FEBRUARY 05 2024

Data-driven decomposition of crowd noise from indoor sporting events

Mitchell C. Cutler ; Mylan R. Cook; Mark K. Transtrum; Kent L. Gee

J. Acoust. Soc. Am. 155, 962–970 (2024)
https://doi.org/10.1121/10.0024724

JASA ARTICLE

Data-driven decomposition of crowd noise from indoor sporting events

Mitchell C. Cutler,[a] Mylan R. Cook, Mark K. Transtrum, and Kent L. Gee
Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602, USA

ABSTRACT:
Separating crowd responses from raw acoustic signals at sporting events is challenging because recordings contain complex combinations of acoustic sources, including crowd noise, music, individual voices, and public address (PA) systems. This paper presents a data-driven decomposition of recordings of 30 collegiate sporting events. The decomposition uses machine-learning methods to find three principal spectral shapes that separate various acoustic sources. First, the distributions of recorded one-half-second equivalent continuous sound levels from men's and women's basketball and volleyball games are analyzed with regard to crowd size and venue. Using 24 one-third-octave bands between 50 Hz and 10 kHz, spectrograms from each type of game are then analyzed. Based on principal component analysis, 87.5% of the spectral variation in the signals can be represented with three principal components, regardless of sport, venue, or crowd composition. Using the resulting three-dimensional component coefficient representation, a Gaussian mixture model clustering analysis finds nine different clusters. These clusters separate audibly distinct signals and represent various combinations of acoustic sources, including crowd noise, music, individual voices, and the PA system. © *2024 Acoustical Society of America*. https://doi.org/10.1121/10.0024724

(Received 2 June 2023; revised 3 January 2024; accepted 12 January 2024; published online 5 February 2024)
[Editor: Con Doolan] Pages: 962–970

I. INTRODUCTION

At collegiate sporting events, there is a complex mixture of sound sources. Among these are music and announcements over the public address (PA) system, music from live bands, noise from individuals, and various responses from the crowd to events on the court. In this paper, a data-driven method to decompose the spectral shapes present in the games is shown to aid in separating the mixture of sound sources into crowd noise and other acoustic sources. This decomposition may facilitate automated interpretation of crowd noise in a variety of applications, including in public gatherings where emotions may escalate quickly,[1,2] as well as in social psychology.[3] The entertainment industry may also find it useful for identifying potentially profitable venues and events, as well as event optimization and contract negotiation.[4]

Related analyses of crowd noise have been done to determine the presence of crowds,[5] the level of crowd involvement,[4] the general sentiment of a crowd,[6,24] and the effect of crowd noise on others.[7,8] Extensive work has been done particularly in identifying key moments where crowd involvement is indicative of notable events during sporting events.[9] Additionally, studies have predicted the noise emissions from crowds at events based on the directivity of the crowd, the synchronization of the source, the number of people in the crowd, and the voice effort of individuals in the crowd.[10,11] Many of these studies use machine learning, including neural networks and clustering algorithms, to identify periods of interest and use combinations of spectral and low-level acoustic features to make these predictions.

Feature selection is required to analyze these acoustic signals to determine which acoustic features are relevant. In machine learning, a feature vector is a combination of measured values used as inputs to create a model, which can then predict outcomes for other measured feature vectors.[12] For example, some have found that acoustic features such as Mel-frequency cepstral coefficients, spectral flatness measure, short-time energy, and zero-cross-rate can be used to identify the presence of crowds as well as their elevated involvement due to observable events.[5,9] Others have used image-based analysis of 1-s overlapping spectrograms covering the full frequency range of human hearing to predict approval, disapproval, and neutral crowd noise.[6]

Previous work by Butler *et al.*[4,13] used spectral shape analysis (1/12-octave bands over non-overlapping half-second intervals), k-means clustering, and elbow and jump analyses to identify six different clusters in acoustic data from men's basketball games.[14] Four of these clusters contained significant levels of crowd noise and had similar spectral shapes. The other two clusters were correlated with music. Todd *et al.*[13] also used hierarchical clustering of spectra and low-level features to identify focused crowd involvement in parade noise.

This study expands on previous work by describing a low-dimensional decomposition of 24 1/3-octave band signals at 30 men's and women's basketball and volleyball games. This decomposition identifies patterns that distinguish crowd noise from other noise sources and could be used to interpret noise produced by the crowd. The acoustic

[a]Email: mitchellccutler@gmail.com

https://doi.org/10.1121/10.0024724

TABLE I. Length of audio and number of games for each sport type.

Sport	Hours	Games
MBB[a]	17.2	10
WBB[b]	12.0	8
MVB[c]	12.5	7
WVB[d]	6.0	5

[a]Men's basketball (MBB).
[b]Women's basketball (WBB).
[c]Men's volleyball (MVB).
[d]Women's volleyball (WVB).

signal decomposition characterizes them by taking linear combinations of three principal spectral shapes, which are found using principal component analysis (PCA).[15] Using only these three principal spectral shapes, one can create approximations of all acoustic spectra at sporting events with a median 1.9 dB error for any 1/3-octave band. The spectral shape coefficients define a three-dimensional space wherein natural clusters are formed by regions of closely spaced points. Clustering using Gaussian mixture models (GMMs)[15,16] reveals nine different clusters with distinct combinations of acoustic sources.

II. DATA PROCESSING

Recordings were taken at Brigham Young University (BYU) basketball and volleyball games, as described by Butler *et al.*[4] Recordings used include audio from game start to game end, totaling more than 47 h of audio data. A breakdown of the distribution of data is given in Table I.

Several types of feature vectors could be created using the acoustic data, depending on the processing method used, such as using different combinations of spectral and temporal resolution. Various combinations were investigated. For this paper, 1/3-octave bands[7] with a temporal resolution of one-half second[4,13] have been used because it was found that higher spectral resolutions do not contain additional useful information, and lower temporal resolutions are not fine enough to capture changes in crowd responses. For this study, the frequency bands from 50 to 10 000 Hz were used since most of the noise produced by the crowd was expected to be in this interval. The resulting spectral feature vectors are unweighted half-second sound pressure levels [dB(Z) re 20 μPa] at 1/3-octave center band frequencies.

This study focuses on the use of spectral data, but it should be noted that acoustic spectra are influenced by several factors that are not considered here. Some of these factors include the venue reverberation time, the effect of microphone location, voice and PA speaker directivities, the PA speaker dynamic range, and the crowd composition. Studies investigating these and other factors may yield useful insights revealing better ways of processing the data.

Spectrograms of the first hour from four of the games (one from each subdivision in Table I) are shown in Fig. 1 and show SPL(f) using half-second time intervals. The recordings are generally loudest in the 630 Hz–1 kHz 1/3-octave bands, similar to what Barnard *et al.* [7] found for crowds at football games. Only part of the recorded sound was produced by the crowd. Other noise sources include the PA system, band, individuals, and various other non-crowd noise sources. These other sources add to the crowd noise above 630 Hz and are often the primary source of noise below 630 Hz, especially at and below 100 Hz.

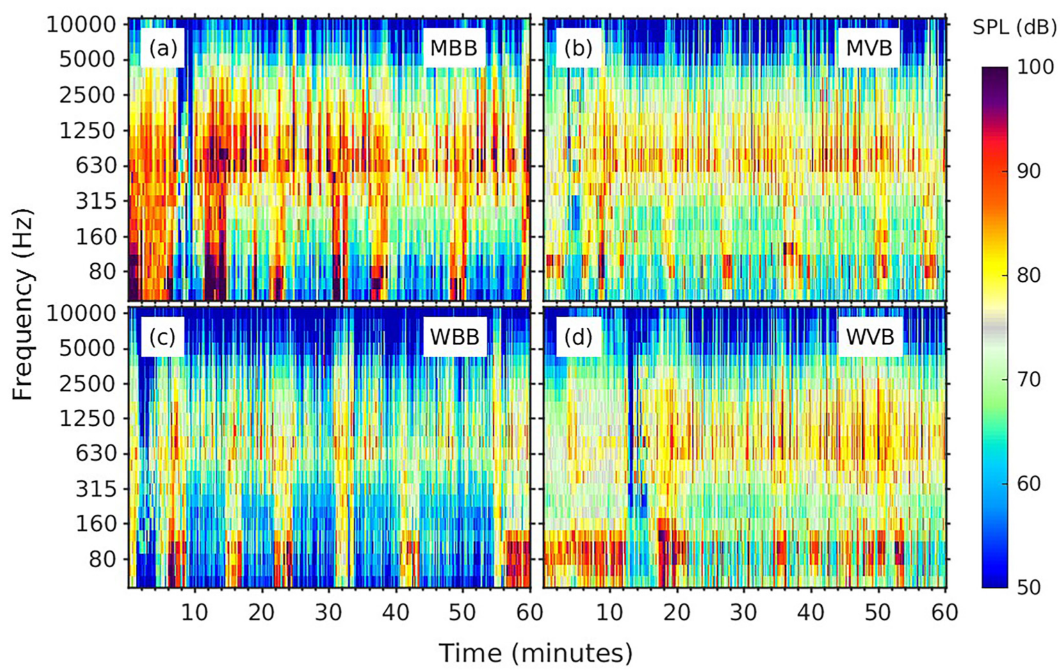

FIG. 1. (Color online) Spectrograms of the first hour of characteristic men's and women's basketball and volleyball games. The 1/3-octave half-second sound pressure level [SPL(f)] is shown by color.

Two other things that were considered when deciding how to create feature vectors were the physical venues where the games were played and the attendance at the games. The basketball games were played at the BYU Marriott Center, which is an indoor stadium seating up to 18 987 people,[17] while the volleyball games were played at BYU's Smith Fieldhouse, which has much smaller dimensions and seats up to 5637 people.[18] The venue reported that men's basketball games had audiences ranging from 10 179 to 16 456 people, while the women's basketball games had audiences between 672 and 1341 people. Both men's and women's volleyball games had comparable numbers of attendees, ranging from 2109 to 3921 people, except for one women's game against a rival team with 5472 attendees.[19]

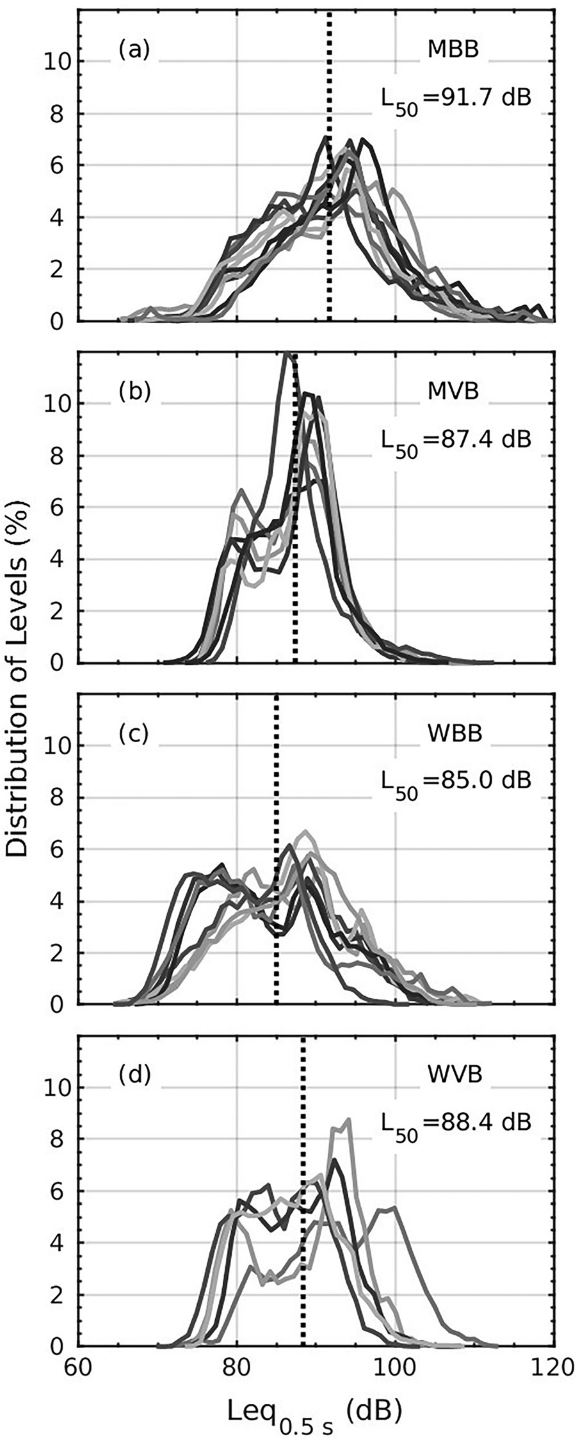

FIG. 2. The distribution of $Leq_{0.5s}$ for each game is shown in each subplot, separated by sport type. The L_{50} exceedance level is marked with a vertical dashed line. The data are grouped into bins with a 1 dB width.

III. ONE-HALF-SECOND EQUIVALENT SOUND LEVELS

Before beginning the analysis of the acoustic spectra, the distribution of unweighted one-half-second equivalent continuous sound levels, $Leq_{0.5s}$ [dB(Z)], is examined. The $Leq_{0.5s}$ is calculated as the sum of energy across all frequencies in a half-second interval. The distributions of $Leq_{0.5s}$ for each game and sport are shown in Fig. 2.

Within each sport, games generally have similar distributions of $Leq_{0.5s}$, especially in men's volleyball [see Fig. 2(b)]. Some of the differences in the distributions of games from a single sport can be explained simply. For example, one women's volleyball game was much louder than the others because it was against a rival team and had many attendees.

The differences in $Leq_{0.5s}$ distributions between the different sports can be explained by the nature of the sporting event and crowd. Figures 2 and 3 show that the $Leq_{0.5s}$ distribution varies by sport and is correlated with attendance levels. Both men's and women's volleyball games had audiences ranging from 2109 to 3921 people (except for the women's volleyball rivalry game) and have similar L_{50} levels. The L_{50} for men's and women's basketball can also be explained partially by attendance levels since average attendance at men's games was over 13 times greater than at women's games. It is important to note that at smaller events, the crowd was situated near the court and the microphone,[4] so sound level differences across events are more compressed than would be expected by merely accounting for crowd size.

Because the L_{50} exceedance level varied from game to game, methods to normalize the sound pressure levels by game were considered. However, normalizing sound pressure levels relative to each game's L_{50} exceedance level had

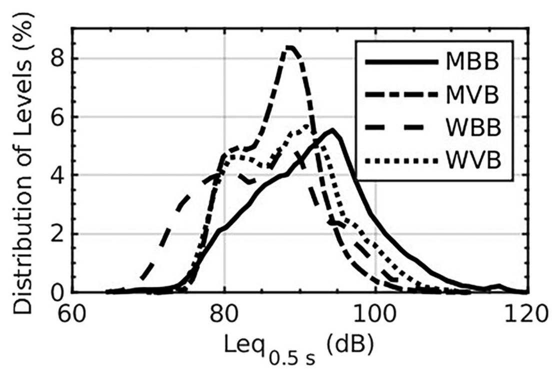

FIG. 3. The distribution of the $Leq_{0.5s}$ by sport type. The data are grouped into bins with a 1 dB width.

little effect on further analyses, so no normalization was ultimately implemented for the results in this paper.

Despite what information can be learned or verified by the distribution of $Leq_{0.5s}$, there is not enough information in them to infer crowd responses. For example, a half-second interval with a high $Leq_{0.5s}$ could be caused by a loudly cheering crowd, but it could also be caused by a crowd making distraction noise just before an opposing team's free throw. It could even be caused by an acoustic source other than the crowd, such as the PA system or an individual shouting near the microphone. Spectral data are used to further characterize acoustic events and try to separate these acoustic phenomena.

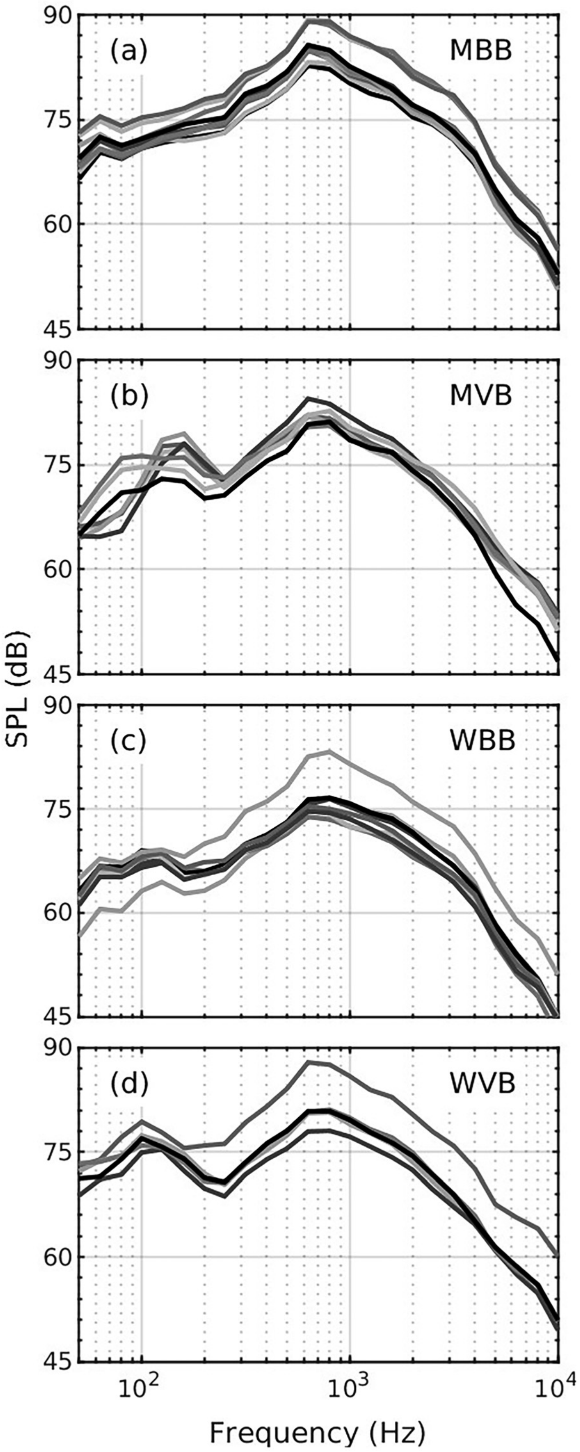

FIG. 4. The mean 1/3-octave spectrum for each game, separated by sport type.

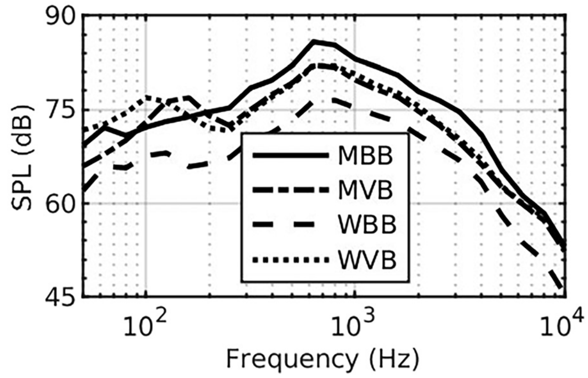

FIG. 5. The mean spectrum across all games in each sport type.

IV. AVERAGE SPECTRA

While typically either median or energetic mean values are used to characterize average spectra, PCA finds variations about arithmetic mean values. Therefore, the average spectra reported herein are obtained using the arithmetic mean. The mean spectra are analyzed by game and by sport (as with the $Leq_{0.5s}$ distributions). This analysis reveals that the mean spectra have the same general characteristics (see Figs. 4 and 5). All have peaks in the 630 Hz band and have steep roll-offs starting at the 800 Hz band. Additionally, the volleyball games have significant secondary peaks between 100 and 160 Hz. These general characteristics are also true for spectral data processed by any combination of quarter-second intervals or 1/12-octave bands, the only difference being that data processed in 1/12-octave bands produce less smooth spectra.

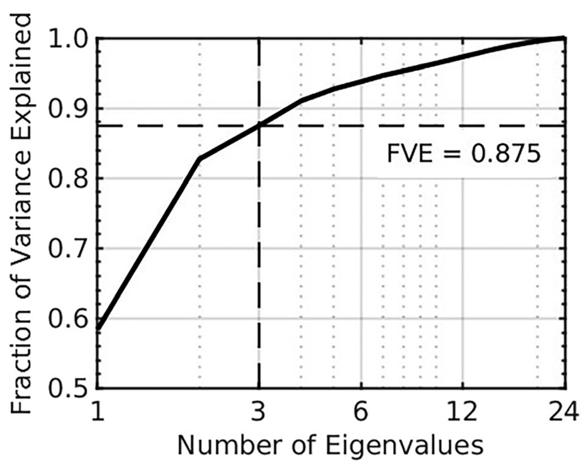

FIG. 6. The cumulative fraction of variance explained (FVE) for a given number of PCs. Larger cardinality eigenvalues indicate that there was less variance in the data in the directions specified by the corresponding PC vector.

V. PRINCIPAL SPECTRAL SHAPES

To decompose the acoustic spectra into the principal spectral shapes that make up the data, a PCA is performed. The PCA[15] is performed on all spectral data together, using each half-second as a 24-dimensional feature vector. The first three principal components (PCs) of the acoustic spectra explain 87.5% of all variance in the data, as seen in Fig. 6. Linear combinations of these first three PCs produce approximate spectra with a median error of 1.9 dB for any given 1/3-octave band center frequency. More PCs are not used because using one more PC increases the percentage of explained data by only about 4%; using more PCs also has the disadvantage of making the spectral shapes begin to correspond with tonal sounds (such as individuals, buzzers, and whistles) rather than the crowd noise. The effects of the first three PCs on the mean spectrum are shown in Fig. 7.

The first PC [see Fig. 7(a)] adjusts the level of the entire spectrum and is highly correlated with the $Leq_{0.5s}$. It does not, however, distinguish between samples that are loud due to the crowd and samples that are loud due to the announcer or music over the PA system.

The difference between spectra containing large amounts of high-frequency content and low-frequency content is characterized by the second PC. This vector changes sign between the 250 and 315 Hz bands [see Fig. 7(b)]. Frequencies higher than 315 Hz are increased by positive coefficients, while frequencies below 250 Hz are decreased by positive coefficients. This spectral shape is correlated with crowd noise, which is verified by listening to samples from the dataset with that characteristic. Multiplying the

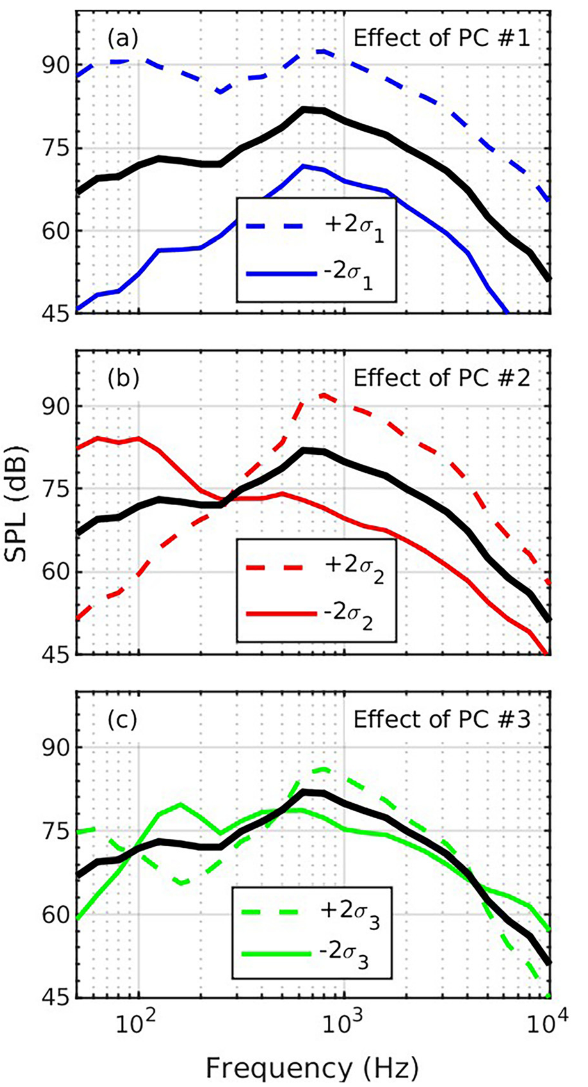

FIG. 7. (Color online) The first three PC vectors and their effect on the mean spectrum (black). The colored lines are the sums of the mean spectrum and one of the first three PCs multiplied by a positive or negative coefficient. The coefficient values used are two standard deviations from the mean coefficient value (zero for PCA). The first PC [Fig. 7(a)] corresponds to the $Leq_{0.5s}$, the second [Fig. 7(b)] corresponds to the sounds primarily produced by the crowd, and the third [Fig. 7(c)] corresponds to more peaked spectra.

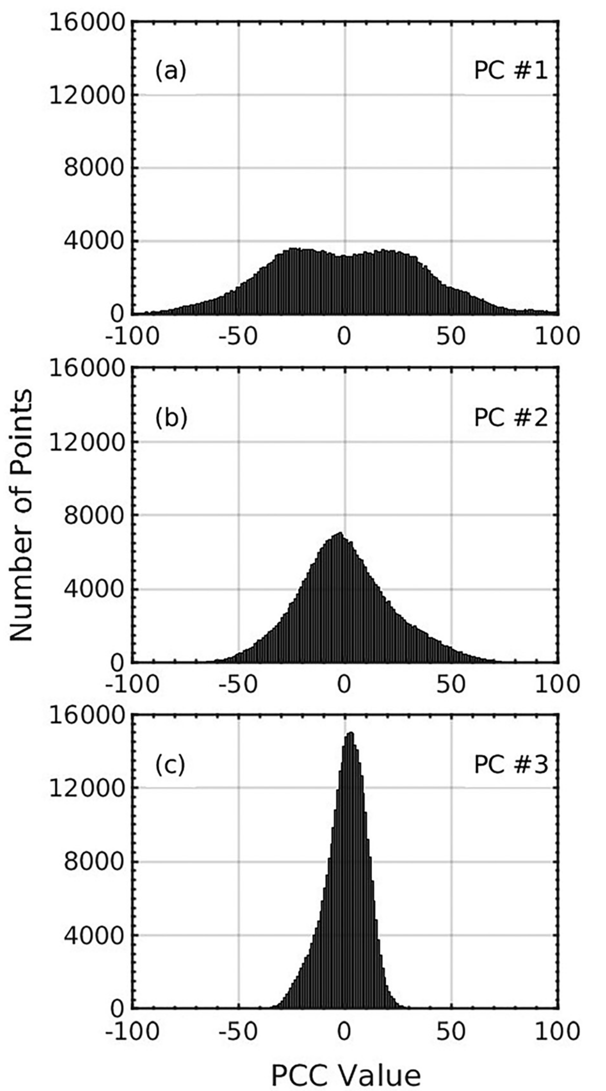

FIG. 8. Histograms of the PC coefficients, derived from a PCA of the 1/3-octave band spectral data from collegiate sporting events. Notice the different scales for each distribution, reflecting the relative scale of the variance in each PC. The first PC [Fig. 8(a)] has a bimodal distribution, the second [Fig. 8(b)] exhibits a large kurtosis (i.e., peakedness), and the third [Fig. 8(c)] is very skewed.

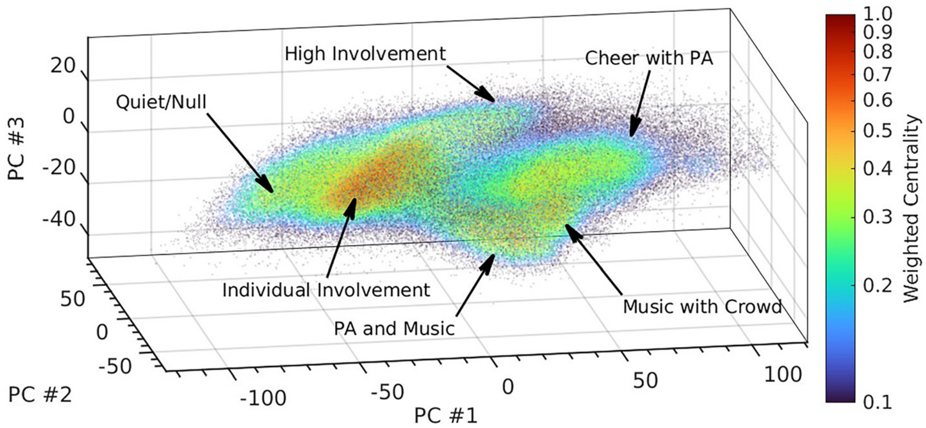

FIG. 9. (Color online) The weighted centrality measure of the PC coefficient data (Refs. 20 and 21). The weighted centrality is calculated from the nearest 0.01% of points to the point of interest and is a type of point density metric. The different parts of the point cloud are labeled by the types of sounds to which they correspond.

second PC vector by a negative coefficient gives the spectral shape corresponding to the PA system.

The effect of the third PC is most prominent in volleyball games. Negative coefficients correspond to samples with flatter spectra, while positive coefficients correspond to more peaked spectra [see Fig. 7(c)].

VI. PC COEFFICIENTS

In addition to the PC vectors, the PC coefficients themselves also reveal useful information. Histograms of the first three PC coefficients are shown in Fig. 8. Many of the statistical features of these histograms are related to different sources of noise. In Fig. 8(a), the first PC coefficients are bimodal. The right peak corresponds to an acoustic event, such as the crowd cheering, music, or the announcer over the PA system. The left peak corresponds to the ambient crowd noise. The second distribution [Fig. 8(b)] has a kurtosis of 3.11 and a positive skewness of 0.28. The third PC has a large negative skewness of -0.49, corresponding to the relatively few flatter spectra at the volleyball games. Most of the negative third PC coefficients come from when the PA system was active during the men's volleyball game.

When the first three PC coefficients are represented in a three-dimensional space, two main lobes are visible, as seen in Fig. 9. The logarithmically spaced color bar shows the weighted centrality[20,21] of each point, which is a type of point density. Different locations in the space represent different PC coefficients, which in turn represent different spectral shapes. The lobe with the labels "High Involvement," "Quiet/Null," and "Individual Involvement" generally corresponds to noise generated only by people, while the lobe with the labels "Cheer with PA," "PA and Music," and "Music with Crowd" corresponds to noise that is at least partially generated by non-crowd sources, such as the band or the PA system. Recordings from various parts of this point cloud, which correspond with the various sport types and clusters identified further on, are given in the supplementary material.

In addition to the two main lobes of points, the centrality colormap shows additional smaller clusters of points. A GMM (Ref. 15) is used to separate the data into clusters. GMMs assume data come from several Gaussian-distributed clusters and fit a Gaussian distribution to each of these clusters. Each point can have a non-zero probability of belonging to every cluster in the model. Unsupervised clustering of data is done by iteratively assigning points to the cluster to which they are most likely to belong and then updating the parameters of each cluster to best fit its assigned data. This was done using MATLAB's[22] GMM clustering routine. Because the initialization of the clusters can affect the final model, many random initial parameters should be tested, and the best-fitting model should be used.[23]

Determining the optimal number of clusters to fit to Gaussian distributions is done by performing an elbow analysis on the negative log-likelihood of the number of clusters. The likelihood is defined by the conditional probability of the data occurring given a predetermined set of parameters, namely, the means and covariances of the Gaussian distributions used to fit the data. The maximum likelihood is monotonically increasing in the number of parameters given that the optimal fit for a given number of clusters is found. This, however, is dependent on the cluster initializations. Taking the negative log of the likelihoods gives the characteristic elbow shape, as seen in Fig. 10. The number of clusters is

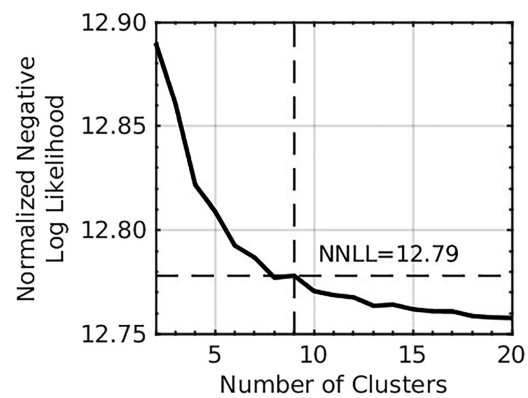

FIG. 10. Elbow analysis of the normalized negative log-likelihood (NNLL), used in determining the number of clusters in the GMM of the PC coefficients of spectral data from recordings of sporting events.

heuristically chosen at the elbow of the plot, where the negative log-likelihood is small but not overfit. In addition to doing the elbow analysis, one thousand different initializations for the cluster means and covariances show similar results, with nine clusters being optimal.

The results of clustering the data using the GMM are shown in Fig. 11(a). (See supplementary material for a gif of the clusters rotating in the PC space.) Each of the clusters is the result of assigning points to one of nine multivariate Gaussians to which it is most likely to belong. The clusters in Fig. 11(a) reflect the clusters seen in Fig. 9 with the weighted centrality measure. The empirically calculated means and covariances for each of the clusters in Fig. 11(a) are also similar to those given by the model parameters. The supplementary material also provides a video displaying the clusters in the three-dimensional PC space. Figures 11(b) and 11(c) show the spectral shape corresponding to the centroid of each of the clusters. The spectra corresponding to cluster centroids that do not contain significant amounts of crowd noise are shown in Fig. 11(b), while those in Fig. 11(c) contain crowd noise along with other noise sources. (See supplementary material for audio clips from each of the clusters.)

VII. CLUSTER INTERPRETATIONS

Brief interpretations and qualitative summaries of the acoustic sources in each cluster were given by listening to audio clips from each cluster. These interpretations and summaries are listed in Table II. Each column in Table II requires an explanation. Because the study's primary purpose is to determine how to separate synchronized crowd noise from other acoustic sources, the "Crowd involvement" column describes the perceived amount of synchronized involvement of the crowd,[11] as opposed to the perceived noise level of the crowd. This is especially relevant to the pink cluster, which had many instances of loud but asynchronous crowd noise during halftime. There were many ways that crowds at the sporting events showed synchronized involvement, and some were assigned to different amounts of crowd involvement. Clapping, chanting, booing, and singing are considered moderate or mid-level involvement. The noise made by the crowd to distract players on the opposing team right before free throws and serves, as well as ecstatic cheering, are considered high crowd involvement.

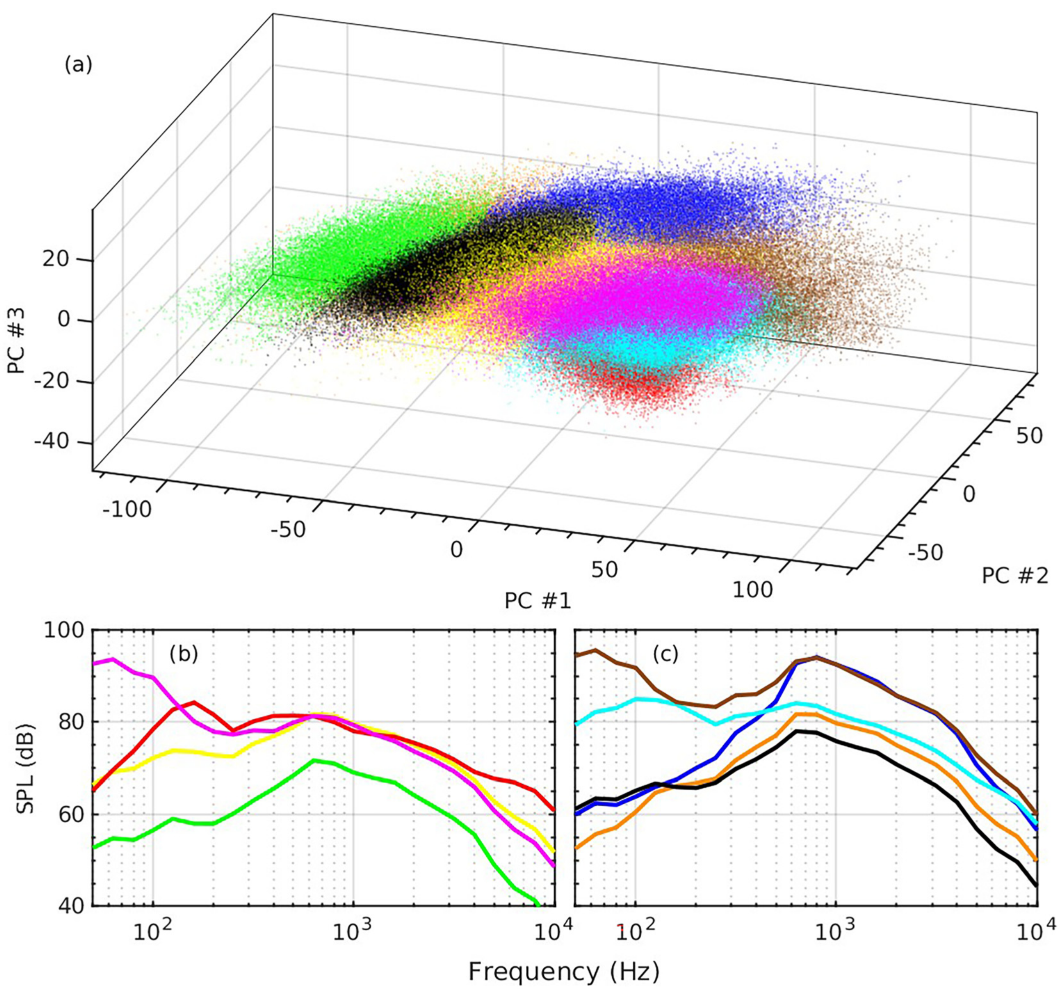

FIG. 11. (Color online) The results of the cluster analysis using the GMM. Points are assigned to the cluster to which they have the highest likelihood of belonging. The clusters can be seen in Fig. 11(a). Figure 11(b) shows the mean acoustic spectra for clusters of data with low amounts of synchronized crowd noise, while Fig. 11(c) shows the mean acoustic spectra for clusters with high amounts of synchronized crowd noise.

https://doi.org/10.1121/10.0024724

TABLE II. Summaries of the clusters and their primary acoustic sources.

		Other noise sources					
Cluster	Crowd involvement	PA announcer	PA music	Live music	Individual involvement	Description	PC space
Green	Low	None	None	None	Low	Minimal noise	(−,0,0)
Pink	Low	Low	High	Low	Low	Music	(+,−,+)
Yellow	Low	High	Mid	Low	Mid	PA/Individual noise	(0,0,0)
Red	Low	High	Mid	High	None	PA/Music	(0,0,−)
Black	Mid	None	None	None	Mid	Individual noise	(−,0,0)
Orange	Mid	None	None	None	Low	Moderate crowd noise	(−,+,0)
Cyan	Mid	High	High	High	Low	Music/Moderate crowd noise	(+,0,0)
Blue	High	Low	None	None	None	High crowd noise	(0,+,0)
Brown	High	None	High	None	None	Music/High crowd noise	(+,0,+)

Other acoustic sources present in the games are the announcer over the PA system ("PA announcer"), music over the PA system and from the band ("PA music" and "live music," respectively), and noise from individuals ("individual involvement"). This last source considered is primarily composed of individuals who could be heard and understood distinctly from the rest of the crowd.

The last column shows where the cluster lies approximately in PC space, with each "+," "−," and "0" referring to whether the PC coefficient value of the cluster centroid was significantly higher or lower, or about equal to zero. For example, the orange cluster has "(−,+,0)" because its cluster centroid had a large, negative first PC, a large, positive second PC coefficient, and a third PC near zero. Note that although the green and black clusters both have (−,0,0), the centroid of the green cluster is at (–57.6, –8.19, 3.78) in PC space, while the black cluster's centroid is at (–25.9, –2.52, 3.03).

Manual classification of audio clips from each cluster reveals that the clusters contain human-interpretable information, although they were identified solely on 1/3-octave, half-second spectral levels. While the included audio clips (see supplementary material) generally confirm this, it should be noted that trends reported in Table II become apparent only after listening to many samples from each of the clusters. Additionally, the half-second intervals used to create the data do not necessarily align with the natural transitions in the noise created by the crowd, the PA system, and other sources. The model is unable to account for transitions from one primary acoustic source to another that occur during a half-second interval. Using additional acoustic features in further analyses may mitigate this error and improve cluster interpretability.

VIII. CONCLUSION

This work on classifying crowd noise at sporting events developed a method for separating multiple acoustic sources in recordings. The first step taken was to analyze the $\mathrm{Leq}_{0.5\mathrm{s}}$ during the games, which led to the conclusion that differences in $\mathrm{Leq}_{0.5\mathrm{s}}$ across games were likely caused by differences in the venue and crowd size at the event.

The primary method for analyzing the spectra was PCA. The first three PC coefficients retained 87.5% of the variance while creating a three-dimensional coefficient space. This significantly simplified the data representation. The data were split into nine clusters representing different combinations of acoustic sources using a GMM, where each data point was assigned to the Gaussian cluster from which it was most likely to be. These sources included crowd noise, music, individual voices, and PA system noise. Acoustic signals were classified by their position in PC space, and this revealed patterns between noise source types and data points in the frequency domain.

The method of using a PCA on spectral data to determine principal spectral shapes can be used as a starting point for future research. For example, sporting events could be used as a model for more complex systems, such as riots, protests, other entertainment venues, public hearings, and other events, where the crowd may be in motion or signals may be more subtle. This data-driven decomposition of the acoustic spectra of crowd noise could also be used in supervised machine-learning models to identify crowd responses. The data-reduction portion of this study is especially relevant since it provides only the most relevant data to these machine-learning models.

SUPPLEMENTARY MATERIAL

See the supplementary material for audio clips from each of the clusters and a gif of the clusters rotating in the PC space.

ACKNOWLEDGMENTS

This work was supported by funding from the Brigham Young University College of Physical and Mathematical Science through the College High Impact Research Program. We also thank the editors for accommodating the tables, many figures, and supplementary material included in this paper.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

[1]N. Singh, N. Roy, and A. Gangopadhyay, "Analyzing the sentiment of crowd for improving the emergency response services," in *Proceedings of the 2018 IEEE International Conference on Smart Computing*, Taormina, Italy (June 18–20, 2018) (IEEE, New York, 2018), pp. 1–8.
[2]S. Reicher, C. Stott, P. Cronin, and O. Adang, "An integrated approach to crowd psychology and public order policing," Policing **27**, 558–572 (2004).
[3]G. Baranowski-Pinto, V. L. S. Profeta, M. Newson, H. Whitehouse, and D. Xygalatas, "Being in a crowd bonds people via physiological synchrony," Sci. Rep. **12**, 613 (2022).
[4]B. A. Butler, K. Pedersen, M. R. Cook, S. G. Wadsworth, E. Todd, D. Stark, K. Gee, M. Transtrum, and S. Warnick, "Classifying crowd behavior at collegiate basketball games using acoustic data," Proc. Mtgs. Acoust. **35**, 055006 (2018).
[5]V. R. Reddy, A. Sinha, and G. Seshadri, "Fusion of spectral and time domain features for crowd noise classification system," in *Proceedings of the 13th International Conference on Intelligent Systems Design and Applications*, Salangor, Malaysia (December 8–10, 2013) (IEEE, New York, 2013), pp. 1–6.
[6]V. Franzoni, G. Biondi, and A. Milani, "Emotional sounds of crowds: Spectrogram-based analysis using deep learning," Multimed. Tools Appl. **79**, 36063–36075 (2020).
[7]A. Barnard, S. Porter, J. Bostron, R. Termeulen, and S. Hambric, "Evaluation of crowd noise levels during college football games," Noise Control Eng. J. **59**, 667–680 (2011).
[8]K. Hummel, E. Ryherd, X. Cheng, and B. Lowndes, "Relating clustered noise data to hospital patient satisfaction," J. Acoust. Soc. Am. **154**, 1239–1247 (2023).
[9]M. Sano, H. Sumiyoshi, M. Shibata, and N. Yagi, "Generating metadata from acoustic and speech data in live broadcasting," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Philadelphia, PA (March 23, 2005) (IEEE, New York, 2005), pp. 1145–1148.
[10]M. J. Hayne, J. C. Taylor, R. H. Rumble, and D. J. Mee, "Prediction of noise from small to medium sized crowds," in *Proceedings of ACOUSTICS 2011*, Gold Coast, Australia (November 2–4, 2011) (Australian Acoustical Society, Darlinghurst, Australia, 2011), pp. 133–138.
[11]M. J. Hayne, R. H. Rumble, and D. J. Mee, "Prediction of crowd noise," in *Proceedings of ACOUSTICS 2006*, Christchurch, New Zealand (November 20–22, 2006) (Australian Acoustical Society, Darlinghurst, Australia, 2006), pp. 176–188.
[12]M. A. Hall and L. A. Smith, "Practical feature subset selection for machine learning," in *Proceedings of the 21st Australian Computer Science Conference*, Perth, Australia (February 4–6, 1998) (Springer, Berlin, 1998), pp. 181–191.
[13]E. Todd, M. R. Cook, K. Pedersen, D. S. Woolworth, B. A. Butler, X. Zhao, C. Liu, K. Gee, M. Transtrum, and S. Warnick, "Automatic detection of instances of focused crowd involvement at recreational events," Proc. Mtgs. Acoust. **39**, 040003 (2019).
[14]C. A. Sugar and G. M. James, "Finding the number of clusters in a dataset: And information-theoretic approach," J. Am. Stat. Assoc. **98**, 750–763 (2003).
[15]M. J. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M. A. Roch, S. Gannot, and C. Deledalle, "Machine learning in acoustics: Theory and applications," J. Acoust. Soc. Am. **146**, 3590–3628 (2019).
[16]M. S. Yang, C. Y. Lai, and C. Y. Lin, "A robust EM clustering algorithm for Gaussian mixture models," Pattern Recognit. **45**, 3950–3961 (2012).
[17]Information available at https://byucougars.com/story/m-basketball/1285668/ ncaa-ranks-marriott-center-no-6-toughest-home-court (Last viewed October 10, 2023).
[18]Information available at https://www.ncaa.com/news/volleyball-women/article/2020-07-24/womens-college-volleyball-9-largest-arenas (Last viewed October 10, 2023).
[19]Information available at https://byucougars.com/news/2018/09/13/no-1-byu-sweeps-no-24-utah-in-front-of-record-crowd (Last viewed October 10, 2023).
[20]L. C. Freeman, "A set of measures of centrality based on betweenness," Sociometry **40**, 35–41 (1977).
[21]S. Segarra and A. Ribeiro, "Stability and continuity of centrality measures in weighted graphs," IEEE Trans. Signal Process. **64**, 543–555 (2016).
[22]The Mathworks Inc., MATLAB, version 9.12.0 (MathWorks, Natick, MA, 2022).
[23]B. G. Lindsay, *Mixture Models: Theory, Geometry and Applications* (Institute of Mathematical Sciences, Hayward, CA, 1995), pp. 65–66.
[24]S. Proksch, M. Reeves, K. Gee, M. Transtrum, C. Kello, and R. Balasubramaniam, "Recurrence quantification analysis of crowd sound dynamics," Cognit. Sci. **47**, e13363 (2023).