Exploring Instrument Blending as a Function of Timbre Saliency

Song Hui Chon,*[1] Stephen McAdams*[2]

*CIRMMT, Schulich School of Music, McGill University, Canada
[1]songhui.chon@mail.mcgill.ca, [2]smc@music.mcgill.ca

ABSTRACT

Background

Timbre is what enables us to distinguish two instruments playing the same note at the same pitch and loudness. The main focus of the timbre research for the last 30 years has been to determine the acoustic correlates of timbre dissimilarity space dimensions, using isolated sounds of instrument timbres. However, understanding timbre spaces of isolated tones does not help much with real-world cases – the perception of combined timbres. There are orchestration handbooks with much-used examples, although their underlying principles have not yet been scientifically studied. We conjecture that timbre saliency, the attention-capturing quality of timbre [Chon & McAdams, 2012], might be a factor behind some of these examples.

Aims

Saliency refers to the character of being able to draw people's attention. If something is salient, it implies that it does not blend well with its surroundings. In this paper, we explore the perception of blending of two instrument timbres as a function of timbre saliency.

Method

A rating experiment was carried out to understand the relationship between blending and timbre saliency. The stimuli were generated from 15 Western orchestral instrument sounds from the Vienna Symphonic Library, all of which were equalized in terms of pitch (C4), loudness and effective duration to minimize the impact of these parameters in the experimental task. The instruments are CL (clarinet), EH (English horn), FH (French horn), FL (flute), HA (harp), HC (harpsichord), MA (marimba), OB (oboe), PF (piano), TB (tubular bells), TN (trombone), TP (trumpet), TU (tuba), VC (violoncello) and VP (vibraphone). There were 105 composite sounds generated for the experiment by combining all pairs of the 15 above-mentioned instrument sounds.

Listeners were presented with a composite of two simultaneous, unison instrument sounds varying in degree of timbral and saliency difference and were asked to rate the degree of blending on a continuous scale between "very blended" and "not blended".

The experiment was carried out in a sound-attenuated booth in the Music Perception and Cognition Laboratory of the Schulich School of Music of McGill University. Mono sound signals were presented to both ears of the participants via Sennheiser HD 280 headphones, to further eliminate any undesired cue for separation. Sixty participants were recruited from a classified ad on the McGill University website. There were equal numbers of males and females, and musicians and non-musicians, all of whom passed the preliminary hearing test. They were compensated upon completion of the experiment.

Results

A repeated-measures analysis of variance (ANOVA) on the blending judgment data from 60 subjects shows that there was no effect of gender, musicianship or age on blending judgments. This means that the only factor affecting the responses was the instrument combination ($p < 0.00001$). The blending judgment data for each instrument pair were averaged across all subjects for further analysis. Figure 1 shows the distribution of average blending judgments for all pairs involving each instrument on the horizontal axis. Notice that the average blending judgments varied quite a bit across instruments. In particular, the instruments with sharper attacks (HA, HC, MA, PF, TB, VP) tended to have less varied ratings (i.e., narrower boxes) as well as lower medians (red bars inside boxes) than others. Lower variability suggests that these timbres tended to affect the overall blending, while greater variability implies that the degree of blend depended more on the other timbre. The two groups (sharper attacks – HA, HC, MA, PF, TB, VP and slower attacks – CL, EH, FH, FL, OB, TN, TP, TU, VC) show a significant difference in averages of median values ($p < 0.0001$), suggesting that the attack patterns play an important role in perceived degree of blending.

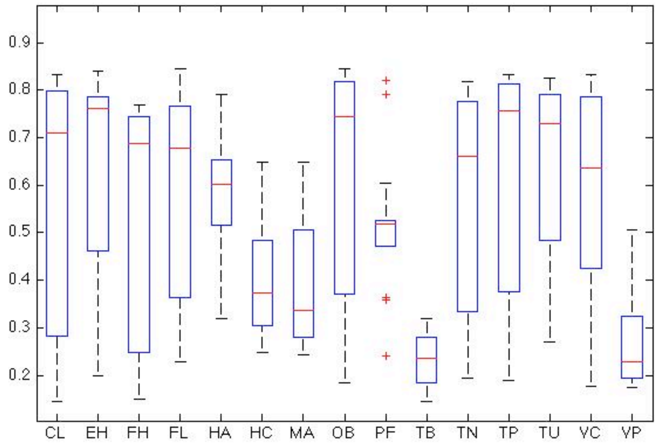

Figure 1: Boxplot of Average Blending Judgments

Mild negative rank correlations are observed between the average degree of blending and the sum ($\rho = -0.34$, df = 103, $p < 0.001$), minimum ($\rho = -0.26$, df = 103, $p < 0.01$) and maximum ($\rho = -0.30$, df = 103, $p < 0.01$) of saliency values of two individual timbres. Timbre saliency values were taken from a previous study [Chon & McAdams, 2012]. Saliency differences show a non-significant negative correlation ($\rho = -0.12$, df = 103, $p = 0.22$). These indicate that a highly salient sound will not blend well, as one might expect intuitively. In addition, it is the individual sound's saliency level and the sum of the saliencies of the sound pair that determine the overall degree of perceived blending, rather than the difference in saliency between individual sounds.

Among all acoustic descriptors from the Timbre Toolbox [Peeters et al., 2011], the best acoustic correlate to describe the average blending is the minimum attack time of the two individual timbres, which alone explains 57% of the variance. This suggests that the instrument sound with a sharper attack tends to dominate the perceived degree of blending. Judging from the positive beta value ($\beta = 13.9$) of the minimum attack time in the regression model, larger minimum attack times contribute to greater degree of blending. The minimum attack time also shows a moderate rank correlation with the average blending ratings ($\rho = 0.76$, df = 103, $p < 0.00001$), which means the larger the minimum attack time is, the better its blending is. This confirms Tardieu & McAdams' (2011) observation that a sound with a longer attack tends to blend better. In terms of spectral centroids, the average blending ratings are mildly rank correlated with the maximum ($\rho = -0.46$, df = 103, $p < 0.000001$), the sum ($\rho = -0.44$, df = 103, $p < 0.000001$), the difference ($\rho = -0.35$, df = 103, $p < 0.001$) and the minimum ($\rho = -0.22$, df = 103, $p < 0.05$) of the two centroids, which supports previous findings by Sandell (1995) and Tardieu & McAdams (2011) that sounds with lower spectral centroids are likely to blend better.

Conclusions

As a first step towards understanding orchestration, the perception of blending of two concurrent unison timbres was studied in terms of timbre saliency, which is the attention-capturing quality of timbre. Something salient must not blend with its surroundings by definition. Therefore, if timbre saliency was indeed a measure of attention-capturing quality of timbre, it should exhibit a negative relationship with the perceived degree of blending.

In a blend-rating experiment with pairs of 15 instrument sounds, no significant effect of gender, musicianship or age was found. The average blending judgments showed negative correlations with sum, minimum and maximum of individual saliency values, confirming the validity of timbre saliency from our previous work. A salient timbre tended not to blend well indeed.

Although we verified the hypothesis that timbre saliency would have a negative relationship with the degree of blending, the correlation was moderate at best. This implies that there are other more significant factors that affect the blending of two unison instrument sounds, such as the attack time and spectral centroids. These factors could be changed by performance techniques on an instrument, which are often indicated on the score.

The instrument sounds we considered were all on the same pitch, loudness and effective duration, where the perceived blending is artificially maximized. The blending judgments in our data then reflect the best degree of blending between two instruments playing the same pitch, duration and loudness; combining instruments in non-unison contexts may lead to a lesser degree of blending due to pitch differences. The next step will consider various pitches in a more musical setting to study the effect of timbre saliency on voice perception.

Keywords

Timbre, saliency, blending, orchestration, perception.

REFERENCES

Chon, S.H. & McAdams, S. (2012). Investigation of Timbre Saliency, the Attention-Capturing Quality of Timbre. *Proceedings of the Acoustics 2012 Hong Kong*.

Peeters, G., Giordano, B.L., Susini, P., Misdariis, N. & McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *Journal of the Acoustical Society of America*, **130**, 2902-2916.

Sandell, G.J. (1995). Roles for spectral centroid and other factors in determining "blended" instrument pairings in orchestration. *Music Perception*, **13**, 209-246.

Tardieu, D. & McAdams, S. (2011). Perception of dyads of impulsive and sustained sounds. *Proceedings of the 2011 Meeting of the Society for Music Perception and Cognition, Rochester*.