

Michael Pisman - EECS

Audio Resampling for Bird Call CBIR

1. Introduction

Bird-call content-based image retrieval (CBIR) systems help identify bird species from recorded calls by analyzing audio features. These systems significantly depend on the quality of the audio processing methods used. Currently, many bird call recordings are captured at a high sampling rate of 44.1 kHz, even though the frequency of bird calls typically does not exceed 9 kHz. According to Nyquist's theorem, sampling audio at more than twice the highest frequency of the signal (in this case, 18 kHz) is unnecessary. Higher sampling rates increase file sizes and computational complexity without necessarily improving performance. This report investigates whether lowering the sampling rate (downsampling) from 44.1 kHz to 22.05 kHz can enhance the efficiency of bird-call CBIR systems without negatively impacting their accuracy.

2. Methods

2.1 Resampling Techniques

The study evaluated three resampling approaches to determine their impact on CBIR performance:

1. **Original (44.1 kHz)**: This method served as the baseline and involved no changes to the audio files.
2. **Naïve**: This method involved simple integer decimation, where the audio signal was downsampled by a factor of 2 (from 44.1 kHz to 22.05 kHz) without any sophisticated resampling techniques.
3. **Librosa**: The Librosa library's resampling method was utilized, which is commonly used in audio processing for machine learning.

Each method was applied systematically to the dataset:

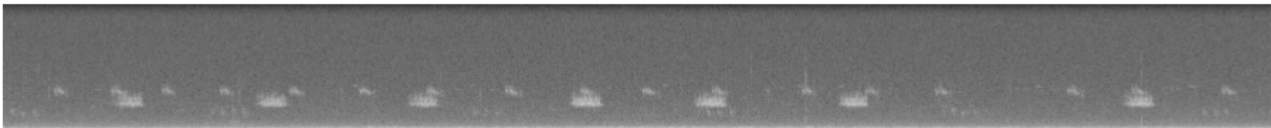
```
for recording in recording_list:
    if resample_method == 'librosa':
        # simple librosa load+resample
        x, sampling_rate = librosa.load(recording_filename, sr=target_sr)
    else:
        # none or naive (integer decimation)
        x, orig_sr = librosa.load(recording_filename, sr=None)
        if resample_method == 'naive':
            x = x[::2]
            sampling_rate = orig_sr // 2
        else:
            sampling_rate = orig_sr
```

2.2 Analysis of Spectrograms and Regions of Interest (ROI)

Spectrograms, which visually represent audio frequencies over time, were compared between the original and downsampled audio signals to ensure essential bird call characteristics were preserved. The ROI analysis targeted specific frequency bands relevant to bird calls, evaluating whether important information was retained after downsampling.

- **Spectrogram Comparison:** Visual inspection of spectrograms at 44.1 kHz vs. 22.05 kHz (Librosa) to confirm preservation of key harmonics.

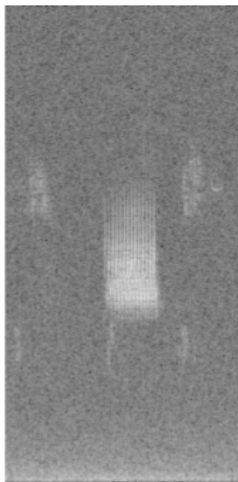
Spectrogram comparison



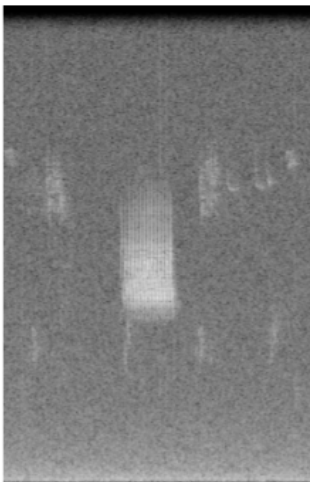
44.1 KHZ



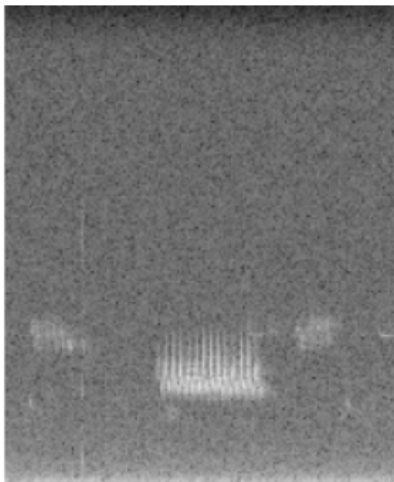
Librosa resampled 22.05 KHZ



Naive resampled 22.05



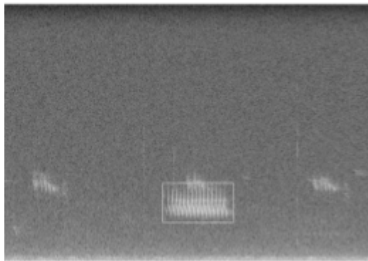
Librosa resampled 22.05



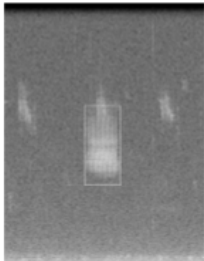
44.1 KHZ

- **ROI (Region-of-Interest) Extraction:** Computed ROIs for dominant frequency bands in both original and resampled signals to check for loss of salient features.

ROI Comparison



Original 44.1 KHZ



Librosa resampled 22.05 KHZ

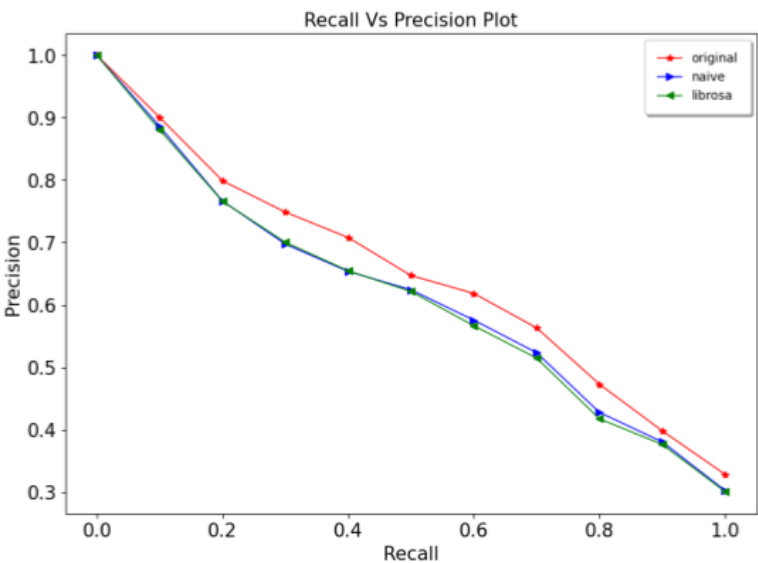
3. Experimental Results

3.1 Performance Comparison of Resampling Methods

Librosa vs Naive vs Original

Average Precision Values

- original = 0.653
- naive = 0.621
- librosa = 0.618

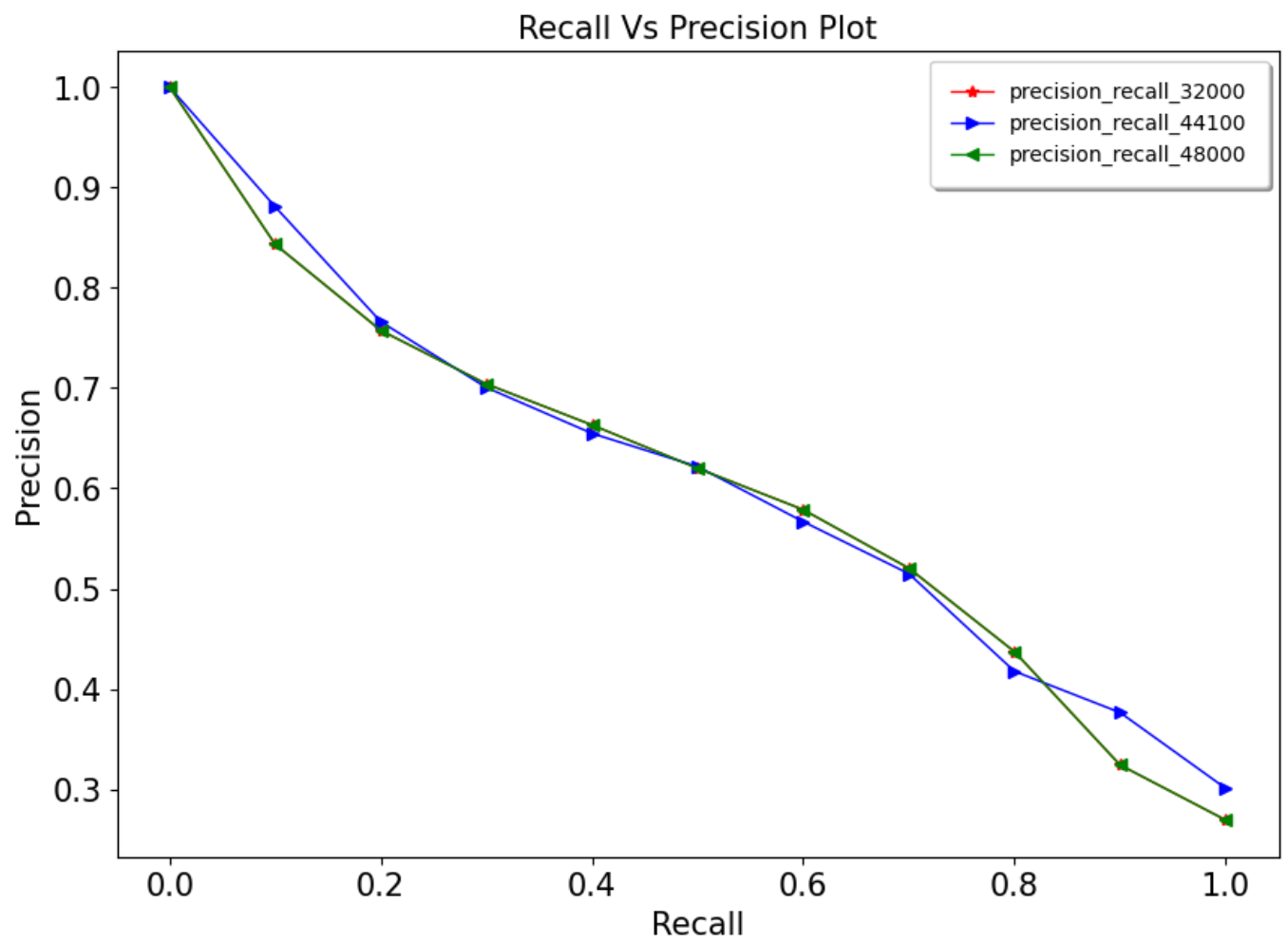


The CBIR system performance was assessed using Average Precision (AP), a measure that summarizes how accurately the system retrieves relevant bird calls.

Method	Average Precision
Original	0.653
Naïve	0.621
Librosa	0.618

The results indicated that downsampling to 22.05 kHz decreased performance by approximately 5%, suggesting that valuable acoustic information may be lost at lower sampling rates.

3.2 Impact of Varying Sampling Rates with Librosa



Different resampling rates were tested to explore the relationship between sampling rate and CBIR performance:

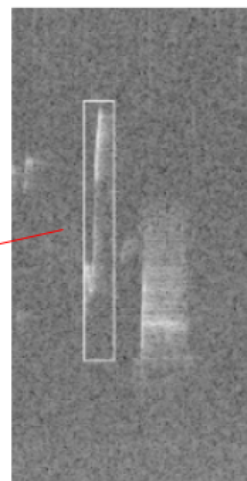
Sampling Rate	Average Precision
32000 Hz	0.610
41000 Hz	0.618
48000 Hz	0.610

No improvement in performance was observed with sampling rates higher than 22.05 kHz, suggesting limited benefit from increasing or moderately decreasing the original sampling rate.

3.3 Issues in ROI Calculation

Issues

- Could not calculate ROI for 00141.wav



- Could not calculate for custom FQs - 18k/32k Frequency

The ROI calculation encountered problems with certain audio files, indicating potential issues in the analysis pipeline:

- **File 00141.wav:** Calculation failed.

4. Discussion

Downsampling offers clear advantages, such as reduced file sizes and potentially faster processing. However, the observed decrease in average precision indicates that critical acoustic features might be compromised at lower sampling rates. Neither the Naïve nor Librosa methods effectively maintained performance compared to the original sampling rate.

5. Conclusion and Future Directions

This study found no performance benefit in reducing the sampling rate from 44.1 kHz to 22.05 kHz using standard resampling methods. Future research should explore advanced preprocessing techniques, including audio normalization and robust noise reduction, to optimize both storage efficiency and retrieval accuracy. Additionally, comprehensive testing across larger datasets and various preprocessing settings will be essential in refining CBIR systems for practical bird-call identification applications.