



The Research of Controlling Loudness in the Timbre Subjective Perception Experiment of Sheng

S. Zhu, P. Ji, W. Kuang and J. Yang

Institute of Acoustics, CAS, NO.21, Bei-Si-huan-Xi Road, 100190 Beijing, China

jipeifeng@mail.ioa.ac.cn

Timbre is the attribute of sound that allows human to distinguish among different sound sources. How to map the objective properties of timbre with its subjective properties of timbre perception has been a core issue of timbre research. It is known that when the objective properties of timbre are changed to study the mapping relationship, its loudness will be influenced, which may affect the subject perception on timbre. To remove such effect, the sound whose objective properties have been changed should be properly adjusted. The general experimental method is to adopt the additive synthesis model with a variety of sound synthesis experiments and subjective perception experiments. This approach has two problems. One is that the study only is on one note from different instruments, and it cannot represent the adjustment methods of entire range, especially for the loudness, which is significantly influenced by frequency. The other is that the study ignores the impact on the synthesis model. To solve the above two problems, we adopted both the additive synthesis model and the filter separation method to synthesis sound and considered the influence on the entire range of loudness with one traditional Chinese instrument called "Sheng". The effects on experimental results of the sound synthesized by these two different synthetic models were investigated and the corresponding loudness adjustment curves were obtained. The necessity of adjusting the corresponding loudness between the timbre property relationships is verified through the multidimensional scale model.

1 Introduction

Timbre is a complex and multidimensional attribute of sound which allows subjects to distinguish between two sounds with the same pitch and loudness [1]. However, it doesn't have an accurate definition. Previous research about timbre focuses on how to map the objective properties of timbre with its subjective properties on timbre perception. The general method is to investigate the influence on subjective properties by changing the objective properties of sounds. The multidimensional scaling (MDS) model is a common approach to analyze the relationship between perception timbre space and objective properties of timbre [2, 3]. By rating the sounds based on a standard category-rating paradigm, the MDS model can allow subjects to group the stimuli and to identify the perceptual dimensions. It is convenient for us to interpret the relationship between varied objective timbre properties and subjective timbre properties by the distance in N -dimensional spaces.

It is known that when the objective properties of timbre are changed to study the mapping relationship, its loudness pitch and duration will be influenced, which may affect the subjective perception on timbre. To remove such effect, the sound whose objective properties have been changed should be properly adjusted [2]. The general method is that the stimuli will be equalized in an on-line experiment [3] 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. However, these two methods are too simple to provide a specific adjusting method. Caclin et al. proposed an equation to balance loudness and duration time [4]. Marozeau et al. [5] pointed out that a dimension is correlated with the fundamental frequency when the spectral centroid is changed. Subsequently Labuschagne et al. [6] proposed a systemic approach for balancing loudness, pitch and duration that integrates Caclin' balancing equation with subjective experiment aimed at 13 western musical instruments.

Previous balancing methods aim at one tone [4, 12]. For example, Labuschagne et al. studied pure tone of 262 Hz only. It is known that the sensitivity of human ear for loudness is significantly influenced by frequency. Hence, balancing methods that aim at one single tone may not represent the adjustment methods of entire range, which

includes multiple tones with different frequencies. Is it necessary to study different tones of balancing method for loudness? This is the first problem we want to solve in this study.

Before investigating the relationship between objective property and subjective property on timbre, the sound with varied objective property of timbre is needed to be synthesized. Synthesis models should satisfy two conditions. One is that the model should control the variation on objective property of sound. The other is that the model should change a single objective property of without affecting other objective properties. In order to satisfy above two conditions, simple additive synthesis models are usually adopted in synthesizing sound with a specific timbre property to be varied [4, 7]. Due to the limit of simple additive synthesis model, a temporal envelop from original recording should be multiplied by the signal resulting from the simple additive synthesis model [6]. It is noted that previous study ignores the impact on the synthesis model. This is the second problem we also want to investigate in this study.

To solve the above two problems mentioned, in this paper we use both the additive synthesis model and the filter separation method to synthesis sound and consider the influence of loudness in an octave. The filter separation method is recently introduced into the synthesis model without a temporal envelop compared to the simple additive synthesis model. A quasi-harmonic instrument recording called Sheng is chosen to study the loudness balancing method, which is a traditional Chinese instrument with a big difference in structure and tone compared with the western instruments. The effects on experimental results of the sound synthesized by these two different synthetic models are investigated to obtain the corresponding loudness adjustment curves. The adjusting methods on loudness aimed at one tone and different tones are verified through the MDS model.

2 Sounds synthesis

2.1 Selection of timbre properties

To compare with Labuschagne's balancing method for loudness, we choose the same two spectral properties, i.e., the spectral centroid and irregularity, which are usually regarded as salient timbre properties in the most timbre perception studies [3, 5, 8].

Each of these two timbre properties was calculated from the original recordings. The spectral centroid of Sheng is evaluated as

$$T_b = \sum_{k=1}^N k a_k / \sum_{k=1}^N a_k . \quad (1)$$

where a_k is the amplitude of the k th harmonic and N is the total number of harmonics [8].

The irregularity of Sheng is evaluated as

$$IRR = \sum_{k=1}^N (a_k - a_{k+1})^2 / \sum_{k=1}^N a_k^2 . \quad (2)$$

The $(N + 1)$ th partial is assumed to be zero [9].

2.2 Synthesis method

In order to compare simple additive synthesis with filter separation method, sounds are synthesized by both these two methods. Simple additive synthesis (Eq.(3)) adds sinusoids of amplitudes a_k at harmonic frequencies $f_k = m f_0$ where f_0 is the fundamental frequency and m is the m th harmonic. The amplitudes a_k can be used to control the variation of the spectral centroid and irregularity.

$$S(t) = \sum_{k=1}^N a_k \sin(2\pi k f_0 t) . \quad (3)$$

Filter separation method directly separates the harmonic components of the original recording [10]. It utilizes an inverse comb filter and a resonator to obtain sounds (Figure 1). The harmonic components are separated by filtering the original recording with a fractional delay inverse comb filter (Figure 2) and with a resonator that picks up single harmonic [11]. The filter passes one partial and attenuates the others from the signal. The transfer function of the resonator is given by

$$H_r(z) = \frac{1}{1 - 2R \cos(\theta) z^{-1} + R^2 z^{-2}} . \quad (4)$$

where R is the radius and θ is the angle of the poles of the resonator in the z -plane.

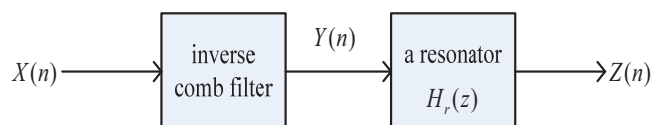


Figure 1: A diagram of filter separation method.

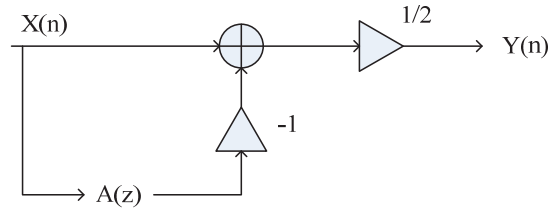


Figure 2: The implementation of the fractional delay inverse comb filter.

2.3 Boundary values of timbre properties

According to Labuschagne's paper the timbre property ranges are summarized in Table 1. The original recording is served as a reference tone with the reference value of timbre properties. The maximum of spectral centroid is set to twice spectral centroid value of the reference tone. The extreme values of irregularity column and the minimum of spectral centroid are calculated by the formula of the spectral centroid and irregularity [6].

Table 1: Boundaries of the values used for the balancing experiment.

	IRR	min	max	Tb	min	max
c1	1.28	0.04	1.35	4.78	2.50	9.56
#c1	1.47	0.04	1.55	4.53	2.35	9.06
d1	1.99	0.04	2.15	4.24	2.75	8.48
#d1	1.51	0.04	1.65	4.26	2.35	8.52
e1	1.61	0.04	1.75	4.55	2.50	9.10
f1	1.71	0.04	1.85	4.06	2.35	8.12
#f1	1.59	0.04	1.65	4.42	2.35	8.84
g1	1.05	0.04	1.25	4.98	2.35	9.96
#g1	1.00	0.04	1.15	5.26	2.35	10.52
a1	1.11	0.04	1.15	3.98	2.35	7.96
#a1	0.57	0.04	1.15	4.39	2.35	8.78
b1	0.96	0.04	1.15	3.74	2.35	7.48

3 Loudness balancing

3.1 Stimuli

The stimuli are derived from a traditional Chinese instrument called Sheng 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 the experiment of Labuschagne, the total experiment duration is between 5 and 7 hours. Too many experiments may affect the listeners' patience. So we reduce the value of one timbre property from 8 to 5 increments for one tone. In the loudness balancing experiment stimuli are synthesized by varying the value of one timbre property in five increments, while another property is kept constant. So each set contains 20 tones recreated by two synthesis methods and two objective timbre properties. A total of 240 sounds are used in the experiment of loudness balancing.

Subjective experiment is conducted in a listening room at the Institute of Acoustics, Chinese Academy of Sciences (China). Listeners 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. Tone presentation was controlled by a Matlab procedure on a personal computer. Tones were presented through an AKG K550 headphone via an RME Fireface UC external soundcard. Listener responses are recorded automatically by the Matlab procedure.

3.2 Listeners and procedure

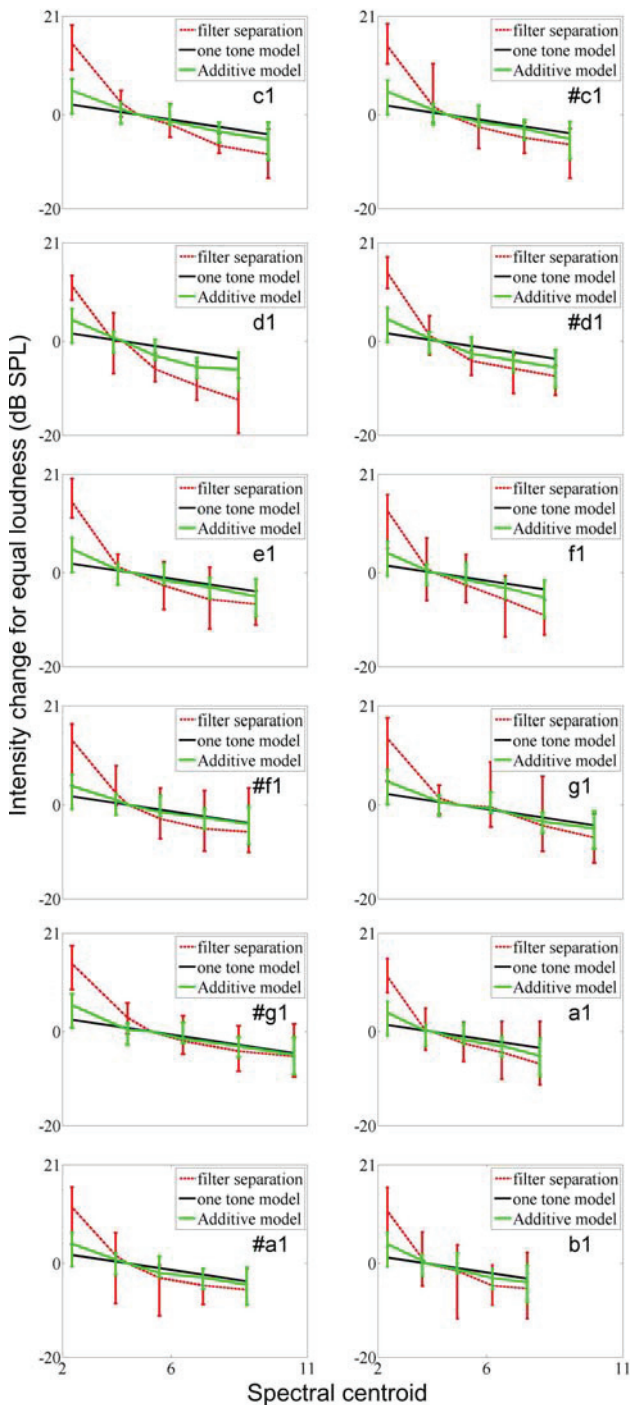


Figure 3: Loudness balancing data for variation in spectral centroid.

Ten listeners (5 females and 5 males aged between 24 and 28 yrs with an average age of 25) with normal hearing [pure tone thresholds 20 dB hearing level (HL) for 250, 500, 1000, 2000, 4000, and 8000 Hz] are employed to obtain perceptual data. The total experiment duration is between 4 and 5 h for each listener.

Specific procedure is almost similar to Labuschagne's [6]. Listeners need to match the loudness of a test tone

changed timbre properties with a reference synthesized tone by adjusting the volume of the test tone. In order to avoid influencing the accuracy of experiment, a MATLAB GUI is designed. It is controlled by listeners to adjust the volume including the threshold value which can offer a more comfortable environment to listeners. The total experiments consist of 12 parts. Each part stands for one tone.

3.3 Results and discussion

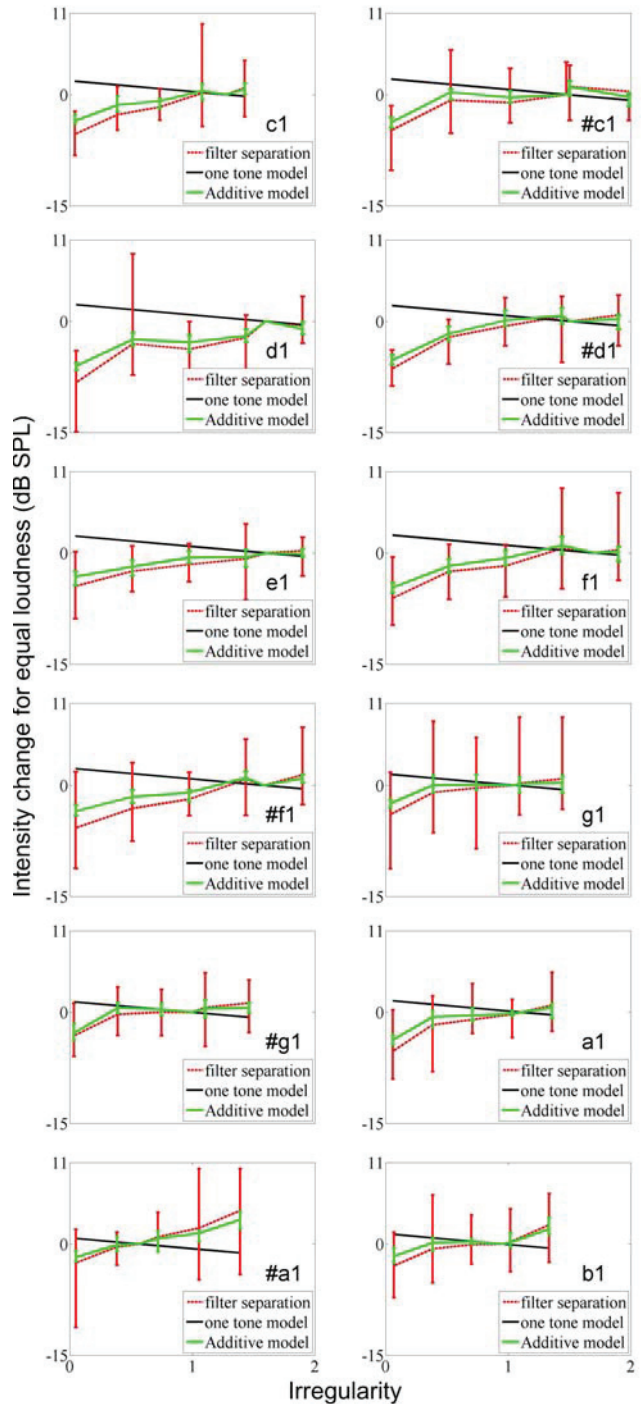


Figure 4: Loudness balancing data for variation in irregularity

The range of intensity changes required for equal loudness differs by several dBs across listeners, which is larger than typical intensity discrimination thresholds, so that balancing methods derived from one set of listeners may not be applicable to another.

The relationship to the intensity change for equal loudness among these three methods is shown in Figures 3

and 4, where the red dotted line and the black line stand for the average loudness balancing data synthesized by the filter separation method and the method aims at one tone from Labuschagne's loudness balancing equation, respectively. The green line symbolizes that the average loudness balancing data synthesized by the simple additive synthesis model. The standard deviation across listeners is also illustrated in Figures 3 and 4. For the spectral centroid, there is big difference in the lower value between three methods. In the lower value range, the gradient of filter separation method is the highest, followed by additive synthesis model. The gradient of one tone model is the lowest. It can be concluded that the adjusting method aims at one tone may not be suitable to balance the spectral centroid in the low value range. It also suggests that there is a big difference between the additive synthesis model and the filter separation method on the balancing data in the low value range of spectral centroid.

As for the irregularity, along with the pitch of tone increases the difference between three balancing methods for loudness becomes substantial in the lower value, especially to g1-b1. There is a big difference in the whole range among these three methods. An opposite tendency on loudness balancing is found in the method aims at one tone. It illustrates that the adjusting method aims at one tone may not be suitable to balance the irregularity. Figure 4 also shows that synthesis sound methods have little influence on the irregularity.

We can see that when the balancing method for loudness aims at one tone is applied into other tones, there are big differences. The verification of adjusting the corresponding loudness between the timbre property relationships is carried out through the MDS model in Section 4.

4 Multidimensional scaling of loudness balancing

4.1 Stimuli

The stimuli are the results of experiment of loudness balancing being equalized by 3 balancing methods for loudness. The first stimuli are equalized by the adjusting method for loudness obtained in the experiment of loudness balancing. The second stimuli are equalized by Labuschagne's balancing methods aims at one tone. The third stimuli are 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 is divided into 12 groups. Each group consists of 6 sets and each set includes 5 sounds.

4.2 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.

4.3 Results and discussion

Figures 5(a) and 6(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. It is expected that if the proper method for loudness balancing is used, the adjacent points of each curve with equal interval variation on objective property of timbre should be approximately equal. In Figures 5(a) and 6(a) it can be observed that the red line standing for adjusting method for each tone is best among these three methods, since its distances of the adjacent points are the most equal than other methods, which can be directly observed in the Figure 5(b) and 6(b). The error bars indicate the standard deviation across different tones.

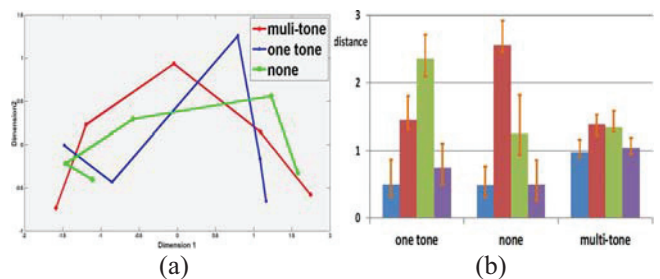


Figure 5: (a): Two-dimensional spatial solution for equal interval variation in spectral centroid for 3 loudness balancing methods, (b): A histogram shows interval distance in each of the two adjacent points for 3 loudness balancing methods.

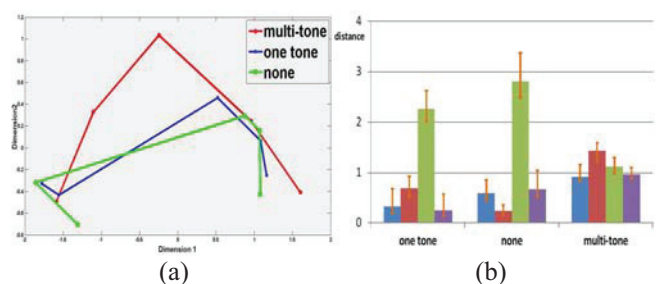


Figure 6: (a): Two-dimensional spatial solution for equal interval variation in irregularity for 3 loudness balancing methods, (b): A histogram shows interval distance in each of the two adjacent points for 3 loudness balancing methods.

Though the method for loudness balancing aims at different tones is the most equal between each two adjacent points, there is also some discrepancy. The possible reason is that the sounds varied timbre properties are also affected by the pitch and duration [12, 13].

The stress and squared correlation (RSQ) in the MDS are shown in Table 2. Stress values stand for the fitting degree for the MDS. The smaller the stress values are, the better the fitting degree is. RSQ values are the proportion of variance of the scaled data (disparities) in the partition (entire data) which is accounted for by their corresponding distances. The higher the RSQ values are, the better the distances can explain the differences between different tests.

It is clear that the fitting degree is suitable and the distances can describe the differences between different tests from the table.

Table 2: Stress and RSQ in MDS

	Tb		IRR	
	Stress	RSQ	Stress	RSQ
Multi-tone	0.00273	0.99997	0.00312	0.99992
one tone	0.0025	0.99987	0.00452	0.99991
none	0.00213	0.99997	0.00241	0.9994

5 Conclusion

The research on loudness adjustment in the timbre subjective perception experiment of Sheng was carried out in this study using two synthesis models. The corresponding loudness adjustment curves were obtained by the subjective tests for variation in spectral centroid and irregularity. The results demonstrate that there is a big difference between the additive synthesis model and the filter separation method on the balancing data in the low value range of spectral centroid. However, there is little influence on the whole range of irregularity. The necessity of adjusting the corresponding loudness between the timbre property relationships is verified through the MDS model. Compared with the balancing method for loudness aims at one tone by the MDS model, the method aims at different tones is verified to be necessary for the balancing test.

Acknowledgments

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

References

- [1] ANSI. ANSI S1.1-1994 (R 1999), American National Standard Acoustical Terminology, Acoustical Society of America, New York (1994)
- [2] J. M. Grey. An exploration of musical timbre, Ph.D. thesis, Stanford University. (1975).
- [3] J. M. Grey. Multidimensional perceptual scaling of musical timbres, *J. Acoust. Soc. Am.* 61, 1270-1277(1977)
- [4] A. Caclin, S. McAdams, B. K. Smith and S. Winsberg, Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones, *J. Acoust. Soc. Am.* 118, 471-482. (2005).
- [5] J. Marozeau and A. D. Cheveigne, The effect of fundamental frequency on the brightness dimensions of timbre, *J. Acoust. Soc. Am.* 121, 383-387. (2007).
- [6] I. B. Labuschagne and J. J. Hanekom, Preparation of stimuli for timbre perception studies, *J. Acoust. Soc. Am.* 134, 2256-2267. (2013).
- [7] K. Jensen, Timbre models of musical sounds, Ph.D. thesis, University of Copenhagen. (1999).
- [8] J. Krimphoff, S. McAdams and S. Winsberg, Characterizing the timbre of complex sounds. II. Acoustic analyses and psychophysical quantification, *J. Phys.* (Paris) (1994).
- [9] K. Jensen, Envelope model of isolated musical sounds, in Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects, Trondheim, Norway, pp. W99-91-W99-94.4,625-628. University of Copenhagen, Denmark, pp. 51-98. (1999).
- [10] M. Ilmoniemi, Modification and brain recordings of musical instrument tones, Master's thesis, Helsinki University of Technology. (2007).
- [11] V. V. Timo, J. Laakso, M. Karjalainen and U. K. Laine, Splitting the unit delay tools for fractional delay filter design, *IEEE Signal Processing Mag.*, pp. 30-60. (1996).
- [12] J. Grey, and J. Gordon, Perceptual effects of spectral modifications on musical timbres, *J. Acoust. Soc. Am.* 63, 1493-1500. (1978).
- [13] S. McAdams, S. Winsberg, S. Donnadieu, G. D. Soete and J. Krimphoff, Perceptual scaling of the synthesized musical timbres: Common dimensions, specificities, and latent subject classes, *Psychol. Res.*, 58, 177-192. (1995).