	1.028571	1.353571	0.342857	0.325000
73	1.371428	1.859285	0.342857	0.487857
80	1.885713	2.373570	0.514285	0.487857
73	2.399999	2.724999	0.514285	0.325000

MIDI NOTE is a converted version of MIDI. I changed the MIDI file into a note using three variables: pitch, step, and duration. Start and end also represent duration. To train for the model effectively, I changed MIDI to note. If it was converted, the results are in Table 2. The model will be trained by this MIDI NOTE and converted to MIDI NOTE again so humans can listen to the generation output.

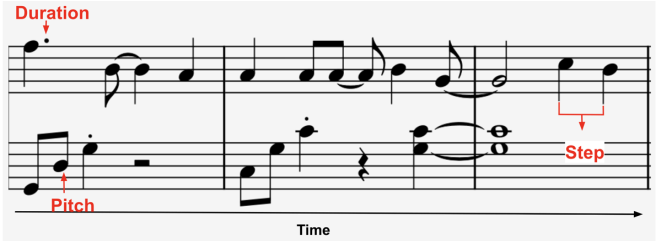

Figure 1: JungKook's "Still With You" music sheet is labeled with duration, pitch, and step. Each note represents duration, pitch, and step; the rest represents only duration.

Pitch:

Like in Figure 6 above, the pitch in MIDI NOTE is pitch. In the table, the pitch is expressed as an integer. It can be converted as 'A4' or 'G4'. In Figure 1 (music sheet), the position of the note represents the pitch.

Duration:

The duration of a MIDI note shows the duration of the pitch. A note is a whole note with four counts of duration; if it is a quarter note, it has one of the four counts as duration. In Figure 1, the type of note represents the duration.

Step:

The Steps represent the length of the previous note and the following note. It can calculate using the table by the previous starting point and subtract the start point.

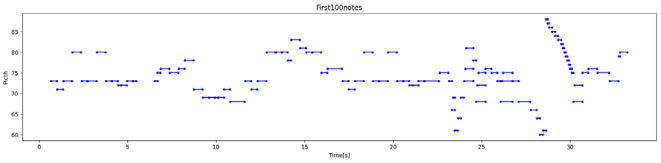

Figure 2: It is JungKook's "Still With You" first 100 notes floating graph. It had a wide range of pitch classes.

Each point on the graph represents each note in the music sheet. The y-axis shows the pitch, and the x-axis shows the times. The length of the line is the duration of the note. As the line is short, it means it has a short duration, and as the line is long, it means it has a long duration (Figure 2).

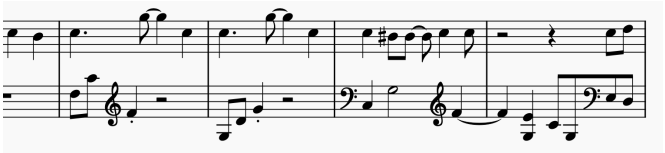

Figure 3: JungKook's music sheet for "Still With You" music sheet shows the song's first few seconds. In the bottom part of the sheet, there are several changes between G glef and F clef.

Figure 3 shows that staccato is represented as a concise line and two points on the graph. The floating graph makes it easier to see how a song will play than people who don't know how to read a music sheet. It is because, as one picture, it tells everything, such as when this pitch starts and where it ends. The floating graph provides a more expeditious comprehension.

DOI: 10.36838/v7i7.14

Deep Learning Model:

K-POP train data:

The train progressed throughout MIDI NOTE. After centralization, like Table 1, convert MIDI to MIDI NOTE with three columns using step, duration, and pitches. It has three columns. Figure 4 is one example of a sequence. The model will train 154 K-POP songs.

```
[[0.5703125  0.         0.32499986]
 [0.5546875  0.342857   0.32499986]
 [0.5703125  0.342857   0.48785694]
 [0.625      0.5142855  0.48785694]
 [0.5703125  0.5142855  0.32499986]
 [0.5703125  0.342857   0.48785694]
 [0.625      0.5142855  0.48785694]
 [0.5703125  0.5142855  0.32499986]
 [0.5703125  0.342857   0.32499986]
 [0.5625     0.342857   0.16214279]]
```

Figure 4: Example of a sequence of "Still With You" by JungKook MIDI NOTE format. All values are between 0 and 1.

LSTM:

LSTM was invented by Sepp Hochreiter in 1997. It stands for Long Short-Term Memory. It is invented to solve vanishing gradient problems. Construction of the original LSTM is input gate, output gate, forget gate and cell. The forget gate intends to teach to remove resources by resetting.[29] The architecture of LSTM is composed of sub-network (memory blocks). Three gates (input, output, and forget) adjust the flow of information related to the cell while the model remembers values over time intervals.[30]

It is a neural network architecture. This deep learning model is used in many fields. It is useful for training time series data. It is an advanced step of RNN. It prevents forgetting previous data.[31] The author uses LSTM. Most of the References and related work use LSTM to train MIDI files in composition, and LSTM is valuable data for time series.

Model:

```
Model: "model"
_________________________________________________________________________________________
 Layer (type)                Output Shape              Param #   Connected to
=========================================================================================
 input_1 (InputLayer)        [(None, 25, 3)]           0         []

 lstm (LSTM)                 (None, 128)               67584     ['input_1[0][0]']

 duration (Dense)            (None, 1)                 129       ['lstm[0][0]']

 pitch (Dense)               (None, 128)               16512     ['lstm[0][0]']

 step (Dense)                (None, 1)                 129       ['lstm[0][0]']

=========================================================================================
Total params: 84354 (329.51 KB)
Trainable params: 84354 (329.51 KB)
Non-trainable params: 0 (0.00 Byte)
```

Figure 5: The model summary of learning MIDI file. Every dense is connected to LSTM.

It also helps the model remember previous data to learn MIDI NOTE more productively. Three variables are added as inputs: step, pitch, and duration (Figure 5). I gave temperature parameters 1 and 2 to demonstrate the effectiveness of temperature on the music composition. The temperature parameter tells the model when it composed the music using only model knowledge or credibility.

Results and Discussion

K-POP data organize:

Usually, a song is about 3 minutes long. However, since each song has a different length, I cut it into 60 seconds to make it the same length for productive training. The model will learn every song for 60 seconds. Sixty seconds is the minimum time that includes the song's verse and chorus. In addition, 60 seconds also consists of the hook.

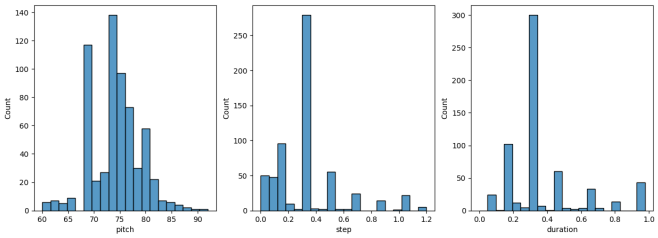

Figure 6: It is "Still With You" by JungKook's plot distribution. It has three graphs: pitch, step, and duration. Step and duration have the same highest count, at 0.3.

The Figure 6 graph shows the song "Still With You" as a composition of pitch, step, and duration. The major pitch is between 70 to 80. The average step is about 0.3. The average duration is also about 0.4. The song features the same step and duration length.

Temperature Parameter and Train Model:

The temperature parameter is used to manage the shape of the probability distribution. The low temperature intends to advance the quality of generation results. On the other hand, it creates repeating issues.[32]

With 50 epochs, the model did train. While model training MIDI notes, to prevent overfitting, it is coded to stop learning if a certain amount of loss continues to rise (Figure 7).

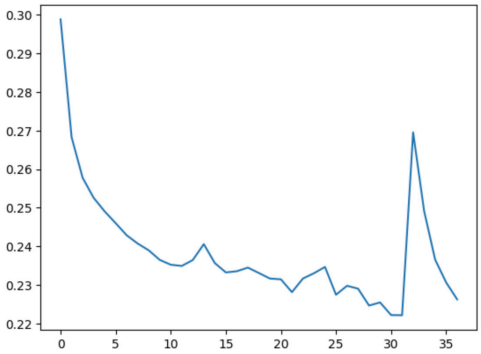

Figure 7: Train the loss history of the model that stops learning. Due to the loss increase at epoch 30, it stops learning at epoch 35.

Result of composed song:

The model's output is MIDI NOTE. The generated song should be converted to MIDI to hear it. Compared to the original song, the melody became simple. Before training the model, when MIDI files were converted to Note and changed into MIDI, there was a similar problem. The sound was more direct. Moreover, even though all songs didn't have a high pitch, the resultant music had a high pitch. Like Figures 8 and 9, most notes had a high pitch and kept repeating.

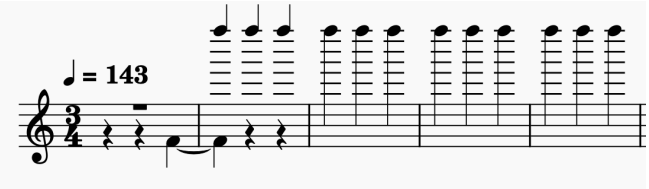

Figure 8: MIDI output music sheet when temperature is 1. It has high pitches.

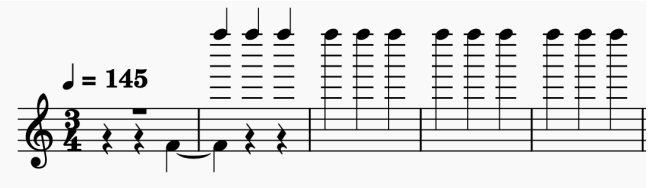

Figure 9: MIDI output when temperature is 2. It shows notes similar to those in Figure 8 but has two more notes in the sheet.

Figures 8 and 9 are music sheets of output MIDI files from the Model. Those didn't make a big difference in the beginning part.

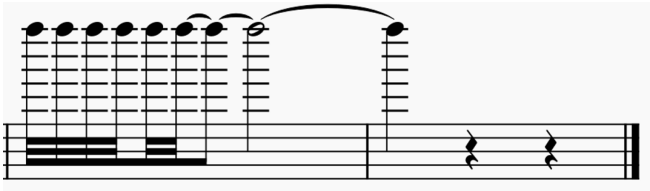

Figure 10: End of MIDI output's music sheet when temperature is 1. The notes are repetitive.

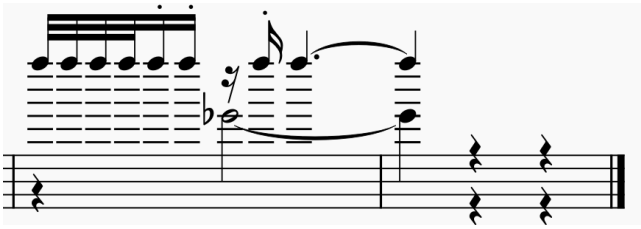

Figure 11: The end of the MIDI output's music sheet when the temperature is 2. It has an extra two rests and notes.

However, the bottom of the music sheet and the end of the music remained the same. In temperature 2, music has modulation (Figure 10, 11).

One K-POP Idol group music composition:

When the model learned 154 K-POP songs, the melody was too simple, and it didn't feel like catching the features of K-POP. There is a hypothesis because it learned too many different genres in K-POP, such as dance-pop or R&B. Thus, the hypothesis is that if only one group is trained, the song will show its own group's characteristics and K-POP characteristics more. Therefore, the model was learned using Black Pink and BTS group songs.

BLACKPINK:

The color of the group's music is evident; most of BLACKPINK's songs were written by the same writer, TEDDY, so the similarity of the songs is higher than that of other groups. In the case of (G)-IDLE, similar to BLACKPINK, Jeon So-Yeon, leader of (G)-IDEAL, wrote and composed all the title songs, so each music has a similar color, but there is a limit to the data that can be collected, so BLACKPINK was chosen. Nineteen songs were learned in the same condition and method as all K-POP songs.

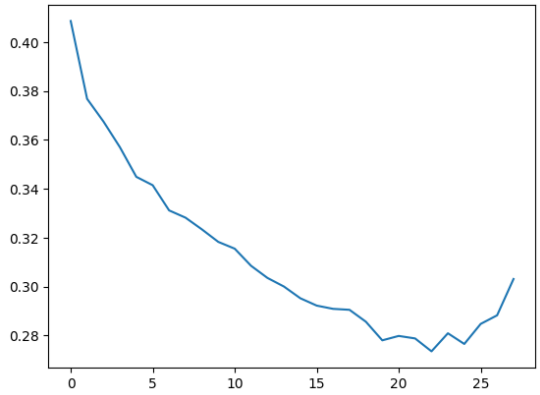

Figure 12: Model train loss history when train BLACKPINK data. After about 22 epochs, the loss increased, and when the epoch is over 27, it stops learning.

While training the songs, when epochs are about 23, it shows the loss is rapidly rising. As the code is set, it stops learning due to the rising of the loss (Figure 12).

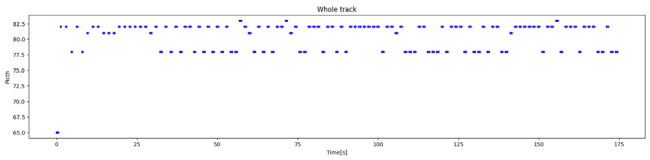

Figure 13: BLACKPINK model output plotting graph when the temperature is 1. It has high-pitch classes.

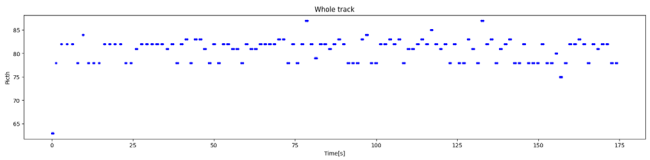

Figure 14: BLACKPINK model output plotting graph when the temperature is 2. It has different pitch classes compared with Figure 21.

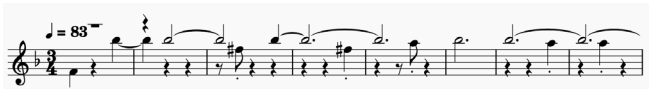

Figure 15: The BLACKPINK model outputs a music sheet when the temperature is 1. It has more rest than when the temperature is 2.

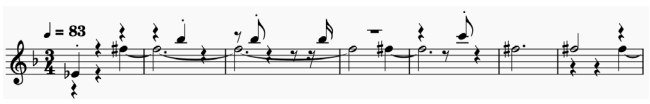

Figure 16: BLACKPINK model output music sheet when the temperature is 2. It has the same number of notes as when the model temperature is 1.

Figures 13 and 14 are MIDI files that the model composed. Unlike when training 154 K-POP songs, there is a difference between temperatures of 1 and 2, and the melody is more abundant. In Figure 14, the note has a higher pitch and less

DOI: 10.36838/v7i7.14

notes than in Figure 13. Figure 14 has the highest pitch, which is about 84. The song's beginning melodies resemble BLACKPINK's "How You Like That." Figures 15 and 16 are music sheets of Figures 13 and 14. They have similar melodies and the same number of notes in the sheets.

BTS:

BTS has a clear music color. Their recent music genre is dance pop. I collected more MIDI files than other groups. BTS was a group that had enough data to train. It has 15 songs. The method and condition of learning are the same as training all 154 K-POP songs.

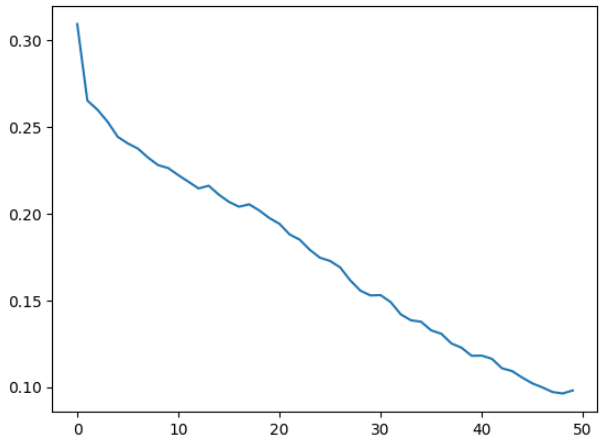

Figure 17: Model train loss history when training BTS data. It is an ideal train loss history. When the epoch is 2, the loss dramatically drops.

While training for a BTS song, there was no stopping. The loss kept decreasing (Figure 17).

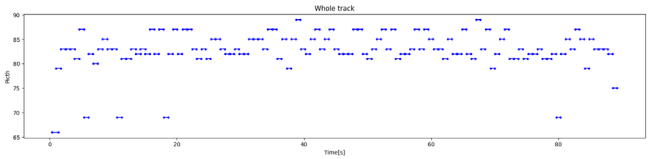

Figure 18: BTS model output plotting graph when the temperature is 1. It is a similar melody when the temperature is 2.

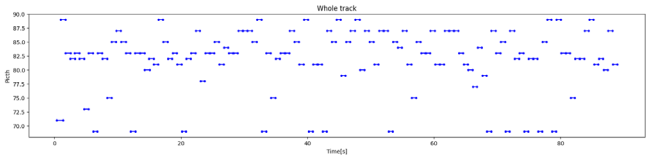

Figure 19: BTS model output plotting graph when the temperature is 2. It shows the whole track of model output. It has a smaller pitch range than when the temperature is 1. The pitch class range is 70 to 90 compared to 65 to 90 when the temperature is 1.

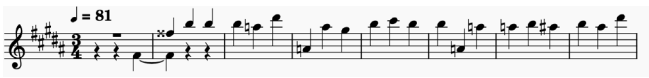

Figure 20: BTS model output music sheet when temperature is 1. It is four-thirds beat.

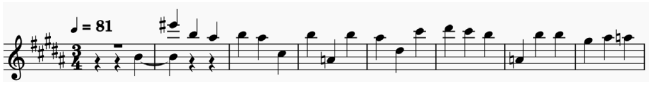

Figure 21: The BTS model outputs a music sheet when the temperature is 2. It has a similar melody to when the temperature is 1. They have the same number of rests and notes.

Figures 18 to 21 show a music model generated depending on the temperature. Figure 19 has more notes than Figure 18. The main melodies and duration of the notes are mostly similar, but the range of pitches and number of notes were different. Figure 18 has a wide range of pitches from 65 to 90.

When the temperature parameter is high, the melody and notes are more plentiful and feel like real music. Rather than training too many diverse genres simultaneously, training the same or similar genres simultaneously is more effective in writing music than humans.

Like collaborative generative AI technology and K-POP music, it is expected that other fields can be incorporated. With these technologies, society can expect to develop by combining each field.

However, commercializing this technology would take much work due to data limitations. While converting the MIDI file to NOTE and converting MIDI again, the song became simple.

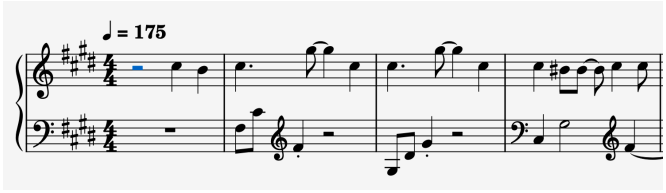

Figure 22: Music sheet of the original MIDI ("Still With You" by JungKook). It is double-handed.

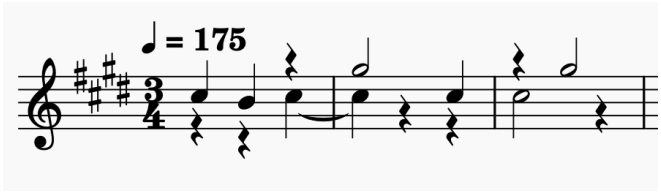

Figure 23: Music sheet after MIDI NOTE to MIDI ("Still With You" by JungKook). It has the same number of notes after it changed into single-hand sheet music.

Like Figure 22, the original MIDI file has two hand sheets of Music, but after converting the MIDI note and changing it into MIDI format, it became single-hand sheet music (Figure 23). Only the main melody of the song was extracted. Moreover, even though the model studies songs for 60 seconds, sometimes, the result is much shorter than that.

■ Conclusion

This paper created a generative K-POP music composition model by LSTM. However, when converting to a MIDI file, music becomes simplified. The MIDI file that the model generates comprises only one instrument, the Piano. Thus, if a song has more instruments, the song becomes more complete. Due to the limited availability of MIDI files for K-pop songs, I was only able to train the model on 154 songs. However, in the future, I plan to increase the size of the dataset by convert-

ing commercially available K-pop songs into MIDI format. In addition, I will improve the learning algorithm to maximize training efficiency even with a relatively small number of songs.

References

1. Guzman, A. L. "Making AI safe for humans: A conversation with Siri." In Socialbots and their friends; Routledge, 2016, pp 85-101.
2. Yoo, C.-R.; Kim, S.-H.; Kim, J.-W. "A Comparative Study of the Use of Intelligent Personal Assistant Services Experiences: Siri, Google Assistant, Bixby." Sci. Emot. Sensibility 2020, 23(1), 69-78.
3. Sarah. Help US Keep Comments Respectful – New Community Guidelines Comment Reminders - Youtube Community. https://support.google.com/youtube/thread/86685658?hl=en (accessed 2024-01-31).
4. How Instagram Uses Artificial Intelligence to Moderate Content. https://help.instagram.com/423837189385631 (accessed 2024-01-31).
5. Midjourney Quick Start Guide. https://docs.midjourney.com/docs/quick-start (accessed 2024-01-31).
6. OpenAI. Introducing Chatgpt. https://openai.com/blog/chatgpt#OpenAI (accessed 2024-01-31).
7. Alexa.https://www.amazon.com/b?node=21576558011&amp=&ref_=alxcom_lrnmore_btn_23 (accessed 2024-01-31).
8. Peter Langston. "Six techniques for algorithmic music composition." Proceedings of the International Computer Music Conference. 1989, Vol. 60, Citeseer.
9. De Mantaras, R. L.; Arcos, J. L. "AI and music: From composition to expressive performance." AI Magazine 2002, 23(3), 43-43.
10. Zong Woo Geem, Jeong-Yoon Choi. "Music composition using harmony search algorithm." Workshops on Applications of Evolutionary Computation. 2007, Springer Berlin Heidelberg.
11. Richard Fox, Adil Khan. "Artificial intelligence approaches to music composition." Proceedings of the International Conference on Artificial Intelligence (ICAI). 2013, The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
12. Jin Ha Lee et al. "K-pop genres: A cross-cultural exploration." Proceedings of the 14th Conference of the International Society for Music Information Retrieval (ISMIR). 2013, The International Society for Music Information Retrieval (ISMIR).
13. Chien-Hung Liu, Chuan-Kang Ting. "Computational intelligence in music composition: A survey." IEEE Transactions on Emerging Topics in Computational Intelligence 2016, 1(1), 2-15.
14. Vasanth Kalingeri, Srikanth Grandhe. "Music generation using deep learning." arXiv preprint arXiv:1612.04928 2016.
15. Keunwoo Choi, George Fazekas, Mark Sandler. "Text-based LSTM networks for automatic music composition." arXiv preprint arXiv:1604.05358 2016.
16. Patrick A. Messerlin, Wonkyu Shin. "The success of K-pop: How big and why so fast?." Asian Journal of Social Science 2017, 45(4-5), 409-439.
17. Jean-Pierre Briot, François Pachet. "Music generation by deep learning-challenges and directions." arXiv preprint arXiv:1712.04371 2017.
18. Moruzzi, Caterina. "Creative AI: Music composition programs as an extension of the composer's mind." Philosophy and Theory of Artificial Intelligence 2017; Springer International Publishing, 2018.
19. Docevski, M.; Zdravevski, E.; Lameski, P.; Kulakov, A. "Towards music generation with deep learning algorithms." 2018.
20. Sunghoon Lee. "Artificial Intelligence Applications to Music Composition." The journal of the convergence on culture technology 2018, 4(4), 261-266.
21. Zulić, Harun. "How AI can change/improve/influence music composition, performance and education: three case studies." INSAM Journal of Contemporary Music, Art and Technology 2019, 1 (2), 100-114.
22. Emma Frid, Celso Gomes, Zeyu Jin. "Music creation by example." Proceedings of the 2020 CHI conference on human factors in computing systems. 2020.
23. Hernandez-Olivan, Carlos; Beltran, Jose R. "Music composition with deep learning: A review." Advances in speech and music technology: computational aspects and applications 2022, 25-50.
24. Bian, W.; Song, Y.; Gu, N.; Chan, T. Y.; Lo, T. T.; Li, T. S.; Wong, K. C.; Xue, W.; Trillo, R. A.; "MoMusic: A motion-driven human-AI collaborative music composition and performing system." Proceedings of the AAAI Conference on Artificial Intelligence 2023, 37 (13).
25. Song. "A Study on the Improvement of Korean Learning Using K-POP." Master Degree Thesis, Dongshin University, 2014.
26. Zölzer, U., Ed. DAFX--Digital Audio Effects; John Wiley and Sons, LTD: West Sussex, England, 2002; http://www.dafx.de/.
27. O'Shaughnessy, D. Speech Communication; Addison-Wesley: Reading, MA, 1987.
28. Jang, Yoonhee. "Analysis Features of Famous K-POP Songs by Librosa." Korea Computer Congress 2023, 2010-2012.
29. Gers, Felix A.; Schmidhuber, Jürgen; Cummins, Fred. "Learning to forget: Continual prediction with LSTM." Neural computation 2000, 12 (10), 2451-2471.
30. Van Houdt, Greg; Mosquera, Carlos; Nápoles, Gonzalo. "A review on the long short-term memory model." Artificial Intelligence Review 2020, 53, 5929-5955.
31. Kim, Jae-Chun. "Detection of the Optimum Spectral Roll-off Point using Violin as a Sound Source." Journal of the Korea Society of Computer and Information 2007, 12 (1), 51-56.
32. Holtzman, A.; Buys, J.; Du, L.; Forbes, M.; Choi, Y. "The curious case of neural text degeneration." arXiv preprint arXiv:1904.09751, 2019.

Author

Yoonhee Jang is a grade 11 Saint Johnsbury Academy Jeju student interested in K-POP, AI, and education. She is interested in STEM education, which significantly improves students' concentration in rapidly developing societies, such as LLM. She hopes to learn multiple subjects related to teaching at university.

DOI: 10.36838/v7i7.14