INTRODUCTION

Timbre is defined as “that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar” (ANSI 1960). It is what enables us to distinguish a note (for example, C4) in mezzo forte played on a piano from the same note played on a clarinet. As inferred from the ANSI definition, timbre is multidimensional because it refers to all characteristics of a sound that are not related to pitch or loudness.

Ever since Helmholtz started research on timbre (Helmholtz, 1877), researchers have spent much effort on scientifically understanding the multidimensionality of timbre. In our earlier work, we proposed the concept of timbre saliency, the attention-capturing quality of timbre, which is a new perceptual property that is a function of this multidimensionality of timbre (Chon & McAdams, 2012).

How then is timbre saliency different from timbre dissimilarity? What is the relationship between them? How can we explain the relationship in terms of acoustic features? These are the questions we examine in this paper.

Timbre Saliency

![Figure 1: The one-dimensional saliency scale of 15 instrument sounds (from Chon & McAdams, 2012).](image)

Timbre saliency is a new concept proposed to consider the attention-capturing quality of timbre (Chon & McAdams, 2012). Its acoustic correlate turned out to be the odd-even harmonic ratio among those calculated by the Timbre Toolbox (Peeters et al., 2011) on the pairwise saliency judgment data from 40 participants for 15 instrument sounds – Clarinet (CL), English Horn (EH), French Horn (FH), Flute (FL), Harp (HA), Harpsichord (HC), Marimba (MA), Oboe (OB), Piano (PF), Trombone (TN), Trumpet (TP), Tuba (TU), Tubular Bells (TB), Violoncello (VC), and Vibraphone (VP). A tapping technique was used to measure saliency of two timbres A and B in a perceptually isochronous ABAB sequence, the pitch, loudness and effective duration of which were all equalized. Participants were instructed to tap to the timbre that sounded like a strong beat, either A or B but not both. The hypothesis was that the more salient a timbre is, the more attention it will draw from the participants, and hence be tapped to more often. Figure 1 shows the one-dimensional saliency scale obtained from CLASCAL (Winsberg & De Soete, 1993), which is a multidimensional scaling (MDS) algorithm that models dissimilarity ratings with a distance model.
including weights on the shared dimensions and specificities for each item in the space for latent classes of subjects. Although the saliency scale is one-dimensional, it is presented in two dimensions in Figure 1 because of the seven instruments closely positioned around 0. This figure suggests that there are three saliency regions – lower (EH, TN, CL, FH), middle (VC, HC, FL, TU, VP, OB, HA) and higher (PF, TB, TP, MA).

**TIMBRE DISSIMILARITY EXPERIMENT**

A classic timbre dissimilarity judgment experiment was carried out in a sound-attenuated booth in the Music Perception and Cognition Laboratory of the Schulich School of Music at McGill University. The same set of stimuli from the saliency experiment was used. Listeners were first presented with all stimuli one after another so that they could establish the (dis)similarity range of the stimuli. Then for each pair of stimuli presented, participants were required to indicate the dissimilarity of the pair on a continuous scale given in the graphic user interface, the extremes of which were marked “identical” on the left and “very dissimilar” on the right (shown in Figure 2). They could repeat listening to the pair of stimuli as many times as they liked, using the “play again” button. The order of stimulus pairs was randomized. Twenty participants from ages 19 to 39, balanced in gender and musicianship, took part in the study.

![Screenshot of the timbre dissimilarity experiment](unnamed.png)

**FIGURE 2.** Screenshot of the graphic user interface of the timbre dissimilarity experiment

The continuous scale in the graphic user interface was converted to a number between 0 and 1, where 0 corresponds to the “identical” end and 1 to the “very dissimilar” end. Repeated-measures ANOVA using SPSS did not show any significant effect of gender or musicianship. Dissimilarity judgments from 20 individuals were then formed into 20 individual lower triangle matrices, which were analyzed with the CLASCAL multidimensional scaling algorithm (Winsberg & De Soete, 1993) to obtain the timbre space in which dissimilarities are modeled as distances. The best solution turned out to be a two-dimensional space with five latent classes and specificities. The two common dimensions are presented in Figure 3.

Notice that the instruments with more percussive attacks are located above the $y = 0$ line. This makes us suspect that the second dimension may be correlated with attack time. Among the timbre descriptors from the Timbre Toolbox (Peeters et al., 2011), the first dimension shows a strong correlation with spectral centroid [$r(13)=.845$, $p<.0001$] and spectral spread [$r(13)=.855$, $p<.0001$], both based on the ERB-FFT model spectrum. The second dimension is moderately correlated with the attack time [$r(13)=.692$, $p=.004$] and power spectral crest [$r(13)=.732$, $p<.005$]. This confirms spectral centroid and attack time as two major acoustic correlates of timbre dissimilarity, as reported in earlier timbre dissimilarity literature (Grey, 1977; Grey & Gordon, 1978; Krumhansl, 1989; McAdams et al., 1995; Lakatos, 2000; Caclin et al., 2005). The five latent classes exhibited little correlation with gender, musicianship or age, which also agrees with the existing literature.

The saliency dimension in Figure 1 is moderately correlated with the second dimension [$r(13)=.578$, $p=.024$] of the timbre space but not with the first dimension [$r(13)=.182$, $p=.52$], which may suggest that the saliency might be more related to the temporal characteristics of timbre. A further investigation may be necessary to study the relationship of the timbre saliency and timbre dissimilarity as timbre saliency research is in its infancy.
A classic timbre dissimilarity experiment was carried out to study its relationship with timbre saliency, a new concept referring to the attention-capturing quality of timbre. The result is a two-dimensional timbre dissimilarity space with five latent classes and specificities. The two dimensions show moderate to strong correlations with spectral centroid and attack time, which confirms previous reports in the literature that they are two of the most important descriptors of timbre dissimilarity. Also, no effect of gender or musicianship was found in the dissimilarity judgment data, which is consistent with the existing literature.

The timbre saliency dimension showed a moderate correlation with the second dimension of the timbre space and an insignificant correlation with the first dimension. This may suggest that the perception of saliency may be attack-dependent, i.e., the shorter the attack time the more salient. This was reflected in participants’ remarks after the saliency experiment that they tended to tap to timbres with a faster attack. They also reported that they often tapped to timbres with lower spectral centroids. This was due to the unforeseen confound with rhythmic saliency in the task, where the ABAB sequences form a duple rhythm, which is often used in rock music where the strong beat is accompanied with the bass drum (hence lowering the spectral centroid) and the weak beat with the snare.

Due to the lack of correlation between the first dimension of the timbre space and the saliency dimension, it was not possible to do any further statistical analysis, such as regression analysis of saliency dimension in terms of the two dissimilarity dimensions. As we had hoped to find a common ground between timbre saliency and timbre dissimilarity, it is disappointing to see the lack of consensus between the saliency dimension and the first dimension of timbre space. This made us wonder, however, about context effects and the transfer of contexts. Our measure of timbre saliency reflects the saliency judgment in the tapping context. This is different from the dissimilarity context and the lack of statistical agreement between two spaces may indicate that the two contexts may not be transferrable. A more detailed explanation would require research on the context effects in auditory saliency, which does not yet exist to our knowledge.

The fact that the saliency space has only one dimension and the dissimilarity space has two dimensions, one of which is moderately correlated with the saliency dimension, may hint that saliency is a sufficient condition to
dissimilarity while dissimilarity is a necessary (but not sufficient) condition to saliency. But the correlation is moderate at best, suggesting that there may be little in common between saliency and dissimilarity and that there are other factors outside of the acoustic correlates of timbre saliency and timbre dissimilarity spaces affecting these two aspects of timbre perception.

ACKNOWLEDGMENTS

This work was funded by grants from the Canadian Natural Sciences and Engineering Research Council (NSERC) and the Canada Research Chairs (CRC) program to Stephen McAdams and by the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) to Song Hui Chon. A special thanks to Bennett Smith for technical assistance with the experiment.

REFERENCES