Analyzing Frequency Spectra of Crowd Speech

Sam Greenwell
Physics 492R Capstone Project Report
Dr. Transtrum
April 13, 2022

Abstract

The purpose of this project is to determine significant acoustic features or conducive analyzation methods for crowd speech detection for subsequent extrapolation to crowd sentiment detection. This project found that crowd speech when treated as a general noise is differentiable from other prevalent crowd noises and that speech is easily overpowered by these other noises. However, differentiating singular letters may not be possible by using frequency spectra alone.

Introduction

Speech recognition has been a common technology for the past decade, reaching accuracies of around 90-95%. It uses linguistic and acoustical properties such as phonemes, formants, fundamental frequencies, and harmonics to analyze someone's speech and determine what they are saying. With crowds this becomes the features that make speech recognition possible are distorted with hundreds or even just a few people.

Some work has already been done on crowd noise analysis. Butler et al (2018) found success in using supervised and unsupervised machine learning (ML) methods for classifying averaged crowd signals. Their results showed that it might be possible to analyze acoustical data with ML methods to determine crowd sentiment. Later, Todd et al (2019) found that using spectral slope and flux might be useful in identifying crowd involvement.

Dr. Transtrum and Dr. Gee's Crowd Noise research group at Brigham Young University (BYU) seeks to find a way to differentiate and recognize crowd sentiment using acoustical data. To aid in this effort I analyzed crowd speech, specifically the "B-Y-U" chant at BYU volleyball games, to find certain acoustical features or analyzation methods that could be later applied to crowd sentiment detection. I found that using frequency spectra can identify crowd speech as a category of noise, but alone cannot discern what the crowd is saying.

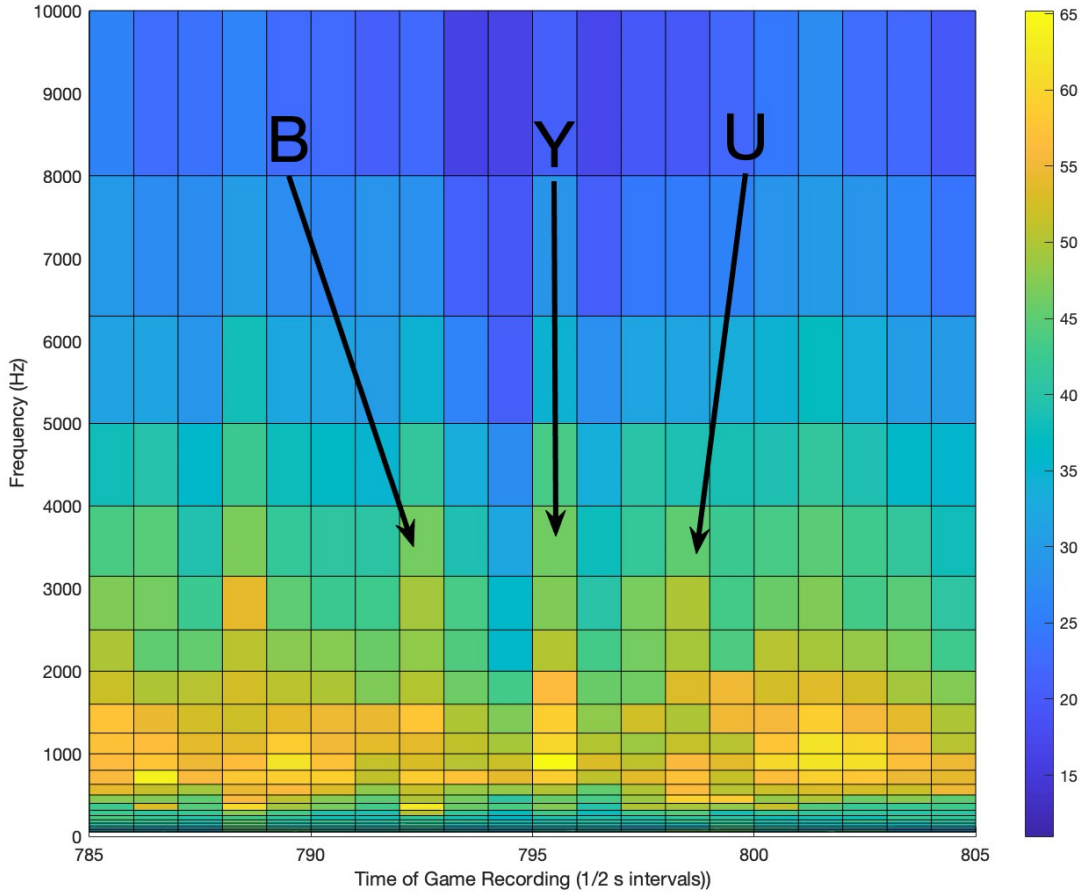
Methods

The acoustical data from BYU volleyball games was already recorded and processed prior to the outset of this project. The data consists of spectra of the volleyball games in 1/3 octave bands at half second intervals. However, the length of some of the processed data files did not match the length of the .wav audio files. This meant that those games could not be used in this analysis. Out of the twelve processed games, three were unusable. To make these games usable, they would need to be reprocessed so that the length of both files is consistent. Of the nine games that were available, I selected two for analysis.

The first step of this project was to find recorded volleyball games with consistent lengths in the .wav audio and processed data files, then listen to them and identify several instances of the "B-Y-U" chant. Since analysis on these specific chants is unprecedented, ideal instances were needed. These ideal instances needed to be audibly distinct and clear. Meaning the letters are highly distinguishable from each other, all three letters were said, and the other background noise levels should be much lower compared to the chant. Depending on many factors relating to the specific game such as length, collegiate division, and outcome I estimate there are approximately 7-12 ideal instances in each game. For each game I analyzed I chose three instances of the "B-Y-U" chant. I chose over the entire length of the game to make my entire

sample more representative of the entire population. After recording the specific times for each letter in each chant instance, the frequency spectra were calculated for them.

This is an example of a spectrogram of the first chant instance in the women's game with labeled peaks for each letter.



The frequency spectra were computed by primarily using a MATLAB script created by the research group, afeLinToFracOct. There is no listed author in the script and a file path of it is listed in Appendix A. This script converts the linear spectrum data into fractional octave band spectral levels using MATLAB's audioFeatureExtractor. The frequency spectra were computed for three "B-Y-U" chant instances in two games, the 9-20-2018 BYU Women's Volleyball game and the 3-14-2019 BYU Men's Volleyball game. These games were chosen because they were average games as far as outcome and attendance are concerned. Both games were BYU wins and had about two to three thousand in attendance.

Results

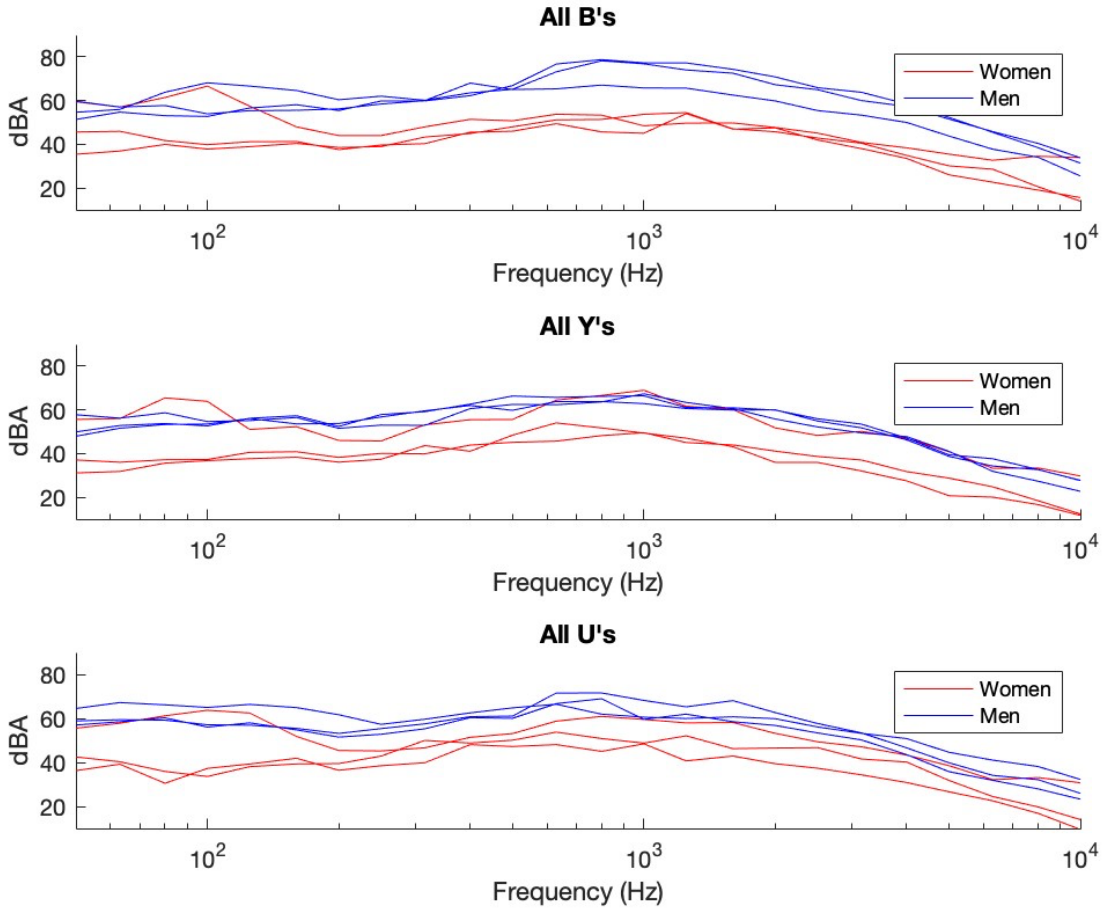
The results are broken down into four main parts.

1. Looking for similarities among the different instances of the same letters.

2. Looking for similarities among the different letters in the same chant instance.
3. Studying the temporal evolution of each letter.
4. Exploring the impact of other noises on crowd chant spectra.

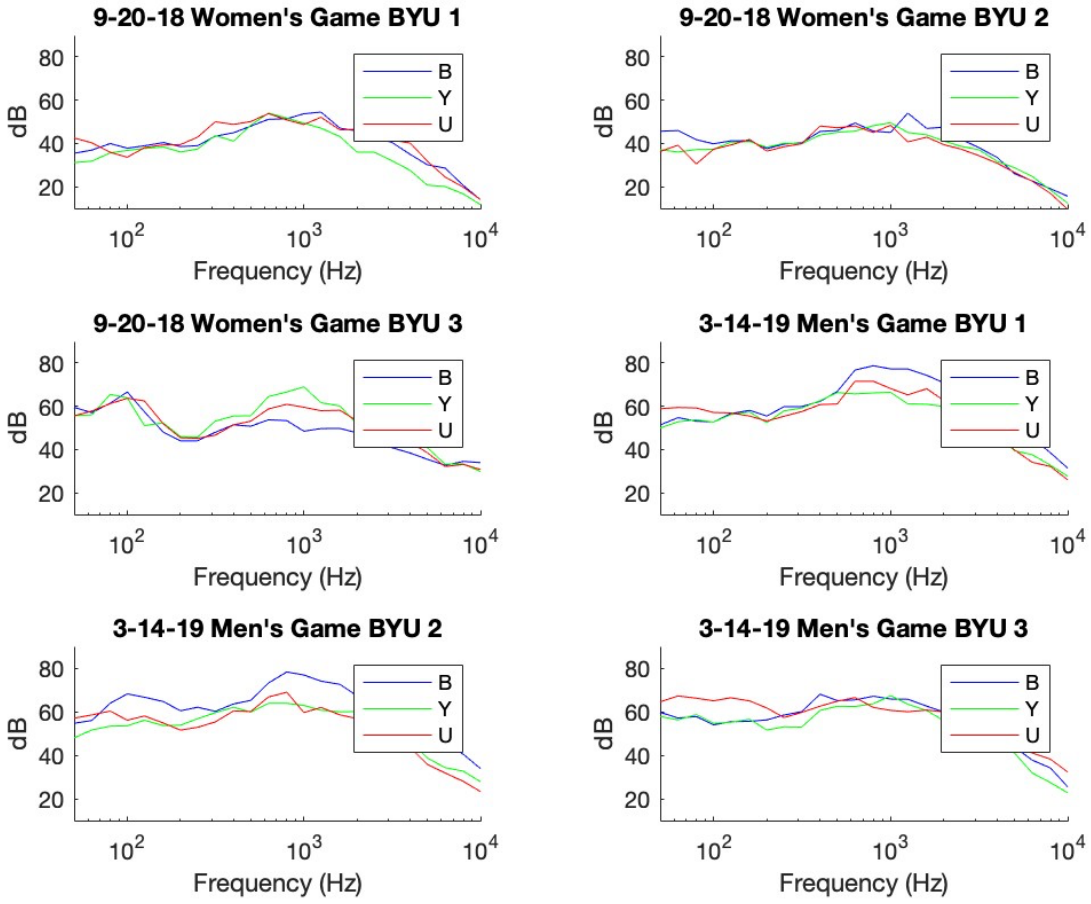
Similarities Among the Same Letters

Figure 2 show the similarities between the instances when grouped by letter. Notice that the women's game spectra have lower decibel levels than the men's game spectra do. However, the general shape for each letter is consistent across almost all instances. There is a women's game instance that deviates severely, but that is due to PA music and will be explained more in depth later in the report.



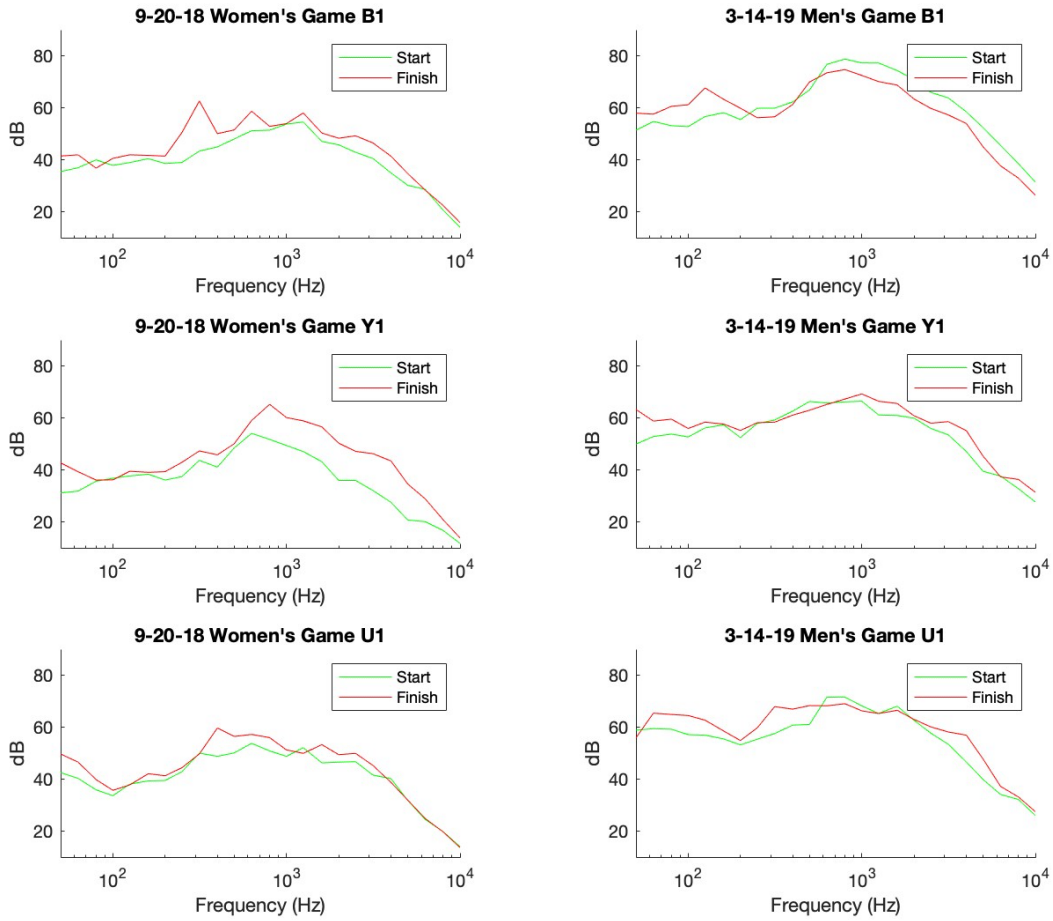
Similarities Among Within Each Chant

Figure 3 shows each instance of the "B-Y-U" chant with each line being the frequency spectra of a different letter. For the most part the letters are very similar to each other. However, there are nuances that are important to notice. The main peak of the "B" is consistently at a slightly higher frequency than the other letters. This is most likely because of the ē phoneme in the pronunciation of "B". Also, at frequencies of 100 Hz and lower the "U" consistently rises in decibel level. This is most likely due to the ū phoneme in the pronunciation of "U".



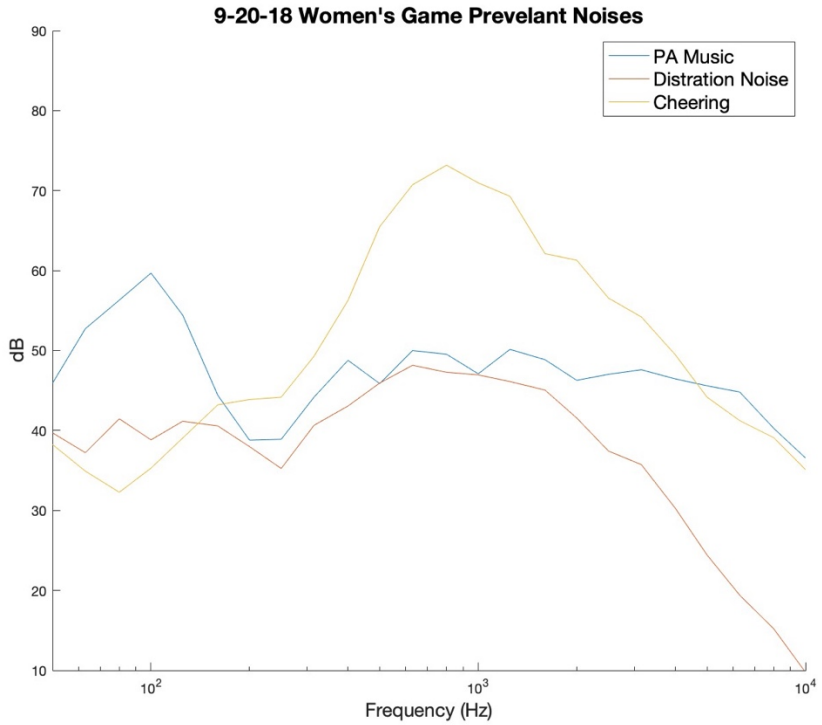
Temporal Evolution

Each letter takes place over about a second. Figure 4 shows the change in spectra from the beginning of each letter to the end. We can see that the "B" has little to no change in decibels or frequency, the "Y" increases slightly in frequency and decibels, and the "U" stays relatively at the same decibel level and decreases slightly in frequency.

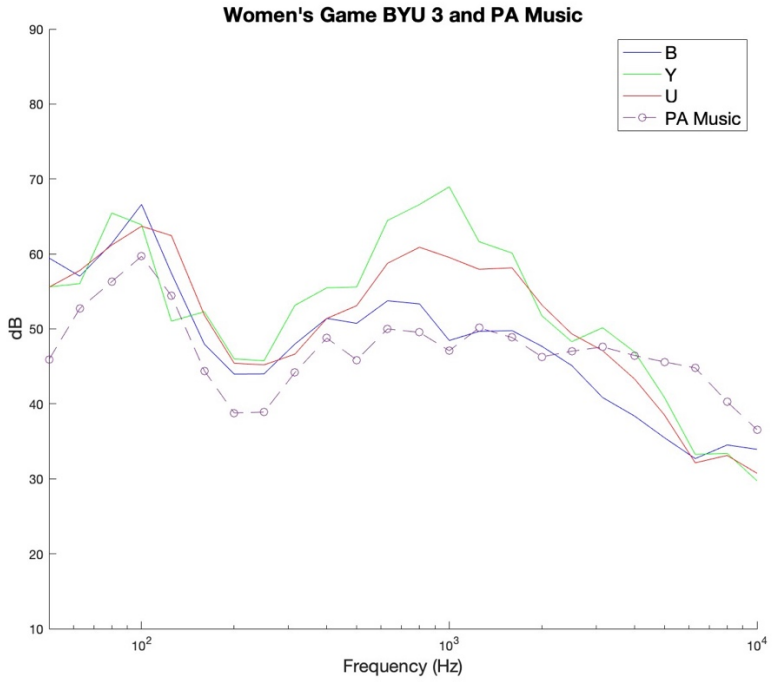


Impact of Other Noises on Crowd Speech Spectra

The last aspect that I investigated was how common noises at BYU volleyball games effect the spectra of the crowd speech. Specifically, PA music, distraction noise, and cheering. The frequency spectra of those noises are found in figure 5. The PA music has a peak in the lower frequency range most likely due to the bass capabilities of the stadium speakers. The distraction noise has a slight peak in the mid-high frequency range and cheering has a large peak in the high frequency range. It is important to note that these spectra are very differentiable from crowd speech spectra.



The third instance of the “B-Y-U" chant in the women’s game is much different from the rest because there is PA music in the background during it. As you can see in figure 6, the frequency spectra of the letters are very similar to the spectra of the PA music. Therefore, some background noises will entirely change the frequency spectra of crowd speech.



Conclusion

In conclusion, crowd speech can be treated as a general noise and the frequency spectra of it can most likely be used to identify it. However, the spectra of different letters from a crowd are very similar although they have nuances that can possibly be used differentiate them. Therefore, we can tell that the crowd is talking, just not what they are saying. Also, other predominant simultaneous crowd noises will most likely have a great effect on the resulting spectra of crowd speech.

In addition to investigating frequency spectra, other acoustic features such as the overall sound pressure level (OASPL) should be looked at in these chant instances to gain greater insight into methods of crowd speech detection. Combining data from multiple features will most likely give much more understanding of how to detect speech from a crowd.

References

Butler, B. A., Paré, P. E., Transtrum, M. K., & Warnick, S. (2021, September 11). *Modeling live crowd emotion dynamics for state estimation and prediction*. IEEE Xplore. Retrieved April 13, 2022, from https://ieeexplore.ieee.org/document/9658725

Butler, B. A., Pedersen, K., Cook, M. R., Wadsworth, S. G., Todd, E., Stark, D., Gee, K. L., Transtrum, M. K., & Warnick, S. (2018, November 5). *Classifying crowd behavior at collegiate basketball games using acoustic data*. Scitation. Retrieved April 13, 2022, from https://asa.scitation.org/doi/abs/10.1121/2.0001061

Todd, E., Cook, M. R., Pedersen, K., Woolworth, D. S., Butler, B. A., Zhao, X., Liu, C., Gee, K. L., Transtrum, M. K., & Warnick, S. (2019, December 2). *Automatic detection of instances of focused crowd involvement at recreational events*. Scitation. Retrieved April 13, 2022, from https://asa.scitation.org/doi/abs/10.1121/2.0001327

Acknowledgments

Dr. Mark Transtrum

Dr. Kent Gee

Mylan Cook

Mitchell Cutler

Michael Pierce

Appendix A

afeLinToFracOct.m File Path in Box

Acoustic Crowd Behavior/1 Data/AFE Feature Analysis/Cutler-Greenwell-Pierce/BYU_Chant/afeLinToFracOct.m