A STATISTICAL ANALYSIS OF TIMBRE DESCRIPTORS FOR MUSICAL INSTRUMENT CLASSIFICATION

Song Hui Chon, Stephen McAdams

Center for Interdisciplinary Research in Music Media and Technology (CIRMMT)
Schulich School of Music, McGill University

BACKGROUND AND AIMS

Classification of musical instrument sounds is an important issue related to other major problems such as fundamental-frequency estimation, timbre recognition and classification, and automatic transcription. The main goal is to maximize the correct classification of the instrument sounds within the given criteria using a set of descriptors with information on pitch, timbre and other aspects. Timbre is what enables us to distinguish sounds from different instruments. Ever since Helmholtz wrote about timbre perception in the latter half of the 19th century, there have been many attempts to define *timbre spaces* that model the human perception of timbre using Multidimensional Scaling (MDS). However, currently there is no unified model that explains human timbre perception across *all* instrument families. Rioux, et al. (2006) collected more than 70 timbre-related descriptors that are used in music research. This paper aims at a performance evaluation of those descriptors in three applications of musical instrument classification: instrument family classification (string, wind or percussion), excitation type classification (impulsive or continuant), and reed type classification (airjet, single reed, double reed, lip reed) in the wind instrument family.

METHOD

A set of 82 stimuli was selected from several well-known timbre perception studies. They were considered for instrument family classification and excitation type classification. For the reed classification in wind instrument family, 42 wind stimuli were considered. The feature values for statistical analyses were obtained by computing each timbre descriptor value on each sound stimulus. Therefore the input data for analysis was in the form of a matrix, where columns correspond to descriptors and rows to stimuli. K-means clustering and correlation analysis were used.

RESULTS

Correlation analyses reveal 14 descriptors to represent 9 feature groups specified by Rioux, et al. (2006). Some groups have two representatives with high intra-subgroup correlations and low inter-subgroup correlations. For the classification problems, K-means clustering was applied to the data for each of the descriptors. The clustering output from 10,000 loops in Matlab program was averaged. The performance ranged from 67% (with an attack-time descriptor) to 36% (with a descriptor from the RMS power and energy group) for instrument family classification, and 79% (with an attack-time descriptor) to 50% (with a spectral spread descriptor and with a RMS power descriptor) for excitation type classification. The instrument family classification result also seems to suggest that the wind family on average is easier to correctly classify than the others. This result supports the conjecture that some instruments may be easier to classify consistently than others. However, we will need more data to verify this conjecture. For the reed type classification, the best mean performance of 54% was achieved by a Spectral Deviation feature and the worst mean performance of 33% by a Spectral Flux feature. The descriptors in Spectral Deviation group seem to be good candidates for the reed

classification task, because the group mean is higher than other groups' means and the intra-group variation is small. The airjet reed type seems to be harder to classify than other types, although we would need more stimuli to confirm this conclusion.

CONCLUSIONS

This paper evaluated the performance of 70 timbre-related descriptors for a few classification problems. Correlation analysis was applied first to figure out the relationships among the groups of descriptors, which revealed 14 descriptors to represent the nine groups. The relationships among them will be useful in various timbre-related tasks. K-means clustering analysis was used for the actual classifications of musical instrument family, excitation types, and the reed types within the wind instrument family. The mean performance ranged from 33 to 79 percent. For both the excitation type and the instrument family classifications, the descriptors in the Spectral Centroid group seemed to yield a consistently good performance, while those in the Attack Time group showed varying degrees of performance, making it less reliable even though the group mean was higher than that for the Spectral Centroid group. It is interesting to notice that the Attack Time group showed the most diverse performance in both classifications, including the best mean performance rates, while the Spectral Centroid group was more or less consistent. The performances of the other groups fall between that of these two groups. This may be related to why many timbre studies have found that the two most important perceptual correlates of timbre to be spectral centroid and attack time.

The reed type classification was the most difficult problem, but a simple K-means clustering yielded a moderate performance of 33-50%. Again, the Spectral Centroid group was one of the most consistent and Attack Time the most varying. However, for this particular classification task, Spectral Deviation seems to be the best measure, because its intra-group variation was very small and the group mean was the highest. The airjet type showed a poorer classification performance, making us wonder if they are harder to classify than others, although more data are required before making this conclusion.

This paper used simple K-means clustering using only one descriptor at a time. It may be worth looking into a combination of different descriptors for some popular classification tasks and gauge the performance. Z-score values instead of raw descriptor values would be useful in this case, so that we can ensure an equal contribution from each descriptor. After an efficient combination of descriptors is found, the next step will be an investigation of efficient algorithms, possibly using machine learning concepts.

REFERENCE

Rioux, V., McAdams, S., Susini, P., Peeters, G. (2006). Psycho-acoustic Timbre Descriptors. (http://recherche.ircam.fr/projects/cuidado/wg/management/deliverables/IRCAM_WP215_M2.pdf)

TOPIC AREAS

Acoustics and psychoacoustics
Musical timbre