

# FINDING TORI: SELF-SUPERVISED LEARNING FOR ANALYZING KOREAN FOLK SONG

Danbinaerin Han<sup>1</sup>

Rafael Caro Repetto<sup>2</sup>

Dasaem Jeong<sup>1</sup>

<sup>1</sup>Department of Art & Technology, Sogang University

<sup>2</sup>Kunstuniversität Graz

naerin71@sogang.ac.kr, rafael.caro-repetto@kug.ac.at, dasaemj@sogang.ac.kr

## ABSTRACT

In this paper, we introduce a computational analysis of the field recording dataset of approximately 700 hours of Korean folk songs, which were recorded around 1980-90s. Because most of the songs were sung by non-expert musicians without accompaniment, the dataset provides several challenges. To address this challenge, we utilized self-supervised learning with convolutional neural network based on pitch contour, then analyzed how the musical concept of *tori*, a classification system defined by a specific scale, ornamental notes, and an idiomatic melodic contour, is captured by the model. The experimental result shows that our approach can better capture the characteristics of *tori* compared to traditional pitch histograms. Using our approaches, we have examined how musical discussions proposed in existing academia manifest in the actual field recordings of Korean folk songs.

## 1. INTRODUCTION

Folk songs are considered as a musical language. Not only that, but they are also regarded as the root of all traditional music. Folk songs are potent embodiments of a region's cultural and linguistic characteristics, serving as a foundation for all artistic and musical developments since their inception. It is believed that research for folk music can provide new inspiration for existing art music research, and facilitate an interdisciplinary approach that encompasses both music and other fields.

In Korean musicology, there have been ongoing discussions aimed at identifying regional characteristics in folk music. Through numerous debates, *tori* is used as the most representative theory in many studies. Even though the use of *tori* as a term to describe the musical characteristics of Korean folk music is widespread and its utility is well acknowledged, the existing *tori* classification methods were unable to fully explain the features of the actual music. Various scholars have defined the different *tori* according to musical characteristics such as scales (intervals between notes), primary notes, ornamentation, ending note and id-

iomatic melodic patterns in folk songs, leading to a refinement process for the classification of *tori*. So far, there are still ongoing opinions, controversies, and discussions concerning the existing *tori* [1, 2].

In order to conduct valuable discussions on analyzing folk songs, the relationships within the audio must be examined and systematically shared. However, checking numerous audio files individually is extremely time-consuming and difficult, so previous research has focused on analyzing small amounts of audio. The most common method used by musicologists has been transcription, which involves listening to the music and notating it in a specific music notation system. However, the task of transcribing orally transmitted music into a music notation system has inherent limitations [3], such as quantizing the pitch and rhythmic features of folk songs.

For these reasons, there have been several research on analyzing folk music empowered with methods derived from music information retrieval (MIR) on audio recording. For example, Indian *raga* music has been analyzed based on first-order pitch distributions [4], and for traditional three-part Georgian singing, F0-based tonal analysis [5] or development of tuning systems [6] has been introduced. [7] also presented analysis using audio-signal processing on Flamenco and Arab-Andalusian vocal music.

In this paper, we take a computational approach to analyze a vast corpus of Korean folk songs, utilizing deep-learning-based methodologies. Our primary aim is to investigate the connection between conventional musicological classifications of Korean folk songs, *tori*, and the actual field recordings. Through a comparison of our analytical results with established musicological frameworks, our objective is to offer insights on identifying meaningful clustering or distinguishable features among Korean folk songs. We also share our code and metadata that contains links to original audio with our manual *tori* labels for 218 songs<sup>1</sup>.

## 2. KOREAN FOLK SONGS AND TORI

### 2.1 Dataset

The “Anthology of Korean Traditional Folksongs” is an audio collection consisting solely of traditional Korean folk songs [8]. As part of a cultural project conducted by Munhwa Broadcasting Corporation (MBC) from 1989



© D. Han, R. Caro Repetto, and D. Jeong. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

**Attribution:** D. Han, R. Caro Repetto, and D. Jeong, “Finding Tori: Self-supervised Learning for Analyzing Korean Folk Song”, in *Proc. of the 24th Int. Society for Music Information Retrieval Conf.*, Milan, Italy, 2023.

<sup>1</sup> <https://github.com/danbinaerinHan/finding-tori>

to 1995, the folk song collection was compiled under the direction of Sang-il Choi. The available audio contains 15,861 songs, with approximately 700 hours in total length.

The audio was field-recorded across 153 city/county and 1,010 villages in South Korea. The metadata accompanying the recordings includes title, machine-readable lyrics, regions, recording dates, and the singer’s name and age. The region represents the administrative districts of South Korea at the time of recording, which consists of nine categories in total. Due to its extensive audio corpus and detailed accompanying metadata, this dataset has become a crucial resource for research on Korean folk music within the domestic academic community [9–11]. We obtained the audio material through crawling the original website,<sup>2</sup> which has been hosted by MBC since April 2022, where every audio and metadata is openly published. We received official approval for using these data for research purposes from MBC.

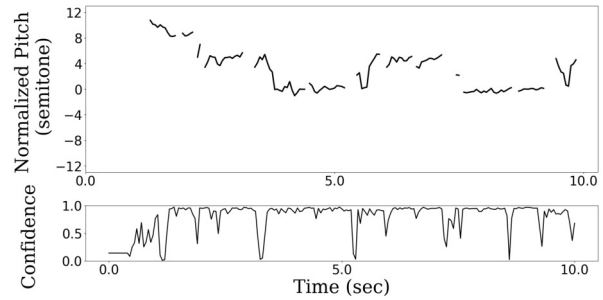
## 2.2 Tori

Tori was proposed by a Korean musicologist named Bo-hyeong Lee in the early 1980s to explain the musical characteristics of regional folk songs [12, 13]. Prior to the development of tori theory, Korean music academia used the terms *jo* and *cheong* to explain the musical features of folk songs, based on the tonal center and intervals between notes [14]. However, Bo-hyeong Lee argued that *jo* and *cheong* are insufficient in capturing the unique musical characteristics of Korean folk music. Therefore, he proposed the tori classification system as an alternative approach, which covers primary notes, ornamentation, ending note and idiomatic melodic patterns.

The most widely-used tori classification method divides songs into four categories. We encourage readers to refer Figure 3, which presents scale of three tori with staff notation. The pitch name in this section is used as a conventional representation, not absolute pitches. **Gyung-tori** utilizes five notes and is mainly sung in the capital region. It often uses a gentle vibrato overall and finessed ornamental melody. **Menari-tori** is widely distributed throughout the eastern regions and the entire Korean Peninsula. When the melody ascends, it leaps through notes, while during the descending melody, it comes down through passing notes. **Yukjabaegi-tori** is commonly found in the southern regions of the Korean Peninsula. It is characterized by a thick and vibrant vibrato in G note and passing briefly through E $\flat$  in a descending melody from E $\flat$  to D. **Sushimga-tori** commonly appears in the northwestern region of the Korean Peninsula, characterized by its unique vibrato inflecting upward.

## 2.3 Tori annotation

As the dataset did not include any tori labels, one of the authors, who has over 10 years of experience in Korean traditional music, selected a subset of 218 high-quality record-



**Figure 1.** Example of pitch contour from the dataset extracted by CREPE. F0 value under confidence of 0.8 was masked out

ings and manually annotated them in terms of tori classification.

To create this tori-subset, we focused on identifying clear musical characteristics present in the tori classification method, such as pitch scale, ornamentation, and idiomatic melodies, rather than considering the audio’s recorded region. We found that there are not many instances of sushimga-tori in the audio dataset, as the dataset predominantly consists of folk songs from the central and southern regions of the Korean Peninsula. Also, we observed that songs belonging to menari-tori appeared not only in the eastern region but also nationwide. In conclusion, we mainly focused on the remaining three types of tori, *gyung-tori*, *yukjabaegi-tori*, and *menari-tori*. Finally, audio recordings that were closer to speech or chanting (non-musical) and did not exclusively belong to any other subdivided tori types were labeled as ‘others’. We have labeled totally 218 songs: 65 gyung, 73 menari, 49 yukja, and 31 others.

## 3. METHODOLOGY

In this study, we take an approach of representing the musical characteristics of a given folk song using a high-dimensional single vector, which can also be called an embedding of the song. If this embedding accurately represents the musical characteristics of tori, it can be utilized for various purposes including tori classification, song similarity searches and clustering songs in the corpus based on tori similarity.

Even though the dataset is provided with audio recordings, we focus on the contour of fundamental frequency (F0) rather than using audio directly. By doing so, our embedding could better capture the melodic features of the songs rather than timbral characteristics such as the dialect of the singer. To extract the F0 contour from each audio recording, we utilized CREPE [15], a widely-used CNN-based model. It extracts F0 value for every 10 ms as well as confidence score ranging from 0 to 1. Figure 1 illustrates an example of an extracted pitch contour.

### 3.1 Pitch Histogram

Since the characteristics of a musical system can be largely studied from the pitch distribution of its melodies, there

<sup>2</sup> <http://urisori.co.kr/urisorien/en/doku.php/>

has been numerous research using pitch histograms to analyze traditional musics from different culture, such as for Indian Carnatic music [16], Turkish Makam [17], Arab-Andalusian music [18], or Iranian *dastgāhi* music [19]. Therefore, we applied a similar approach to Korean folk music.

One of the important features one has to know to exploit pitch histogram is the tonic of the song so that one can normalize the pitch histograms of each song into relative pitch, rather than directly using absolute pitch value. During our preliminary experiment, we found that the tonic of each recording in the dataset is usually the most frequently appearing pitch in the song. Therefore, we determined the tonic by counting the number of appearances of each pitch in terms of time frames, utilizing a 100 cents range to count pitch (i.e. MIDI pitch 60.49 and 59.51 are counted as the same pitch). As the center pitch may differ from the typical 440 Hz tuning, we identified the best matching pitch by adjusting the pitch center by increments of 10 cents and selecting the pitch that showed the maximum pitch count.

### 3.2 CNN contour encoder

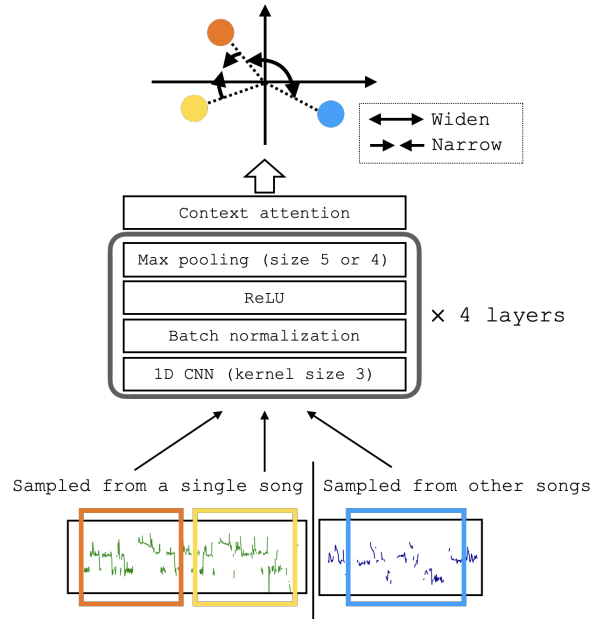
While the pitch histogram can offer a comprehensible representation of each song, it cannot fully capture the concept of tori as it does not consider the relationship between each note in the scale. Additionally, because most of the recordings were sung by non-professional singers without instrumental accompaniment, the intervals of singing and tonic frequency has significant amount of noise. Therefore, we employ a convolutional neural network (CNN) to learn the representation of a given pitch contour. This approach allows us to better capture the underlying musical characteristics of each song in relation to tori classification.

The problem is setting the training objective of the CNN model. One option is to train the model in a supervised manner, as typically done in other MIR tasks, such as classification tasks with an annotated dataset. In our dataset, region labels, which represents the administrative region where the recording took place, can be used for training labels. It can be regarded as a reasonable approach considering that tori has strong correlation with the regional characteristics of each folk song. However, we discovered several different types of tori in a single recorded region, which would make it difficult to learn distinguishable musical characteristics only using the region label.

To address this issue, we propose to adopt a self-supervised representation learning, so that we can obtain a representation of given pitch contour that is consistent within a song without extra annotated labels as shown in Fig. 2. This approach of exploiting intra-song similarity has been widely used for music audio representation learning [20,21]. We use triplet loss with hinge margin as Equation 1,

$$L = \max(0, m - \text{Sim}(v_a, v_p) + \text{Sim}(v_a, v_n)) \quad (1)$$

where  $m$  denotes hinge margin,  $\text{Sim}(v_a, v_{p,n})$  represents cosine similarity between anchor vector  $v_a$  and positive vector  $v_p$  or negative vector  $v_n$ .



**Figure 2.** Self-supervised learning with triplet loss using cosine similarity, with four layers of convolutional neural network

As presented in Fig. 2, the model consists of 4 layers of 1D convolution layer, each with a kernel size of three and channel size of 64, 128, 256, and 256, respectively, each followed by batch normalization layer. We used a max pooling layer between each convolutional layer with size 5 for the first layer and 4 for the second and third layers. On top of the convolutional stack, we used context attention to summarize the arbitrary length of vectors, which was initially proposed in hierarchical document classification [22], with a slight modification of multi-head attention weight [23] instead of single-head. Context attention is a type of attention mechanism that uses an independent learnable parameter named context vector as a query while using the input vectors as key and value. The final embedding size of the model was 256.

The input is a pitch contour in the shape of  $T \times 2$ , where  $T$  denotes the number of time step of pitch contour. Throughout the experiment, we used the frame rate of 20 Hz, which means that 30 seconds of pitch contour is converted to 600 time steps. During the training, we randomly slice the contour to have 30 seconds length, while using the entire contour in the test or visualization. Two channels are tonic-normalized MIDI pitch and confidence of the F0 estimation at that time frame. Since the automatic F0 estimation includes noise, especially in the unsung silent part, we masked the estimated pitch value to zero if the confidence is lower than 0.8.

## 4. EXPERIMENTS

### 4.1 Evaluation Design

In the experiments, we explored the quantitative relation between tori and different embedding schemes to see which type of embedding can explain the concept of tori

better. To evaluate the correlation between tori and the given embedding scheme, we exploit our tori-labeled subset with two metrics.

The first metric is the ranking of cosine similarity. If the embedding shares important characteristics with the concept of tori, we can expect that songs with high cosine similarity in the given embedding space share the same tori with the query song. Therefore, we calculated the cosine similarity between each song in the tori-subset, and obtained normalized discounted cumulative gain (nDCG), which is a frequently used metric to evaluate the quality of search results. If all the other  $n$  songs with the same tori are ranked within  $n$ -th order in the similarity, nDCG becomes one.

The second metric is the tori classification accuracy using a random forest classifier. If the embedding includes essential characteristics that define tori, one can expect that it can be exploited to classify the tori for a given song. Among many options, we used a simple random forest classifier. The random forest classifier with 100 trees was trained with 75% of the tori-set and tested with the remaining 25%. For each embedding scheme, we repeated the procedure 30 times with different train/test split and reported mean and deviation of accuracy to ensure that the results were not dependent on the specific dataset split.

## 4.2 Training

While a pitch histogram can be obtained from a given song without additional training data, CNN models require training procedures. We employ F0 contour extracted from the ‘‘Anthology of Korean Traditional Folksongs’’ to train our model.

Some of the songs in the dataset were recorded with multiple singers or percussion accompaniment. Because these can degrade the performance of F0 estimation, we excluded them while training the CNN model. Instead of manually filtering the dataset, we used a CNN-based sound event detection model [24] to calculate the activation of ‘choir’ and ‘percussion,’ which are included among the model’s event vocabulary. We filtered the song by the maximum activation value exceeding a certain threshold, which was manually decided by observing the activation distribution across the dataset.

The CNN model was trained for 25,000 updates. For the self-supervised learning using triplet loss, we used eight negative samples with one positive sample, and hinge loss using a margin of 0.4. For the region-supervised training, we use cross-entropy loss with class weight to address the class imbalance issue. The architecture of the CNN remained the same, but we added a fully connected layer to project the embedding to logits for nine different region categories. We used the Adam optimizer with an initial learning rate of 0.001 and batch size of 128. The entire training process was conducted on a single RTX A5000, which took approximately 10 minutes to complete 25,000 updates. We validated the model’s training procedure and hyperparameters using a specific split of the tori-subset and chose the simplest setting to avoid overfitting the hyperpa-

Embedding	Similarity Ranking	Random Forest Classifier
	nDCG	Accuracy
Hist. (25 bin)	0.783	0.744 $\pm$ 0.058
Hist. (124 bin)	0.777	0.722 $\pm$ 0.054
CNN (region )	0.792	0.634 $\pm$ 0.055
CNN (SSL)	<b>0.853</b>	<b>0.848</b> $\pm$ 0.039

**Table 1.** Experiment results on the tori subset. Hist. denotes normalized pitch histogram, specified with the number of bins to cover two-octave range. region and SSL denotes the region-supervised and self-supervised learning.

rameters to the small subset.

## 4.3 Results

The evaluation result of each embedding scheme is presented in Table 1. As tori is closely related to the use of pitch, the pitch histogram showed about 74 % accuracy when combined with a random forest classifier, and 0.783 or 0.777 nDCG in the similar song search. The result shows that increasing the resolution of the histogram does not achieve better performance. The performance of CNN embeddings trained with region classification was worse than the pitch histogram in classification accuracy. However, the CNN embedding trained with self-supervised learning showed the best nDCG and classification accuracy, even though both CNN shares the same architecture except the final projection layer.

The performance gain compared to the pitch histogram can be caused by the fact the concept of tori includes not only the pitch scale the song uses but also ornamentation and idiomatic phrases it, which can be easily captured by contour CNN.

The result also shows that the region label did not help to learn tori-related characteristics. Even though the concept of tori is strongly related to each region, folk songs were widely spread nationwide at the time of the recording, which made it difficult to learn coherent musical characteristics from the region label.

It is also worth noting that self-supervised learning did not use any other musical knowledge that is related to tori, except that it was trained to extract constant embedding throughout a given song regardless of the segment position. From a machine learning point of view, the model has to extract an embedding that can explain the entire song or distinguish it from other songs from a given fragment. The high correlation between the trained embedding and tori labels shown in the evaluation implies that the concept of tori is clearly related to the distinguishable characteristics of each Korean folk song.

## 5. ANALYSIS AND DISCUSSION

In this section, we introduce interesting musicological findings that we have found from the pitch histogram and embedding learned from self-supervised training.

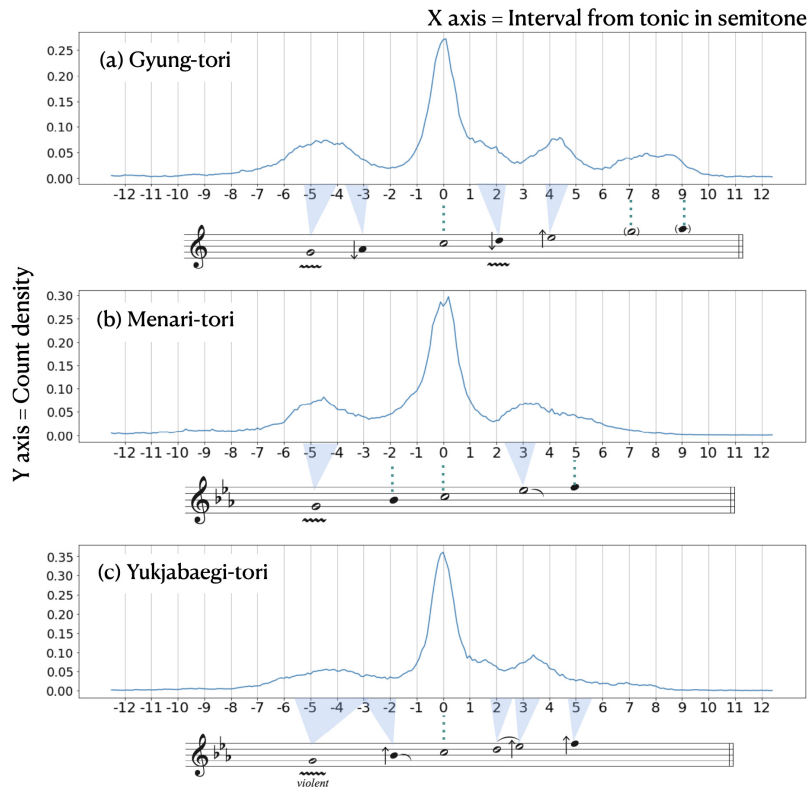


Figure 3. Average pitch histogram of each tori obtained from the tori-subset and corresponding scale description.

### 5.1 Pitch Histogram Analysis

To explain the scale and characteristics alongside with pitch histogram of each tori, we present the staff notation of each tori in Figure 3 based on various studies describing tori [2, 25, 26]. In the score, the *cheong* (a main note that generally appears most frequently) was fixed as the C5. The unfilled notes are primary tones, and the filled notes are passing tones. In cases where there is a vibrato mark below the note, it is called *yoseong*, meaning the pitch generally oscillates up and down. In addition, the arrow on the left side of the note indicates that the pitch is slightly raised or lowered compared to the equal-tempered pitch. The slur mark flowing down to the right of the note represents *twoe-seong*, which means the pitch slides down.

While menari-tori and yukjabaegi-tori remain within a perfect fifth range from the center, in gyung-tori, there is a larger proportion of high notes (G5-A5) or more from the center. This indicates that the distribution of notes from low to high in gyung-tori is evenly spread. This could be due to the fact that folk songs of gyung-tori are more commonly found in popular folk songs than in those that appear nationwide. Perfect 4th below the center tone is common in all Korean music. However for example in gyung-tori, there is vibrato in the G4 note, and the A4 note has a low pitch, resulting in a distinct distribution in that part. In menari-tori, it can be observed that the B $\sharp$ 4 note is used as a passing tone between C5 and G4. Yukjabaegi-tori is characterized by a vibrant vibrato in the G4, resulting in a gentle distribution in the nearby pitch than other tori.

Whether a major third or a minor third is used in the

scale is an important factor in distinguishing the regional musical characteristics of the eastern and western parts of the Korean Peninsula in the traditional method. Subtle pitch differences can be observed in all three tori. In gyung-tori, the D5 note is clearly distributed lower and the E5 note is definitely higher than the line appearing the equal-tempered tune. In yukjabaegi-tori, it is also confirmed that the E $\flat$  note is slightly higher than the minor third note's

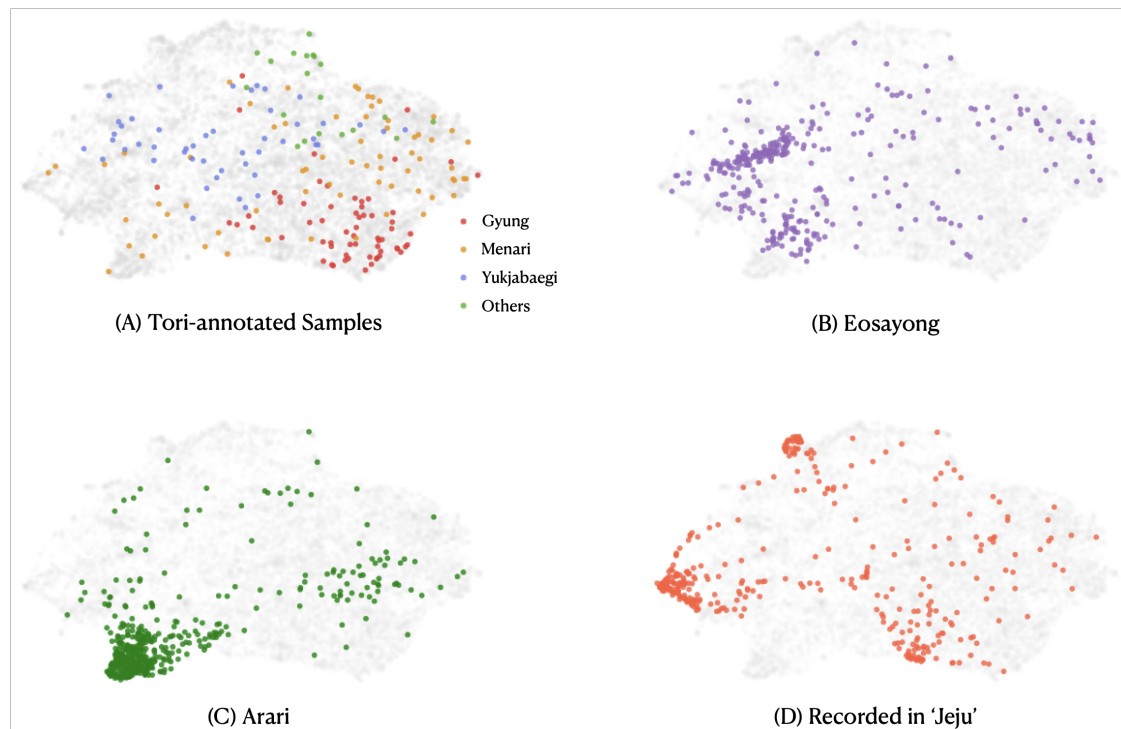
Through the above analysis, we found that pitch histograms can be utilized to roughly identify musical characteristics, such as the range, interval from tones and subtle pitches, etc. However, we also acknowledged that this approach solely exhibits the constituent pitches and their frequencies, without accounting for other aspects like uses of melody and ornamentation and more.

### 5.2 UMAP visualization

We visualized the embeddings learned through self-supervised learning in 2D using UMAP, a dimension reduction technique that is frequently used for visualization of high-dimensional vectors in Figure 4. In the first figure, we demonstrated the distribution of four different tori labels. In addition, by examining the metadata information of the embeddings, we figured out several interesting topics from this visualization. We also implemented a web demo using this visualization where one can directly access the corresponding audio recording<sup>3</sup>.

Menari-tori has a much wider distribution area and en-

<sup>3</sup> <https://danbinaerinhan.github.io/korean-folksong-visualization/>



**Figure 4.** UMAP visualization of folk song embeddings obtained from our self-supervised-trained CNN model. Note that no tori label was used to train the model.

compasses a broader musical range. In Figure 4 (A), the menari-tori distribution can be seen to be wider compared to the other tori. We mentioned that several subdivided tori of it have been identified by scholars. Among them, eosayong-tori is a representative example of refining the musical characteristics of eastern folk songs [11]. Eosayong-tori, a lamenting song sung by lumberjacks, was suggested to have distinct musical characteristics compared to menari-tori. For instance, in eosayong-tori, the lowest pitch is a semitone higher, and it concludes on the lowest pitch of the scale instead of the final pitch and the middle pitch found in menari-tori. We examined embeddings corresponding to eosayong-tori in the metadata’s title information, and they were clearly gathered in a different space than the area annotated as menari-tori, as presented in Figure 4 (B).

Similar results can be observed in the following cases with *arari*. *Arirang*, having repetitive refrain lyrics such as “arirang” or “arari”, is the most representative Korean folk song, with countless versions appearing in various regions. Among them, *arari*, another name for *Jeongseon arirang*, was sung primarily in Gangwon province before spreading nationwide in the 1930s [27,28]. *Arari* is also regarded as a separated tori from menari-tori for some researchers [29], with its own slower tempo, monotonous skeleton tones, and the use of decorative tones. In Figure 4 (C), we can see the clustering of entities with the title *arari* extracted from the metadata.

Another interesting example is songs from Jeju. Folk songs sung from Jeju island, which is the largest island in Korea, have fewer research results compared to other regions. In Jeju folk songs, the interval between the pitches

is narrower compared to other provinces’. Due to these reasons, the traditional method of transcribing into staff notation was unsuitable for capturing the characteristics of Jeju music [30–32]. However, using our embeddings, we can identify three clusters as in Figure 4 (D). It implies the potential usefulness of our approach to uncover the unique characteristics of folk songs in Jeju Island.

We also checked the UMAP of the pitch histograms but could not find clear clustering as in the presented examples.

## 6. CONCLUSIONS

In this paper, we have presented our computational approach to obtaining a high-dimensional embedding vector for a given pitch contour, which allows us to analyze a vast amount of Korean folk songs. Using these embeddings and a manually labeled subset, we have examined how musical discussions proposed in existing academia are manifested in our dataset. As our results cover various music characteristics, the learned embeddings can be utilized as a monitoring aid when dealing with numerous undefined data to review and find the tori. Also, we have discussed the methods and possibilities for utilizing and interpreting experimental results in the future research. Through this material, we can review and refine our understanding of the concept of the tori, and provide easily accessible resources and utilize them as appropriate evidence. Furthermore, there is potential to clarify the musical characteristics of regions that are distinct from other regions, like Jeju, about which existing research has not been as active. Ultimately, we expect this research to be valuable in illuminating the relationships and transformations of folk songs over time.

## 7. ACKNOWLEDGEMENT

We wish to express our deepest gratitude to Sang-il Choi and his team for their invaluable work on the “Anthology of Korean Traditional Folksongs”. Their effort to collect more than ten thousand unique recordings from across Korea has not only significantly contributed to the preservation of the rich tapestry of Korean folk music but has also created a monumental resource for academic research. Without their extraordinary efforts, our work would not have been possible. We sincerely thank them for laying such a robust foundation and for enabling future researchers and us to build upon it. We also appreciate MBC for kindly allowing us to utilize the dataset for this research.

The web demo was implemented with the kind help of Dongmin Kim.

## 8. REFERENCES

- [1] E. J. Shin, “Reconsideration on the modes(tori) of korean folk song(한국 민요 선법 (토리) 의 연구 성과 검토 및 논점),” *The Society Of Korean Folk Song(한국민요학)*, vol. 46, pp. 153–195, 2016.
- [2] G. R. Sung, *Controversial Topics in the Research of Korean Music Theory, 한국음악이론 연구의 쟁점*. Publishing department of the academy of Korean studies(한국학중앙연구원출판부), 2020.
- [3] H. J. Kim, *Transcription and Interpretation of Korean Folk Songs(민요의 채보와 해석)*. Publish Company Minsokwon(민속원), 2013.
- [4] G. K. Koduri, S. Gulati, P. Rao, and X. Serra, “Rāga recognition based on pitch distribution methods,” *Journal of New Music Research*, vol. 41, no. 4, pp. 337–350, 2012.
- [5] S. Rosenzweig, F. Scherbaum, and M. Müller, “Detecting stable regions in frequency trajectories for tonal analysis of traditional georgian vocal music.” in *Proc. of Int. Society of Music Information Retrieval Conf. (ISMIR)*, 2019, pp. 352–359.
- [6] F. Scherbaum, N. Mzhavanadze, S. Rosenzweig, and M. Müller, “Tuning systems of traditional georgian singing determined from a new corpus of field recordings,” *Musicologist*, vol. 6, no. 2, pp. 142–168, 2022.
- [7] N. Kroher, E. Gómez, A. Chaachoo, M. Sordo, J.-M. Díaz-Báñez, F. Gómez, and e.-R. Mora, Joaquin", *Computational Ethnomusicology: A Study of Flamenco and Arab-Andalusian Vocal Music*, bookTitle="Springer Handbook of Systematic Musicology. Berlin, Heidelberg: Springer Berlin Heidelberg, 2018, pp. 885–897.
- [8] S. I. Choi, “Articles on recordings : ‘anthology of korean traditional folksongs’ about the project and the records published(창간 10 주년 기념호: 음반; ‘한국민요대전’사업과 음반 발간),” *Korean Recording Studies(한국음반학)*, vol. 10, pp. 459–480, 2000.
- [9] I. G. Bae, “Musicological study of the tori of chungcheong province : focusing on the rice field weeding folksongs contained in 『mbc hankuk minyo daejeon (the compilation of korean folksongs)』(충청도 토리를 찾기 위한 시론:[mbc 한국민요대전] 수록 논매기소리를 중심으로),” *The Society Of Korean Folk Song(한국민요학)*, vol. 12, pp. 159–184, 2003.
- [10] N. kyong Choi, “Distribution of folk songs and its musical characteristics(1) (민요의 종류와 음악적 특성의 분포 연구 (1)-모심는 소리, 노짓는 소리, 만선 풍장소리를 중심으로),” *Research Institute of Korean Studies(민족문화연구)*, vol. 41, pp. 85–127, 2004.
- [11] Y. W. Kim, ““a study of the yoengnam folk song, eosayong(嶺南民謠 어사용의 音組織 研究),” *The Society Of Korean Folk Song(The Society Of Korean Folk Song(한국민요학)*, vol. 6, pp. 45–131, 1999.
- [12] B. H. Lee, “Music and folk song of central and western region and tori(경서토리권의 무가, 민요),” *Collection of Korean Music Studies(나운영박사 회갑기념 한국음악논총)*, 1982.
- [13] —, “Music culture of the menari-tori folk song(메나리토리 무가 민요권의 음악문화),” *Koanthrol(한국문화인류학)*, vol. 15, pp. 233–249, 1983.
- [14] D. W. Baek, “The definition and interpretation of terminology of korean musicology, 한국음악학에서의 용어 개념 정의와 해석,” *National Folk Museum of Korea(민속학연구)*, no. 14, pp. 193–230, 2004.
- [15] J. W. Kim, J. Salamon, P. Li, and J. P. Bello, “Crepe: A convolutional representation for pitch estimation,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 161–165.
- [16] G. K. Koduri, J. Serrà Julià, and X. Serra, “Characterization of intonation in carnatic music by parametrizing pitch histograms,” in *Proc. of the 13th Int. Society for Music Information Retrieval Conf.* International Society for Music Information Retrieval (ISMIR), 2012.
- [17] A. C. Gedik and B. Bozkurt, “Pitch-frequency histogram-based music information retrieval for turkish music,” *Signal Processing*, vol. 90, no. 4, pp. 1049–1063, 2010.
- [18] N. Pretto, B. Bozkurt, R. Caro Repetto, and X. Serra, “Nawba recognition for arab-andalusian music using templates from music scores,” in *Proc. of the 15th Sound and Music Computing Conference (SMC2018)*. Cyprus University of Technology, 2018, pp. 394–99.
- [19] B. Nikzat and R. Caro Repetto, “KDC: An open corpus for computational research of *dastgāhi* music,” in *Proc. of the 23rd Int. Society for Music Information Retrieval Conf.* International Society for Music Information Retrieval (ISMIR), 2022, pp. 321–328.

- [20] J. Lee, J. Park, and J. Nam, “Representation learning of music using artist, album, and track information,” in *Machine Learning for Music Discovery Workshop, International Conference on Machine Learning*, 2019.
- [21] J. Spijkervet and J. A. Burgoyne, “Contrastive learning of musical representations,” in *Proc. of the 22nd Int. Society for Music Information Retrieval (ISMIR)*, 2021.
- [22] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, “Hierarchical attention networks for document classification,” in *Proc. of the 2016 conf. of the North American chapter of the association for computational linguistics: human language technologies*, 2016, pp. 1480–1489.
- [23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
- [24] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, “Panns: Large-scale pretrained audio neural networks for audio pattern recognition,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2880–2894, 2020.
- [25] B. H. Lee, “Joga jisihaneun seonbeobgwa torieui gaenyeom(조(調)가 지시하는 선법과 토리의 개념),” *한국음악연구*, vol. 51, pp. 245–271, 2012.
- [26] Y. W. Kim, *Introduction to Korean Traditional Music(국악개론)*. Eumaksekye(음악세계), 2020.
- [27] D. H. Kang, “A study on the folksong’s ecology and cultural meaning of chongsong arari in korean folksong arirang(정선아라리의 민요생태와 문화적 의미),” *The Society Of Korean Folk Song(The Society Of Korean Folk Song(한국민요학)*, vol. 23, pp. 257–288, 2008.
- [28] Y. W. Kim, “A study of the formation of arirang(<아리랑> 형성과정에 대한 음악적 연구),” *한국문학과 예술*, vol. 7, pp. 5–55, 2011.
- [29] —, “Reexamination of eastern folk song tori -necessity to set the lower category of menari-tori and proposal(동부민요 토리의 재검토-동부민요 하층위 설정의 필요성을 중심으로),” *Studies in Korean music(한국음악연구)*, vol. 61, pp. 35–64, 2017.
- [30] Y. B. Cho, “A study of the musical style of jeju folk songs, 濟州道民謠의音樂樣式研究,” Ph.D. dissertation, the academy of Korean studies(한국정신문화연구원 한국학대학원), 1997.
- [31] B. H. Lee, “Research methodology of jejudo tori based on traditional tori features(전통적 토리 특성에 의한 제주도 토리 연구 방법론),” *Korean Recording Studies(한국음반학)*, vol. 28, pp. 105–130, 2018.
- [32] E. J. Shin, “A review and proposal of studies of the mode of the folk songs in cheju island(제주 민요 음조직에 대한 연구 검토 및 제언),” *The Society for Korean History Musicology(한국음악사학회)*, vol. 62, pp. 169–206, 2019.