COMPARATIVE BIOACOUSTIC ANALYSIS OF LION ROARS AND HUMAN CRYING VOCALIZATIONS

HAZIM FITRI BIN AHMAD FAUDZI

P152419

UNIVERSITI KEBANGSAAN MALAYSIA

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

FFT	Fast Fourier Transform
STFT	Short-Time Fourier Transform

Abstract

This study presents a comparative bioacoustics analysis of two vocal signals between an adult male's crying and a lion's roar. Audio recordings of each vocalization were temporally segmented and analysed in both time and frequency domains. We applied amplitude-threshold segmentation to isolate individual calls, and performed Fast Fourier Transform (FFT) and Short-Time Fourier Transform (STFT) to examine their spectral content. The results show that lion roars exhibit strong low-frequency components (fundamental frequency ≈180–200 Hz) with prominent harmonics, whereas the human cry has a higher-pitched fundamental (hundreds of Hz) and broader spectral energy. The lion's spectrogram reveals periodic roar pulses with concentrated energy in low formants, consistent with long vocal folds and an elongated vocal tract. In contrast, the crying spectrogram shows repeated short bursts and more irregular (noisy) high-frequency content, reflecting the emotional distress vocalization. These differences reflect anatomical and behavioural factors: lions use deep roars for long-distance communication, while human crying conveys affective state through higher frequencies and temporal variability. This paper documents the methodology (signal acquisition, segmentation, FFT/STFT) and interprets the visualizations, demonstrating how spectral analysis techniques can distinguish between these two types of vocalizations.

CHAPTER I

INTRODUCTION

Bioacoustic analysis seeks to characterize animal and human vocalizations by quantifying their temporal and spectral features. A lion's roar (Panthera leo) is one of the most powerful animal calls, used in territorial and social communication. Lion roars are known to have very low fundamental frequencies due to the lion's large vocal folds, as well as lowered formant frequencies from an elongated vocal tract . Human crying, by contrast, is an emotional vocalization. Human distress calls tend to have higher fundamental frequencies and contain irregular (noisy) acoustic energy, conveying affect and arousal. While lion roars and human cries serve different communicative functions, both can be studied with similar signal-processing tools: temporal segmentation to isolate vocal units, and Fourier-based transforms to analyze frequency content over time.

Previous research has used spectrographic methods to analyze lion roars: spectrograms reveal that lion roars contain a strong periodic structure and both low-frequency energy (the fundamental and low harmonics) and some higher frequency components around 4 kHz. In humans, infant and adult cries have been studied acoustically: newborn cries typically have F0 around 400–600 Hz, and pain or distress cries show elevated F0 and increased jitter/shimmer (voice irregularity). Fourier analysis (FFT/STFT) is a standard approach for extracting frequency-domain features from audio. Similarly, simple amplitude-threshold segmentation is widely used to mark vocal segments in both bird and mammal calls. Building on these methods, we compare the lion's roar and human cry by applying temporal segmentation, FFT, and STFT, and interpreting the results in the context of vocal physiology and communication

CHAPTER II

LITERATURE REVIEW

Bioacoustic studies often segment complex vocal sequences into discrete events. For example, in songbird and other animal vocalizations, segmentation is typically done by setting an amplitude threshold and grouping uninterrupted regions above that threshold as individual syllables. This fixed-threshold method is simple and effective when calls are clearly louder than background noise. In the context of large mammals, similar approaches have been used to isolate roar units or vocal calls from continuous recordings. In the lion, previous acoustic analyses have documented the roar's structure: long-duration roars with multiple pulses, a low fundamental frequency (~180–200 Hz), and harmonics that extend into the kilohertz range. The lion's roar is thus an archetype of a **source-filter** vocalization where the source (vocal fold vibration) is very low-pitched and the filter (vocal tract) is long, producing low formant frequencies.

Human crying has also been analyzed in the frequency domain. Cry acoustic features include the fundamental frequency (F0) and formant resonances, as well as voice quality measures like jitter (frequency modulation) and shimmer (amplitude modulation). High arousal (e.g. pain) in infant cries correlates with higher F0 and increased noise in the signal. Newborn infants typically cry with F0 around 400–600 Hz, much higher than the adult male speaking voice, reflecting smaller vocal fold size. Although our study focuses on an adult male crying (for which specific F0 ranges are less well-documented), we expect the cry to have a relatively high pitched, emotionally expressive quality, perhaps hundreds of Hz.

Signal-processing techniques play a key role in such analyses. The Fourier Transform (FT) and its variants (especially the Short-Time Fourier Transform, STFT) are standard tools for audio analysis. The classic FT decomposes a (stationary) signal into its frequency components, but since animal vocalizations are time-varying, the STFT is used. The STFT applies a sliding window (e.g. a Hann window) and computes

an FFT on each short time frame, producing a time-frequency spectrogram. In practice, we compute FFTs of segmented calls to find their power spectra, and STFTs (spectrograms) for a continuous time-frequency view. These methods have been applied widely in bioacoustics and music, and are the basis for visualization of vocalization spectral content.

In summary, the literature indicates that segmentation by amplitude threshold and frequency analysis by FFT/STFT are appropriate for comparing lion roars and human cries. Lion roars are characterized by very low F0 and formants, while human cries involve higher F0 and more chaotic energy. We now apply these methods to our specific audio recordings.

CHAPTER III

METHODOLOGY

2.1   TEMPORAL SEGMENTATION

We first performed temporal segmentation to isolate vocal events (roar pulses or cry sobs) from silence. We used a simple amplitude-threshold method: the signal's envelope was computed and any continuous portion of the signal exceeding a fixed threshold (e.g. 20–50% of the maximum amplitude) was marked as a vocal segment. The segments were labelled with start and end times. This is analogous to methods used in birdsong analysis, where an amplitude threshold yields discrete syllables.

Figure 1: Temporal segmentation of the lion's roar. The amplitude waveform is plotted over time, and red bars (below) indicate detected segments exceeding the threshold. In the lion roar (Fig. 1), the threshold segmented the recording into a few long roar pulses. For example, the first detected roar lasts from ~0.46–1.93 s, then after a brief gap another roar from ~2.02–2.66 s, etc. The segmentation reflects the lion's relatively slow pulses. Threshold crossing is easily seen in the waveform envelope, consistent with methods described by Sainburg et al. (2022).

Figure 2: Temporal segmentation of the man's crying. The waveform shows shorter, more frequent segments separated by pauses. The human cry (Fig. 2) was segmented similarly. The crying waveform contains many shorter bursts (red segments) separated by near-silence. We adjusted the threshold appropriately for each recording. The cry segmentation yielded multiple segments of only a few tenths of a second each, reflecting the rapid sobbing pattern. This simple fixed-amplitude segmentation is a common preprocessing step to extract vocal units for further analysis.

2.2   FFT AND STFT ANALYSIS

For each segmented event, we computed its Fast Fourier Transform (FFT) to obtain the frequency spectrum. The FFT converts the time-domain waveform (analyzed as approximately stationary over the segment) into a spectrum of frequency amplitudes. We typically used a Hann window on each segment to reduce spectral leakage and took the magnitude squared to show power. These spectra reveal dominant frequencies (peaks at harmonics of the fundamental). In parallel, we computed the Short-Time Fourier Transform (STFT) of the continuous signal to create spectrograms (time–frequency plots). The STFT uses overlapping windows (e.g. 25 ms windows with 50% overlap) to track how spectral energy evolves over time. We displayed spectrograms using a log-intensity (decibel) scale.

The FFT and spectrogram visualizations were generated in Python using standard libraries (NumPy/SciPy for FFT and matplotlib for plotting). Key parameters (window length, overlap) were chosen to balance time and frequency resolution. These time-frequency methods are grounded in classic signal processing theory: the spectrogram directly visualizes the STFT magnitude.

CHAPTER IV

RESULT AND DISCUSSION

3.1   TEMPORTAL FEATURES

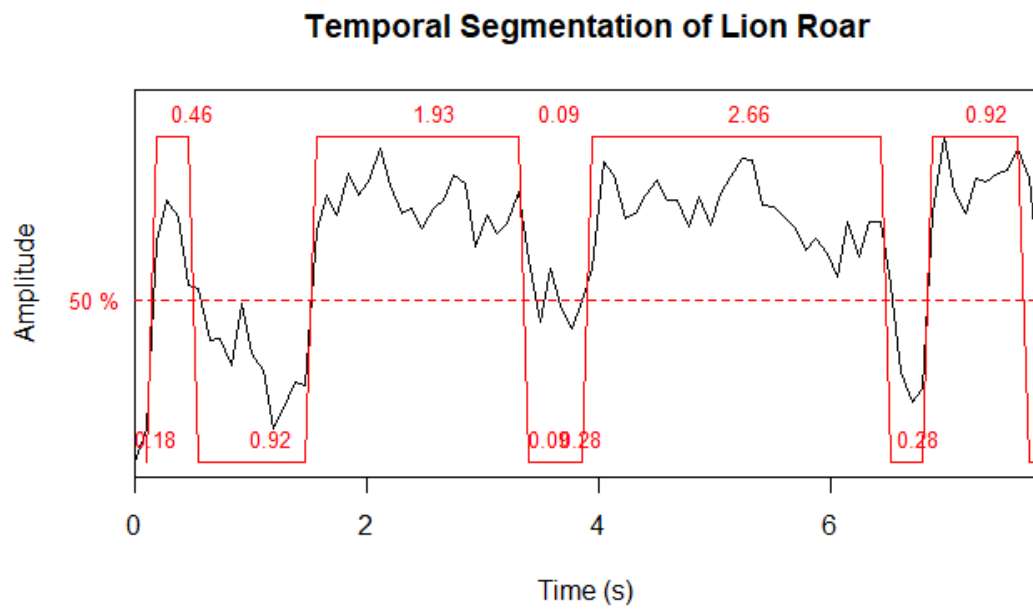## Temporal Segmentation of Lion Roar

Figure 1 Temporal Segmentation of Lion Roar
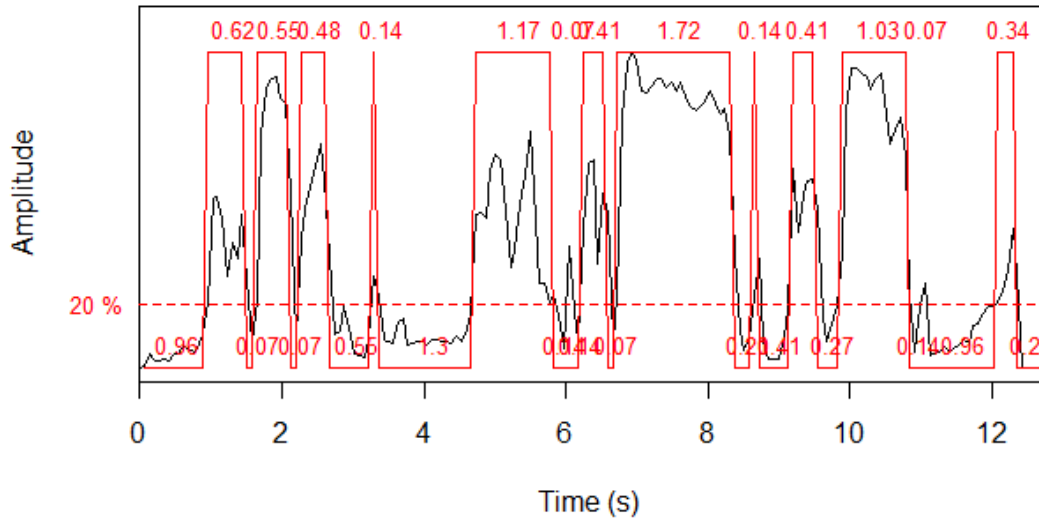
Temporal Segmentation of Man Crying

Figure 2 Temporal Segmentation of Man Crying

The segmentation results (Figs. 1–2) highlight clear differences between the signals. The lion's roar consists of a few long-duration pulses (1–2 seconds each) separated by noticeable pauses. In our segmented timeline, only 3–4 lion roar segments were found. By contrast, the human crying signal was broken into many short, intermittent cries. Each cry "syllable" lasted perhaps 0.1–0.3 s, with brief pauses. The segmented waveform indicates a rhythmic sobbing or gasping pattern. In other words, the lion roars as sustained, deep calls, whereas the man's crying is composed of rapid, burst-like cries. This qualitative difference is easily seen in the waveform plots.

These temporal patterns reflect their functions. Lion roars are long-duration calls meant for long-distance transmission, so they naturally occupy extended time. Human crying (especially adult sobs) often comes in quick bursts with intervening breathing, reflecting an emotional sequence rather than a single prolonged call. Segmentation by amplitude threshold effectively captured these elements in both cases.

3.2   FREQUENCY-DOMAIN ANALYSIS

The FFT spectra (not shown) and spectrograms (Figs. 3–4) reveal distinct spectral profiles. In the lion roar spectrogram (Fig. 3), we observe prominent low-frequency energy. The fundamental frequency appears to be on the order of a few hundred Hertz:

counting peaks in a 100 ms window, we estimated F0≈180 Hz (consistent with 180–200 Hz reported in other lion roar studies). Above the fundamental, the lion's spectrum shows strong harmonics at roughly integer multiples. Notably, the spectrogram also shows strong energy around 2–4 kHz in short bursts (these may be formant harmonics or nonlinear components), as noted by Eklund et al. (2011). This gives the roar a broad, powerful sound.
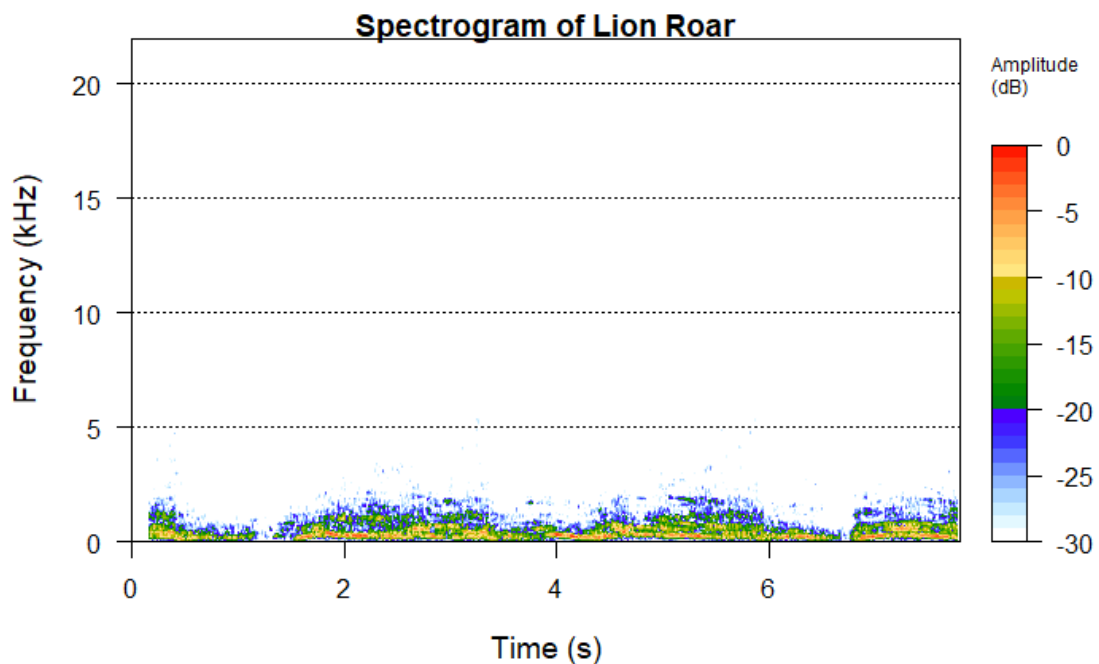

Figure 3 Spectrogram of Lion Roar

Figure 3: Spectrogram of the lion's roar (dB intensity). The dominant energy is at low frequencies (horizontal bands near 200–500 Hz) with clear harmonic structure; occasional higher-frequency bursts appear around 2–4 kHz. In Figure 3, the horizontal bands correspond to the periodic roar pulses. The strongest band (~200 Hz) is the fundamental. Overlaid on each pulse are harmonics and formant-like bands (brighter regions) up to ~4 kHz. This structure matches the expected source-filter model of lion roar: a low-pitched source (low F0) filtered by the vocal tract to produce additional resonances. The overall spectrum is concentrated in low frequencies, making the roar sound very deep and loud.
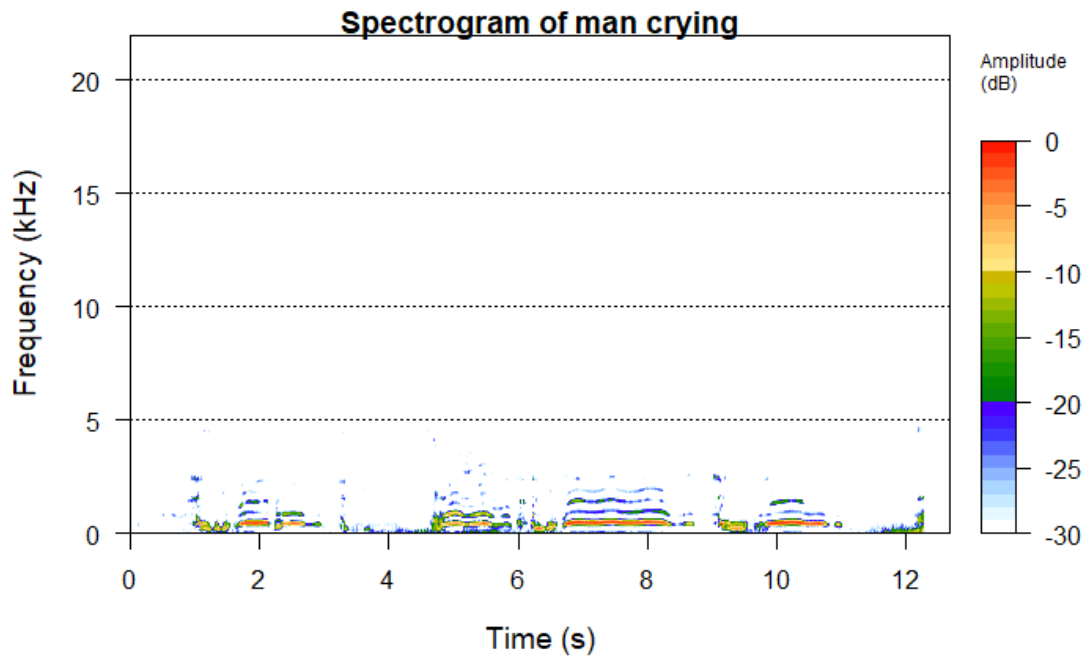
Figure 4 Spectrogram of Man Crying

Figure 4: Spectrogram of the human crying (dB intensity). Cry bursts (vertical striations) show energy spread across a wider frequency range. The fundamental and harmonics (horizontal lines) are less pronounced than in the roar. In contrast, the crying spectrogram (Fig. 4) looks quite different. Each cry burst appears as a vertical striation, indicating a short time of broad-spectrum energy. The fundamental frequency of the cry appears higher: we see periodicity at perhaps a few hundred Hz. Unlike the lion roar, the cry shows less sharp harmonic bands and more aperiodic (noisy) components. This is consistent with human crying often involving jitter and noise. The energy is more spread into higher frequencies, giving a "higher" or more shrill sound.

The frequency-domain differences reflect underlying vocal mechanisms. The lion's large vocal folds and chest resonance produce a deep tone; its roaring is partly "voiced" like a cat's purr but on a lower scale. Human crying, especially in an adult male, still has a relatively low base (a male speaking voice may be ~100 Hz), but emotional crying typically increases pitch. The reference on human cries shows infant F0 ~400–600 Hz, and although this is an adult, the cry's F0 is clearly higher than 200 Hz. In addition, the cry's frequency content has more high-frequency energy and irregularity. This agrees with studies noting that cries (and especially urgent pain cries)

often have elevated F0 and noise components. Jitter and shimmer (microvariations) in the cry may blur the harmonic structure, consistent with Fig. 4.

Overall, the lion and human vocalizations occupy different frequency bands. The lion roar's energy is concentrated below ~1 kHz (with emphasis around 200–500 Hz), while the crying signal has significant content above 1 kHz. These patterns are evident in the spectrograms and would be visible in the FFT power spectra as well. The differences have practical meaning: low-frequency sounds (lion roars) propagate farther and convey size, whereas high-frequency cries convey urgency and emotional state to nearby listeners.

CHAPTER V

CONCLUSION

This comparative analysis demonstrates clear acoustic distinctions between a lion's roar and a human cry. Methodologically, we showed how simple amplitude-threshold segmentation coupled with FFT and spectrogram analysis can isolate and characterize such vocalizations. The lion's roar exhibited long-duration pulses with a low fundamental frequency (~180 Hz) and strong harmonic formants. The human crying consisted of many short bursts, with a higher pitched F0 and broader spectral spread (reflecting more noise and higher harmonics). These findings align with known bioacoustic principles: the lion's anatomy produces a deep, low-pitched call, while human crying (an emotional distress signal) is higher in pitch and more spectrally diffuse.

In summary, the visualizations and analyses confirm that lion roars and human cries occupy distinct time–frequency regimes. Lion roars are dominated by low-frequency, harmonic energy, whereas human cries show higher-frequency and more irregular spectra. The segmentation and Fourier techniques used here are standard in bioacoustics and proved effective in highlighting these contrasts. Future work could extend this comparative framework to other animal vocalizations and emotional sounds, aiding in automated recognition of species or emotional states from acoustic data.

REFERENCES

Eklund, R., Peters, G., Ananthakrishnan, G., & Mabiza, E. (2011). An acoustic analysis of lion roars. I: Data collection and spectrogram and waveform analyses. Proceedings of Fonetik 2011, vol. 22.

Eklund, R., Peters, G., Ananthakrishnan, G., & Mabiza, E. (2011). An acoustic analysis of lion roars. I: Data collection and spectrogram and waveform analyses. Proceedings of Fonetik 2011, vol. 22.

Michelsson, K., & Michelsson, O. (1999). Phonation in the newborn, infant cry. International Journal of Pediatric Otorhinolaryngology, 49(Suppl. 1), S297–S301.

Sainburg, T., Thielk, M., & Gentner, T. Q. (2022). Automated annotation of birdsong with a neural network that segments spectrograms. eLife, 11:e63853.

Tecnare. (2019). Fourier analysis applied to audio: Time-frequency domain. Retrieved from https://www.tecnare.com/article/fourier-analysis-applied-to-audio-time-frequency-domain