INSTRUMENT IDENTIFICATION IN CONCURRENT UNISON DYADS: THE EFFECT OF TIMBRE SALIENCY

Song Hui Chon, David Huron

School of Music
Ohio State University

ABSTRACT

Saliency refers to the character of an object standing out, hence easily capturing someone's attention. This concept, when applied to music, can be interpreted in a way that a highly salient instrument sound will more likely draw a listener's attention than other concurrent instruments. That is, we might expect in turn that a highly salient instrument timbre will be more easily identified from other concurrent sounds, which will be also reflected by listeners' higher confidence ratings. In order to test these hypotheses, an experiment was carried out where twenty-seven listeners heard concurrent unison dyads of different instruments and were asked to identify one instrument present. Presented with a list of possible instruments, listeners identified one instrument they heard in a stimulus and rated their confidence that the chosen instrument was present. The results proved not to be consistent with the hypothesis that listeners are better at identifying those instruments that exhibit greater timbre saliency (Chon, 2013). Possible explanations for this failure are discussed. Still, a mild yet statistically significant correlation was observed between the timbre saliency values and average confidence ratings, which is consistent with one of our hypotheses that listeners would be more confident of identifying more salient instrument sounds.

1. INTRODUCTION

Timbre is widely regarded as the principal correlate of *sound source identity*. In the musical application of orchestration, timbre plays an important role in the phenomenon of *blend*. In the theory that motivates this research, both sound source identity and blend are conjectured to be influenced by a third concept, *timbre saliency*.

Timbre saliency refers to the attention-capturing quality of timbre (Chon & McAdams, 2012a). This definition can be interpreted in a way that a highly salient sound will more likely draw a listener's attention than other concurrent sounds. In the case of music, one might suppose that the degree of instrument saliency might have been an implicit factor in orchestration practices. Chon (2013) carried out a series of experiments to define timbre saliency and to examine its effect on perceived blend and voice recognition. Chon and McAdams (2012b) reported that the perceived degree of blend was negatively correlated with the timbre saliency values measured by Chon (2013), which was in agreement with the hypothesis that a salient timbre would not blend well.

There has been only a handful of research on timbre blend (Kendall & Carterette, 1993; Sandell, 1995; Tardieu & McAdams, 2011; Chon & McAdams, 2012b). Unlike most timbre research that focus on isolated sounds, blend studies focus on the interaction of concurrent sounds. In these studies, similar timbres were observed to blend better (Kendall & Carterette, 1993). The degree of perceived blend has been reported to be negatively correlated with the identifiability of the underlying timbres (Kendall & Carterette, 1993). Lower spectral centroids (i.e., a darker sound) and a slower attack have been shown to be important in increasing the degree of blend (Sandell, 1995; Tardieu & McAdams, 2011; Chon & McAdams, 2012b).

In this study, the theory of timbre saliency will be used to examine the identification of an instrument in concurrent unison dyads. Specifically, we have two hypotheses:

H1. It is easier for listeners to identify timbres that are more salient. This ease of identification will be reflected in the average correct identification rate across listeners.

H2. Listeners are more confident of their choice of instrument identification with more salient instruments. This implies that there exists a positive correlation between the estimated timbre saliency scores of different timbres from Chon (2013) and the confidence ratings for listener identification of those sounds.

2. EXPERIMENT

2.1 Stimuli

The sounds used in the study were the same as those used in Chon and McAdams (2012a/b). These sounds include recorded samples of tones produced by 15 Western orchestral instruments from the Vienna Symphonic Library (VSL, 2011): Clarinet (CL), English Horn (EH), Flute (FL), French Horn (FH), Harp (HA), Harpsichord (HC), Marimba (MA), Oboe (OB), Piano (PF), Trombone (TN), Trumpet (TP), Tuba (TU), Tubular bells (TB), Violoncello (VC) and Vibraphone (VP). Each single-instrument tone was equalized in terms of pitch (C4, 261.6 Hz), loudness and effective duration. For this experiment, some stimuli consisted of single instrument tones. However, most stimuli consisted of combinations of two different timbres. In total, 120 stimuli were tested.

2.2 Participants

Three groups of participants were recruited based on self-identification: 10 professional musicians, 8 amateur musicians, and 9 non-musicians. Participants were undergraduate and graduate students and a faculty member from the Ohio State University community, including 20 females and 7 males. All of them reported normal hearing.

Table 1: The confusion matrix from single-instrument timbre identification trials. On the row is the indicated instrument, whereas on the column is the actual played instrument. The instruments are listed according to the instrument families. (CL=clarinet, FL=flute, OB=oboe, EH=English horn, TP=trumpet, FH=French horn, TN=trombone, TU=tuba, MA=marimba, VP=vibraphone, TB=tubular bells, HC=harpsichord, HA=harp, PF=piano, VC=violoncello)

	CL	**FL**	**OB**	**EH**	**TP**	**FH**	**TN**	**TU**	**MA**	**VP**	**TB**	**HC**	**HA**	**PF**	**VC**
CL	**0.85**	0.08	0.04	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FL	0.07	**0.83**	0.01	0.03	0.01	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OB	0.00	0.03	**0.57**	0.33	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01
EH	0.01	0.00	0.21	**0.72**	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
TP	0.00	0.00	0.04	0.04	**0.69**	0.11	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FH	0.00	0.01	0.00	0.03	0.08	**0.42**	0.39	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TN	0.00	0.03	0.00	0.03	0.21	0.14	**0.58**	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TU	0.00	0.01	0.00	0.01	0.01	0.10	0.07	**0.74**	0.00	0.01	0.00	0.00	0.03	0.01	0.00
MA	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	**0.92**	0.06	0.00	0.00	0.00	0.00	0.00
VP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	**0.79**	0.02	0.04	0.00	0.00	0.00
TB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	**0.90**	0.02	0.00	0.00	0.00
HC	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.04	0.00	**0.92**	0.02	0.00	0.00
HA	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.02	0.00	0.00	**0.85**	0.04	0.00
PF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	**0.98**	0.00
VC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**1.00**

2.3 Procedure

Prior to the experiment, participants received an introduction to the experiment and gave a verbal consent. They then read the instructions on a printed sheet of paper. Participants were tested individually in an Industrial Acoustics Corporation sound attenuated room. Stimuli were heard through Sennheiser HD 280 Pro headphones at a comfortable listening level. The order of stimuli was randomized for each participant.

The entire experiment was implemented in MATLAB including the graphic user interface. The experiment consisted of three parts: training, preliminary test and main test. A "timbre training" display was provided for the training part with buttons whose labels identified each of the 15 individual timbres. After exploring these timbres at leisure, the participant proceeded to the preliminary test, where he/she was asked to identify each of the 15 timbres. Since the wind instruments were known in advance to be more easily confused with each other, the eight wind instrument sounds were tested twice, resulting in 23 trials. All trials were randomized in order.

Feedback was provided during the preliminary test. Participants were forced back to training and preliminary test until they had achieved a minimum of 65% correct identification to advance to the main experiment. This criterion was set a priori, prior to any data collection.

In the main test, participants were presented with either a single instrument timbre or concurrent unison dyads of two instrument timbres. The stimuli were presented in two blocks of 60 stimuli each, with a brief break available between the blocks. The task was to identify one of the instruments present in the stimulus and to provide the confidence rating for the indicated instrument, on a five-point scale from "not sure" to "very confident."

In indicating the instrument heard in each stimulus, participants made a selection from a pull-down menu listing all 15 instruments. We were concerned that an alphabetic ordering of instruments might encourage participants to select instruments earlier in the alphabet, so we randomized the order of instruments on the menu for each stimulus. After selecting an instrument, the participant then selected the confidence level using one of five radio buttons. Stimuli were presented just once with no opportunity for the participant to replay a stimulus. This was done in order to avoid any unbalanced training effect from multiple exposures to more difficult (or more blended) stimuli.

3. RESULTS

Before starting on any analysis, we established an inclusion criterion. Recall that in the main experiment there were 15 single-instrument timbre trials, separate from 105 trials with concurrent unison dyads. If participants had difficulty identifying a single instrument in isolation, then results for identifying an instrument in paired stimuli would be much less reliable. Accordingly, a priori, we decided to include data for only the participants who scored better than 50% instrument recognition for single-instrument stimuli across the main experiment (a chance level would be 6.7%). As a result, 24 of 27 participants satisfied the inclusion criterion.

First, the responses for the single-instrument stimuli were examined. A confusion matrix has been computed from all of the single-instrument stimuli, including the results of 23 trials from the final preliminary test plus the 15 single-instrument stimuli dispersed through the main experiment (38 trials X 24 participants = 912 responses). Since each participant tested wind instruments twice and non-wind instruments once in the preliminary test, the values in Table 1 have been normalized accordingly. The timbre on each row is the instrument identified by the participants. On each column is the instrument timbre that was actually played. Table 1 is organized according to instrument families: the first four rows and columns are woodwinds; the next four are brass; the following three are percussions; the last four are string instruments, where HC and HA are plucked, PF is struck and VC is bowed. For each row or column, the largest value is observed on diagonal entries, indicating that the played instrument was identified correctly more often than not. However, the diagonal entries range from 0.417 corresponding to FH to 1 corresponding to VC, reflecting different degrees of ease in identifying 15 instruments.

The greatest confusion can be seen between FH-TN and between OB-EH, which happen to be the most similar timbre pairs according to the timbre dissimilarity study by Chon (2013). The confusion patterns exhibit a strong tendency to confuse instruments within a family. These results are comparable to the report by Giordano (2005) that listeners often confuse instrument tones produced by similar physical structure.

Since musicians are more likely to be familiar with different instruments, we might anticipate that the accuracy of instrument identification is associated with self-declared musicianship status. Accordingly, we calculated the per-group average identification rate of 15 timbres in isolation. There is a significant performance difference according to the group, *F*(2,21)=10.672, *p*=.001. A Bonferroni post-hoc test reveals that the difference between professional/amateur versus non-musician was found to be significant (*p*=.001 and *p*=.020, respectively). There was no significant difference between self-declared professional and amateur musicians (*p*=.963). This result suggests that self-declared musicians in general performed better in the given task, which is in agreement with participants' post-experiment interviews.

Addressing the main experimental hypothesis, we next consider whether timbre saliency influences instrument identification for isolated tones. Using the timbre saliency values from Chon (2013), we calculated the correlation between saliency values and average percent correct identification rates across all participants. Contrary to the experimental hypothesis, no statistically significant relationship is evident, *r*(13)=.39, *p*=.1562 for timbres in isolation; *r*(13)=.27, *p*=.3372 for timbres presented with others.

Our main experimental hypothesis predicts that timbre saliency would contribute to instrument identification. Conceptually, we might suppose that a number of factors influence the ability of a listener to identify the presence of an instrument in a unison dyad. First, we might expect that the more salient an instrument, the more likely it would be identified. However, the ability to identify the more salient instrument might be influenced by the saliency of the other instrument in the pair. If the two instruments have quite similar saliency values, then the competition is apt to be high. Conversely, if there is a large difference in the saliency of the two instruments, one might predict that the more salient instrument will benefit, and therefore be easier to identify. That is, the *difference* in saliency might be expected to influence listener judgments.

In order to test our main hypothesis, we constructed an a priori formal model that includes these proposed effects and interactions. Specifically, our model includes four factors, which can be interpreted as predictor variables in a multiple regression analysis:

$$p(X \mid X+Y) = c_1 p(X) + c_2 S_X + c_3 S_{diff}$$

where $p(X|X+Y)$ is the probability of identifying instrument X when presented concurrently with instrument Y, $p(X)$ is the probability of identifying the target instrument X in isolation, S_X is the saliency value of the target instrument X, S_{diff} is the absolute difference in saliency with respect to the other instrument. S_X was obtained by adding 0.5 to the saliency values from Chon (2013), because the original values ranged from -0.5 to +0.5. By shifting all values to the [0, 1] range, S_X now has the same range with other variables in the equation above.

Using the ENTER method of multiple regression, the adjusted R-squared value was determined 0.322, with the regression model being statistically significant to the accounted variance, *F*(3,206)=34.067, *p*<.000001. However, a close examination revealed that *p*(*X*) was the only significant factor in predicting the correct identification of an instrument timbre in dyads, *beta*=.575, *p*<.000001. Neither timbre saliency values nor saliency differences had a significant impact.

One of the hypotheses was that listeners would be more confident of their identification with more salient timbres. In order to address this hypothesis, we calculated the correlation between the saliency values from Chon (2013) and the average confidence ratings across all participants. The result was a mild but statistically significant correlation, *r*(208)=.2405, *p*<.001, which is in agreement with our hypothesis.

4. DISCUSSION

In this paper, we examined the identification of an instrument sound in concurrent unison dyads. As a salient timbre is defined to be the one that captures listeners' attention easily and tends not to blend well with concurrent sounds (Chon & McAdams, 2012a/b), we can then logically expect that a salient timbre will be easily identified. In that sense, this experiment can be considered as a comparison with timbre saliency previously measured by Chon and McAdams (2012a).

In the earlier study, Chon and McAdams (2012a) used a tapping technique to a perceptually isochronous ABAB sequence where A and B were two different timbres with their pitch, loudness and effective durations equalized. The result was a one-dimensional timbre saliency scale where the distance between a pair of timbres would be proportional to the saliency difference between them. This scale exhibited a negative correlation with the perceived blend of concurrent unison dyads in a later study (Chon & McAdams 2012b).

As a highly salient timbre was reported to show little blend with a concurrent sound (Chon & McAdams 2012b), we hypothesized that a highly salient sound will also be easy to identify. Hence, the identification task can be considered as a measure of timbre saliency. Consequently, we expected result that would be confirmative of earlier measured saliency values by Chon (2013). Contrary to our expectation, the measured identification result failed to show a statistically significant correlation with the previous timbre saliency values or the saliency differences. The only significant factor in predicting the correct identification of a timbre in concurrent dyads turned out to be the correct identification score in isolation. In other words, a sound that was easy to identify in isolation was also easy to identify when presented with another concurrent sound, and vice versa. This may sound trivial, but to our knowledge this is the first study to notice this correlational relationship.

Our data were consistent with the other hypothesis, namely, that listeners would be more confident of their identification of a more salient timbre. However, if we grant that the measured identification score is another measure of saliency, the correlation of measured saliency (i.e., the average identification score) and the confidence rating becomes very strong, $r(208)=.9899$, $p<10^{-10}$. This much higher correlation seems to make sense, as this is a result of two dependent variables obtained from the same experiment, whereas the timbre saliency values were collected from another experiment. This difference in correlation magnitudes might suggest a context effect. It is not known yet how a measure of saliency in one context can be applied to another context. Nevertheless, in summary, regardless of which measure of saliency to use, there is likely a statistically significant relationship between saliency and ease of identification.

In post-experiment interviews many participants mentioned a common set of timbres that were easy to identify and a different set of timbres that were difficult to identify. Easily identifiable instruments included the VC (the only bowed string instrument that had some uncharacteristic nasal quality due to a high pitch on a cello), PF, HC and "the bells" (MA, VP, TB). Participants were unanimous in reporting the brass instruments as the most difficult to identify. FH-TN and OB-EH pairs were the most challenging to distinguish, even by musicians who were trained on those instruments.

In comparison with the experiment by Chon and McAdams (2012a) to measure saliency, our participants reported different instruments to be more easily identifiable. Most noticeable is the case of VC. VC, being the only bowed string instrument, was identified more easily than any other instrument in this experiment, whereas it was one of the least preferred ones in the tapping task. This discrepancy might come from the multidimensionality of timbre, in the sense that different contexts bring out different characters of timbre, which participants unconsciously latch on to.

The correct identification score in dyads also showed a significant and very strong correlation with the confidence ratings, which additionally showed a mild yet statistically significant correlation with the timbre saliency values. If we consider the identification scores as another measure of saliency, both significant correlations are in agreement with our hypothesis that listeners would be more confident of identifying more salient timbres.

5. REFERENCES

Chon, S.H. (2013). Timbre Saliency, The Attention-Capturing Quality of Timbre. Unpublished Ph.D. Dissertation. McGill University, Canada.

Chon, S.H. & McAdams, S. (2012a). Investigation of Timbre Saliency, the Attention-Capturing Quality of Timbre, *Journal of Acoustical Society of America*, *131*(4), 3433 – 3433.

Chon, S.H. & McAdams, S. (2012b). Exploring Blending as a Function of Timbre Saliency, In *Proceedings of the 12th International Conference of Music Perception and Cognition (ICMPC)*, Thessaloniki, Greece (pp. 222 – 223).

Giordano, B.L. (2005). Sound source perception in impact sounds. Unpublished Ph.D. Dissertation. University of Padova, Italy.

Kendall, R., & Carterette, E. (1993). Identification and blend of timbres as a basis for orchestration. *Contemporary Music Review*, *9*, 51–67.

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. (2012). Perception of dyads of impulsive and sustained sounds. *Music Perception*, *30*(2), 117–128.

Vienna Symphonic Library GmbH. (2011). Vienna symphonic library. Available from http://vsl.co.at.