Volume 50 http://acousticalsociety.org/

183rd Meeting of the Acoustical Society of America

Nashville, Tennessee

5-9 December 2022

Psychological and Physiological Acoustics: Paper 1pMU5

Uncovering the differences between the violin and erhu musical instruments by statistical analysis of multiple musical pieces

Wenyi Song and Andrew Brian Horner

Computer Science and Engineering, The Hong Kong University of Science and Technology School of Engineering, Hong Kong, HONG KONG; wsongak@cse.ust.hk; horner@cse.ust.hk

The violin and erhu are two of Western and Chinese music's main bowed string instruments. Recent work has compared the different emotional characteristics between the violin and erhu on the Butterfly Lovers Concerto. In our study, we examine several hypotheses to investigate whether the previous studies' results hold generally. Four musical excerpts were extracted from four famous Chinese and Western classical pieces, and the excerpts were divided into four emotional categories: Happy, Sad, Agitated, and Calm. Based on the Butterfly Lovers results, we expected that: (1) the violin has a more Happy emotional characteristic than the erhu, while the erhu is comparatively more Sad, and (2) the violin is better at conveying high-Arousal excerpts. We used the Bradley-Terry-Luce (BTL) paired-comparison model to obtain the ranking scores and identify statistically significant differences between the two instruments. The erhu was consistently perceived as sadder than the violin for all Sad excerpts, while the violin was generally calmer and more agitated for those categories. Further study with more listeners and excerpts is needed to verify whether these results generally hold and at a statistically significant level.

Published by the Acoustical Society of America

 https://doi.org/10.1121/2.0001710

1. INTRODUCTION

Music emotion studies help us understand the psychological relationship between human affect and music. Much work has been done on music emotions with structural features such as tempo, mode, loudness, melody, rhythm, and timbre [1]. Moreover, these fundamental musical elements exert strongly different emotional characteristics on the sound of an instrument [2]. Correspondingly, several studies have correlated the timbre of musical instruments with various emotional characteristics [3–10]. Others have investigated the effects of different instruments, pitches, and dynamics on various emotional characteristics [11–15].

Previous studies have focused on single notes of varying levels of pitch and dynamics for different families for Western classical musical instruments, including the piano [11, 12], pitched percussion [15], woodwinds [16], brass [17], and the Western bowed strings [13, 14]. The string family is most related to our current study. In this previous work [13, 14], listeners compared the same note played at different octaves and with different dynamics by the violin, viola, cello, and double bass, and judged how their emotional characteristics varied with pitch and dynamics. The idea is familiar to composers and singers who often select the key of a song that best brings out the desired emotional shading, knowing that slightly different pitch registers bring out slightly different emotional nuances due to shifts in the vocal resonances. It is valuable to systematize this intuition to whatever degree possible. Moreover, aside from a few preliminary studies, the Chinese bowed strings remain relatively unexplored in this respect [18–20].

The pitch ranges of the violin and erhu are similar, but the constructions are different. With the structural differences between the violin and erhu, there are expected to be striking emotional differences as well. It is a natural to compare the violin and erhu since Chinese instruments are increasingly popular in film and computer game music for martial arts and mystical scenes, especially genre mashups of traditional Western orchestral music with an added solo erhu. For example, some western films add solo erhu to introduce a Chinese character. Similarly, The popular video game *Genshin Impact* uses the western orchestra and solo violin to create the *Liyue Harbor Main Theme (OST 1)* (Liyue city represents China in the game), and is widely performed in the violin community.

People generally consider that the violin sounds bright while the erhu is nasal, and the erhu is especially sadder in slow and sorrowful pieces. However, very few music emotion studies have considered Chinese instruments, which naturally widen the palate of possible timbres and emotional shades available to composers and sound designers. It is desirable to compare the emotional characteristics of instruments from different cultural traditions in a multifaceted way [21, 22].

Musical excerpts have been traditionally used in emotion recognition and response research [23–26]. One advantage of using musical excerpts rather than single notes is that they represent how emotional characteristics ebb and flow in a musical piece. This approach is usually not intended to compare differences between instruments. Our research group has performed two preliminary tests comparing the violin and erhu on the *Butterfly Lovers Concerto* [21, 27]. This piece is especially interesting since it strongly references both classical Chinese bowed string music as well as romantic Western violin concertos. Although originally written for the violin, the piece is frequently played on the erhu, and is widely known. The story it paints is a classic tragic romance similar to *Romeo and Juliet* with some extra twists that include wild swings in emotion.

In these two previous studies [21, 27], absolute and relative tests were used and listeners were asked to judge four excerpts from *Butterfly Lovers* comparing five different performances from both the violin and erhu. The four excerpts selected were from the emotional categories Romantic, Joyful, Agitated, and Bittersweet. Subjects were asked to rate the Arousal (energy-level), Valence (positiveness), and emotional intensity of these four excerpts. Figure 1 shows the 2D valence-arousal model of emotion. In this study, high-Arousal excerpts would be more Joyful and more Agitated. The limitations of the absolute test were that the absolute scale made comparisons rather approximate [21]. The number of subjects was also relatively small (about 30).

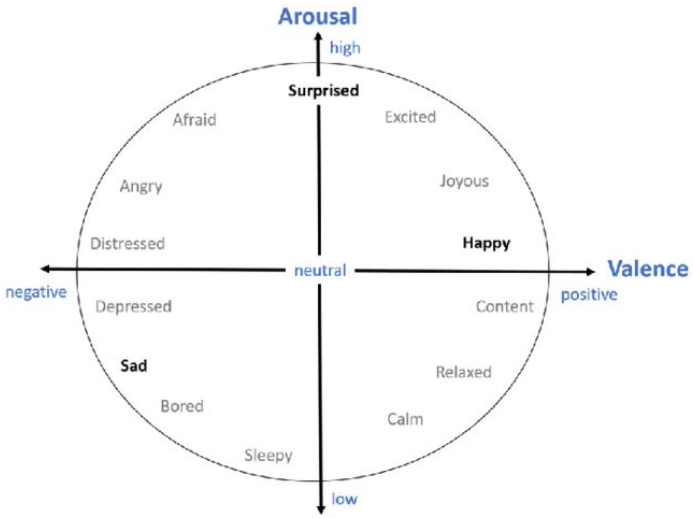

Figure 1: The 2D valence-arousal model of emotion proposed by Russel [28].

A relative test used paired-comparisons and more subjects (about 90) to obtain more precise and statistically significant results [27]. Again, however, the limitation was the focus on a single musical piece. It is uncertain whether the results hold generally for erhu and violin.

In the next section (Section 2), we describe our new experiment to compare the violin and erhu on our selected musical pieces. Section 3 gives the results of our statistical analysis. Section 4 discusses the limitations of our results, and other future work.

2. METHODOLOGY

In order to further explore the study of Chinese instruments and the emotional characteristics of instruments from different cultural traditions, we continue the study of comparing the erhu and violin. In our current study, we expand the number of pieces to determine the general relationship between the two instruments. Since *Butterfly Lovers* was initially written for the violin and yet the style is Chinoiserie, we selected four different pieces, two written for the erhu in traditional Chinese style, and the other two for violin and from the Western classical music tradition. The Chinese pieces are frequently performed on the violin, and the Western pieces are frequently performed on erhu, giving us a common repertoire for comparison of the results.

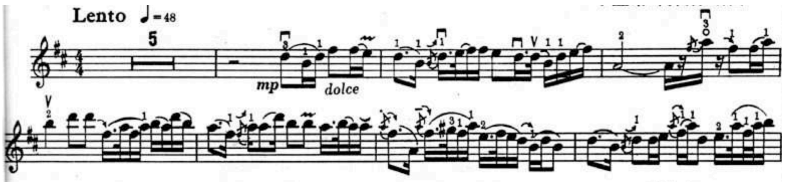

Figure 2: Example of erhu piece Erquan Yingyue in Western musical notation [29].

Chinese traditional music often uses a pentatonic scale to emphasize the sense of line in the melody with a strong tradition of musical story-telling while Western music makes more use of harmony and rhythm to create an atmospheric aesthetic. For example, Figure 2 shows the erhu score of *Erquan Yingyue*, which uses a pentatonic scale in D Major, and tells the story of the composer A-Bing himself suffering from bitter humiliation at the hands of old society [31]. Figure 3 is the violin piece *Zigeunerweisen* which uses fast

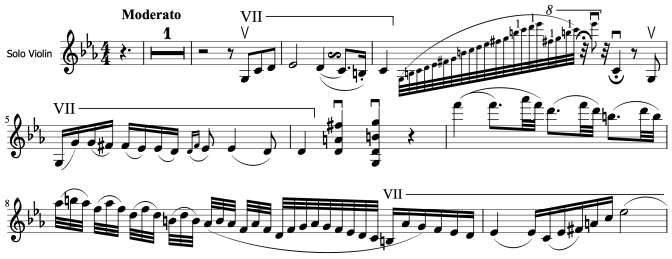

Figure 3: Example of violin piece Zigeunerweisen, Op. 20 [30].

scales to agitate the overall flow of the music over moody chords, bringing about sudden shifts in mood and to glue together ascending and descending progressions in a minor key.

We changed the four emotional categories from Romantic, Joyful, Agitated, and Bittersweet in our previous studies to Happy, Sad, Agitated, and Calm. The former categories made sense for the particular emotional shades of *Butterfly Lovers*, but the later categories are more general since we want them to apply to other pieces of music not just *Butterfly Lovers*. The Bradley–Terry–Luce (BTL) statistical model [32–34] can be used once again, and the score results will allow us to compare the emotion intensity over a variety of musical pieces among four emotional categories. Finally, we will also introduce a Preference Tree model [35, 36] to compare with the BTL model results.

Our study will help quantify and visualize how the emotional characteristics of the violin and erhu vary in shading. It is valuable in applications such as orchestration, arrangement, and blending and balancing instruments in symphonic film and computer game music. Basic orchestration is usually taught in an example-based way, and seeing the underlying graduations in emotional characteristics can assist student orchestrators and arrangers to internalize the emotional potentialities inherent in the instruments. Of course, a great composer can always use melodic and harmonic devices to bring out angry characteristics from even an instant that is usually associated with a calm and peaceful character. In this work, we are not trying to be prescriptive or over-simplistic, but rather to simply expose the shadings for musicians to use in whatever ways are musically useful.

A. HYPOTHESES

Before conducting the listening test, we had several hypotheses to test to investigate the generality of the results from previous erhu and violin studies.

Hypothesis 1 *The violin has a more Happy emotional characteristic, while the erhu is more Sad.*

The violin has what is generally regarded as a warm romantic tone, while the erhu has a distinctly nasal tone which gives it an undertone of nostalgia and sadness [37]. Therefore, we might expect the violin to have a more Happy emotional characteristic and the erhu to have a more Sad emotional characteristic. The previous study using an absolute test verified this hypothesis for the *Butterfly Lovers concerto*.

Hypothesis 2 *The violin is better at conveying high-Arousal excerpts (i.e., the violin is more Happy and more Agitated than the erhu).*

In general, the violin tone is brighter, and multi-string techniques can be used on the violin but not the erhu. Therefore, we might expect the violin to be better at conveying high-Arousal excerpts. Both absolute and relative tests in previous studies supported this hypothesis for the *Butterfly Lovers concerto*.

Hypothesis 3 *The violin can more intensely represent all four emotional categories and has an overall wider range of emotional expressions (i.e., it would be more Happy, more Sad, more Agitated, and more Calm).*

This hypothesis might be true because the violin is widely expressive as a heart-on-sleeve instrument that can scale the heights of joy and depths of despair. Nevertheless, at the same time, the erhu might be more restrained in its emotional outpourings, perhaps as a result of the traditional Chinese ideal of emotional restraint. The findings of the previous relative test study for the *Butterfly Lovers concerto* supported this hypothesis, and we propose to see if it is true in general for other musical pieces.

B. LISTENING TEST

The listening test was designed with relative comparisons to compare the emotional characteristics of the erhu and violin. Excerpts were extracted from four pieces, and two were originally written for the erhu in traditional Chinese style, while the other two were for the violin from Western classical music. Paired comparison was used for the violin-erhu excerpt pairs from the same piece, and the instrument order was set at random. For each violin-erhu or erhu-violin pair, listeners answered simple "Yes/No" questions to indicate which excerpt had the higher emotional intensity.

i. Stimuli

We selected four famous pieces commonly played by both erhu and violin: *Singing the Night Among Fishing Boats* and *Wonderful Night* originally for the erhu, and *Flight of the Bumblebee* and *Czardas* originally for the violin. The extracted excerpts were complete phrases. Furthermore, each excerpt had a different key, dynamics, tempo, and time signature. Happy excerpts were extracted from *Singing the Night Among Fishing Boats*, Sad excerpts from *Czardas*, Agitated excerpts from *Flight of the Bumblebee*, and Calm excerpts from *Wonderful Night*.

Ideally, the erhu and violin excerpts should have minimum differences (e.g., audition environments) to focus the comparison on the solo instruments. A musical solo can be performed in many ways depending on the performer's interpretation, so it helps to include multiple performances in the comparisons. We selected five different performances of each excerpt from each instrument, and tried to minimize the accompaniment instruments in the comparison.

The forty excerpts were from commercial CD recordings [38, 39]. The excerpts covered a wide range of performers (i.e., gender, age, and nationality). For each excerpt, the average duration was 8.65 sec. We slightly processed the excerpts by adding a short fade in and fade out of 1 sec to prevent distracting pops. The volume adjustment ranged from -5 to $+15$ dB to minimize the effect of loudness between the excerpts.

ii. Test Procedure

Before taking the online listening test, listeners were required to: (1) Read brief instructions on how to run the listening test on the computer and understand the definitions of the four emotional categories, (2) Be seated in a quiet room with minimal background noise, (3) Use a good-quality wired headset for the test, and (4) Set up the screen recording and turn the camera, audio, and screen sharing on.

Listeners were allowed to ask questions before the test, but were required to be alone with no interruption during the test. Moreover, there were some example trials before the test, allowing listeners to get used to the environment and minimize the effects of learning.

The listening test took about 40 minutes with four emotional categories (including relaxing time), and each emotional category was an independent section. Before starting each section, listeners received a notification about which emotional category was going to be tested. Also, they could take some time to

relax if they felt tired or bored. Then in each section, 25 violin-erhu pairs (from five different performances of each instrument) were formed into 25 questions. Each question included two trials, one trial for the violin, the other for the erhu. Each violin-erhu pair was played once in this section. After finishing listening to the two performances, the question was asked, *"Do you think the second is more Happy?" (assuming the current emotional category was Happy)*. Listeners were only required to make a *"Yes/No"* binary decision without remembering other details, which allowed answering to be generally quick.

The emotional categories, excerpt pairs, and instruments were unordered and randomized. For example, one listener could hear the emotion category sequence as Sad, Calm, Happy, Agitated, and the first question could be any of the 25 violin-erhu pairs. Also, the order of the first question could be violin-erhu or erhu-violin. In the complete listening test, there were 100 questions (4 emotional categories $\times$ (5×5 instrument pair combinations)), and listeners could take a short rest after finishing 25 questions of each emotional category.

After finishing all 100 questions, listeners submitted video recordings and the listening test results.

iii. Participants

The subjects were students in the *Creative Sound Design* course at HKUST, who volunteered to participate in our listening test experiments to satisfy part of the course requirements. All were fluent in English and without hearing issues.

46 subjects participated in this listening test in total. There were 30 males and 16 females, aged 21.6 years on average (range: 20–24). 29 out of 46 subjects could play at least one instrument and had musical training from 1 to 17 years.

3. RESULTS

We summarize the listening test results into a frequency table for each emotion category, and Table 1 is an example of the emotional category Happy. The stimuli consisted of 10 excerpts with 25 violin-erhu pairs, including 5 Erhu performances (Er1-5) and 5 Violin performances (Vln1-5). Table 1 indicates how often the row stimuli was chosen over the column stimuli. To fulfill the requirements for paired-comparison square matrix of the Bradley–Terry–Luce (BTL) statistical model [32–34] and keep the listening test from being too long, we manually set the same instrument pair value to 23 (i.e., erhu-erhu pair $(Er1, Er2) = 23$), which is half of the number of subjects.

The excerpts were ranked by the number of votes they received for each emotional category, and the scale values were derived using the BTL model. The sum of all BTL values of each excerpt for each emotional category is 1. The BTL value is the probability that listeners will select the excerpt for that emotional category. For example, if all 10 excerpts are equal in emotional intensity, the BTL value for each excerpt would be $1/10 = 0.1$. Bradley's method was used to estimate the scaled BTL values and calculate the corresponding 95% confidence intervals [33].

Figure 4 shows the bar charts for the BTL scale values and the corresponding 95% confidence intervals for each excerpt and emotional category. For the excerpts of the same emotion category and instrument, there were some variations between the BTL scale values, especially for Agitated. Also, for each estimated BTL scale value, $P_r(> |z|) < 2e - 16$, the result was individually significant, and there were detectable variations in listeners' choices. We observe that the erhu was consistently sadder than the violin for all Sad excerpts (i.e., the smallest erhu BTL value 0.099 was about as large as the largest violin BTL value 0.101). Moreover, most of the violin's BTL values were larger than the largest erhu BTL value.

Figure 5 shows the BTL scale values for each instrument and the corresponding 95% confidence intervals for each emotional category. We observe that the erhu had a lager BTL value than the violin for the Sad

Table 1: Aggregate choice frequency table for emotion category Happy.

Excerpt ID	Er1	Er2	Er3	Er4	Er5	Vln1	Vln2	Vln3	Vln4	Vln5
Er1	0	23	23	23	23	21	28	20	24	24
Er2	23	0	23	23	23	24	24	22	25	19
Er3	23	23	0	23	23	25	25	26	19	17
Er4	23	23	23	0	23	24	16	22	23	26
Er5	23	23	23	23	0	27	27	23	20	27
Vln1	25	22	21	22	19	0	23	23	23	23
Vln2	18	22	21	30	19	23	0	23	23	23
Vln3	26	24	20	24	23	23	23	0	23	23
Vln4	22	21	27	23	26	23	23	23	0	23
Vln5	22	27	29	20	19	23	23	23	23	0

Note: The number indicates how often the row stimuli was chosen over the column stimuli.

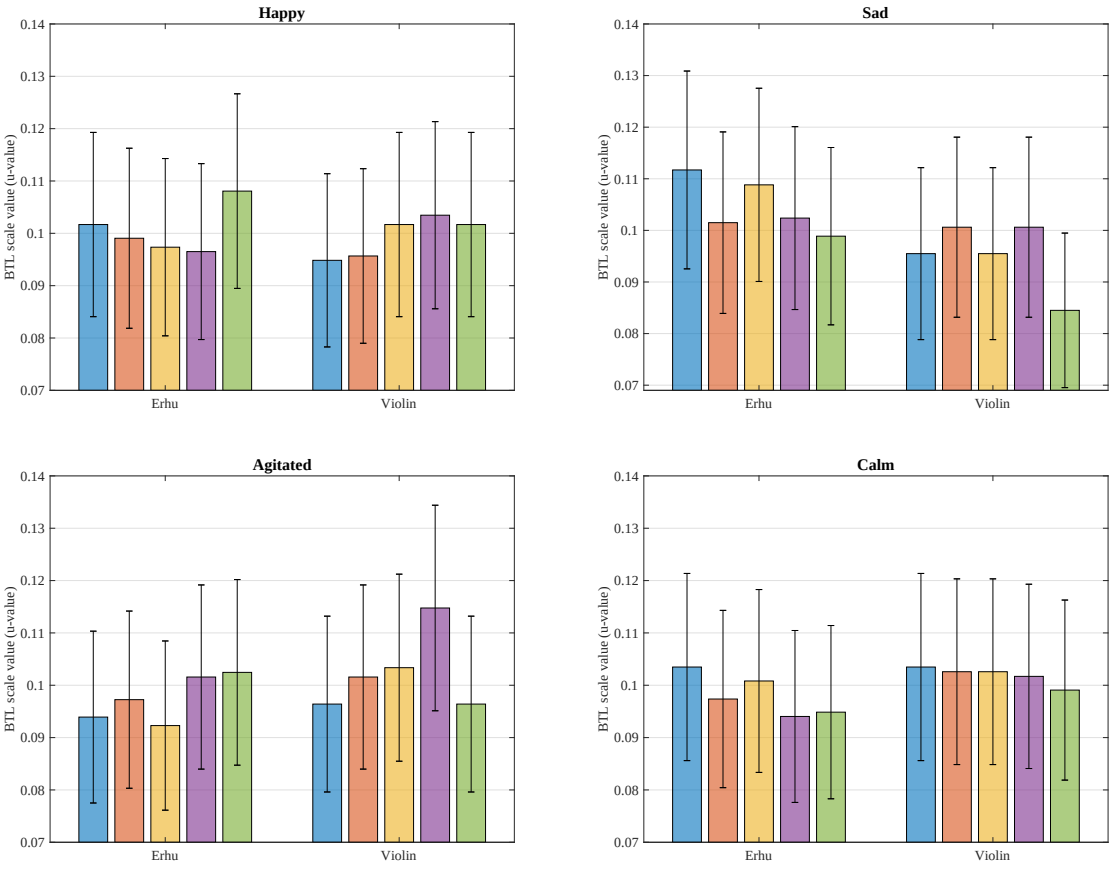

Figure 4: BTL scale values of all excerpts for each emotion category and the corresponding 95% ***confidence intervals.***

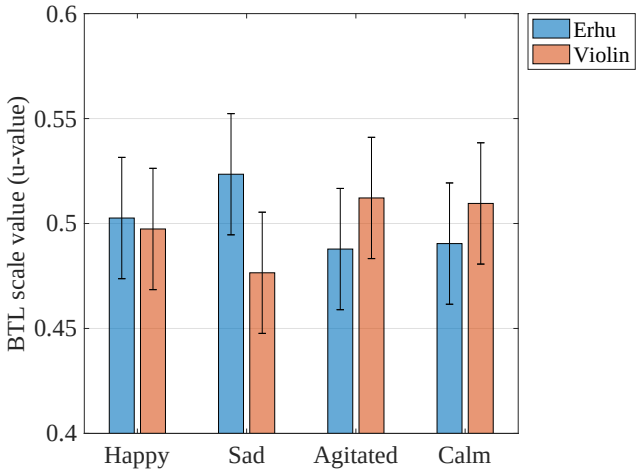

Figure 5: BTL scale values of each instrument for each emotion category and the corresponding 95% ***confidence intervals.***

Table 2: Parameter estimation for the four emotion categories.

Emotion	Label	Number	Estimate (u)	Std. Error	z value	$Pr(>\|z\|)$
Happy	Erhu	1	0.50261	0.01474	34.09	$< 2e-16$
	Violin	2	0.49739	0.01474	33.73	$< 2e-16$
Sad	Erhu	1	0.52365	0.01473	35.54	$< 2e-16$
	Violin	2	0.47668	0.01473	32.35	$< 2e-16$
Agitated	Erhu	1	0.48787	0.01474	33.1	$< 2e-16$
	Violin	2	0.51222	0.01474	34.75	$< 2e-16$
Calm	Erhu	1	0.49046	0.01474	33.27	$< 2e-16$
	Violin	2	0.50959	0.01474	34.57	$< 2e-16$

emotion category, while the violin was larger for Agitated and Calm. The largest difference was for Sad, while Happy was the smallest.

Table 2 shows the estimated BTL values for the two instruments in each emotion category. Although both instruments have strongly significant individual results $P_r(>|z|) < 2e{-}16$, the overlapping confidence intervals in Figure 5 don't show significant differences between the erhu and violin for the four emotion categories, though it is closest for Sad.

In most studies using the BTL model, the paired comparison judgements are context-independent. In our study, the performer is a factor that affects the judgement between the same instrument playing the same melody and the listener's concentration on the differences between the two instruments. Ideally, the performer factor should be as small as possible. Therefore, we considered using the Preference Tree model [35, 36], which can group the different performances by the two instruments in a hierarchical tree structure. It can obtain the branch attribute and use the category information as the overall evaluation index to compare the erhu and violin classes, instead of simply summing the BTL values. Figure 6 shows a Pretree structure for the emotion category Happy.

We compared the BTL and Pretree results for each emotion category, and Table 3 shows the likelihood-

Table 3: The goodness of fit of BTL and Pretree (PT) Models for the four emotion categories.

Emotion	Model	χ^2	df	p-value
Happy	BTL	20.1159556	36	0.9849672
	PT	20.63723	34	0.965361
Sad	BTL	19.706051	36	0.987504
	PT	20.9878415	34	0.9605682
Agitated	BTL	14.5793692	36	0.9994253
	PT	15.6706082	34	0.9969665
Calm	BTL	26.5431572	36	0.8749137
	PT	26.8401518	34	0.8038721

Note: The EBA model is rejected if the p-value is less than 10%.

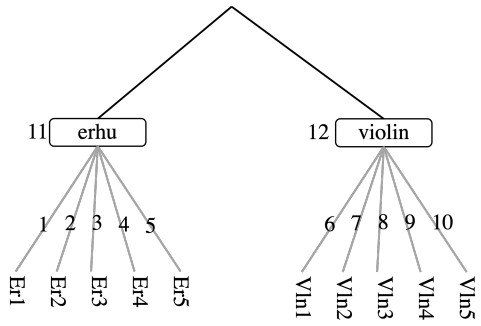

Figure 6: Preference tree models hierarchical structure for the emotion category Happy.

ratio test on the goodness of fit for the two models. The model is rejected if the p-value is less than 10% [33]. According to Table 3, both models are fit to the ratings.

To select the better model for representing the dataset, we calculated Akaike's information criterion (AIC) [33] for both the BTL and Pretree models as shown in Table 4. A lower AIC value indicates a better fit. The result shows that they are similar, though the BTL model resulted in a slightly better fit for each category.

We also tested the differences between the violin and erhu by the Pretree model and found that the differences were not statistically significant for all the emotion categories (the Wald-test [35, 36] showed that for all four emotion categories p was greater than 0.05).

4. DISCUSSION AND CONCLUSIONS

A. RELATIONSHIP BETWEEN THE ERHU AND VIOLIN

Before conducting the listening test we proposed three hypotheses regarding the general relationship between the violin and erhu in emotional characteristics. We expected the results to perhaps support consistent and significant differences between the two instruments.

For each estimated BTL value, the result was individually significant, and there were detectable variations in listeners' choices. The erhu was consistently sadder than the violin for all Sad excerpts, while the violin was generally calmer and more agitated in those categories. However, the results were not large

Table 4: Model selection by Akaike's information criterion (AIC) for all four emotion categories.

Emotion	Model	AIC
Happy	BTL	230.7404
	PT	235.2617
Sad	BTL	230.2407
	PT	235.5224
Agitated	BTL	225.274
	PT	230.3653
Calm	BTL	237.0334
	PT	241.3304

enough to be statistically significant to support the proposed hypotheses.

In contrast to the previous study on *Butterfly Lovers*, we did not find decisive evidence to show that the erhu and violin have differences in emotional expression for other musical pieces in general. A possible reason is that the number of subjects was relatively small in our listening test (46 versus 95 in previous studies). The consistency of the results for Sad, Calm and Agitated is an encouraging reason to expand the listening test with additional subjects in search of a more decisive result.

B. LIMITATIONS

We resolved some limitations from the previous study by including more musical pieces and performers for each instrument. However, there are obvious areas for further investigation.

i. Paired-comparison Data

We only included pairs of two different instruments for comparison in the listening test and assumed that the difference between the same instrument were minimal. Therefore, we manually set the frequency value of the same instrument pairs (violin-violin/erhu-erhu) to half the number of subjects. While reducing the number of pairs needed to be tested and shortening the listening test, it may have also had the undesirable effect of diluting the results.

Moreover, this shortcut might affect the result of BTL and Pretree model selection. According to the AIC comparison, they were similar, though the BTL model slightly better represented that the dataset model. However, previous studies have shown that Pretree can improve on BTL if the dataset has a natural hierarchy [33, 35, 36, 40]. Therefore, including the same instrument paired comparisons in the listening test can help improve the accuracy of the comparisons.

ii. Generality of Excerpt Selection

We used only four different excerpts from four pieces in our listening test, with two originally written for the erhu and two for the violin. However, ideally more should be used to test the violin and erhu. An improvement is that we could include at least two excerpts from two different pieces for each of the four emotional categories, one originally written for the erhu, and the other for the violin. This would increase the number of excerpts for each category and instrument from five to eight.

Nevertheless, the preliminary experiment we have done deepens our general understanding of the instruments, and gives us some clear clues on how to further pin down their salient differences. These insights are also useful in future cross cultural studies of other instrument pairs, such as the flute and its Chinese counterparts, the Dizi and Xiao.

REFERENCES

[1] K. R. Scherer and M. R. Zentner, "Emotional effects of music: Production rules." 2001.

[2] W. L. Sin, B. Y. Chang, X. Ma, and A. Horner, "The acoustics features and their relationship to the emotional characteristics of rain sound effects," Ph.D. dissertation, Hong Kong University of Science and Technology, 2019.

[3] T. Eerola, R. Ferrer, and V. Alluri, "Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds," *Music Perception: An Interdisciplinary Journal*, vol. 30, no. 1, pp. 49–70, 2012.

[4] B. Wu, S. Wun, C. Lee, and A. Horner, "Spectral correlates in emotion labeling of sustained musical instrument tones." in *ISMIR*, 2013, pp. 415–420.

[5] B. Wu, A. Horner, and C. Lee, "Musical timbre and emotion: The identification of salient timbral features in sustained musical instrument tones equalized in attack time and spectral centroid," in *ICMC*, 2014.

[6] ——, "Emotional predisposition of musical instrument timbres with static spectra." in *ISMIR*, 2014, pp. 253–258.

[7] ——, "The correspondence of music emotion and timbre in sustained musical instrument sounds," *Journal of the Audio Engineering Society*, vol. 62, no. 10, pp. 663–675, 2014.

[8] C.-j. Chau, B. Wu, and A. Horner, "Timbre features and music emotion in plucked string, mallet percussion, and keyboard tones," in *ICMC*, 2014.

[9] ——, "The emotional characteristics and timbre of nonsustaining instrument sounds," *Journal of the Audio Engineering Society*, vol. 63, no. 4, pp. 228–244, 2015.

[10] ——, "The effects of early-release on emotion characteristics and timbre in non-sustaining musical instrument tones," in *ICMC*, 2015.

[11] C.-j. Chau and A. Horner, "The effects of pitch and dynamics on the emotional characteristics of piano sounds," in *ICMC*, 2015.

[12] C.-j. Chau, R. Mo, and A. Horner, "The emotional characteristics of piano sounds with different pitch and dynamics," *Journal of the Audio Engineering Society*, vol. 64, no. 11, pp. 918–932, 2016.

[13] S. J. Gilburt, C.-j. Chau, and A. Horner, "The effects of pitch and dynamics on the emotional characteristics of bowed string instruments," in *International Computer Music Conference (ICMC), Utrecht*, 2016, pp. 405–410.

[14] C.-j. Chau, S. J. Gilburt, R. Mo, and A. Horner, "The emotional characteristics of bowed string instruments with different pitch and dynamics," *Journal of the Audio Engineering Society*, vol. 65, no. 7/8, pp. 573–588, 2017.

[15] C.-j. Chau and A. Horner, "The emotional characteristics of mallet percussion instruments with different pitches and mallet hardness," in *Proceedings of the International Computer Music Conference*, 2016, p. 401.

[16] H. T. Chan, B. Y. Chang, and A. Horner, "The emotional characteristics of woodwind musical instruments with different pitch and dynamics," in *Proc. Int. Comp. Music Conf.(ICMC)*, 2019.

[17] H. T. Chan, R. Mo, C. Keyes, and A. Horner, "The emotional characteristics of brass musical instruments with different pitch and dynamics," in *Proc. Int. Comp. Music Conf.(ICMC)*, 2019.

[18] A. De Ritis, "Thomas d. rossing: Science of percussion instruments," *Computer Music Journal*, vol. 26, no. 2, pp. 97–98, 2002.

[19] T. D. Rossing *et al.*, *The science of string instruments*. Springer, 2010.

[20] Y. S. Siao, W. L. Chang, and A. W. Su, "Analysis and transynthesis of solo erhu recordings using additive/subtractive synthesis," in *120th Convention Spring Preprints 2006*, 2006, pp. 1–7.

[21] D. Lee, "A comparison of the emotional characteristics of the violin and erhu on the butterfly lovers concerto," Ph.D. dissertation, The Hong Kong University of Science and Technology, 2020.

[22] X. Wang, Y. Wei, L. Heng, and S. McAdams, "A cross-cultural analysis of the influence of timbre on affect perception in western classical music and chinese music traditions," *Frontiers in Psychology*, vol. 12, p. 732865, 2021.

[23] S. Filipic, B. Tillmann, and E. Bigand, "Judging familiarity and emotion from very brief musical excerpts," *Psychonomic Bulletin & Review*, vol. 17, no. 3, pp. 335–341, 2010.

[24] J. Grekow, *From content-based music emotion recognition to emotion maps of musical pieces*. Springer, 2018, vol. 747.

[25] J. Bhattacharya and J. P. Lindsen, "Music for a brighter world: brightness judgment bias by musical emotion," *PloS one*, vol. 11, no. 2, p. e0148959, 2016.

[26] S. Vieillard, I. Peretz, N. Gosselin, S. Khalfa, L. Gagnon, and B. Bouchard, "Happy, sad, scary and peaceful musical excerpts for research on emotions," *Cognition & Emotion*, vol. 22, no. 4, pp. 720–752, 2008.

[27] D. Lee, W. Song, and A. B. Horner, "A head-to-head comparison of the emotional characteristics of the violin and erhu on the butterfly lovers concerto," 2020.

[28] Y. Tsalamlal, M.-A. Amorim, J.-C. Martin, and M. Ammi, "Modeling emotional valence integration from voice and touch," *Frontiers in Psychology*, vol. 9, p. 1966, 2018.

[29] Abing, Z. Ding, Z. He, and Z. Ding. (1950) Erquan yingyue. IMSLP. [Online]. Available: https://s9.imslp.org/files/imglnks/usimg/3/3c/IMSLP33861-PMLP76659-200511320245918.pdf

[30] P. de Sarasate. (1878) Zigeunerweisen, op. 20. IMSLP. [Online]. Available: https://s9.imslp.org/files/imglnks/usimg/0/00/IMSLP106743-PMLP04860-Zigeunerweisen,_Op_20_Solo_Violin.pdf

[31] R.-g. HUANG, "The exploration of the root-source of the chinese music "er quan ying yue"," *Canadian Social Science*, vol. 3, no. 5, pp. 97–98, 2010.

[32] R. A. Bradley, "14 paired comparisons: Some basic procedures and examples," *Handbook of statistics*, vol. 4, pp. 299–326, 1984.

[33] F. Wickelmaier and C. Schmid, "A matlab function to estimate choice model parameters from paired-comparison data," *Behavior research methods, instruments, & computers*, vol. 36, no. 1, pp. 29–40, 2004.

[34] D. Firth, "Bradley-terry models in r," *Journal of Statistical software*, vol. 12, no. 1, pp. 1–12, 2005.

[35] A. Tversky and S. Sattath, "Preference trees." *Psychological review*, vol. 86, no. 6, p. 542, 1979.

[36] A. Tversky, "Elimination by aspects: A theory of choice." *Psychological review*, vol. 79, no. 4, p. 281, 1972.

[37] S. W. Shengmiao. Bowed stringed instruments - 9 erhu. Atlas Ensemble. [Online]. Available: https://www.atlasensemble.nl/assets/files/instruments/Erhu/ErhubySamuelWongShengmiao.pdf

[38] I. NetEase. (1997-2022) Netease cloud music. NetEase, Inc. [Online]. Available: https://music.163.com

[39] T. H. Ltd. (1998-2022) Qq music. NetEase, Inc. [Online]. Available: https://y.qq.com

[40] A. Leschanowsky, S. Das, T. Bäckström, and P. P. Zarazaga, "Perception of privacy measured in the crowd-paired comparison on the effect of background noises." in *Interspeech*, 2020, pp. 4651–4655.