



LOUDNESS EFFECT OF THE DIFFERENT TONES ON THE TIMBRE SUBJECTIVE PERCEPTION EXPERIMENT OF ERHU

Siyu Zhu, Peifeng Ji, Wei Kuang, Jun Yang

Institute of Acoustics, Chinese Academy of Sciences, Beijing, China 100190

e-mail: jyang@mail.ioa.ac.cn

The loudness of sound sources should be properly adjusted to remove its influence on timbre perception when the objective properties are changed. A comprehensive adjusting method is proposed which aims at the variations of loudness between different tones, and it is verified by the multi-dimensional scale model. The adjusted curves of loudness in an octave against the objective properties are obtained according to subjective experiments using a traditional Chinese instrument called “Erhu”. Compared with a traditional adjusting method that only aims at one tone, the results show that it is necessary to adjust all the tones' corresponding loudness when the timbre properties are changed.

1. Introduction

Timbre perception is important in the study of music perception. The research on timbre perception focuses on how to build quantitative correlations between the position along a perceptual dimension and a value along property dimension¹. Many researchers attempt at explaining all perceptual dimensions of a given timbre space by correlating timbre properties with perceptual dimensions². The most common approach to timbre perception is multidimensional scaling analysis (MDS). The MDS approach can determine how a subject has grouped the stimuli and identified the perceptual dimensions by rating the sounds based on a standard category-rating paradigm³. It is convenient for us to interpret the similarity to sounds with varied timbre properties by the distance in N -dimensional spaces.

Before studying the relationship between property dimension and perceptual dimension, we need to synthesize sound. Simple additive synthesis models are usually used in instrument synthesis sound with a specific timbre property to be varied. During the synthesizing, we should make sure that a specific timbre property is varied independently from other timbre properties. Apart from the influence from one timbre property to another, the influence on loudness, pitch, duration must also be eliminated⁴. Before experimentation the stimuli can be equalized in an on-line experiment⁵ or by adjustment independently at first and then by consensus in the case of differences in adjustment from two listeners for loudness, pitch, and perceived duration. These two equalizing methods are very simple and don't require a specific adjusting method. Caclin et al. present an equation which can balance loudness and duration time⁶. But they don't provide an equation for pitch balancing. Subsequent work shows that it is necessary to balance stimuli for pitch⁷ when the spectral centroid of the stimuli is changed. Subsequently Iise et al. present a systemic balancing method for loudness,

pitch and duration⁸ which combines Caclin' loudness and duration balancing equation with subjective experiment by using 13 western musical instruments.

In this paper, we focus on loudness balancing for the sound of Erhu, which is a traditional Chinese string instrument with a relatively large difference in structure and tone compared with the western instruments. If we directly use the method Iise presents to balance sound of Erhu for loudness when the timbre property of sound has been varied for the research on timbre perception, there may be two problems. One is that whether the Iise's adjusting method aimed at 13 western instruments is suitable for other instruments. Another is that since Iise's method for loudness balancing points at one tone is probably inaccurate, it is probably inaccurate for us to balance the loudness of other tones by using the same method.

Considering the potential problems directly using Iise's method for loudness balancing on Erhu's perceptual research, a further investigation is carried out in this study on (i) whether the balance method for western instrument is suitable for Erhu, (ii) a comprehensive adjusting method which aims at different tones.

2. Problem Statement and improvement

There are two possible problems on Iise's balancing method for loudness. (i) Iise provides a further research on basis of Caclin's equation on loudness balance. In the Eq. (1) T_r and T_a stand for the reference and intensity-adjusted tones, respectively. TP and A stand for the timbre property varied and the intensity of the sound. G is the gradient which is calculated by linear regression fitting to the loudness response data of 13 western instruments. Since G is a statistical value aimed at 13 western instruments, other instruments like Erhu may be unsuitable. (ii) Iise's original recordings are obtained from one tone selected by the University of Iowa Electronic Music Studios⁹. Since the sensitivity of human ear for loudness is influenced by frequency, adjusting method that only aims at one tone is unsuitable for other tones possibly.

$$A(T_a) = A(T_r) \times 10^{G \times (TP(T_r) - TP(T_a)) / 20} \quad (1)$$

To avoid the potential problems in the Iise's balancing method for loudness during the study on Erhu, our research consists of two steps. In the first step, we examine whether Iise's balancing method for loudness is suitable for Erhu. Specific methods are as follows: (i) Obtain an adjusting method for Erhu according to subjective experiments, (ii) Utilize multidimensional scaling analysis of dissimilarity ratings on pairs of Erhu's sound differing different balancing methods. In the second step, we compare balancing methods for loudness aimed at one tone with different tones. Specific methods are that through subjective experiments we obtain balancing method for loudness for 12 tones in an octave, and make a comparison between two methods.

3. Experimental method

Experiment 1 and experiment 2 are conducted in a listening room at the Institute of Acoustics, Chinese Academy of Sciences (China).

3.1 Experimental 1.loudness balancing

3.1.1 Stimuli

The stimuli are derived from a traditional Chinese instrument called Erhu played 12 tones in an octave. The tones are performed and recorded in an anechoic chamber at the Institute of Acoustics, Chinese Academy of Sciences (China). In order to compare Iise's balancing method for loudness with the adjusting method aimed at Erhu, we obtain the adjusting method aimed at Erhu according to Iise's procedure including selection of timbre properties and sound synthesis method.

In order to obtain sound varied specific timbre properties, we should select the most salient timbre properties to serve as the basis for sound synthesis. Most timbre perception studies confirm that dimensions correlating to spectral centroid (Tb)¹⁰, spectral flux and attack (characterized by the LRT). The local spectrum variation and the shape of the spectral envelope have also been recognized as an important property¹¹.

Spectral flux has been applied in timbre perception studies and instrument sound synthesize techniques. But Caclin observes that the contribution of spectral flux to dissimilarity ratings will decrease when co-varies with the spectral centroid (Tb) and attack time. This shows that spectral centroid and attack time are more salient than the spectral flux. Studies on musical instrument recognition find that decay/sustain as one of the most basic levels in the hierarchy of the categorization of musical instruments¹². Therefore, four timbre properties are regarded as the most salient to serve as the basis for sound synthesis. The four timbre properties consist of spectral centroid (Tb), spectral irregularity (IRR), attack time (LRT) and a property describing the remainder of the temporal envelope, called sustain/decay or SD.

Each of the four timbre properties was calculated from each of recordings. Specific definition, calculation and sound synthesis methods and procedures can be referred to Iise. We also make some changes. Since too many experiments will influence the accuracy of subjects, it is necessary to improve accuracy of subjective experiment. So we reduce the value of one timbre property from 8 to 5 increments for one tone.

3.1.2 Listeners and procedure

Ten listeners (5 females and 5 males aged between 24 and 28 yrs with an average age of 25) with normal hearing are employed to obtain perceptual data. The total experiment duration is between 1 and 2 h for each listener.

Specific procedure is almost similar to Iise. Subjects need to match the loudness between a test tone changed timbre properties and a reference synthesized tone by adjusting the volume of the test tone. Iise selects a fix threshold value for loudness which subjects can only adjust the volume under 6dB SPL. In order to avoid above condition influencing the accuracy of experiment, we design a MATLAB GUI controlled by subjects to adjust the volume including the threshold value which can offer a more comfortable environment to subjects. The total experiments consist of 12 parts. Each part stands for one tone. After each part, we will check the reliability to examine if the subjects can continue to test.

3.1.3 Results and discussion

In Fig. 1 the red line stands for the balancing method for loudness aimed at present tone instead of the black line which from other tone. The slope of the black line stands for the value of gradient (G in the Eq. (1)). The relationship between methods aimed at present tone and other tone can be observed in Fig. 1. For the spectral centroid, the value of gradient aimed at present tone is a constant. There is a big difference between two methods in the lower value to the spectral centroid. Obviously, the value of gradient aimed at present tone is bigger than the value of gradient aimed at present tone in the lower value to the spectral centroid. Though the pitch of tone increases, the difference between two balancing methods for loudness becomes substantial in the higher value, especially to #a1-#c2. As for irregularity, when the pitch of tone increases, both in the lower value and higher value the difference become larger.

From Fig. 1, we can see that the balancing method for loudness that only aims at one tone can't be suitable to other tones essentially. But whether our method aimed at each tone is better than aimed at one tone in the research on timbre perception needs to be proved.

Since the result balancing method for loudness balancing aimed at attack time (LRT) and sustain /decay (SD) shows human ear can hardly distinguish loudness of changing the timbre property

and keeping original timbre property, we can consider that changing attack time (LRT) or sustain /decay (SD) will not influence loudness of Erhu.

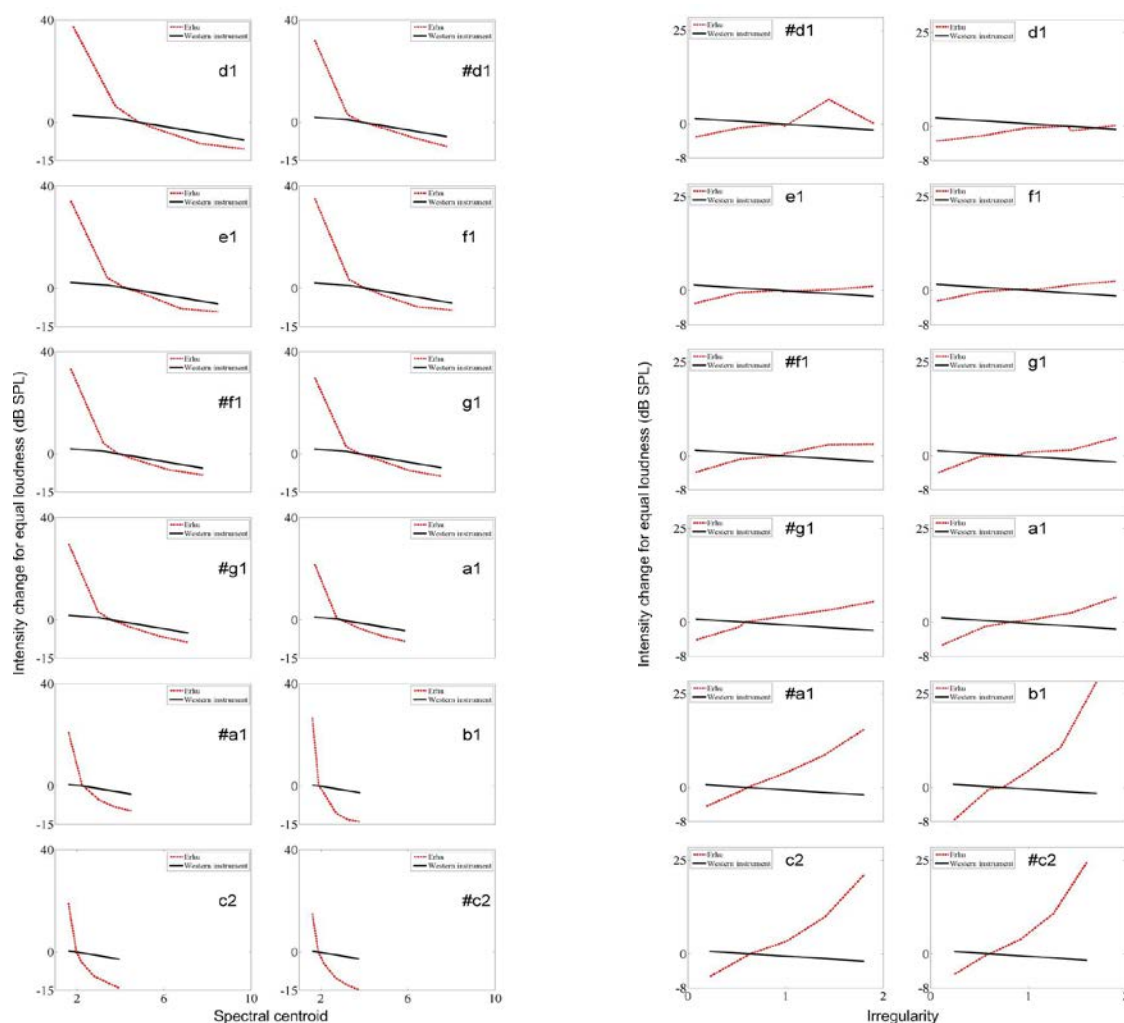
Stimuli in the loudness balancing experiment were synthesized by varying the value of one timbre parameter [Tb, IRR, LRT, or log (SD)] in five increments, while the other three parameters were kept constant. The timbre property range is calculated by the value of timbre property from original recording data and synthesis model. Since the value of timbre properties are changed by tones, the timbre property range will vary from tones. For example, Fig. 1(a) shows that when the pitch gets lower the spectral centroid range used for the loudness balancing experiments decrease.

3.2 Experimental 2. Multidimensional scaling of loudness balancing of different methods

3.2.1 Stimuli

The stimuli are the stimuli of experiment 1 being equalized by 3 balancing methods for loudness. One of the stimuli is equalized by the adjusting method for loudness obtained in the experiment 1. One is equalized by Iise's balancing methods for loudness. The other is not equalized.

A total of 360 sounds equalized by 3 balancing methods for loudness are selected. These sounds consist of 12 groups standing for 12 tones. Each group includes 3 parts standing for 3 balancing methods for loudness. Each part consists of 2 sets standing for 2 timbre property of the sounds. So this experiment divides into 12 groups. Each group consist of 6 sets and each set includes 5 sounds.



(a) (b)
Figure 1. Loudness balancing data for variation in (a): spectral centroid, and (b): irregularity

3.2.2 Listeners and procedure

Ten listeners are employed for this experiment. The total experiment duration is between 1 and 2 h for each listener.

The experiment consists of two parts. One is training phase and another is experiment phase. There are a total of 720 pairs for comparison. Pairs are presented in random order. In the training phase, listeners are told to be familiar with the 360 tones that they can change their rating strategies during the time. In the experiment phase, listeners are required to rate the similarity of the two tones. The similarity rating is made on a scale of 1 to 30, including 3 ranges, very dissimilar, average level of similarity and very similar.

3.2.3 Results and discussion

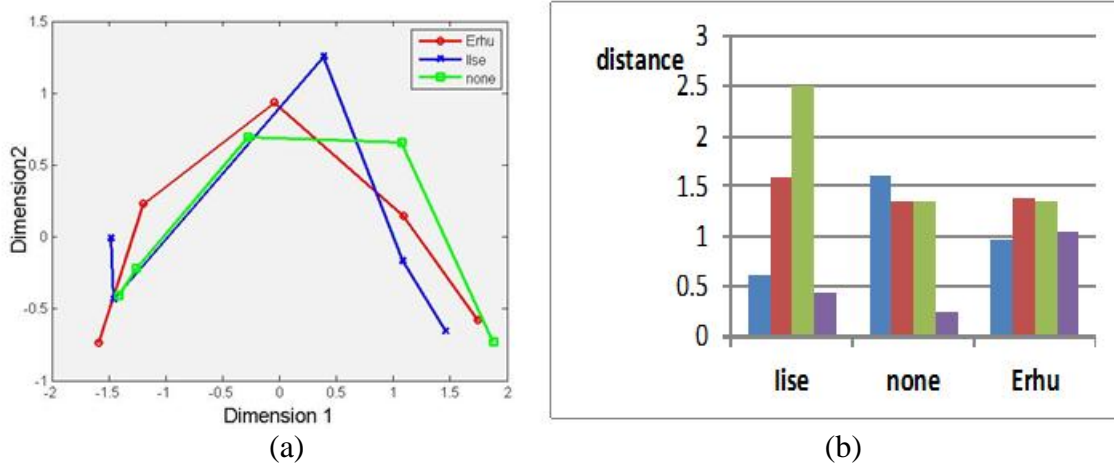


Figure 2. (a): Two-dimensional spatial solution for equal interval variation in spectral centroid for 3 loudness balancing methods, (b): A histogram shows interval distance in each of the two adjacent points for 3 loudness balancing methods.

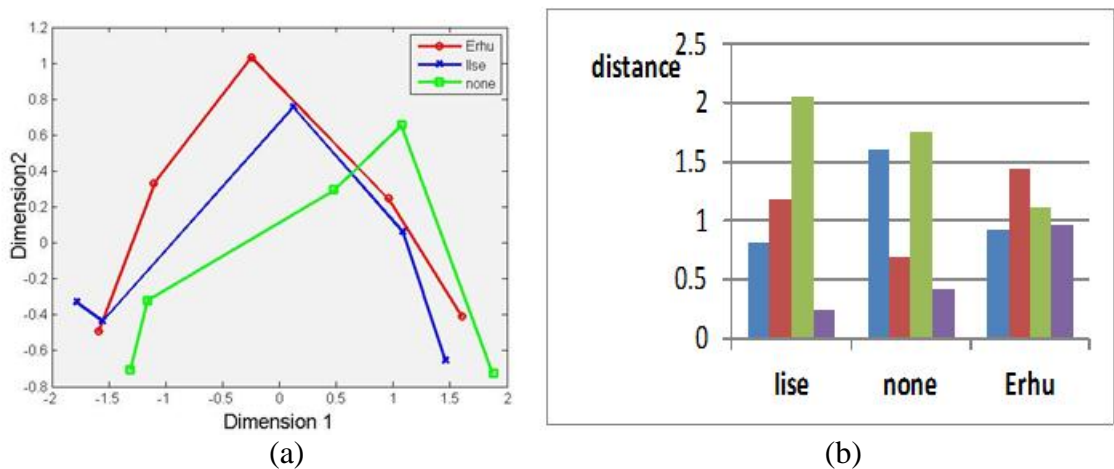


Figure 3. (a): Two-dimensional spatial solution for equal interval variation in irregularity for 3 Loudness balancing methods, (b): A histogram shows interval distance in each of the two adjacent points for 3 loudness balancing methods.

Fig. 2(a) and Fig. 3(a) show a two-dimensional spatial figure for spectral centroid and irregularity, respectively. The distance between points (each point stands for a sound) represents the difference on the timbre. If the proper method for loudness balancing is used, the adjacent points of each curve with equal interval variation on timbre property should be approximately equal^{2,4}. In Fig. 2(a) and Fig. 3(a) we can see that the red line standing for adjusting method for each tone presented by us is best between three methods. Distances of the adjacent points are the most equal than other methods. In the Fig. 2(b) and Fig. 3(b), we can directly find that the interval distance between two adjacent points is the most similar by the loudness balancing method for Erhu. Ilse's balancing method for loudness shouldn't be suitable for Erhu well.

4. Conclusion

According to subjective experiments, a comprehensive adjusting method on loudness for Erhu aimed at each tone is presented. By analysing the data, we can see that there are large differences between two methods aimed at one tone and each tone. In order to confirm the assumption we present, we compare with two methods for loudness balancing by multidimensional scaling analysis of dissimilarity ratings on pairs of Erhu's sound. The result shows that the adjusting method for each tone presented by us is better than Ilse's balancing method for loudness. The result also demonstrates that Ilse's balancing method for loudness aimed at western instruments shouldn't be suitable for Erhu well.

Acknowledgments

This work was supported by National Natural Science Fund of China under Grant No. 11174317.

REFERENCES

- ¹ McAdams, S., Windsberg, S., Donnadiou, S., DeSoete, G., and Krimphoff, J., Perceptual scaling of synthesized musical timbres: Common dimensions specificities and latent subject classes, *Psychol. Res.*, **58**, 177–192, (1995).
- ² Grey, J., and Gordon, J. Perceptual effects of spectral modifications on musical timbres, *J. Acoust. Soc. Am.* 63, 1493–1500. (1978).
- ³ Plomp, R., Timbre as a multidimensional attribute of complex tones, *Frequency Analysis and Periodicity Detection in Hearing*, 397–414. (1970)
- ⁴ Grey, J. M. *An exploration of musical timbre*, Ph.D. thesis, Stanford University, California, (1975).
- ⁵ Grey, J. M. Multidimensional perceptual scaling of musical timbres, *J. Acoust. Soc. Am.* 61, 1270–1277. (1977).
- ⁶ Caclin, A., McAdams, S., Smith, B. K., and Winsberg, S., Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones, *J. Acoust. Soc. Am.* 118, 417–482. (2005).
- ⁷ Marozeau, J., and De Cheveigne, A. The effect of fundamental frequency on the brightness dimensions of timbre, *J. Acoust. Soc. Am.* 121, 383–387. (2007).
- ⁸ Ilse, B. L and Johan, J. H., Preparation of stimuli for timbre perception studies, *J. Acoust. Soc. Am.* 134, 2256–2267. (2013).
- ⁹ Fritts, L. The University of Iowa Electronic Music Studios [Online.] available: <http://theremin.music.uiowa.edu/index.html>. (1997).

- ¹⁰ Marozeau, J., De Cheveign_e, A., McAdams, S., and Winsberg, S. The dependency of timbre on fundamental frequency, *J. Acoust. Soc. Am.* 114, 2946–2957. (2003).
- ¹¹ Gunawan, D., and Sen, D. Spectral envelope sensitivity of musical sounds, *J. Acoust. Soc. Am.* 123, 500–506. (2008).
- ¹² Martin, K. D. Sound-source recognition: A theory and computational model, Ph.D. thesis, Massachusetts Institute of Technology, Massachusetts, 1–172. (1999).