Are Lullabies Cross-Culturally Soothing?
A Comparison Between Chinese and German Lullabies

Jocelyn Kavanagh and Rose Sun

1 Introduction

Music is considered a universal language, and analyses have revealed melodic features that predicted adults' identification of lullabies cross-culturally (Unyk et al., 1992). To our knowledge, no research has been carried out to determine the existence of universal soothing features in world lullabies. In this project, we report a study of Chinese and German lullabies to analyze their musicological features that may have cross-culturally soothing effects on a phrase level.

We state our hypotheses as follows:

H1. Chinese and German lullabies have similar melodic and phrase structures.

H2. These melodic and phrase structures have soothing effects.

A total of 89 phrases from 12 notated German lullabies and a total of 82 phrases from 12 notated Chinese lullabies were examined. We define lullabies as children's songs sung for soothing or sleep, and phrases as musical sections marked by breath points. The melodic and phrase structures are examined in terms of interval step to leap ratio, note density per measure, long to short note duration ratio, phrase length, and melodic contour.

To anticipate our results, we find statistically significant similarity in Chinese and German lullabies on all features.

2 Methods

2.1 Materials

This study utilized our previously developed corpus of Chinese and German lullabies. The corpus contains symbolic data representing 12 Chinese and 12 German lullabies for the analysis of musicology features that have cross-culturally soothing effects. The lullabies were collected from the Mama Lisa website and the album "Chinese Folksong - Chinese Lullabies". The lullabies fell under the categories of German and Chinese depending on the language of the song or on the composer's nationality (if available).

The 24 sampled lullabies were transcribed from their original sources into .krn files using the Humdrum Toolkit. Transcriptions were made from existing notation files on the Mama Lisa site, available jianpu notations of the Chinese lullabies, and by ear via the top audio recordings found on YouTube. Only the melodic vocal line was transcribed, and only one section of repeated material was recorded per song. Beyond basic transcription, we also notated the phrases within each song as defined as breath points.

2.2 Procedure

To determine similarity in melodic structure between the two groups, we first defined the elements of melodic structure. We determined these to be the number of steps and leaps in a song, the note density of a song, the ratio of long and short notes in a song, the phrase lengths found in a song, and the contour of phrases found in a song. We will define each of these concepts and outline how we extracted them from the corpus below.

2.2.1 Step to Leap Ratio

Supporting the hypothesis that lullabies would have more stepwise motion than leapwise motion, we investigated the ratio of steps and leaps in each song. We defined a step as melodic movement up or down by 1-3 semitones, and we defined a leap as any movement greater. We chose to include minor thirds (3 semitones) as steps due to the pentatonic nature of some of the Chinese songs in the corpus.

To find the step to leap ratio for each song, we used the Music21 interval object to extract the melodic intervals for each song. Then, we iteratively counted the number of step and leap intervals according to our operational definition. A ratio was assigned to each song by dividing the number of steps by the number of leaps.

2.2.2Note Density

We hypothesized that lullabies would be slower to induce more soothing qualities, however, our manually transcribed corpus lacked accurate tempi information. As such, we decided to evaluate the temporal complexity by finding the note density. The note density is defined as the average number of notes per measure for each song. In this definition, rests are not counted as notes, and ties are counted as one note.

We found the note density for each song by finding the number of notes and the number of measures in the song through the Music21 note and parts object. We then divided the number of notes by the number of measures for the note density.

2.2.3 Long to Short Note Duration Ratio

Again, due to our interest in the pace of the lullabies but without accurate tempo information, we used note duration and the ratio between long and short notes as another marker for temporal complexity. The definition of long and short notes depends on the given songs meter because it relies on the number of beats per measure. We defined the long notes of a song as any note with a beat value of greater than or equal to half of the beats per measure. For example, in 4/4 a long note is a half note (2 beats) or greater.

We found the long to short note ratios with the Music21 getTimeSignatures function and by recursing through the Music21 Note objects in a song. We calculated the center number of beats by dividing the numerator of the time signature by two. Then, following our operational definition, we iteratively counted the number of long and short notes in the song by comparing the note duration quarterlength values to the measure center. The ratio for each song was calculated by dividing the number of long notes by the number of short notes.

2.2.4 Phrase Length

To investigate the structural similarities of the Chinese and German songs, we extracted the length of each phrase in each song. Phrase length is defined as the number of notes in between two breath points, where rests do not count as notes. Phrase length was calculated by extracting and iterating through each song's Humdrum spine. The beginning and end of phrases were marked by "{" and "}" in the .krn files.

2.2.5 Contour

Contour was the most difficult melodic feature to define. We deliberated between using intervallic distances or maximum and minimum note placements. Ultimately, we defined contour as the shape of a phrase as determined by the position of the highest and lowest note in the phrase, along with the pitch of the first and last note of the phrase. We developed 7 different contour categories: flat, ascending ("/"), descending ("\"), V-shaped ("V"), arch ("/\"), N-shaped ("N"), and reverse N ("\/\").

To collect the contour types, we extracted the Humdrum spines with the notes listed as their frequency values, and we iterated through the phrases. Following the logic outlined below, we categorized each phrase as a contour type. For the contour assignment algorithm, a midpoint frequency was calculated for each phrase by finding the average between the highest and lowest frequency.

Flat: The frequencies of the highest and lowest notes are equal.

Ascending:

1) The lowest note is the first note and the highest note is the last note

2) The first note's frequency is lower than the midpoint frequency, the lowest note appears before the highest note, and the last note is higher than the midpoint

Descending:

1) The highest note is the first note and the lowest note is the last note
2) The first note's frequency is higher than the midpoint frequency, the highest note appears before the lowest note, and the last note is lower than the midpoint

V-Shaped: The first and last note is higher than the midpoint frequency, so the lowest note is somewhere in the middle.

Arch: The first and last note is lower than the midpoint frequency, so the highest note is somewhere in the middle.

N-Shaped: Multiple non-consecutive lowest and highest notes are found in the phrase, the first note's frequency is lower than the midpoint, and the last note's
frequency is higher than the midpoint

Reverse N: Multiple non-consecutive lowest and highest notes are found in the phrase, the first note's frequency is higher than the midpoint, and the last note's frequency is lower than the midpoint

2.2.5 Statistical Tests

Because we are testing similarity across 2 groups with both categorical and continuous data, the statistical tests were difficult to choose and interpret. For simplicity's sake, we decided to run separate tests on each feature, and we will look to accept the null hypotheses of T-tests and reject the null hypothesis of the Chi-Square test.

We chose to run individual student's T-tests on the long to short ratio, step to leap ratio, and phrase length. We chose this test because it compares the means of two independent variables. We chose to run a Chi-Square test on the contour results. We chose this because it suggests the independence between two or more categorical variables. The null and alternative hypotheses stated below.

Note Density:
$H0$: $\mu1 = \mu2$ The note densities of the Chinese and German groups are equal
$H1$: $\mu1 \neq \mu2$ The note densities of the Chinese and German groups are not equal
Long to short Ratio:
$H0$: $\mu1 = \mu2$ The long-to-short ratios in the Chinese and German groups are equal
$H1$: $\mu1 \neq \mu2$ The long-to-ratios in the Chinese and German groups are not equal
Step to leap ratio:
$H0$: $\mu1 = \mu2$ The step-to-leap ratios in the Chinese and German groups are equal
$H1$: $\mu1 \neq \mu2$ The step-to-leap ratios in the Chinese and German groups are not equal
Phrase Length:
$H0$: $\mu1 = \mu2$ The phrase lengths in the Chinese and German groups are equal
$H1$: $\mu1 \neq \mu2$ The phrase lengths in the Chinese and German groups are not equal
Contour:
H0: Chinese contour is not associated with German contour
H1: Chinese contour is associated with German contour

3 Results

The figures below display the comparison of step to leap ratio, note density, long to short note duration ratio, and phrase length between the Chinese and German group in boxplots. At face value, the level of similarity is consistent with the hypothesis that Chinese and German lullabies have similar melodic and phrase structures. In order to test the significance of this similarity, we calculated the correlation using indicated statistical tests in the procedure section.

The similarity in step to leap ratio remains statistically significant with a calculated p-value of 0.890 (t-statistic = -0.140, $p > 0.05$, two tailed test), therefore we accept the null hypothesis. There is not enough evidence to suggest that there is a difference between the Chinese and German step to leap ratios.

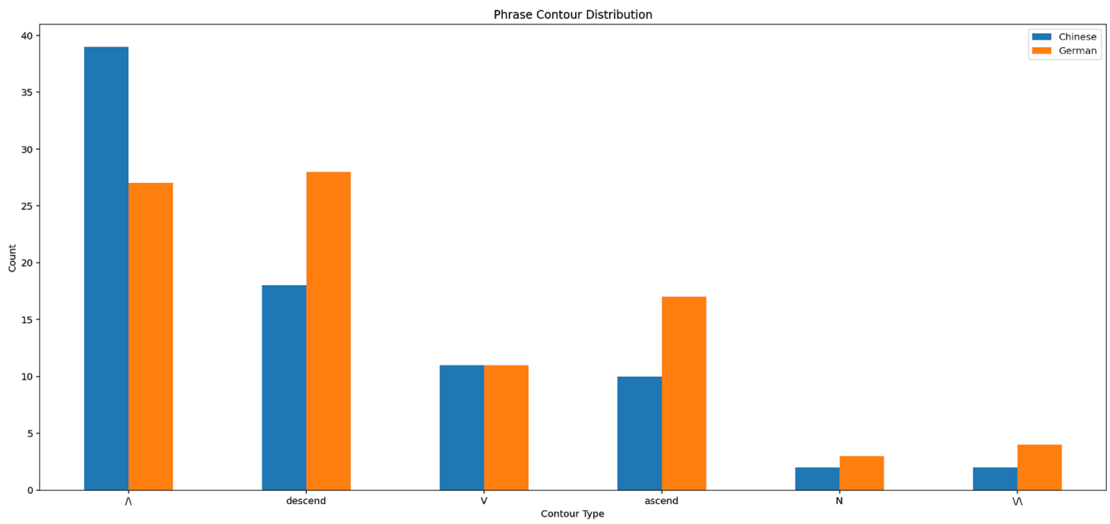
The similarity in note density remains statistically significant with a calculated p-value of 0.187 (t-statistic = 1.363, $p > 0.05$, two tailed test), therefore we accept the null hypothesis. There is not enough evidence to suggest that there is a difference between the Chinese and German note densities.

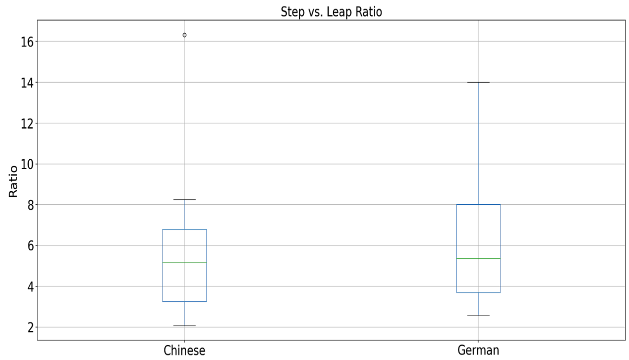
The similarity in long to short note duration ratio remains statistically significant with a calculated p-value of 0.579 (t-statistic = 0.563, $p > 0.05$, two tailed test), therefore we accept the null hypothesis. There is not enough evidence to suggest that there is a difference between the Chinese and German long to short duration ratios.

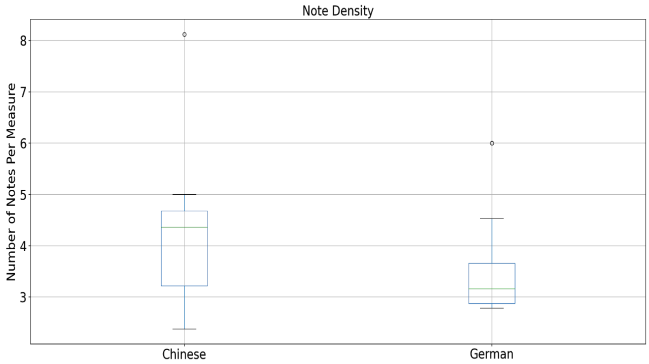
The similarity in phrase length remains statistically significant with a calculated p-value of 3.934 (t-statistic = 7.871, $p > 0.05$, two tailed test), therefore we accept the null hypothesis. There is not enough evidence to suggest that there is a difference between the Chinese and German phrase lengths.

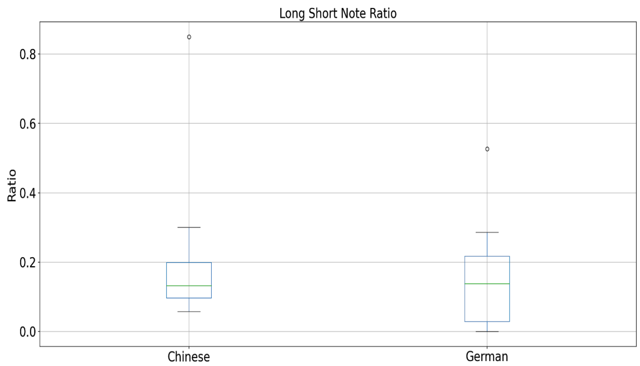
The similarity in contour remains statistically significant with the p value being 0.245, therefore we can reject the null hypothesis. We can conclude that there is an association between Chinese and German contour.

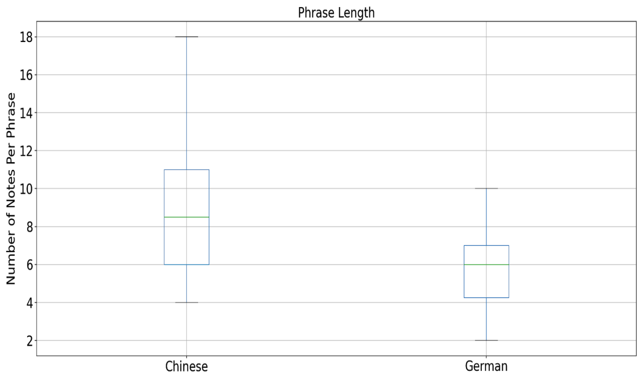
No statistical tests have been carried out to effectively determine the level of anticipated soothing effects of these melodic and phrase structures. However, traits such as relatively low note density and short phrase length, more stepwise motion than leaps, more descending motion than ascending towards the end of a phrase, remains true for all samples.











4 Discussion

Although our results showed a consistency with the hypothesis that Chinese and German lullabies have similar melodic and phrase structures, it may very well remain that the observed similarity is confined within our small sample size, and only within these two cultural groups. For a broader conclusion, more cross-cultural samples need to be examined at a greater scale.

Our definition of melodic structure might have overlooked certain contributing features. We might be able to obtain a better result with even clearer definitions that account for all possible contingencies. Potential similarity in rhythm could also be detected to strengthen our hypothesis. More analysis could be carried out in terms of the soothing effects but we could not attempt all these analysis within the class time frame.