Analyzing Polyphonic Vocal Patterns of Ross Seals

Julian Härtel[1], Ilse van Opzeeland[2], Volker Hohmann[1], Mia Wege[3], Thomas Brand[1]
[1] CvO Universität Oldenburg, Dep. Medizinische Physik und Akustik, Oldenburg, Germany
[2] Ocean Acoustics Lab, Alfred-Wegener Institute, Bremerhaven, Germany
[3] University of Canterbury, School of Earth and Environment, Canterbury, New Zealand

Introduction

The Ross seal (*Ommatophoca rossii* is a rarely observed, solitary species living in the Antarctic. Ross seals produce exceptionally complex vocalizations whose origin is not fully understood. It was hypothesized that these sounds are generated by an as-yet unexplained anatomical structure, possibly involving two coupled oscillators within the vocal tract [1]. This hypothesis so far remained unconfirmed, and the underlying mechanisms of Ross seal vocal production are still largely unknown.

The aim of this study is to develop and apply an acoustic model capable of analyzing and replicating the vocal patterns of the Ross seal. Our goal is to understand the signal structure, characterize the present nonlinear phenomena, and make assumptions about possible production mechanisms.

The dataset analyzed in this work is based on a rare aerial recording of a female Ross seal named Callas. An example audio segment can be accessed via [2] under the filename "99_original_audio.wav". It contains a section from a one-hour air recording with much vocal activity by Callas. The recording includes substantial wind and some engine noise but still reveals clearly audible vocalizations. Most notably, the sound contains polyphonic elements - two simultaneously present tonal components - highlighting the unusual and rich structure of Ross seal calls.

Vocalization Components

The vocalizations of the Ross seal can be decomposed into different acoustic complex tones and resulting combination tones that arise from their interaction. To better understand the structure of these calls, we analyzed their frequency spectrograms. Figure 1 shows a spectrogram of a typical sound sequence.

A low-frequency harmonic complex tone (LFHCT), labeled c, consists of a fundamental frequency c_0 and its integer-multiple overtones (harmionics) with frequencies c_1 to c_{n1}. These harmonics are evenly spaced in frequency, producing a highly structured and tonal sound. As this complex tone appears very constant throughout time it is referred to as c.

A high-frequency harmonic complex tone (HFHCT), referred to as m, is similarly composed of a fundamental frequency m_0 and its harmonics with frequencies m_1 to m_{n2}. Like c, it is a harmonic complex tone with regularly spaced frequency components. Unlike c, the frequencies of m are frequency-modulated over time. This modulation is clearly visible in Figure 1, for instance, between 1.7 s and 1.9 s, where m appears isolated and is easily identifiable.

Furthermore, the inharmonic complex tone (ICT), arises from interactions between c and m. It consists of distortion products, which are not integer multiples of any single fundamental frequency. These include both subtraction tones of the form $m_i - c_j$ (with $i = 0, ..., n1$ and $j = 0, ..., n2$), shown in dark-green, and addition tones $m_i + c_j$ (with $i = 0, ..., n1$ and $j = 0, ..., n2$), shown in black. The ICT is less structured and has a more dissonant spectral content. This inharmonicity produces a sound character comparable to that of some analog synthesizers, and is typical of nonlinear processes such as coupled oscillations in biological or physical systems.

Spectral Distribution

To analyze the spectral behavior of the vocal components, our model estimates the fundamental frequencies using a two-oscillator pattern-matching algorithm. This method identifies harmonic structures in the spectrogram and assigns them to either the low-frequency component c or the high-frequency component m, depending on their harmonic alignment. Figure 2 shows the analyzed fundamental frequencies of both components across a sequence, with $m0$ in red and $c0$ in blue.

For the spectral distribution we differentiate between two presence groups of m:

- m co-occurring with c ($m \cap c$), and
- m occurring alone ($m \setminus c$).

This distinction allows us to examine how the spectral characteristics of m change in the presence of c.

Figure 3 shows the spectral distribution of both components across all analyzed sequences, separately for the two presence groups. $c0$ is always around 300 Hz. When m appears without c, the median of $m0$ is approximately 750 Hz. In contrast, when m co-occurs with c, the median of $m0$ rises to around 1.5 kHz, indicating a systematic influence of c on m's frequency behavior and suggesting possible acoustic coupling or modulation effects between the two components.

Acoustic Phenomena

Now that we identified a polyphonic structure in Ross seal vocalizations, also acoustic phenomena were found in the calls from Callas. Among these are the presence of subharmonics and abrupt jumps in the fundamental frequency, both of which suggest the involvement of complex, nonlinear mechanisms. It is further assumed that deterministic chaos may occur in connection with these frequency instabilities. Despite their irregular appearance, these phenomena follow identifiable patterns. In this section, we take a closer look at these acoustic fea-

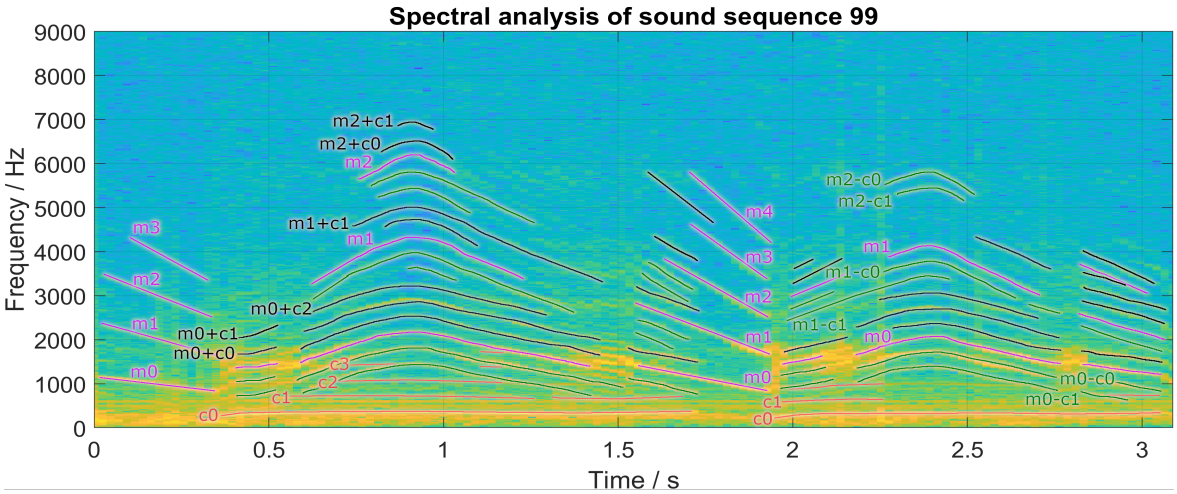

Figure 1: Spectral analysis of a sound sequence with highlighted sound components and combination tones

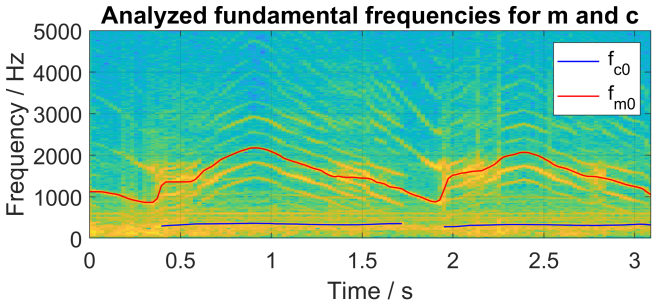

Figure 2: Fundamental frequencies of the detected sound components analyzed by the model

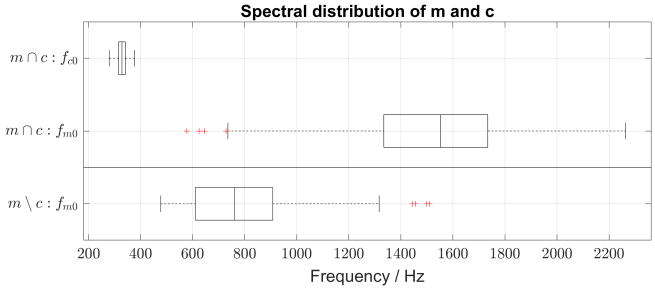

Figure 3: Spectral distribution of the fundamental frequencies of the sound components across all sound sequences

tures to better understand their structure and potential origins.

Subharmonics

Subharmonics are an evidence of a nonlinear vocal production mechanism. They emerge when multiple oscillators interact in a coupled or irregular manner, often as a result of changes in airflow, tension, or other vocal tract parameters.

In nearly every sequence analyzed, subharmonics were observed. Figure 4 highlights a representative example: a roughly 200 ms segment marked by a red rectangle, in which clear subharmonic patterns appear below the primary harmonic components.

To automatically detect such segments, our model uses a Harmonic-to-Subharmonic Ratio (HSR). This measure

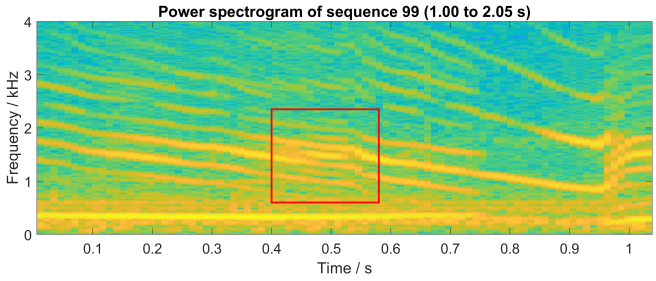

Figure 4: Power spectrogram of a sound sequence, red rectangle indicates an approximately 200 ms area with subharmonics

quantifies the ratio between the summed amplitudes at harmonic positions and those at half-integer positions (where subharmonics are expected). The HSR is calculated using the following equation 1:

$$\mathrm{HSR} = \frac{\sum_{k=1}^{n} A_{kf_0}}{\sum_{k=1}^{n} A_{(k-\frac{1}{2})f_0}} \tag{1}$$

Where, A_{kf_0} denotes the amplitude at the k-th harmonic of the fundamental frequency f_0, and $A_{(k-\frac{1}{2})f_0}$ refers to the amplitude at the corresponding subharmonic positions. A lower HSR indicates a higher presence of subharmonics relative to regular harmonics. When a certain threshold level is reached, the analysis window is considered to have detected subharmonics.

We also examined the spectral distribution of the component m_0 within the subharmonic segments. As shown in figure 5, $m0$ in these subharmonic areas clusters strongly around 1450 Hz, with most values within a 200 Hz range. This suggests that subharmonic generation is linked to a specific frequency-dependent mechanism in the modulated component m.

Fundamental Frequency Jumps

Another acoustic phenomenon observed in the Ross seal vocalizations is the occurrence of abrupt shifts in the fundamental frequency of the modulated high-frequency component m, referred to as fundamental frequency jumps. These events represent sudden transitions in $m0$. Figure 6 illustrates this phenomenon: the fundamental

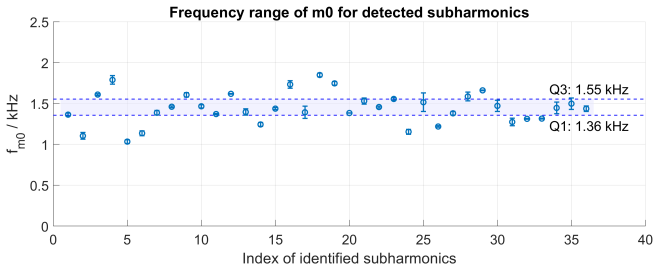

Figure 5: Spectral distribution of the fundamental frequency of sound component m ($m0$) in detected subharmonic areas

frequency of m was manually analyzed across the sequence. In regions where m appears without c, such as around 2 seconds, $m0$ is clearly identifiable. When we follow this trace backward and compare it with the forward trace beginning at the preceding clearly identifiable $m0$, we see that the two traces do not merge into each other. This indicates that the tracking of the $m0$ component is ambiguous and that the $m0$ component can be hidden between the other components of m. Small frequency steps are visible, e.g. around 1.6 seconds and 4.25 seconds. But these steps do not connect seamlessly, indicating discontinuity. The sequence reveals a repetitive structure, suggesting the presence of an underlying pattern.

Notably, these frequency jumps often occur nearly to the onset or offset of oscillator c, suggesting a possible interaction between the two components. Regions close to the jumps such as around 0.5 s and 3 s often exhibit evidence for deterministic chaos. That and the above-mentioned frequency jumps makes the tracking of $m0$ particularly challenging and we couldn't solve this problem, so far.

These findings point to a complex and dynamic vocal mechanism. While further investigation is necessary, the observed behavior may be comparable to register transitions in human singing, where shifts between different vibratory regimes (e.g. chest to head voice) result in discontinuous frequency jumps.

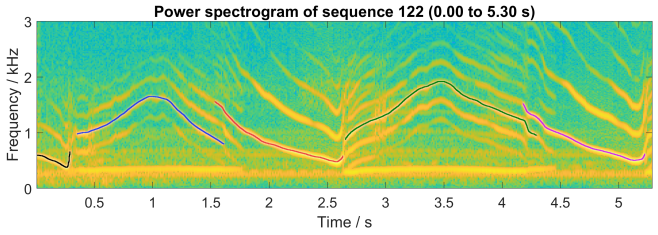

Figure 6: Power spectrogram of a sound sequence, with highlighted manually analyzed fundamental frequencies by tracking for- and backwards

Modelling

To further analyze and interpret the vocalizations, we developed a signal model that serves two main purposes. First, a matched-filtering algorithm is applied to isolate the vocal components from the original recording, effectively suppressing background noise. Second, the model includes a synthetic signal generation function, allowing us to recreate Ross seal vocalizations based on hypothesized production mechanisms. By reproducing the observed signal characteristics, the generator model provides a valuable tool to evaluate assumptions about how these complex sounds are produced.

Matched Filtering

The matched filtering component of our model is designed to highlight the vocal signal. The improvement can be described by the estimated noise suppression (NS). It operates by applying a binary mask to the FFT bins of the signal, with a frequency resolution of 11 Hz per bin. This selective filtering approach retains only the spectral regions associated with the modeled vocal components.

Figure 7 provides an overview of the matched filtering process. The top plot shows the original, unprocessed signal and the middle plot displays the result after matched filtering. The signal components are clearly enhanced, while interfering noises such as wind, engine hum, and background speech are effectively removed. The bottom panel illustrates the estimated noise suppression level over time. Across all analyzed sequences, the mean estimated noise suppression was -13 dB. The resulting enhanced signal can be heard via [2] under the filename "99_filtered_audio.wav".

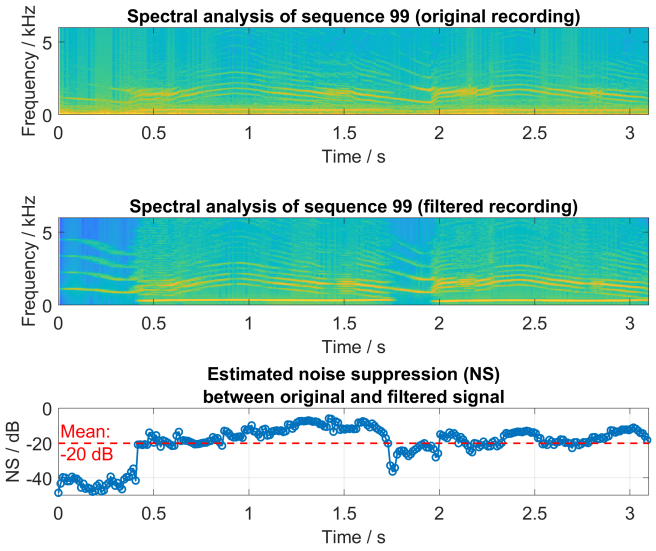

Figure 7: Overview of the matched filtering algorithm using a sound sequence. Top: Original recorded signal; Middle: Filtered signal; Bottom: Estimated noise suppression over the analysis frames

Synthetic Signal Generation

To validate the assumed vocal production mechanism, we used our model to generate synthetic Ross seal vocalizations. These synthetic signals confirm the model's accuracy in reproducing key acoustic features, such as the harmonic structure, modulation patterns, combination tones and the presence of subharmonics. The generation process is based on a two-oscillator interaction, described by the following equation 2:

$$s_k(t) = c_k(t) \cdot (1 + m_k(t)) + m_k(t) \qquad (2)$$

Here, $s_k(t)$ represents the signal within a single analysis frame k. The carrier signal $c_k(t)$, corresponding to

component c, is amplitude-modulated by the modulator $m_k(t)$, corresponding to component m, through the term $c_k(t) \cdot (1 + m_k(t))$. Additionally, m is added directly via the $+m_k(t)$ term. This structure models a plausible biological mechanism in which a low-frequency oscillator modulates a high-frequency tone through airflow variations. This process is shown in the block diagram in figure 8. This type of modulation suits with the limited anatomical knowledge of Ross seal vocal production [1].

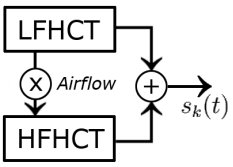

Figure 8: Block diagram of the generator model for synthetic signal generation

The complete synthetic signal $S(t)$ is then constructed by summing the overlapping, windowed signal frames as:

$$S(t) = \sum_k s_k(t) \cdot w_k(t), \tag{3}$$

where $w_k(t)$ is a Hann-window applied to each frame, and the signals are recombined using the overlap-add technique. An example of the synthetic signal is shown in figure 9 and can be accessed via [2] under the filename "99_synthesized_audio.wav".

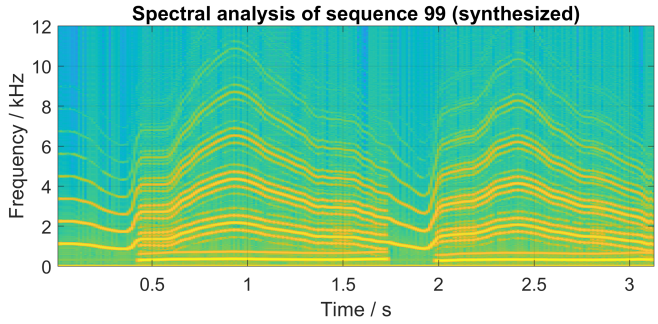

Figure 9: Spectral analysis of the synthesized signal for the filtered sequence from figure 7

Comparative Analysis

The complex acoustic features found in Ross seal vocalizations, like subharmonics and deterministic chaos, are also found in other species. Similar nonlinear phenomena have been observed in other mammals as well as in musical instruments, suggesting shared underlying mechanisms across biological and artificial sound production systems.

Subharmonics are a typical result of coupled oscillations within the vocal tract, where multiple vibratory sources interact to produce frequency components at fractional intervals of a fundamental. These have been documented in marine mammals, notably in whales [4], and also in woodwind instruments, where they arise due to complex airflow dynamics and resonances [5].

Deterministic chaos arises when a control parameter, such as subglottal pressure or airflow, exceeds a critical threshold. The result is a vocal pattern that is structured yet unpredictable, displaying broadband energy and irregular fluctuations. Such chaotic dynamics have been observed in whale calls and in rhesus monkey vocalizations [4].

Conclusion

This study has demonstrated that Ross seal vocalizations shows a distinct polyphonic structure, consisting of a LFHCT, a HFHCT, and resulting ICT. Within this structure, we identified key nonlinear phenomena such as subharmonics, combination tones, and deterministic chaos, all of which point to a complex and dynamic vocal production mechanism.

Our acoustic model successfully analyzed and reproduced these components and phenomena. Due to its flexible design and numerous input parameters, the model is highly adaptable to different vocal types, making it a tool for investigating animal vocalizations and their underlying mechanisms.

Future work will focus on improving signal clarity through contrast-based pre-filtering, which may enhance frequency tracking performance.

Annotations

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project ID 352015383 - SFB 1330 A1.

References

[1] Judith E. King. Some aspects of the anatomy of the Ross seal. Techn. Ber. British Museum of Natural History, 1969

[2] Julian Härtel. Audio examples for analysis, filtering and synthesization with a developed two-oscillator pattern-matcher model for polyphonic vocal patterns of Ross seals. March 2025. doi: 10.5281/zenodo.15092436. url: https://doi.org/10.5281/zenodo.15092436.

[3] Justin Hofman, Alamy Stock Photo. Ein Ross-Seal (Ommatophoca rossii) auf einem Eisberg in der Antarktis.

[4] W. Tecumseh Fitch, Jürgen Neubauer und Hanspeter Herzel. „Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production“. In: Animal Behaviour 63 (2002), S. 407–418. doi: 10.1006/anbe.2001.1912.

[5] Douglas H. Keefe. „Acoustical wave propagation in cylindrical ducts: Transmission line parameter approximations for isothermal and thermoviscous boundary conditions“. In: The Journal of the Acoustical Society of America 75.5 (1991), S. 1754–1763. doi: 10 . 1121 / 1 . 398711.

[6] Lionel Velut, Christophe Vergez und Joël Gilbert. „Measurements and time-domain simulations of multiphonics in the trombone“. In: The Journal of the Acoustical Society of America 140.4 (2016), S. 2876-2887. doi: 10.1121/1.4964634.