

Visualization and Interpretation of Mel-Frequency Cepstral Coefficients for UAV Drone Audio Data

Mia Y. Wang¹^a, Zhiwei Chu²^b, Conner Entzminger³, Yi Ding⁴ and Qian Zhang⁵^c

¹Department of Computer Science, College of Charleston, Charleston, SC, U.S.A.

²Department of Computer and Information System, Purdue University, West Lafayette, IN, U.S.A.

³Department of Biology, College of Charleston, Charleston, SC, U.S.A.

⁴Department of Business Administration, Anhui Polytechnic University, China

⁵Department of Engineering, College of Charleston, Charleston, SC, U.S.A.

Keywords: Unmanned Aerial Vehicles Classification, UAV Audio Dataset, Audio Analysis, Mel-Frequency Cepstral Coefficients (MFCCs), Spectrogram, Waveform, Mel Filterbank, Feature Extraction.

Abstract: Unmanned Aerial Vehicles (UAVs) have become a focal point in various fields, prompting the need for effective detection and classification methodologies. This paper presents a thorough investigation into UAV audio signatures using Mel-Frequency Cepstral Coefficients (MFCCs). We meticulously explore the influence of varying MFCC quantities on classification accuracy across diverse UAV categories. Our analysis demonstrates that employing 30 MFCCs produces promising outcomes, characterized by reduced variance and heightened discriminatory capability compared to alternative configurations. Moreover, we introduce a novel image-based dataset derived from our existing audio dataset, encompassing waveform, spectrogram, Mel filter bank, and MFCC plots for 26 UAV categories, each comprising 100 audio files. This dataset facilitates comprehensive analysis and the development of multimodal UAV detection systems. Our research highlights the significance of leveraging diverse datasets and identifies future paths for UAV detection and classification research.


1 INTRODUCTION


Unmanned Aerial Vehicles (UAVs), commonly referred to as drones, have emerged as versatile tools for various applications, including surveillance, environmental monitoring, and search and rescue operations. In these applications, audio data captured by UAV drones plays a crucial role in understanding environmental conditions, detecting anomalies, and monitoring acoustic events of interest. However, analyzing UAV drone audio data poses several challenges due to its complex and dynamic nature, requiring sophisticated signal processing techniques for effective interpretation and understanding.


One widely used technique for analyzing audio signals is the extraction of Mel-Frequency Cepstral Coefficients (MFCCs), which provide a compact and informative representation of the spectral character-

istics of the audio signal. MFCCs have been extensively employed in speech recognition, music analysis, and sound classification tasks due to their ability to capture key acoustic features while reducing the dimensionality of the feature space (Muda et al., 2010)(Zheng et al., 2001)(Wang et al., 2022b)(Wang et al., 2024). In recent years, there has been growing interest in exploring the visualization and interpretation of MFCCs for UAV drone audio data analysis. Visualizing MFCC features enables researchers and practitioners to gain insights into the underlying acoustic properties of the audio recordings, identify distinct sound events, and understand the temporal and spectral dynamics present in the data (Stylianou, 2001)(Le et al., 2021).

This paper presents a comprehensive exploration of techniques for visualizing and interpreting MFCCs in the context of UAV drone audio data analysis. We investigate various visualization methods, including spectrograms, feature distributions, time series plots, and cluster analysis, to elucidate the acoustic char-

^a <https://orcid.org/0000-0003-2954-0855>

^b <https://orcid.org/0000-0002-8283-5831>

^c <https://orcid.org/0009-0009-3362-8420>

acteristics captured by MFCC features. By leveraging these visualization techniques, we aim to enhance our understanding of UAV drone audio data and facilitate the development of more effective signal processing algorithms and machine learning models for audio event detection, classification, and environmental monitoring tasks.

The remainder of this paper is organized as follows: Section 2 provides an overview of related work in the field of audio signal processing and MFCC analysis. Section 3 describes the methodology used for extracting, visualizing, and interpreting MFCC features from UAV drone audio data. In Section 4, we present experimental results and insights gained from our analysis of MFCC visualizations. Finally, Section 4 concludes the paper and outlines future research directions in the field of UAV drone audio data analysis and visualization.

2 LITERATURE REVIEW

The analysis of audio data captured by Unmanned Aerial Vehicles (UAVs) has garnered significant attention in recent years due to its relevance in various fields such as surveillance, environmental monitoring, and search and rescue operations. In this context, Mel-Frequency Cepstral Coefficients (MFCCs) have emerged as a popular choice for analyzing UAV drone audio data. Previous research, such as that conducted by (Stylianou, 2001), (Muda et al., 2010) and (Wang et al., 2022b), has demonstrated the effectiveness of MFCCs in capturing the spectral characteristics of audio signals while reducing dimensionality. These coefficients offer a compact representation of audio signals, making them well-suited for processing large volumes of data collected by UAV drones in real-world scenarios.

In addition, recent studies by (Wang et al., 2022b) and (Kim et al., 2022) have delved into the classification of drones carrying payloads and the detection of target drones using deep learning models, respectively. These studies underscore the evolving landscape of UAV drone audio data analysis and the significance of innovative methodologies in advancing the field.

Visualization techniques, including spectrograms, time series plots, and feature distributions, play a crucial role in interpreting MFCC features and understanding the underlying acoustic properties of UAV drone audio data. (Altes, 1980) have highlighted the importance of spectrograms in providing a time-frequency representation of audio signals, aiding in the identification of distinct sound events. Addition-

ally, cluster analysis techniques, as discussed by (Le et al., 2021), have been employed to group similar MFCC feature vectors together, enabling the identification of common acoustic patterns and clusters within UAV drone audio data.

In our prior work (Wang et al., 2024), we introduced a novel approach for UAV drone detection and classification using audio signals, leveraging MFCCs as key features. However, a key challenge identified was the limited availability of publicly accessible datasets tailored for audio-based UAV detection and classification systems. To address this gap, we curated a comprehensive dataset comprising a diverse range of UAVs and developed a convolutional neural network (CNN) model for UAV classification tasks. The results demonstrated impressive accuracy, highlighting the efficacy of the proposed methodology and dataset. Building upon this foundation, our current study aims to explore additional modalities and enhance classification accuracy, contributing to the advancement of audio-based UAV detection and classification systems.

In recent years, there has been growing interest in exploring the visualization and interpretation of MFCCs and other audio features for UAV drone audio data analysis. While existing research has demonstrated the effectiveness of MFCCs for analyzing UAV drone audio data, there remains a need for further exploration of visualization and interpretation techniques to gain deeper insights into the acoustic characteristics of the recordings. By leveraging advances in signal processing, machine learning, and data visualization, researchers can develop novel approaches for analyzing and interpreting MFCC features in the context of UAV drone audio data analysis. Notably, recent studies by (Wang et al., 2022b) and (Kim et al., 2022) have delved into the classification of drones carrying payloads and the detection of target drones using deep learning models, respectively. These studies underscore the evolving landscape of UAV drone audio data analysis and the significance of innovative methodologies in advancing the field.

In the subsequent sections, we outline the methodology employed for the extraction, visualization, and interpretation of Mel-Frequency Cepstral Coefficients (MFCCs) from UAV drone audio data. Additionally, we introduce a novel dataset derived from the original audio dataset and provide detailed insights into the experimental outcomes and observations derived from our analysis of MFCC features.

Table 1: UAV Audio Dataset 26 Classes.

Manufacture	Model	Drone Type	Number of Files	Duration (sec)
Self-build	David Tricopter	Outdoor	102	510
Self-build	PhenoBee	Outdoor	116	580
Autel	Evo 2 Pro	Outdoor	100	500
DJI	Avata	Outdoor	117	585
DJI	FPV	Outdoor	121	605
DJI	Matrice 200	Outdoor	100	500
DJI	Matrice 200 V2	Outdoor	105	525
DJI	Matrice 600p	Outdoor	118	590
DJI	Mavic Air 2	Outdoor	131	655
DJI	Mavic Mini 1	Outdoor	120	600
DJI	Mini 2	Outdoor	116	580
DJI	Mavic 2 Pro	Outdoor	100	500
DJI	Mavic 2s	Outdoor	115	575
DJI	Phantom 2	Outdoor	106	530
DJI	Phantom 4	Outdoor	100	500
DJI	RoboMaster TT Tello	Indoor	118	590
Hasakee	Q11	Indoor	108	540
Syma	X5SW	Indoor	110	550
Syma	X5UW	Indoor	105	525
Syma	X8SW	Indoor	108	540
Syma	X20	Indoor	112	560
Syma	X20P	Indoor	104	520
Syma	X26	Indoor	138	690
Swellpro	Splash 3 plus	Outdoor	120	600
Yuneec	Typhoon H Plus	Outdoor	113	565
UDI RC	U46	Outdoor	101	505
Total	-	-	2842	14210

3 METHODOLOGY

3.1 Audio Data Collection

In the data collection process, audio data was gathered from 26 different UAVs, detailed in Table 1. Each UAV contributed at least 100 audio entries, with each entry consisting of 5 seconds of flying drone audio data. In total, 2678 audio files were collected, amounting to 13390 seconds. UAVs were sourced from various manufacturers, including DJI, Autel, Syma, Yuneec, UDI, Hasakee, and self-built models. Of the 26 UAVs, 24 were quadcopters. DJI, Autel Robotics, Yuneec, and self-built UAVs were operated and recorded in outdoor environments, while Syma, UDI, Hasakee, and one DJI UAV were used for indoor recordings.

The dataset also includes 1 tricopter and 1 hexacopter. Additionally, two self-built UAVs are part of the dataset: David Tricopter and PhenoBee (Chen, 2023). David Tricopter, designed and built by David Windestal, is equipped with AfroFlight Naze32 flight control, weighs 2.6 lbs with the battery, and has a diameter of 34 inches. PhenoBee, constructed by Ziling Chen, is the largest UAV in the dataset, weighing approximately 23kg, with a height and diameter of 1.35

meters. PhenoBee operates on Ardupilot framework with Cubepilot Cube Orange hardware.

No post-processing was conducted on the original data, resulting in the presence of background noises like wind, birds, and traffic in the outdoor audio recordings. These recordings were also influenced by changing weather conditions, spanning from sunny and cloudy days with low wind speeds to foggy and windy days. Wind speeds varied between 5mph to 13mph, falling below the threshold of a "Moderate Breeze" as classified by the National Weather Service (Service, 2023). Additional weather metrics, including temperatures ranging from 39 to 79 degrees Fahrenheit and humidity levels averaging around 64% during recording days, were also documented.

3.2 Mel-Frequency Cepstral Coefficients Extraction

Mel-Frequency Cepstral Coefficients (MFCCs) are widely used in audio signal processing for capturing the spectral characteristics of audio signals. MFCCs are derived from the Short-Time Fourier Transform (STFT) of the audio signal and provide a compact representation of the spectral envelope.

3.2.1 Preprocessing

Before extracting MFCCs, the audio signal is preprocessed to enhance its suitability for analysis. Common preprocessing steps include noise reduction, normalization, and windowing to improve the signal-to-noise ratio and minimize artifacts introduced during signal acquisition and processing.

3.2.2 Short-Time Fourier Transform (STFT)

The Short-Time Fourier Transform (STFT) is computed to obtain a time-frequency representation of the audio signal. The STFT decomposes the audio signal into short overlapping windows, each of which is Fourier transformed to obtain the frequency content of the signal within that window.

Mathematically, the STFT of a signal $x(t)$ at time t and frequency ω is given by:

$$X(\omega, t) = \int_{-\infty}^{\infty} x(\tau)w(\tau - t)e^{-j\omega\tau}d\tau$$

where $w(\tau - t)$ is a window function centered at time t .

3.2.3 Mel Filterbank

The Mel Filterbank is used to approximate the non-linear human auditory system's frequency response. It divides the audio spectrum into a set of overlapping triangular filters, each corresponding to a specific frequency range.

The transformation from linear frequency scale (Hz) to Mel scale is given by:

$$M(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

where $M(f)$ is the frequency in Mel scale and f is the frequency in Hz.

3.2.4 Discrete Cosine Transform (DCT)

The Discrete Cosine Transform (DCT) is applied to the logarithmically-scaled filterbank outputs to decorrelate the filterbank energies and obtain a compact representation of the spectral envelope.

The n -th MFCC coefficient c_n is computed as:

$$c_n = \sum_{k=0}^{N-1} \log(A(k)) \cos\left[\frac{\pi n(k+0.5)}{N}\right]$$

where $A(k)$ represents the magnitude spectrum obtained after applying the Mel filterbank, and N is the number of filterbanks.

3.2.5 MFCC Coefficients

After applying the DCT, the resulting coefficients represent the Mel-Frequency Cepstral Coefficients (MFCCs). Typically, a subset of these coefficients is selected for further analysis and feature extraction, depending on the specific application requirements.

In this study, we extract M MFCC coefficients from each audio segment to capture the relevant spectral characteristics and reduce the dimensionality of the feature space.

3.3 MFCC Feature Visualization, Interpretation, and Analysis

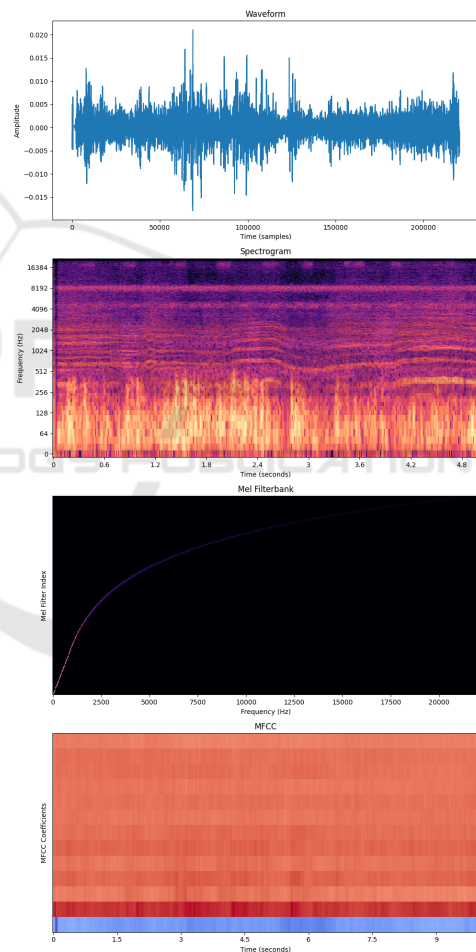


Figure 1: Visualization of DJI Mini 2 Audio Recording Illustration.

Mel-Frequency Cepstral Coefficients (MFCCs) serve as fundamental features in the analysis of UAV drone audio data. In this subsection, we delve into the visualization, interpretation, and analysis of MFCCs to unveil meaningful insights and patterns within the au-

dio dataset.

We adopted a variety of visualization techniques to explore the spectral characteristics and temporal dynamics inherent in the audio signals represented by MFCCs. Leveraging Python libraries such as *librosa* and *matplotlib*, we utilized spectrograms and heatmaps to visualize the intensity of MFCC coefficients across both time and frequency bins. These visualizations provide a comprehensive overview of the spectral content and temporal variations present within the audio signals. Figure 1 showcases four distinct plots extracted from a 5-second audio recording captured by a DJI Mini 2 drone, representing the audio signal in waveform format, the corresponding spectrogram, the Mel Filterbank analysis, and the MFCC coefficients.

The interpretation of MFCC features involves understanding the acoustic properties embedded within individual coefficients and their contextual significance in diverse sound phenomena and environmental settings. Through a meticulous analysis of MFCC coefficient distributions and magnitudes across an array of audio samples and environmental contexts, our objective is to discern unique patterns and characteristic signatures emblematic of distinct UAV drone activities and acoustic occurrences. To achieve this, we extracted 10, 20, 30, and 40 MFCCs, as well as 32, 64, and 128 coefficients, from the audio data. We computed statistical metrics including mean and variance for MFCCs extracted from each audio file, storing the computed statistics in a CSV file. This approach facilitates further analysis and visualization to unearth deeper insights into the underlying characteristics of the audio data.

4 EXPERIMENTS AND RESULTS

4.1 UAV Audio Data Visualization Dataset

We have extracted and curated a new image-based dataset derived from our existing audio dataset (Wang et al., 2024)(Wang et al., 2022a), expanding the scope and versatility of our research efforts in UAV detection and classification. The new dataset represents a significant augmentation of our previous audio dataset, encompassing waveform, spectrogram, Mel filter bank, and MFCC plots of 26 distinct categories of UAV drones. Each category comprises 100 audio files, resulting in a total of 2600 audio recordings. Each category comprises 100 audio files, resulting in a total of 2600 images, allowing for a comprehensive analysis of acoustic features across various UAV plat-

forms. The inclusion of multiple audio representations enables a deeper exploration of the spectral and temporal characteristics of UAV audio signals, providing valuable insights into the underlying patterns and nuances present in different drone types.

The importance of the new dataset lies in its potential to enrich our understanding of UAV audio signatures and enhance the capabilities of audio-based UAV detection and classification systems. By leveraging a more extensive and diverse dataset, researchers and practitioners can develop more robust algorithms and models for UAV detection, classification, and tracking tasks. The expanded dataset also facilitates the exploration of advanced machine learning techniques, such as deep learning and transfer learning, which can further improve the accuracy and reliability of UAV audio analysis systems. Moreover, the availability of a comprehensive dataset fosters collaboration and knowledge sharing within the research community, encouraging the development of standardized evaluation metrics and benchmark datasets for UAV audio analysis. Ultimately, the new dataset serves as a valuable resource for advancing research in UAV acoustics and contributing to the development of more efficient and reliable UAV monitoring and surveillance technologies.

4.2 Experiment

In the process of analyzing UAV drone audio data, we embarked on a systematic procedure to extract distinct sets of Mel-Frequency Cepstral Coefficients (MFCCs) tailored to each UAV drone category. Initially, we organized the audio data into categorized folders, each representing a unique UAV drone model. Subsequently, we meticulously traversed through these folders, meticulously loading individual audio files utilizing the powerful *librosa* library. Employing the *librosa.feature.mfcc()* function, we extracted MFCC features from the audio signals, affording us the flexibility to specify the number of coefficients to extract. Through this methodical approach, we computed essential statistical descriptors such as mean and variance across the audio files for each UAV drone category and for varying numbers of MFCC coefficients. This meticulous analysis provided us with valuable insights into the acoustic profiles of the different UAV drone models, aiding in discerning distinctive acoustic signatures pertinent to each model's operation and functionality.

4.3 Statistical Analysis

In Figure 2 and Figure 3, we analyzed the mean and variance of Mel-Frequency Cepstral Coefficients (MFCCs) features extracted from various UAV drone audio data categories. We calculated the mean ranges for each number of MFCC using $\text{mean} \pm \text{variance}$ and plotted the ranges as bar plots. The results from the same drone category are stacked in the same bar with different colors representing different numbers of MFCC. Figure 2 compares four different coefficient numbers: 10, 20, 30, and 40, represented by bars in red, green, blue, and yellow, respectively. Meanwhile, Figure 3 examines three distinct coefficient counts: 32, 64, and 128, depicted by bars in red, green, and blue, respectively.

The plotted results from Figure 2 show that the calculated mean ranges decrease as the number of MFCC increases from 10 to 30. A narrower interval of mean values and fewer overlaps among the computed mean of different types of drones can more accurately reflect the distinct categories. Therefore, using 30 coefficients appears to be the optimal choice for representing the different drone categories in our dataset. In the first plot, the mean range for 30 MFCCs tends to exhibit a balance between capturing relevant information and mitigating the effects of noise. The variance of MFCCs tends to decrease as the number of coefficients increases, indicating a more robust representation of the audio features. This observation is further supported by the second plot, where the mean and variance of MFCCs for 30 frames show a favorable performance compared to other frame sizes.

Figure 3 plotted the calculated feature mean range for 32, 64, and 128 MFCCs. Beyond 32 MFCCs, there are minimal changes observed in the plotted mean ranges for both 64 and 128 MFCCs. Hence, we can infer that utilizing 30 MFCCs effectively conveys comparable information to 64 and 128.

Thus, based on the plots' observations, 30 coefficients seem to offer a promising balance between feature richness and noise resilience in capturing the acoustic characteristics of UAV drone audio data. This choice may facilitate more effective classification or analysis tasks leveraging MFCC-based features. However, further experimentation and validation could provide additional insights into the optimal configuration for feature extraction in this context.

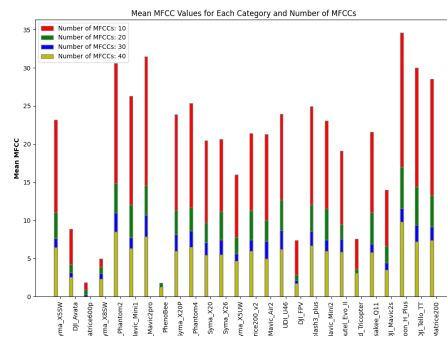


Figure 2: Range Comparison with 10, 20, 30, 40 Numbers of MFCCs.

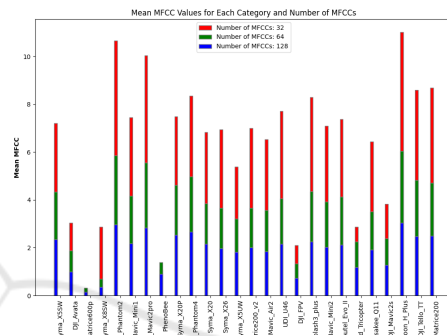


Figure 3: Range Comparison with 32, 64, 128 Numbers of MFCCs.

5 CONCLUSION

In conclusion, our study has provided valuable insights into the realm of UAV audio analysis, particularly focusing on the extraction of Mel-frequency cepstral coefficients (MFCCs) with varying numbers of coefficients. The comprehensive exploration of multiple UAV categories, each represented by distinct audio datasets, has offered a nuanced understanding of the impact of different MFCC configurations on mean and variance statistics. Our findings suggest that the selection of 30 coefficients in the MFCC extraction process may be a promising choice, exhibiting smaller variances across various UAV categories compared to configurations with fewer or more coefficients.

The visual representation of these results through plots showcasing mean MFCC values for different UAV categories and coefficients further aids in understanding the trends and variations. Notably, the use of sub-figures illustrates the impact of the number of coefficients (10, 20, 30, and 40) and the number of frames (32, 64, and 128) on mean MFCC values. The identified trends and potential optimal configurations contribute valuable knowledge for researchers and practitioners engaged in UAV audio analysis.

Looking ahead, future research directions include delving into advanced machine learning techniques for feature selection, exploring alternative audio feature extraction methods, and addressing real-time processing capabilities for UAV audio systems. Collaboration between researchers, industry stakeholders, and regulatory bodies is crucial to advancing the field and translating these findings into practical applications. While our study presents a substantial foundation, acknowledging the need for further validation across diverse UAV platforms, environmental conditions, and real-world deployment challenges is imperative for realizing the full potential of UAV audio analysis.

REFERENCES

- Altes, R. A. (1980). Detection, estimation, and classification with spectrograms. *The Journal of the Acoustical Society of America*, 67(4):1232–1246.
- Chen, Z. (2023). *PhenoBee: Drone-Based Robot for Advanced Field Proximal Phenotyping in Agriculture*. PhD thesis, Purdue University Graduate School.
- Kim, J., Lee, D., Kim, Y., Shin, H., Heo, Y., Wang, Y., and Matson, E. T. (2022). Deep learning based malicious drone detection using acoustic and image data. In *2022 Sixth IEEE International Conference on Robotic Computing (IRC)*, pages 91–92. IEEE.
- Le, N., Nguyen, N. V., and Dang, T. (2021). Real-time sound visualization via multidimensional clustering and projections. In *The 12th International Conference on Advances in Information Technology*, pages 1–6.
- Muda, L., Begam, M., and Elamvazuthi, I. (2010). Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques. *arXiv preprint arXiv:1003.4083*.
- Service, N. W. (2023). Estimating Wind Speed.
- Stylianou, Y. (2001). Applying the harmonic plus noise model in concatenative speech synthesis. *IEEE Transactions on speech and audio processing*, 9(1):21–29.
- Wang, M. Y., Chu, Z., Ku, I., Cho Smith, E., and Matson, E. T. (2024). A 15-category audio dataset for drones and an audio-based uav classification using machine learning. *International Journal of Semantic Computing*, pages 1–16.
- Wang, Y., Chu, Z., Ku, I., Smith, E. C., and Matson, E. T. (2022a). A large-scale uav audio dataset and audio-based uav classification using cnn. In *2022 Sixth IEEE International Conference on Robotic Computing (IRC)*, pages 186–189. IEEE.
- Wang, Y., Fagiani, F. E., Ho, K. E., and Matson, E. T. (2022b). A feature engineering focused system for acoustic uav payload detection. In *ICAART (3)*, pages 470–475.
- Zheng, F., Zhang, G., and Song, Z. (2001). Comparison of different implementations of mfcc. *Journal of Computer science and Technology*, 16:582–589.