Proceedings of Meetings on Acoustics

SEPTEMBER 08 2026

Acoustic characterization of football games in support of feature selection for estimating crowd engagement with machine learning

Jason D. Bickmore ; Kent L. Gee

Check for updates

Proc. Mtgs. Acoust. 56, 022003 (2025)
https://doi.org/10.1121/2.0002385

View Online
Export Citation

Connected Content

This is a companion to: Principal component analysis applied to feature selection at collegiate basketball and football games

Articles You May Be Interested In

Explainable machine learning models for classifying reactions within crowd noise during men's collegiate basketball games

J. Acoust. Soc. Am. (October 2025)

Open set classification strategies for long-term environmental field recordings for bird species recognition

J. Acoust. Soc. Am. (June 2022)

30 September 2026 17:36:01

ASA
ACOUSTICAL SOCIETY OF AMERICA

ASA
LEARN MORE
Advance your science and career as a member of the
Acoustical Society of America

Volume 56 http://acousticalsociety.org/

ASA/ICA 2025 New Orleans

188th Meeting of the Acoustical Society of America
joint with
25th International Congress on Acoustics

New Orleans, Louisiana
18-23 May 2025

Computational Acoustics: Paper 1aCA5

Acoustic characterization of football games in support of feature selection for estimating crowd engagement with machine learning

Jason D. Bickmore and Kent L. Gee
Department of Physics and Astronomy, Brigham Young University, Provo, UT, 84602, USA; jdb387@byu.edu; kentgee@byu.edu

30 September 2026 17:36:01

Machine-learning models that estimate crowd engagement at college football games could improve the home-field advantage, help stadiums serve fans better, and become a tool for other scientific studies. To prepare to create these models, college football games were acoustically characterized with feature identification and selection in mind. Data were collected from six Brigham Young University (BYU) football games and analyzed for patterns in level, spectra, and principal components. The probability distribution of sound pressure levels from an entire game was found to contain three distinct peaks that were found to be associated with (1) quiet during offense, (2) music being played over the stadium loudspeakers, and (3) distraction noise when BYU was on defense. Crowd-generated sounds were found to have the most energy in the 630 and 800 Hz bands, consistent with other crowd noise previously studied. Furthermore, a principal component analysis on the football games' one-third octave spectra revealed that the principal components were similar to those from BYU basketball and volleyball games. Based on these acoustic characteristics, the levels of the 100 Hz, 800 Hz, and 2.5 kHz bands are proposed as features for estimating crowd engagement at college football games.

Proceedings of Meetings on Acoustics, Vol. 56, 022003 (2026); https://doi.org/10.1121/2.0002385.

*©2026 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 4.0 International (**CC BY**) License (https://creativecommons.org/licenses/by/4.0/).*

1. MOTIVATION AND BACKGROUND

College football games attract millions of spectators each year. In 2024, total attendance at college football stadiums exceeded 39 million fans.[1] Each gathering includes tens of thousands of people or more, with the ten largest college football stadiums in the United States seating between 90,000 and 107,000 people.[2] Because a college football stadium brings so many people together at once, it provides a unique opportunity to study the noise of the crowd at sporting events.

One interesting aspect of the crowd's noise is its link to crowd engagement. Because making noise is the primary way the crowd participates in the game, this noise may be used to estimate their engagement. Estimating the crowd's engagement through noise has several useful applications. By measuring engagement and seeking to improve it, a stadium could build the fans' energy and give the home team an extra home-field advantage. It could be a tool for further scientific study, such as looking for a link between crowd engagement and physiological excitement or heart-rate synchronization. Ultimately, it could serve attendees by helping to make the fan experience better.

There has been a limited number of previous studies on crowd noise. Work from Brigham Young University has focused on using machine learning to identify crowd engagement and behavior at college basketball and volleyball games. Butler et al.[3] identified specific crowd reactions at basketball games with a random forest model trained on audio features, and found that k-means clustering of one-twelfth octave band spectra grouped data by degree of crowd involvement. Todd et al.[4] found that focused crowd effort at college basketball games is associated with high spectral flux and slope, then checked whether these features also highlight focused crowd effort during a Mardi Gras parade and found that the association was less clear. Proksch et al.[5] found that four quantities derived from recurrence quantification analysis (RQA) of noise at college basketball games could separate crowd behaviors into three groups with different degrees of coordination. They then found that 19 RQA metrics could be used to classify six kinds of crowd behaviors with accuracy between 50% and 96%. Cutler et al.[6] connected the first three principal components of one-third octave band spectra from college basketball and volleyball games to sound sources at the game and found that Gaussian mixture model clusters in this principal component space roughly separated noises by source and degree of crowd involvement. In 2025, Cutler et al.[7] used five feature selection methods to determine that half-second equivalent level, the scaling of the second principal component, and spectral characteristics like flatness were the most useful features for identifying crowd behaviors like cheering, distraction noise, and chanting at college basketball games with a random forest model. Other studies have investigated the acoustic character of on-field noise at college football games to determine how players are impacted. Barnard et al.[8] studied the impact of crowd noise at Penn State football games on speech intelligibility for the opposing team's offensive line. Navvab et al.[9] captured the sound of the crowd across the University of Michigan football stadium to model how reverberation time would change after new skyboxes were added. While previous work has focused on basketball and volleyball games or the impact of noise on team communication, this study examines the acoustic characteristics of college football games to support the development of models that estimate crowd engagement at a football game in real time.

To accomplish this goal, this study focuses on two aspects of noise at a football game that are relevant to identifying crowd involvement: sound sources and crowd behaviors. The sound sources considered are the announcer on the public address (PA) system (announcer), the pep band next to the student section (band), and music played on the PA system. The crowd behaviors considered are cheering, noise to disrupt the visiting team's communication (distraction noise), intentional quiet to help the home team during offense (quiet), and the singing of the fight song (singing). Additional crowd behaviors found in other studies like applause or positive chanting are not studied here,[3] but their acoustic characteristics could be studied if it is determined that they are critical to the overall goal of estimating crowd engagement.

Because a machine learning model is the final goal, this study is also a preliminary search for features, the acoustic details that distinguish crowd engagement. Most of the acoustic characterization is guided by this search. One advantage of doing acoustic characterization and feature selection simultaneously is that the features found are often easily interpretable because they are being studied in context.

The following sections first explain how data were collected. Then, data analysis is presented in three parts. First, the equivalent levels of individual events are identified and used to interpret the entire game. Second, spectra of sound sources and crowd behaviors are studied. Third, the principal components of the entire dataset are interpreted and compared to principal components from a dataset of basketball and volleyball games. Finally,

30 September 2026 17:36:01

findings are summarized, a list of features for estimating crowd engagement is suggested, and the study's implications and limitations are discussed.

2. DATA COLLECTION

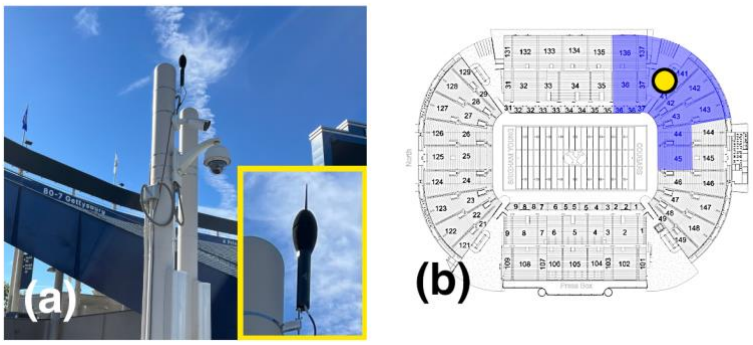


Figure 1: (a) Larson Davis 377B02 1/2" microphone with PRM2103 heated preamplifier and windscreen. (b) Microphone location (yellow) in center of student section (blue).

BYU's football venue, the LaVell Edwards stadium, ranks 40th largest by capacity among college football stadiums in the United States[10] at a maximum capacity of 62,073 spectators.[11] It has an open top, and is constructed mostly of concrete and metal. At its widest, it is about 240 meters long by 190 meters wide. The student "Roar of Cougars" (ROC) section,[12] is in the Southeast corner, as indicated in Figure 1(b).

The microphone used in this study was a Larson Davis 377B02 ½" microphone on a PRM2103 heated preamplifier. The microphone and preamplifier were covered by an outdoor windscreen and secured to a pole about 4 meters above the spectators (Figure 1(a)). This spot was approximately 195 meters away from the PA system. The location of the microphone remained the same throughout the season. It was strategically placed in the student section because students are known to participate most in crowd behaviors and produce the highest sound pressure levels,[8,9] providing plenty of data to study. Pressure was sampled at 51.2k samples per second and 16-bit depth by a Larson Davis 831C class-1 sound level meter with recording capability. The meter was strapped below the microphone inside a weather-resistant case. This microphone placement is different from other studies, which used microphones as close to the field as possible to estimate the level experienced by players.[8,9] Because of this, absolute levels will not be compared with those given in Barnard et al.

Data collection started when the game began and stopped when the game ended, including halftime, all timeouts, and other game stoppages. Noise was recorded at all six home football games during the 2024 football season. During the season, attendance varied between 59,213 and 64,420 people.[13] BYU won five of the six games.

Because a previous study determined that half-second blocks are long enough to capture all relevant spectral information while keeping up with speed at which crowd responses shift,[6] most analyses were done using half-second, nonoverlapping blocks.

3. ANALYSIS

The following section presents and discusses the results of several methods of data analysis. While the data shown make up a small part of the dataset, samples were chosen that represent all the data collected.

A. EQUIVALENT LEVEL OF COMMON EVENTS

I. UNWEIGHTED AND A-WEIGHTED EQUIVALENT LEVEL

First, the half-second equivalent level ($Leq_{0.5\,s}$) of events during the game was considered. Figure 2 shows the unweighted and A-weighted $Leq_{0.5\,s}$ during ten minutes of the third quarter of BYU vs. Arizona on October 12, 2024. This segment includes a BYU offensive drive, a timeout, and an Arizona offensive drive.

30 September 2026 17:36:01

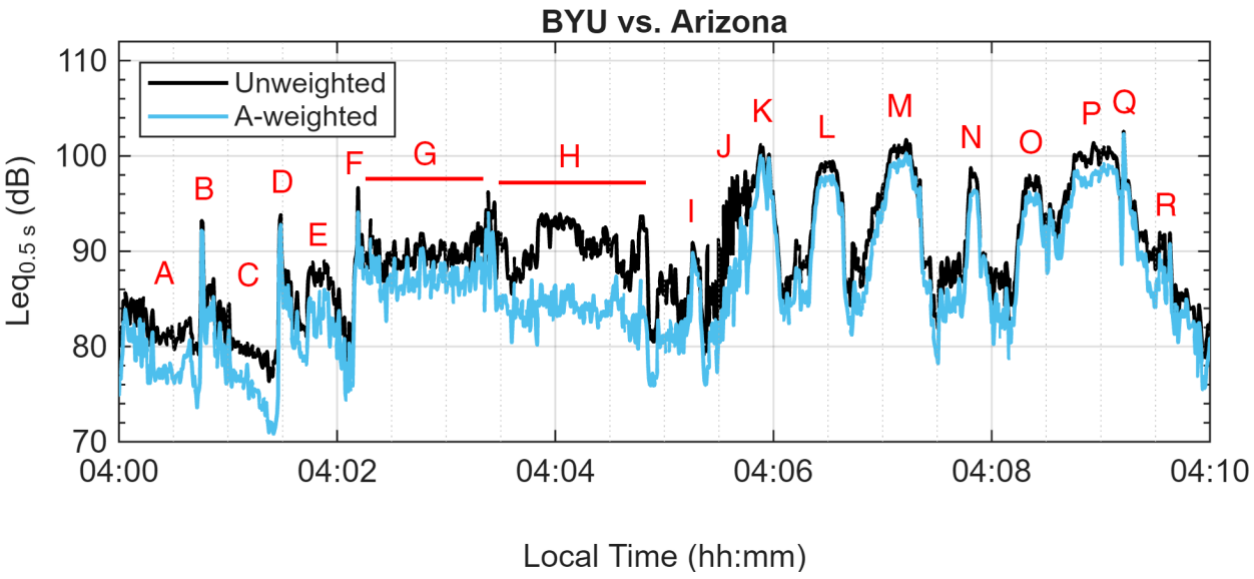


Figure 2: $Leq_{0.5\ s}$ during ten minutes of BYU vs. Arizona third quarter on October 12, 2024. Before BYU's field goal around 4:02 PM, the score was BYU 21, Arizona 7. Unweighted $Leq_{0.5\ s}$ is shown in black, A-weighted in blue.

Table 1: Labels of events in Figure 2.

Label	Description
A	BYU 2nd down
B	BYU incomplete pass
C	BYU 3rd down
D	BYU incomplete pass
E	BYU 4th down, music on PA system
F	BYU field goal, cannon fired
G	Fight song
H	Music on PA system during "selfie cam" activity
I	Arizona kick goes out of bounds
J	*We Will Rock You*, distraction noise grows
K	Arizona 1st down
L	Arizona 2nd down
M	Arizona 3rd down
N	Arizona 1st down
O	Arizona 2nd down
P	Arizona 3rd down
Q	Arizona incomplete pass
R	Arizona 4th down ends with BYU fair catch of Arizona punt

Each drive is visible in the plot. BYU has possession first from 4:00 to 4:02. During BYU offense, level alternates between quiet, around 80 dB, to loud, 94 to 98 dB. The quiet periods A, C, and E occur when the crowd is quiet to help the offensive line focus and communicate. Each loud spike (B, D, and F) is the crowd reacting when a play ends. The smaller spikes, B and D, correspond to incomplete passes. The larger spike, F, is the crowd cheering when BYU makes a field goal.

30 September 2026 17:36:01

A media timeout occurred between 4:02 and 4:05. The crowd sang the BYU fight song during G, then music was played on the PA system for a "selfie cam" activity during H.

From 4:05 to 4:10, Arizona had possession. Before each play, the crowd became very loud to interrupt communication between the offensive line and provoke a false start (K-P). This distraction noise tends to be 5–10 dB louder than cheering and lasts several times longer, up to 20 seconds (M). Once play begins, the crowd becomes quieter, but not as quiet as when the home team is on offense. The crowd reaches its highest level during Arizona's possession when third down ended with an incomplete pass (Q), getting up to 102 dB. This pattern of quiet when the home team is on offense and distraction noise when the home team is on defense was also identified in Barnard et al.[8]

While increases in the unweighted level reflect crowd activity during plays, increases at other times such as before plays and during timeouts are sometimes caused by other sources like the PA system or the band. To estimate crowd engagement, a metric that distinguishes the crowd from non-crowd noise sources would be more useful. One potential solution is A-weighting. According to Barnard et al.,[8] the crowd's peak frequency is about 800 Hz, and the PA system is loudest between 50 and 200 Hz. Assuming these characterizations are accurate, A-weighting should attenuate the PA system frequencies by 20 dB but reduce the level of crowd noise frequencies by only 3 dB. The amount by which A-weighting reduces sound pressure level might be one way to identify crowd-generated noise.

The blue line in Figure 2 shows the A-weighted $Leq_{0.5\,s}$. The A-weighted levels of quiet during offense are approximately 4 dB lower than unweighted levels (A, C, E). However, when the crowd cheers after each play, the A-weighted spikes are only 2 dB less than the unweighted spikes (B, D, F). The most dramatic difference between the A-weighted and unweighted levels is during the "selfie cam" activity (H). During this activity, which acoustically consists of mostly music on the PA system, the levels of each weighting are separated by as much as 8 dB. This period has very little focused crowd involvement. When the crowd is active, however, like when they sing the fight song (G), the difference between the unweighted and A-weighted levels is smaller, around 2–3 dB. A-weighting appears to reduce the level of non-crowd generated sounds more than crowd-generated sound.

Table 2: Mean unweighted and A-weighted $Leq_{0.5\,s}$ of events during the ten minutes shown in Figure 2. Values are given as mean ± standard deviation. Rows are sorted so levels that change the least after A-weighting show up on top.

Event	$Leq_{0.5\,s}$ (dB)	$LAeq_{0.5\,s}$ (dB)	Difference (dB)
Cheering after incomplete pass (Q)	97.8 ± 1.7	97.3 ± 1.8	0.5
Distraction noise (K-P)	97.5 ± 2.8	95.9 ± 2.9	1.6
Fight song (G)	89.8 ± 1.4	87.1 ± 1.5	2.6
Announcer on PA	84.2 ± 1.9	81.4 ± 2.2	2.8
Quiet (A, C, E)	80.2 ± 1.4	76.2 ± 2.2	4
Music on PA (H)	90.9 ± 2.2	84.1 ± 1.3	6.8

The impact of A-weighting on the $Leq_{0.5\,s}$ is summarized in Table 2. Rows are sorted by the difference between the mean unweighted and A-weighted levels, shown in the rightmost column, from smallest to largest difference. A-weighting appears to change the level of non-crowd-generated sounds the most. Crowd-dominated sounds, like cheering and distraction noise, change the least. Of the events considered, music on the PA system is impacted the most. This may be because most of the unweighted level comes from the music's bass.

The differences between the A-weighted and unweighted mean level of some events bring up further questions. Although music and the announcer both come through the PA system, the announcer's voice is affected by A-weighting 4 dB less (Table 2). This suggests that the announcer is more spectrally similar to the crowd than to music. One possible explanation is that both the announcer and crowd noise are based on the human voice, while music includes other frequencies created with instruments or synthesizers. It is also noteworthy that the second most impacted sound was quiet. If A-weighted $Leq_{0.5\,s}$ is to be considered as a possible feature for machine learning, understanding why quiet's level is reduced is important because quiet is a critical part of estimating crowd engagement. One additional observation about A-weighting is that it impacts

30 September 2026 17:36:01

different crowd behaviors unequally. The level of cheering after an incomplete pass is nearly unchanged, but distraction noise is reduced by 1.6 dB. This hints that crowd behaviors may have different frequency content. These questions are explored further in the spectra section.

II. USING DISTRIBUTION OF EQUIVALENT LEVEL TO INTERPRET THE ENTIRE GAME

With the levels of major game events quantified, the distribution of levels during an entire game may be interpreted. Figure 3 shows the distribution of $Leq_{0.5\ s}$ during two games. Each distribution contains everything in the game including timeouts and breaks. The unweighted distribution in Figure 3(a) has three peaks: one around 79 dB, one around 88 dB, and a small peak around 95 dB. Similar peaks exist in in Figure 3(b), but peak 1 is small and peak 3 is more pronounced. The level of peak 3 in this game is 98 dB.

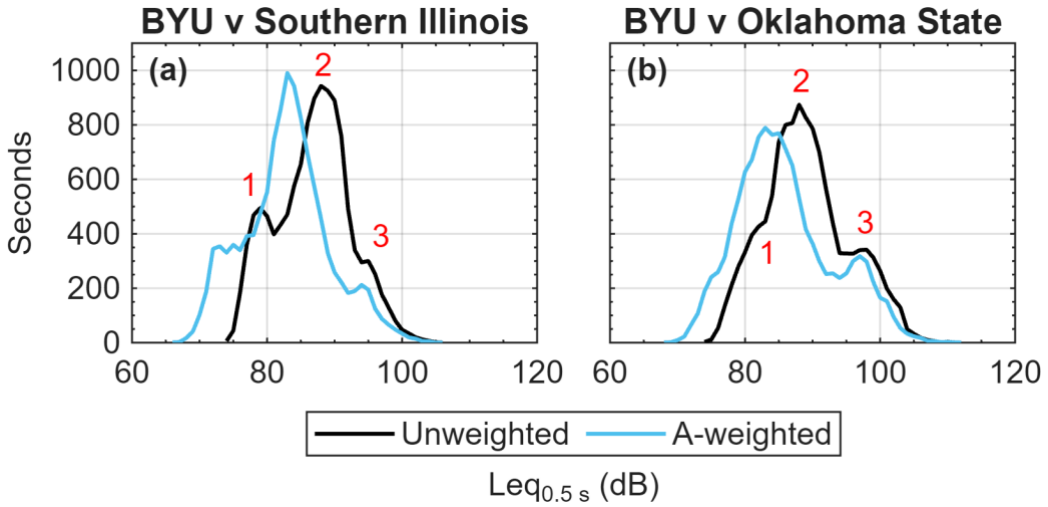


Figure 3: Distribution of unweighted and A-weighted $Leq_{0.5\ s}$ during two football games. Bin width is 1 dB. Annotations in red mark the peaks of the unweighted distributions.

Peaks 1 and 3 are consistent with quiet and distraction noise. The level of quiet, 80.2 dB, is close to the level of peak 1, and the level of distraction noise, 97.5 dB, matches the level of peak 3 (Table 2). Although cheering also frequently reaches approximately 98 dB, it is more likely that peak 3 is associated with distraction noise because distraction noise tends to last much longer than cheering.

At 88 dB, peak 2 most closely matches the level of music played on the PA system. There is no crowd-generated sound that typically has this level. While the level of the crowd singing the fight song is close, the fight song is not sung often enough to occupy this much of the distribution. It is more likely that time spent under this peak is time when the crowd is not engaged, like during timeouts or halftime. Music is often played on the PA system during these times. Because football games have so many stoppages, it seems reasonable that these times would occupy the largest portion of the recording. It is possible that there are other events not considered that contribute to the time under this peak, such as the crowd chanting; however, more game time would need to be analyzed to identify these events. Even still, the amount of time the crowd spends chanting is likely less than the amount of time they spend in an unfocused, low-engagement state.

A-weighting provides additional clues about the content of each peak. After A-weighting, shown in blue, peak 2 of both distributions shifts down by about 5 dB (Figure 3). This difference is most similar to music from the PA system, whose level changes 6.8 dB after A-weighting is applied (Table 2). Peak 1, assumed to be associated with quiet, shifts down between 4 and 6 dB. This is similar to the change in quiet of 4 dB reported in the rightmost column of Table 2. Additionally, quiet is also the only event with a low enough A-weighted level to be a candidate for that peak. Finally, A-weighting changes the level of peak 3 by 1 dB. This matches the small impact of A-weighting on distraction noise and cheering and suggests that peak 3 is mostly composed of crowd-generated noise.

30 September 2026 17:36:01

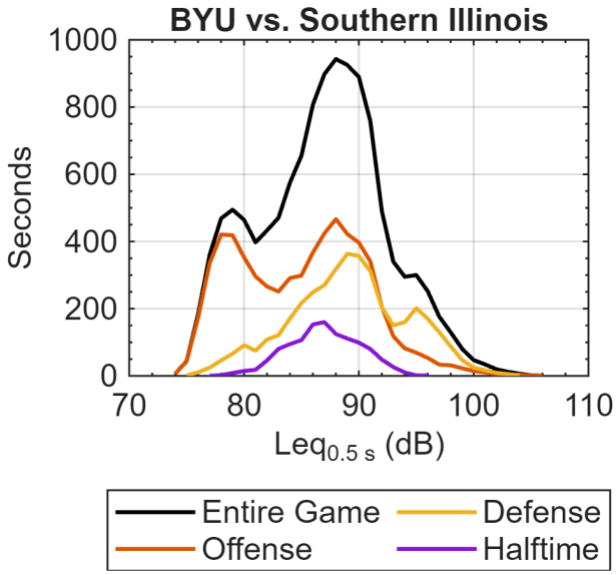


Figure 4: Unweighted $Leq_{0.5\,s}$ distribution for BYU on offense, BYU on defense, halftime, and the entire game.

As a final validation of the meaning of each peak, time spent on offense, defense, and halftime during the BYU vs. Southern Illinois game was labeled and their distributions were plotted (Figure 4). Breaks and timeouts were not included in the offense and defense data but downtime between plays was. The offense and defense distributions each have two clear peaks. In the offense distribution, the peaks are around 78 and 88 dB. In the defense distribution, the peaks are around 89 and 95 dB. These peaks in the offense and defense distributions line up with the peaks in the entire game distribution. Peak 1 is dominated by offense. Peak 2 comes from both offense and defense. Peak 3 is dominated by defense. These patterns are consistent with the interpretation that peak 1 is quiet, which only happens during offense, and peak 3 is distraction noise, which only happens during defense.

The last distribution is from halftime. It peaks around 87 dB, roughly the same as peak 2 in the entire game's distribution (Figure 4). It lacks the additional peaks associated with quiet and distraction noise. The position of its peak right underneath peak 2 is one clue that peak 2 is associated with low crowd engagement. During halftime, the crowd is unfocused and lots of music is played over the PA system. This would also explain the meaning of the peaks around 88 and 89 dB in the offense and defense distributions: they come from downtime between plays when the crowd is not engaged.

If the interpretation of each peak is correct, the shape of the distribution can be used to draw conclusions about each game. In the BYU vs. Southern Illinois game, peak 1 is very distinct and is larger than peak 3 (Figure 3(a)). This could suggest that BYU spent more time on offense than on defense. According to ESPN, BYU did spend almost 50% more time on offense than Southern Illinois (Table 3). In the BYU vs. Oklahoma game, peak 3 is more distinct than in the BYU vs. Southern Illinois game and extends out to higher levels (Figure 3). This suggests that the crowd created more distraction noise during the BYU vs. Oklahoma game than during the BYU vs. Southern Illinois game. This is consistent with the actual game: the teams were repeatedly tied until BYU made a touchdown in the last two minutes of play.[13]

Table 3: Time spent on offense and defense, two games. Information from espn.com.[13]

Game	BYU Offense	BYU Defense	Final Score (BYU, Visiting)
BYU vs. Southern Illinois	39:14	20:46	41, 13
BYU vs. Oklahoma	28:17	31:43	38, 35

30 September 2026 17:36:01

These distributions can also provide information about crowd engagement that is independent from team performance. Overall crowd engagement could be estimated by dividing the time spent under peaks 1 and 3 by the actual time the team spent on offense or defense. While this analysis would be most useful once the game is over, these distributions have implications for real-time modeling, too. The most important events for estimating crowd engagement, quiet and distraction noise, are consistently separated by level. These levels are consistent from game to game. Level is likely level is the only way to identify quiet, one of the most common manifestations of crowd engagement.

B. SPECTRA

To investigate the frequency content of game events, average one-third octave band spectra were created by labeling each sound source and crowd behavior in the ten minutes of Figure 2. Ambient sound, considered to be when the crowd is not engaged and no other sound sources are active, was taken from later in the third quarter. Instances were chosen that did not overlap with other labels, except for distraction noise which often overlaps with music from the PA system before third down. The one-third octave band spectrum from each half-second interval was calculated using the Audio Feature Extractor toolbox from MATLAB[14] and ANSI S1.11.[15] Each average spectrum is shown in Figure 5.

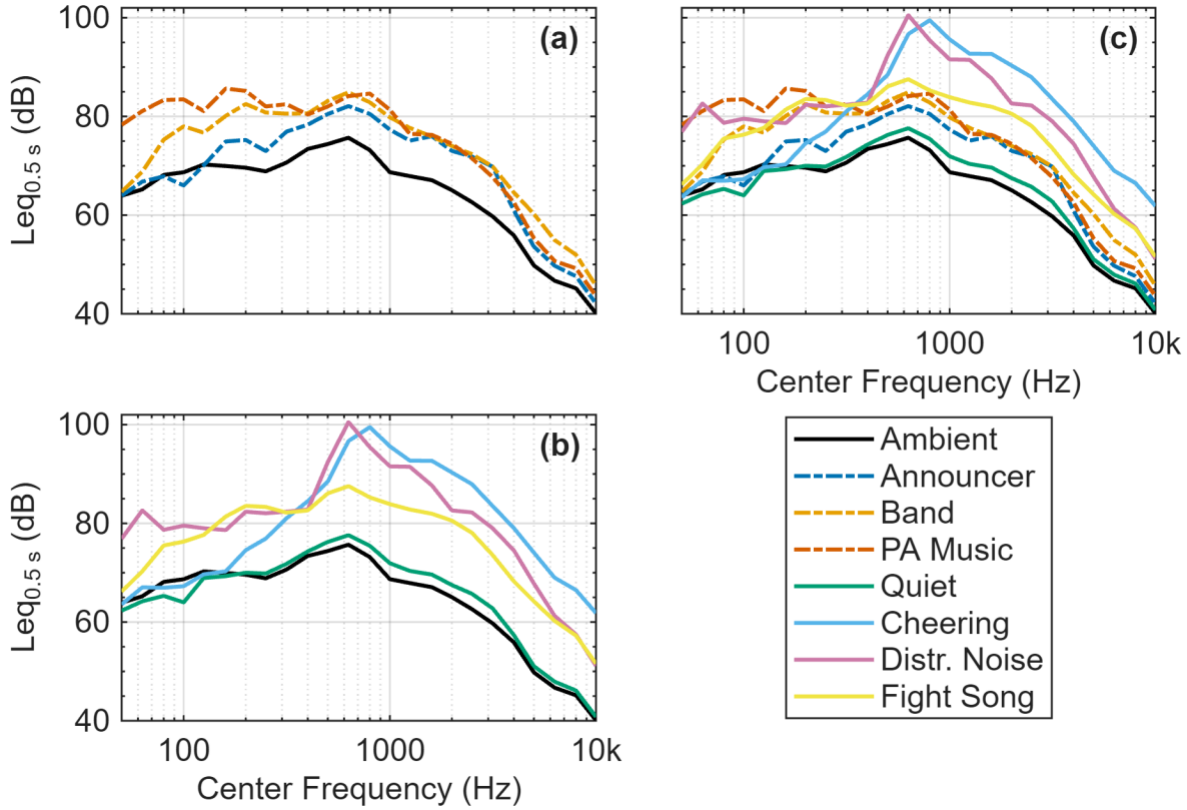


Figure 5: Average one-third octave spectra of sound sources (a) and crowd behaviors (b). (c) is all spectra together.

In the rare moments of the ambient condition, the 630 Hz band has the highest level (Figure 5). This may be related to the spectral peak frequency of human speech in a loud environment. While the fundamental frequency of adult voices is between 100 and 200 Hz, the dominant frequency in the long-term average spectrum of spoken English is between 400 and 500 Hz.[16] When talkers communicate in a loud environment such as a stadium, this frequency shifts up.[17] The energy at 630 Hz may be the result of tens of thousands of people talking in a loud environment.

The announcer's spectrum also peaks at 630 Hz. Below 160 Hz, it matches the ambient spectrum. Its peak and lack of low-frequency energy make it similar to crowd-generated noise such as cheering. However, cheering has a more distinct peak (Figure 5(b)). The announcer's relatively distributed energy could be why the difference between the mean unweighted and A-weighted level of the announcer is 2 dB larger than cheering (Table 2).

30 September 2026 17:36:01

The band and music over the PA system both peak between 630 and 800 Hz but have different amounts of low-frequency content. The band's low frequencies may come from low instruments such as the tuba, which is characterized by energy between 150 and 200 Hz,[18] and bass drum, which has significant energy around 40 Hz.[19] The PA system has so much low-frequency content that its level in the 160 Hz band is higher than the 630 and 800 Hz bands. These low frequencies may come from songs with a significant bass component, which are common at sporting events and were heard throughout this dataset. This is consistent with Barnard et al.[8] which describes the band and PA system's energy as between 50–200 Hz, and Butler et al.[3] which found energy below 315 Hz to be associated with music on the PA system.

The mean spectra of four crowd behaviors are shown in Figure 5(b). The ambient spectrum is also included. Cheering has a higher peak frequency than distraction noise and has more high-frequency energy than all other crowd behaviors. Its peak frequency may be raised because the crowd is expressing positive emotion, which has been connected to a higher mean vocal pitch in dozens of studies.[20] The additional energy above 800 Hz could be caused by clapping; while the frequency content of a clap varies, it is generally significant between 2 and 5 kHz.[21] These factors explain the spectrum of cheering up to 5 kHz, but do not explain the energy between 5 and 10 kHz. Cheering's peak fits well with other studies which have identified the peak frequency of crowd noise to be between 630 Hz and 800 Hz.[3,8]

Distraction noise has a lower peak frequency, less high-frequency energy, and more low-frequency energy than cheering. It peaks at 630 Hz. The value of this peak frequency may be associated with the crowd's vocalization. Distraction noise is often similar to the vowel /o/ (sounds like "ohhhh!"), whose first formant is between 500 and 600 Hz for men, women, and children.[22] Much of this energy would fall in the 630 Hz band where the peak is observed. The next formant is between 900 and 1100 Hz, where the spectrum also has significant energy. In the low frequencies, distraction noise has energy between 50 and 500 Hz that is not found in the ambient condition or other crowd behaviors. These low frequencies are visible in Figure 10 of Barnard et al.[8] Listening to the recordings indicated that this is caused by the crowd stomping their feet. Some of it could also be bassy music on the PA system that is played before third down. This low-frequency energy may be why A-weighting reduces distraction noise more than cheering (Table 2).

The fight song and band spectra are similar below 200 Hz, but the fight song has more energy between 1 and 2 kHz (Figure 5(c)). The low-frequency similarity makes sense because the band plays the fight song while the crowd sings. The high-frequency part of the fight song spectrum is somewhat similar to the shape of the cheering spectrum. It may be associated with positive emotion in people's voices and clapping along to the song. However, it would be best to confirm this with additional instances.

C. PRINCIPAL COMPONENT ANALYSIS

Principal component analysis was performed on the dataset of one-third octave band spectra from all football games. Previous studies have found that the first three principal components of one-third octave band spectra from college basketball and volleyball games are useful for classifying crowd reactions.[6,7] In those studies, the principal components were found to explain 87.5% of the variance in the dataset, and the scaling of each component was found to be associated with overall level, whether noise came from the crowd or PA system, and certain moments in volleyball games.[6] These basketball and volleyball principal components are compared to those from the football dataset in Figure 6. To simplify the comparison, only the first three principal components of football were considered, which explain 85.5% of the variance in the football dataset. The first row is the principal components of football, and the second row is the principal components of the combined basketball and volleyball dataset, referred to as basketball/volleyball. To provide context, each dataset's mean spectrum is shown in black, and the sum of a principal component and the mean is shown in red.

30 September 2026 17:36:01

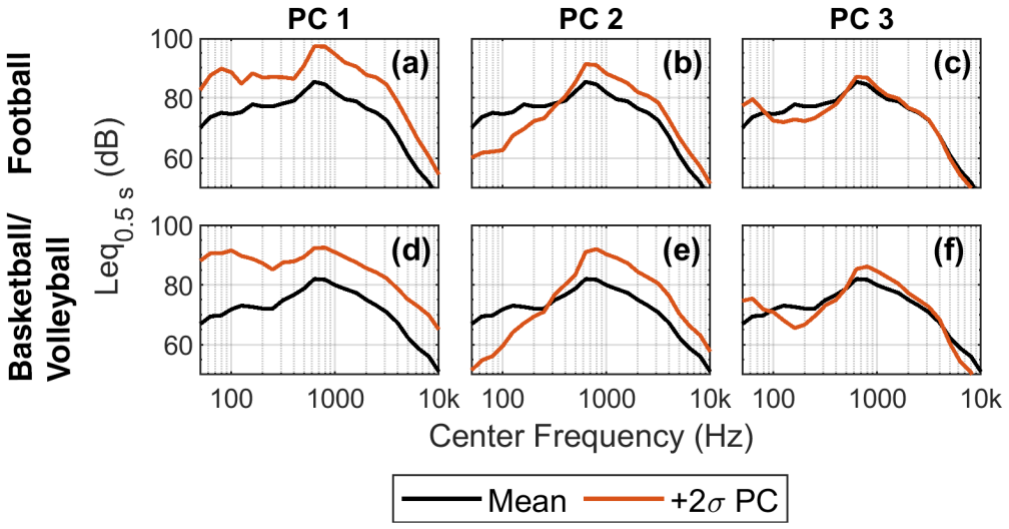


Figure 6: First three principal components of one-third octave band spectra of two datasets, football and combined basketball and volleyball. Each principal component is scaled by two standard deviations from its average scale in the dataset and added to the mean spectrum to show its approximate effect (red).

The mean spectra of the sports differ at high and low frequencies. Both peak at 630 Hz, but basketball/volleyball has more energy above 3 kHz and less energy below 1 kHz (Figure 6). Basketball/volleyball's additional high-frequency energy may be because sound traveled a shorter distance to reach the microphone in the basketball or volleyball stadium and the high frequencies were attenuated less by absorption in the air. The differences below 1 kHz could be attributed to differences between the stadiums such as their PA systems. The details of the mean spectra are not investigated further in this study.

The general effect of the first principal component of both datasets (PC1) is to raise the overall level. Both PC1 peak around 630 and 800 Hz, but football's peak is more distinct and falls off faster in the high frequencies. This peak matches the frequencies of crowd-generated noise identified in the spectra section. Its distinctness may suggest that crowd-generated frequencies are louder than other frequencies at football games more often than at basketball and volleyball games. For example, distraction noise might stand out against music played on the PA system more at a football game than it does at a basketball or volleyball game. This could happen because the football stadium holds more spectators, or could be a consequence of microphone placement. Football PC1's faster high-frequency roll-off may be associated with the larger size of the football stadium as discussed earlier.

In both datasets, the second principal component (PC2) boosts the peak frequency and attenuates the low frequencies. Football PC2 crosses the mean spectrum close to 315 Hz and the basketball/volleyball PC2 crosses the mean spectrum close to 280 Hz (Figure 6(b) and (e)). These crossing frequencies may be significant for describing each dataset. It is interesting that the frequency where the spectra in Figure 5 diverge, around 400 Hz, is close to the frequency where football PC2 crosses the mean spectrum. The crossing point may reflect that low-frequency energy is a significant mode of variation in football game noise. This could be related to the PA system, which frequently plays music with low frequencies.

The third principal components of each dataset (PC3) cross the mean spectrum in similar spots: first near 80 Hz, once near 500 Hz, then again around 4 kHz. This shape is not easily interpretable; however, it is significant that the crossing points of both PC3 are similar. These similarities, along with the similarities between PC1 and PC2, suggest that the datasets are fundamentally similar. While both datasets have similar sound sources and crowd behaviors, they also have significant differences, such as venue size, whether the venue is indoors or outdoors, and the size of the crowd. The resemblance between the principal components suggest that the datasets' similarities outweigh their differences.

4. CONCLUSIONS

Noise at a football game is characterized by patterns in its levels, spectra, and principal components. Events like quiet during offense, music on the PA system, and distraction noise have consistent, distinct levels. The

30 September 2026 17:36:01

difference between the unweighted and A-weighted level of non-crowd-generated sounds is larger than crowd-generated sounds, enabling crowd-generated sounds to be distinguished from other sound sources. The distribution of $Leq_{0.5\ s}$ from an entire game has three peaks that correspond to quiet during offense, low crowd engagement, and distraction noise during defense. The shapes of the A-weighted level distribution and labeled data distributions support this interpretation. Sound sources at the game are distinguished by their low-frequency content below 315 Hz; music on the PA system has the most and the announcer has the least. Cheering and distraction noise peak in similar spots at 800 and 630 Hz, but cheering has high-frequency energy that may be attributed to positive vocalization and clapping while distraction noise has low-frequency energy that may be attributed to foot stomping and music played on the PA system before third down. The principal components of the football dataset are similar to the principal components of a basketball and volleyball dataset: PC1 raises the overall level, PC2 separates high and low frequencies, and PC3 crosses the mean spectrum three distinct points. The resemblance between the principal components suggests that the datasets are fundamentally similar.

D. IMPLICATIONS FOR CHOOSING FEATURES TO ESTIMATE CROWD ENGAGEMENT

The levels, spectra, and principal components of football games provide clues about how to design features that are useful for estimating crowd engagement. The three peaks in the level distribution reveal that the most common crowd behaviors, quiet and distraction noise, can often be identified by level alone. Level may be the only way to identify quiet. Comparing the spectra of crowd behaviors and sound sources reveals that each event has a unique combination of low-, mid-, and high-frequency energy. Low frequencies distinguish music on the PA system, mid frequencies signal high crowd activity, and high frequencies may help identify crowd involvement in certain situations, like when the crowd sings the fight song. Principal component analysis reveals the most common ways the dataset varies, which may have ties to useful features. The first principal component indicates that overall level is important, and the second principal component indicates that the presence of low frequencies is important. This reinforces the significance of level and low-frequency energy identified by the other analyses.

Based on these findings, the level of the 100 Hz, 800 Hz, and 2.5 kHz bands are suggested as features. These features incorporate information about both spectral shape and level. While unlikely, it is possible that $Leq_{0.5\ s}$ contains additional level information that is not represented in the levels of these three bands. To verify that $Leq_{0.5\ s}$ may be excluded, models with and without $Leq_{0.5\ s}$ should be tested. Similarly, the scaling of PC1 and PC2 are excluded because the levels of these frequency bands seem to provide the same information: overall level and the presence of low frequencies. A model with fewer features is desirable because it is interpretable and easier to improve. Because noise throughout the stadium is not uniform, a model must be retrained at each microphone location. However, the features themselves may not need to change. To confirm this, the analysis in this study should be repeated in multiple venues at multiple locations.

The validity of the proposed features is limited by several simplifications. Because these features are based on events' average spectra, they may inaccurately estimate crowd engagement if the spectrum of an event differs from the average. For example, before the bass comes in, the spectrum of music on the PA system would look different from the average spectrum of music shown in Figure 5 and confuse the model. Measuring the standard deviation of spectra from their average spectrum would give an idea of how well these features may be expected to work. A previous study touched on the impact of including a few seconds of feature history when classifying crowd behaviors.[7] This study only investigates a single half-second interval and does not include any history. It is possible that certain crowd behaviors like chanting require several seconds of history to be identified. The evolution of acoustic characteristics of football games over a short period of time could be the topic of another study. While the features proposed are based on the spectra of isolated events, sound sources and crowd behaviors often overlap. The acoustic characteristics of moments of overlap were not studied. It is not clear how a model trained with the suggested features would navigate these moments of overlap. Studying the acoustic characteristics of overlapping sources could reveal what additional features would keep the model from becoming confused.

ACKNOWLEDGEMENTS

The authors thank Eli Farrer, Justin Durfey, Scott Kavanaugh, and Kevin Pelham for providing access to the BYU football facilities and helping maintain the measurement equipment.

30 September 2026 17:36:01

REFERENCES

[1]R. Bachman, "College Football Has Never Been Better --- Player movement and pay disrupted the sport. But those changes have made the on-field product even more popular.," Wall Street Journal

[2]Anthony Chiusano, *The 25 biggest college football stadiums in the country*, Available: https://www.ncaa.com/news/football/article/2018-07-30/25-biggest-college-football-stadiums-country, (date last viewed: 17-Jun-26).

[3]B. A. Butler, K. Pedersen, M. R. Cook, S. G. Wadsworth, E. Todd, D. Stark, K. L. Gee, et al., "Classifying crowd behavior at collegiate basketball games using acoustic data," Proceedings of Meetings on Acoustics **35**, 055006 (2018). doi:10.1121/2.0001061

[4]E. Todd, Mylan, R. Cook, K. Pedersen, D. S. Woolworth, B. A. Butler, X. Zhao, et al., "Automatic detection of instances of focused crowd involvement at recreational events," Proceedings of Meetings on Acoustics **39**, 040003 (2019). doi:10.1121/2.0001327

[5]S. Proksch, M. Reeves, K. Gee, M. Transtrum, C. Kello, and R. Balasubramaniam, "Recurrence Quantification Analysis of Crowd Sound Dynamics," Cogn. Sci. **47**, e13363 (2023). doi:10.1111/cogs.13363

[6]M. C. Cutler, M. R. Cook, M. K. Transtrum, and K. L. Gee, "Data-driven decomposition of crowd noise from indoor sporting events," J. Acoust. Soc. Am. **155**, 962–970 (2024). doi:10.1121/10.0024724

[7]M. C. Cutler, J. Bickmore, M. K. Transtrum, K. Pedersen, S. Proksch, E. Farrer, and K. Gee, "Explainable machine learning models for classifying reactions within crowd noise during men's collegiate basketball games," J. Acoust. Soc. Am. **158**, 3456–3471 (2025). doi:10.1121/10.0039709

[8]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). doi:10.3397/1.3654144

[9]M. Navvab, G. Heilmann, and D. W. Sulisz, "Crowd noise measurements and simulation in large stadium using beamforming.,"

[10]*College Football Stadium Comparisons*, Available: https://www.collegegridirons.com/comparisons-by-capacity/, (date last viewed: 28-Sep-25).

[11]*Football Facilities - BYU Athletics - Official Athletics Website - BYU Cougars*, Available: https://byucougars.com/football-facilities, (date last viewed: 28-Sep-25).

[12]*The ROC*, Available: https://byucougars.com/sports/roc, (date last viewed: 30-Jun-26).

[13]*College Football on ESPN - Scores, Stats and Highlights*, Available: https://www.espn.com/college-football/, (date last viewed: 11-Jun-26).

[14]*MATLAB audioFeatureExtractor*, Available: https://www.mathworks.com/help/audio/ref/audiofeatureextractor.html, (date last viewed: 01-Jul-26).

[15]"Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters.," Retrieved from https://webstore.ansi.org/standards/asa/ansis1112004?srsltid=AfmBOooPZ1pCt0nfAz6J4rMFQSoIDFmZiNsbHAdUeL2MgW8ArMm_winP

[16]D. Byrne, H. Dillon, K. Tran, S. Arlinger, K. Wilbraham, R. Cox, B. Hagerman, et al., "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**, 2108–2120 (1994). doi:10.1121/1.410152

[17]K. S. Pearsons, R. L. Bennett, and S. A. Fidell, *Speech levels in various noise environments*, (U.S. Environmental Protection Agency, Office of Health and Ecological Effects, Office of Research and Development, 1977).

[18]C. Beauregard, *The tuba; a description of the five orchestral tubas and guidelines for orchestral tuba writing.*, (University of Rochester, 1970).

[19]H. Fletcher, I. G. Bassett, and P. A. Mtgs, "Some experiments with the bass drum," J. Acoust. Soc. Am. **64**, 1570–1576 (1978). doi:10.1121/1.382140

[20]R. G. Kamiloğlu, A. H. Fischer, and D. A. Sauter, "Good vibrations: A review of vocal expressions of positive emotions," Psychon. Bull. Rev. **27**, 237–265 (2020). doi:10.3758/S13423-019-01701-X

[21]B. H. Repp, "The sound of two hands clapping: An exploratory study," J. Acoust. Soc. Am. **81**, 1100–1109 (1987). doi:10.1121/1.394630

[22]J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111 (1995). doi:10.1121/1.411872

30 September 2026 17:36:01