Classifying crowd behavior at collegiate basketball games using acoustic data FREE

Brooks A. Butler; Katrina Pedersen; Mylan R. Cook; Spencer G. Wadsworth; Eric Todd; Dallen Stark; Kent L. Gee; Mark K. Transtrum; Sean Warnick

Proc. Mtgs. Acoust. 35, 055006 (2018)
https://doi.org/10.1121/2.0001061

POMA Proceedings of Meetings on Acoustics

Volume 35 http://acousticalsociety.org/

176th Meeting of Acoustical Society of America
2018 Acoustics Week in Canada
Victoria, Canada
5-9 Nov 2018

Signal Processing in Acoustics: Paper 1pSP11

Classifying crowd behavior at collegiate basketball games using acoustic data

Brooks A. Butler
Brigham Young University, Provo, UT 84602; brooks.butler93@gmail.com

Katrina Pedersen and Mylan R. Cook
Department of Physics and Astronomy, Brigham Young University, Provo, UT, 84602; katrina.pedersen@gmail.com; mylan.cook@gmail.com

Spencer G. Wadsworth, Eric Todd and Dallen Stark
Brigham Young University, Provo, UT, 84602; spencergwadsworth@gmail.com; eric.w.todd@gmail.com; drs1108@gmail.com

Kent L. Gee and Mark K. Transtrum
Department of Physics and Astronomy, Brigham Young University, Provo, UT, 84602; kentgee@byu.edu; mkt24@byu.edu

Sean Warnick
Brigham Young University, Provo, UT, 84602; sean.warnick@gmail.com

The relationship between crowd noise and crowd behavioral dynamics is a relatively unexplored field of research. Signal processing and machine learning (ML) may be useful in classifying and predicting crowd emotional state. This paper describes using both supervised and unsupervised ML methods to automatically differentiate between different types of crowd noise. Features used include A-weighted spectral levels, low-level audio signal parameters, and Mel-frequency cepstral coefficients. K-means clustering is used for the unsupervised approach with spectral levels, and six distinct clusters are found; four of these clusters correspond to different amounts of crowd involvement, while two correspond to different amounts of band or public announcement system noise. Random forests are used for the supervised approach, wherein validation and testing accuracies are found to be similar. These investigations are useful for differentiating between types of crowd noise, which is necessary for future work in automatically determining and classifying crowd emotional state.

Published by the Acoustical Society of America

1. INTRODUCTION

Machine learning has become a common data analysis tool across scientific fields. Using machine learning with audio data has proven to be particularly useful in domains such as automatic speech recognition,[1-5] language processing,[6,7] music classification,[8-11] and event detection for urban, sports, and television environments.[12-14]

We are interested in using machine learning and acoustic data to probe crowd dynamics, specifically crowd sentiment. The ability to passively classify crowd emotions using sound has many potential applications, such as alerting peace-keeping officials of potentially unstable crowd environments as well as providing insights for social psychologists studying crowd dynamics. Identifying metrics of crowd sentiment would also be useful to the entertainment industry. The professional sports and live events industries may find real-time analysis of crowd emotions useful for event optimization and organization, contract negotiations, and performance evaluations.

While there is general interest in crowd psychology,[15,16] little is known about modeling and extracting live crowd sentiment. There are, however, examples of analysis of crowd media through text-based sentiment analysis,[17-19] as well as modeling for predicting crowd movement.[20] These studies attempt to model crowd behavior in various environments, such as on Twitter or during a physical emergency, yet have not attempted to identify the emotional state of a live crowd.

In this paper, we explore the possibility of using acoustics as a real-time probe into the emotional state of a crowd. We collect acoustic data of crowd responses from several sporting events. The data are processed to extract a variety of spectral and low-level features which are taken as inputs into a machine-learning analysis. In addition, crowd events are manually labeled based on the audience response (e.g., cheer, boo, etc.) along with the event that triggered the response.

We analyze the data using both supervised (random forest) and unsupervised learning (k-means clustering) methods. We find that supervised classification performs well on averaged crowd signals. The unsupervised algorithm successfully distinguishes between crowd noises and other acoustic signals such as the PA system. Our unsupervised approach further partitions crowd response based on the level of engagement. Our results demonstrate a proof-of-concept for using acoustic data with machine-learning tools to probe the emotional state of a crowd.

2. METHODS

A. DATA COLLECTION

The goal of this study is to relate the acoustic signals of crowds to their emotional state. We use spectators at sporting events as model crowds. Audio recordings were collected at several sporting events at Brigham Young University, including basketball (men's and women's), volleyball (men's and women's), soccer (women's), and football (men's). Recordings totaled over 160 hours of audio from 54 different games; however, for this initial study, results are based on nine basketball games, totaling over 18 hours recorded data. While sampling rates vary between recordings as part of an exploration of optimal collection methodology, all games were recorded at a minimum sampling rate of 25 kHz with a 24-bit system and a Type-1, 12.7 mm diameter free-field microphone.

At sporting events there are many concurrent acoustic sources, including the crowd itself, as well as announcements/advertisements, events on the court, music from the band, etc. We use microphones placed in strategic locations to focus on the crowd noise while minimizing the signal from individual voices. An example of a typical setup from a basketball game is shown in Figure 1, where the microphone was mounted adjacent to the shot clock on the backboard and the data acquisition system was located at the base of the standard.

Using the raw audio data, crowd responses are manually labeled as "cheer", "applause", "distraction noise", "positive chant", "negative chant", "singing" and "silence." A description for each crowd response is included in Table 1. Multiple labels can be applied to the same event. For example, cheer is often accompanied by applause. (Importantly, the converse may not always be true as cheering usually reflects a stronger positive response than applause alone.) For some games, video recordings are also available, and are used to label the game event to which the crowd was responding (e.g., points were scored).

Table 1. Descriptions of crowd responses.

Crowd response	Description
Cheer	Loud, positive crowd vocalization.
Applause	Crowd clapping that can include crowd vocalization.
Distraction noise	Attempts by crowd to draw an opposing team member's attention away from the game, most commonly when the opposing team possesses the ball or is about to shoot a free throw.
Positive chant	Rhythmic crowd shouting, usually directed towards the home team, e.g. "Defense" or "B-Y-U Cougars."
Negative chant	Crowd shouting in anger or distress, usually directed towards referees after a less than ideal call or towards a player from the opposite team.
Singing	Harmonic crowd vocalization accompanied by the pep band or the PA system.
Silence	Little or no crowd vocalization

The nine basketball games used in this study are divided into two categories of men's or women's games. The overall level of the acoustic signal at the men's games is higher than that at women's games, due to the difference in crowd size. In addition to crowd size, microphone placement impacts sound levels. Independent of sound level, we empirically observe that the spectral shapes of individual events (such as cheering, booing, etc.) are similar. We correct for the difference in amplitude by applying a standard scaling to each individual game.

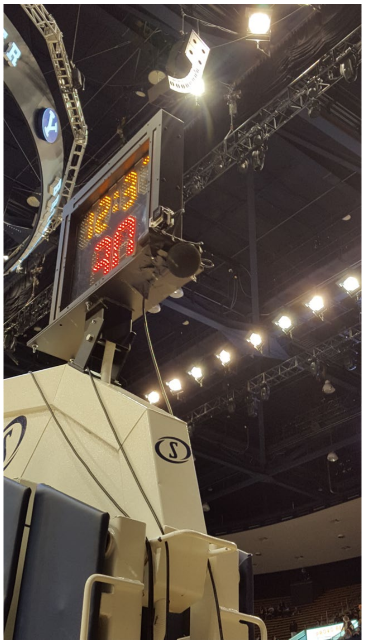

Figure 1. The typical microphone setup used for basketball games.

B. FEATURE EXTRACTION

Machine learning algorithms operate on a "feature vector", i.e., a list of inputs that characterize the signal to be classified. Because audio features require calculation over a given length of time, the pressure waveform data must be split into discrete blocks and sub-blocks of signal and features extracted for each block. Features can be classified into two main categories: spectral features that rely on fractional octave and cepstral analysis, and so-called "low-level" features that have been adopted from audio processing.

We choose 500 ms as the primary time interval for calculating features for machine learning, though we have also started investigating other time intervals. Our rationale is that a half-second interval will provide enough temporal resolution to show rapid changes in crowd behavior while still capturing the overall energy of different crowd reactions.

An Introduction to Audio Content Analysis, by Alexander Lerch,[21] is used as a reference for extracting audio features. In addition to using this literature, we modify MATLAB code from a public repository,[22] also maintained by Lerch, to extract a variety of so-called "low-level" signal parameters and Mel Frequency Cepstral Coefficients[23] (MFCCs).

The processed features are sorted into three categories: (1) spectral features, (2) low-level signal parameters, and (3) MFCCs. Multiple frequency weightings are possible including A-weighting, D-weighting (because of its particular emphasis of high-frequency energy), and flat weighting; however, in this initial analysis only A-weighted features are used.

I. SPECTRAL FEATURES

A discrete Fourier transform (DFT) is used over each half-second interval to calculate the A-weighted level across one-third and one-twelfth octave bands. For this analysis, one-twelfth resolution is primarily used.

A spectrogram for a complete game is shown in Figure 2. Even at this scale, major game events are visible, such as time-outs, half time, etc., that are marked by major changes in crowd involvement. For example, the period during which the national anthem was sung (marked 1) contains very low sound levels across all frequencies compared to periods where the speaker systems are playing music with high levels of low-frequency content (marked 2).

Each spectrum contains 93 one-twelfth octave frequency bands ranging from 50 Hz to 10 kHz. In addition to the spectral energy, we track the change in spectral levels for each half-second time interval and refer to this finite difference as the "delta spectrum", a type of first-derivative estimate. Similarly, the change in the delta spectrum is calculated to create an "acceleration spectrum." The spectral frequency bands, delta spectrum, and acceleration spectrum together constitute a 279-dimensional feature vector.

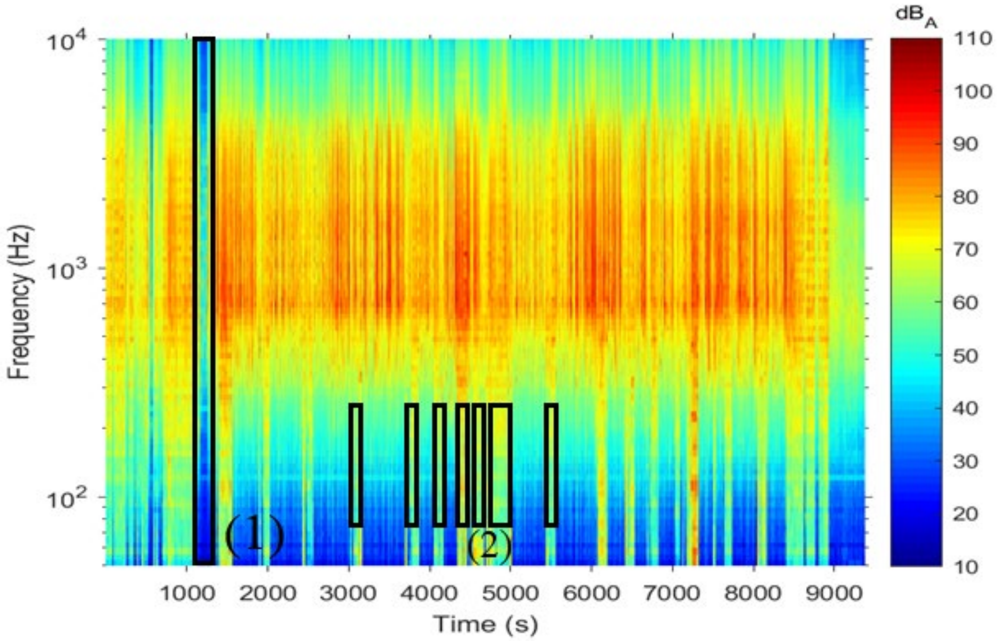

Figure 2. One-twelfth octave spectrogram of a men's BYU basketball game with color values in units of A-weighted decibels (dBA). Certain events are visually distinguishable such as (1) the silence during the national anthem and (2) the strong low-frequency content of the PA system during time-outs.

Table 2. Descriptions of low-level features/signal parameters.

Low-level Feature	Description	Low-level Feature	Description
Centroid	The frequency value of the center of mass of the spectrum.	Tonal power ratio	The ratio of tonal sound power to overall sound power.
Crest factor	A measure of tonality, which compares the maximum of the magnitude spectrum with the sum of the magnitude spectrum.	ACF coefficient	The AutoCorrelation Function (ACF) coefficient for a given length of signal.
Decrease	The steepness of the fall-off of the spectral envelope over frequency.	Max ACF	The absolute value of the overall ACF maximum, which is a simple estimate of the signal's tonality.
Flatness	The ratio of geometric mean and arithmetic mean of the magnitude spectrum.	Predictivity ratio	A measure of how well the audio signal can be predicted by O-order linear prediction.
Flux	The root-mean-square of the change in spectral shape over a given interval.	Root-mean-square	The root-mean-square value of a given block of signal.
Kurtosis	The fourth central moment divided by the fourth power of the standard deviation.	Standard deviation	The standard deviation value of a given block of signal.
Roll off	The frequency bin below which the accumulated magnitudes of the frequency bins reach a certain percentage of the overall sum of magnitudes.	Zero crossing rate	The number of changes of sign in consecutive blocks of audio samples.
Skewness	The third central moment divided by the cube of its standard deviation.	Peak	The absolute maximum per block of audio samples.
Slope	A measure of the linear change of the spectral shape.	Peak program meter	Using a so-called Peak Program Meter (PPM), the peak envelope is extracted on a sample-per-sample basis.
Spread	The concentration of the power spectrum around the spectral centroid.	Pitch chroma	A histogram-like 12-dimensional vector with each dimension representing one pitch class (C, C sharp, D, ..., B).

II. LOW-LEVEL SIGNAL PARAMETERS

In addition to spectral features, we use acoustic signal processing methods to calculate 20 low-level signal parameters. This set of features measures various statistical metrics across each half-second interval including the root-mean-square (RMS) level, spectral centroid, zero-crossing rate (ZCR), and others. For a full list of low-level features see Table 2.

For the calculation of low-level features, each 500 ms time block is subdivided into 20ms sub-blocks. Individual low-level features are calculated across each sub-block of audio. A power spectrum is then calculated for each feature, across the 25 consecutive, non-overlapping sub-block values, resulting in the overall analysis frame of 500 ms. A similar method was used by McKinney et al. to process features used in music classification.[24]

The power spectrum for each low-level feature over an analysis frame characterizes the amount of temporal oscillation in that frame, with the mean energy contained in the 0 Hz bin and subsequent bins characterizing feature modulation energy. For the 13-point single-sided spectrum and the 20 low-level features, this results in a 260-dimensional low-level feature vector.

III. MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCC)

The final type of features we use in this study are the MFCCs, which are calculated for each half-second interval. These coefficients are calculated by taking the discrete cosine transform of the log of the Mel-scaled power spectrum.[25] MFCCs are commonly used in audio analysis of speech and language processing because of their ability to extract important characteristics from recordings of the human voice.[26] This is optimal for speech recognition applications which attempt to recognize vocal formants and vowels independent of time waveform variation between speakers.

Typically, 26 coefficients can be extracted from a standard Mel-scaled spectrum—also containing 26 bins—but since the most relevant information for speech is contained in low-frequency energies,[27] only the first 13 coefficients are used. Future work may explore the use of the MFCC's corresponding to higher frequencies.

C. FEATURE SCALING

To account for numerical range differences between features, standard scaling is applied to each feature for individual games. This process performs an affine transformation such that each feature has a mean of zero and a standard deviation of one for every game. For simplicity, this is the only scaling method we use for our data. However, other scaling methods may prove useful in future research.

An interesting question is whether features should be scaled with respect to the entire length of recorded audio or with respect to individual game recordings. Considering the differences between complete and individual scaling, we choose individual scaling to help compensate for the absolute sound level differences between games of differing attendance and different microphone placement. This individual scaling is used to help reduce biases based on crowd size and make common crowd responses—regardless of size—more easily distinguishable.

D. MACHINE LEARNING METHODS

I. K-MEANS CLUSTERING

Clustering is the process of grouping sets of objects such that objects in one group are more similar to each other, by some numerical metric of similarity, than they are to objects in other groups.[28,29] Clustering is considered a form of unsupervised machine learning.

K-means clustering is commonly used in various applications of data science to explore the underlying structure of data in both high- and low-dimensional spaces.[30] In the case of our acoustic data, each time interval of crowd noise is characterized by a large vector of features—several hundred features when including both spectral and low-level signal parameters as described above, effectively putting our crowd noise events into a high-dimensional space. By clustering the feature vectors extracted from 18 hours of game recordings, we can use clustering to find the natural groupings of acoustic events based on feature values.

One challenge of k-means clustering is the need to specify *a priori* the number of clusters. This poses a problem for high-dimensional data since determining the optimal number of clusters through data visualization is a difficult process. There are, however, numerical methods for validating the number of clusters used in a given analysis.

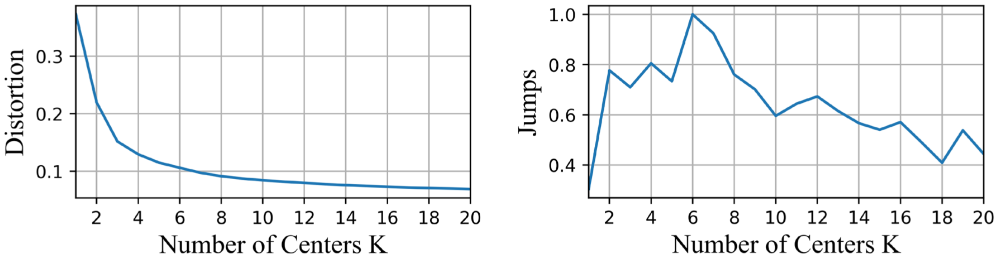

Figure 3. The distortion and jump values calculated for each number of cluster centers in a clustering analysis of crowd audio using the one-twelfth octave spectral frequency ranges as features.

A commonly used method for cluster determination is called an "elbow analysis".[31] This method utilizes the variance of the data—with respect to their closest cluster centroid—to measure the descriptive power of a given number of clusters on a dataset. This variance is also referred to as the "distortion" of a given clustering analysis. We can naively assume that the optimal number of clusters corresponds to the point where the variance of this function begins to level off.

An extension of this analysis, called a jump analysis, is described by Sugar et al.[32] This analysis uses a transformation of the curve used in an elbow analysis to calculate "jumps" between numbers of clusters, with the highest jump value indicating the optimal number of clusters. Using a jump analysis, as shown in Figure 3, we determine six clusters to be optimal when using only the A-weighted spectral levels as features.

II. RANDOM FOREST LEARNING

Supervised machine learning may be characterized as creating a mapping from an input space to an output space based on collected data. Input data are comprised of a set of features and output data are the corresponding labels. An algorithm is used on the data to learn the mapping during the training process. The mapping should be fit well enough so that correct labels can be accurately predicted for new input feature data. The random forest algorithm[33] can be used to learn the mapping of the crowd noise data to our crowd noise labels.

The random forest algorithm builds an ensemble of N decision trees during the training process. To preserve randomness and increase precision, N so-called "bootstrap samples" of size n are drawn from the initial dataset to generate each tree. Every node, or junction, of a tree is then assigned to split on the feature that gives the best split of m randomly selected features.

Following the training process, predictions can be made on new input data. Each of the three decision trees shown in Figure 4 is comprised of several nodes. At each node, some Boolean (yes or no) question is asked, and the resulting answer determines the path that is followed along each tree. The figure shows three of the N total trees that make up the forest. When using the forest to make predictions, we first find the final outcome or prediction of each individual tree. The most popular outcome from all N trees determines the predicted output of the forest for a given instance.

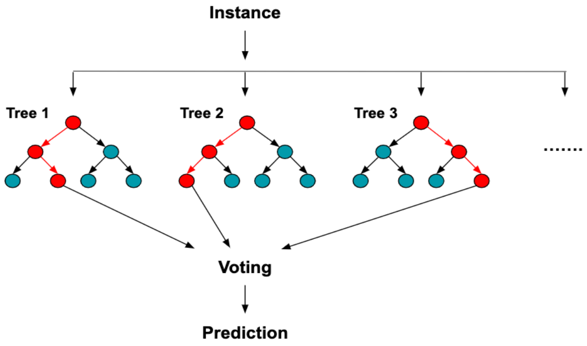

Figure 4. A graphic illustrating how a random forest algorithm makes predictions.

3. RESULTS

A. SUPERVISED RESULTS

Labeled data from the nine collegiate basketball games are used to generate training, validation, and testing data sets. The feature data for each labeled instance are generated by averaging over the duration of the crowd response. Our feature space is composed of all A-weighted features described in Section 2B, excluding pitch chroma and MFCCs. We resample the labeled data four times using a bootstrap method so that half the data are from one of four specific crowd response classes (cheer, applause, positive chant, or distraction noise) and the other half are random instances of all other response classes. This allows for binary classification of the four responses. These initial data sets are split into training, validation, and testing sets using a 60:20:20 split.

Weka machine learning software is used to implement the random forest algorithm described previously.[34] For each of the four responses we tune model hyper-parameters, such as the number of trees and number of features to sample from at each node, to minimize error on each validation set. For each crowd response the accuracies for the validation and testing data sets are shown in Table 3.

Table 3. Supervised Machine Learning accuracies (labels with fewer instances have been excluded).

Crowd Response	**Validation Accuracy**	**Testing Accuracy**
Cheer	65.6	63.6
Distraction Noise	79.2	82.4
Positive Chant	74.2	65.6
Applause	69.9	71.2

B. UNSUPERVISED RESULTS

Using the number of clusters identified by the jump analysis (six), we perform a k-means clustering analysis on our data, wherein each 500 ms audio sample is assigned to a specific cluster. We can represent each cluster with a different color; this allows us to create a video for further analysis.

We create 500 ms video frames by graphing the spectral levels of the audio sample in the color of the assigned cluster. These frames are then combined with the audio data to allow us to hear the audio while seeing to which cluster each audio segment has been assigned. For example, during the National Anthem the spectral levels are mostly shown in green, and when the crowd starts cheering the spectral levels change to red.

The compiled videos allow us to identify cluster labels and descriptions for each colored cluster. The cluster labels are: null, murmur, moderate involvement, high involvement, softer music, and louder music. A summary of these results, with both cluster labels and descriptions, can be seen in Table 4.

Another possible visual representation can be given by examining the centroids from the clustering analysis. Since only spectral features were used for this clustering approach, each cluster centroid can be represented by a one-twelfth octave spectrum between 50 Hz and 10 kHz. The spectral levels for the six cluster centroids can be seen in Figure 5.

The spectra in Figure 5 have two distinctly spectral shapes, with varying energy levels. One shape, which has a large amount of low-frequency content, corresponds to the non-crowd categories identified through the video clips (softer music and louder music, and the PA system). The other shape—which exhibits lower amounts of low-frequency energy, but peaks between 500 Hz and 2 kHz—corresponds to the crowd categories identified in the visual interpretation of the clusters: Null, Murmur, Moderate Involvement, and High Involvement. The results show that the relative amount of crowd involvement in the categories corresponds to the level increase across most frequency bands. For example, the sound level increases consistently from the null cluster, to murmur, to moderate involvement, and to high involvement, as expected.

Table 4. Human interpretation of cluster labels for the spectral clustering analysis.

Cluster Color	**Cluster Label** **Events Present in Cluster**
Green	Null Background noise, low crowd involvement, National Anthem, silence
Blue	Murmur Crowd neither loud nor quiet
Red	Moderate Involvement Applause, cheering
Magenta	High Involvement Loud Cheering, screaming, booing
Cyan	Softer Music Quieter music (band/PA), minimal crowd noise
Yellow	Louder Music Louder music (band/PA), minimal crowd noise

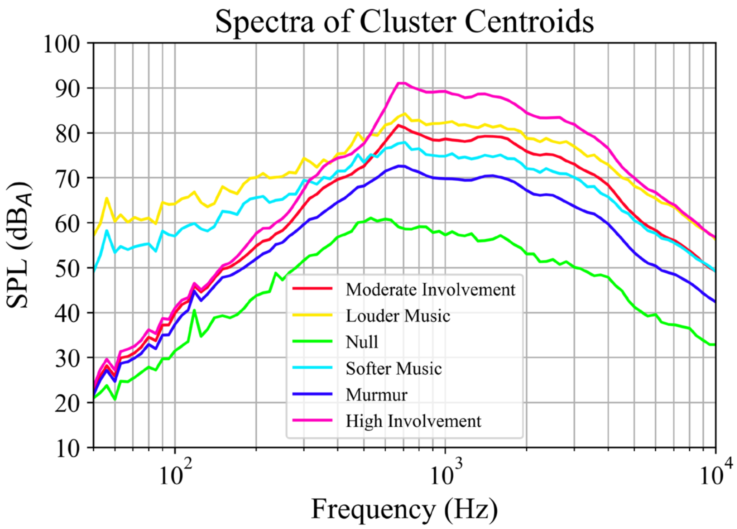

Figure 5. A representation of spectral centroids for each cluster with a human-assigned label for each cluster color. Note the difference in spectral slope for various levels of crowd involvement and high low-frequency energy present in music clusters.

4. CONCLUSION

This paper describes an initial approach to classifying crowd responses at sporting events using audio features. Our results demonstrate a proof of concept for identifying crowd reactions from acoustic data. This is an important first step toward the ultimate goal of extracting crowd sentiment in real time from audio. Future work includes further exploration into supervised and unsupervised classification of crowd noise, as well as feature development and reduction for real-time audio features. Future analysis may include crowd noise at multiple sports games, such as volleyball, soccer, and football. Our approach could also be extended to include other types of crowds such as at political rallies or parades.

ACKNOWLEDGMENTS

This work was supported by a CHIRP grant by the BYU College of Physical and Mathematical Sciences.

REFERENCES

[1]Levinson, S. E., L. R. Rabiner, and M. M. Sondhi. "An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition." *Bell System Technical Journal* 62, no. 4 (1983): 1035-074. doi:10.1002/j.1538-7305.1983.tb03114.x.

[2]Miao, Yajie, Mohammad Gowayyed, and Florian Metze. "EESEN: End-to-end Speech Recognition Using Deep RNN Models and WFST-based Decoding." *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, 2015. doi:10.1109/asru.2015.7404790.

[3]Hinton, Geoffrey, Li Deng, Dong Yu, George Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Brian Kingsbury, and Tara Sainath. “Deep Neural Networks for Acoustic Modeling in Speech Recognition.” *IEEE Signal Processing Magazine* 29 (November 2012): 82-97.

[4]Yu, Dong, and Li Deng. *Automatic Speech Recognition: A Deep Learning Approach*. London: Springer, 2016.

[5]Kim, Samuel, Shrikanth Narayanan, and Shiva Sundaran. “Acoustic Topic Model for Audio Information Retrieval.” *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (October 2009).

[6]Jurafsky, Dan, and James H. Martin. *Speech and Language Processing*. Upper Saddle River, NJ: Prentice Hall, Pearson Education International, 2014.

[7]Yu, Dong, and Li Deng. "Deep Learning and Its Applications to Signal and Information Processing." IEEE Signal Processing Magazine 28, no. 1 (2011): 145-54. doi:10.1109/msp.2010.939038.

[8]Basili, Roberto, Alfredo Serafini, and Armando Stellato. “Classification of musical genre: a machine learning approach.” In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)* (October 2004).

[9]Mandel, Michael I., Graham E. Poliner, and Daniel P. W. Ellis. "Support Vector Machine Active Learning for Music Retrieval." *Multimedia Systems* 12, no. 1 (2006): 3-13. doi:10.1007/s00530-006-0032-2.

[10]Panagakis, Ioannis, Emmanouil Benetos, and Constantine Kotropoulos. “Music Genre Classification: A Multilinear Approach.” In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR),* Philadelphia, Sep. 2008.

[11]Tzanetakis, George, and Perry Cook, "Musical Genre Classification of Audio Signals." *Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293-302, 2002. DOI: 10.1109/TSA.2002.800560.

[12]Salamon, Justin, and Juan Pablo Bello. "Unsupervised Feature Learning for Urban Sound Classification." *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015. doi:10.1109/icassp.2015.7177954.

[13]Kumar, Anurag, Pranay Dighe, Rita Singh, Sourish Chaudhuri, and Bhiksha Raj. "Audio Event Detection from Acoustic Unit Occurrence Patterns." *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012. doi:10.1109/icassp.2012.6287923.

[14]Radhakrishan, R., Ziyou Xiong, A. Divakaxan, and Y. Ishikawa. "Generation of Sports Highlights Using a Combination of Supervised & Unsupervised Learning in Audio Domain." *Fourth International Conference on Information, Communications and Signal Processing*. 2003. doi:10.1109/icics.2003.1292595.

[15]Bon, Gustave Le. *The Crowd: A Study of the Popular Mind*. Singapore: Origami Books Pte., 2018.

[16]Stephen Reicher, Clifford Stott, Patrick Cronin, Otto Adang, "An integrated approach to crowd psychology and public order policing", *Policing: An International Journal of Police Strategies & Management*, Vol. 27 Issue: 4, pp.558-572, https://doi.org/10.1108/13639510410566271

[17]Montejo-Ráez, A., M.c. Díaz-Galiano, F. Martínez-Santiago, and L.a. Ureña-López. "Crowd Explicit Sentiment Analysis." *Knowledge-Based Systems* 69 (2014): 134-39. doi:10.1016/j.knosys.2014.05.007.

[18]Haselmayer, Martin, and Marcelo Jenny. "Sentiment Analysis of Political Communication: Combining a Dictionary Approach with Crowdcoding." *Quality & Quantit*y 51, no. 6 (2016): 2623-646. doi:10.1007/s11135-016-0412-4.

[19]Gratch, Jonathan, Gale Lucas, Nikolaos Malandrakis, Evan Szablowski, Eli Fessler, and Jeffrey Nichols. "GOAALLL!: Using Sentiment in the World Cup to Explore Theories of Emotion." *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2015. doi:10.1109/acii.2015.7344681.

[20]Sime, J.D., “Crowd Psychology and Engineering.” *Safety Science* 21, no. 1 (November 1995): 1-14. https://doi.org/10.1016/0925-7535(96)81011-3

[21] Lerch, Alexander. An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics. Hoboken, NJ: IEEE Press, 2012.

[22] Lerch, Alexander. "Alexanderlerch/ACA-Code." GitHub. March 24, 2018. Accessed June 19, 2019. https://github.com/alexanderlerch/ACA-Code.

[23] Logan, Beth. "Mel Frequency Cepstral Coefficients for Music Modeling." In Proceedings of the International Symposium on Music Information Retrieval (ISMIR), Plymouth, Oct. 2000.

[24]McKinney, Martin F., and Jeroen Breebart, "Features for Audio and Music Classification." In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR)*, Baltimore, Oct. 2003.

[25]Logan, Beth. "Mel Frequency Cepstral Coefficients for Music Modeling." In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, Plymouth, Oct. 2000.

[26]Hasan, Rashidul. “Speaker Identification Using Mel Frequency Ceptral Coefficients.” In *Proceedings of the 3rd International Conference on Electrical & Computer Engineering (ICECE)*, Dhaka, Bangladesh, Dec. 2004.

[27]Tzanetakis, George, and Perry Cook, "Musical Genre Classification of Audio Signals." *Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293-302, 2002. DOI: 10.1109/TSA.2002.800560.

[28]Madhulatha, T. Soni. “An Overview on Clustering Methods.” *IOSR Journal of Engineering*. Vol. 2 no. 4, pp.719-725, April 2012.

[29]Xu, Rui, and Donald Wunch. “Survey of Clustering Algorithms.” *IEEE Transactions om Neural Networks*, Vol. 16, no. 3, May 2005.

[30] Kanumgo, Tapas, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu. “An Efficient k-Means Clustering Algorithm: Analysis and Implementation.” *IEEE Transactions on Pattern Analysis and Machine Intelligence.* Vol 24, no. 7, July 2002.

[31]Kodinariya, Trupti M., and Prashant R. Makwana. “Review on Determining Number of Cluster in K-Means Clustering.” *International Journal of Advance Research in Computer Science and Management Studies.* Vol. 1, no. 6, November 2013.

[32] Sugar, Catherine A., and Gareth M. James. “Finding the Number of Clusters in a Dataset: An Information-Theoretic Approach.” The Journal of the American Statistical Association. Vol. 98 Issue: 463, pp.750-763, https://doi.org/10.1198/016214503000000666.

[33] Breiman, Leo. “Random Forests.” Machine Learning. Vol. 45 no. 1, pp.5-32, 2001.