Vowels and Diphthongs in Sperm Whales

Gašper Beguš*[1,5], Ronald L. Sprouse[1], Andrej Leban[2,6], Miles Silva[3,6], and Shane Gero[4,5,6]

[1]Department of Linguistics, University of California, Berkeley, United States
[2]Department of Statistics, University of Michigan, Ann Arbor, United States
[3]Department of Brain and Cognitive Sciences, MIT, Cambridge, United States
[4]Department of Biology, Carleton University, Ottawa, Ontario, Canada
[5]The Dominica Sperm Whale Project, Roseau, Dominica
[6]Project CETI, New York, United States

January 17, 2024

Abstract

Sperm whale vocalizations are among the most intriguing communication systems in the animal kingdom. Traditionally, sperm whale *codas*, or groups of clicks, have been primarily analyzed in terms of the number of clicks and their inter-click timing. This paper brings a new dimension to the study of sperm whale communication — spectral properties — and argues that spectral properties are likely actively controlled by whales and potentially meaningful in this communication system. We uncover previously unobserved recurrent spectral patterns that are orthogonal to the traditionally analyzed properties. We present a visualization technique that allows us to describe several previously unobserved patterns. We introduce the source-filter analysis of sperm whale codas and argue that they are on many levels analogous to human vowels and diphthongs: vowel duration and pitch correspond to the number of clicks and their timing (traditional coda types), while spectral properties of clicks correspond to formants in human vowels. We identify two recurrent and discrete coda-level spectral patterns that appear across individual sperm whales: the *a*-coda vowel and *i*-coda vowel. Both coda vowels are possible on different traditional coda types. Our discovery thus suggests that spectral (filter) properties are independent of the source properties (number of pulses and timing). We also show that sperm whales have diphthongal patterns on individual codas: rising, falling, rising-falling and falling-rising formant patterns are observed. Finally, we control for whale movement and present several pieces of evidence suggesting that the observed patterns are not artifacts, but are actively controlled by sperm whales. We also show that the two coda vowels (the *a*-vowel and *i*-vowel) are actively exchanged by sperm whales in dialogues. These uncovered patterns suggest that spectral properties have the potential to add to the communicative complexity of codas independent of the traditionally analyzed properties.

1 Introduction

How sperm whales (*Physeter macrocephalus*) might encode information into their communication system is one of the most intriguing questions in animal research. Sperm whales communicate with click vocalizations that they group into units called *codas* (Worthington and Schevill, 1957; Watkins and Schevill, 1977; Whitehead and Weilgart, 1991; Whitehead, 2003). Clicks in such codas

[*] First and corresponding author: Gašper Beguš (begus@berkeley.edu).

Table 1: Parallels in human vowels and sperm whale codas.

	human vowels	**sperm whale codas**
source (vocal folds/phonic lips)	vowel duration F0	number of clicks inter-click interval (ICI)
filter (vocal tract/spermaceti organ)	height & backness (F1 & F2) formant trajectories	coda vowel (*a* vs. *i*) diphthongal codas

are acoustically distinguishable from echolocation clicks (Madsen et al., 2002b,a; Møhl et al., 2003). Coda vocalizations are likely culturally learned (Rendell et al., 2012; Rendell and Whitehead, 2003), and different culturally defined clans feature different coda types (Andreas et al., 2022; Amano et al., 2014; Amorim et al., 2020; Gero et al., 2016b; Huijser et al., 2020; Rendell and Whitehead, 2003). Several coda types have been identified based on two primary characteristics: the number of clicks and the timing between clicks (Weilgart and Whitehead, 1993).

The communication system of sperm whales has thus far been analyzed primarily as a discrete system. Nearly all research on sperm whale codas focuses primarily on the number of clicks and their timing, but not their spectral properties. The two properties — the number of clicks and their relative timing — are used to classify codas into groups traditionally called coda types (Weilgart and Whitehead, 1993; Whitehead, 2003). Recently, Sharma et al. (2023) argued that these traditionally analyzed properties can be further analyzed as four distinct timing/click number features ("rhythm, tempo, rubato, and ornamentation") that can be independently combined. Because our proposal focuses on a different dimension — spectral properties rather than timing/click number features — we use the traditional coda type notation throughout the paper.

This paper proposes that another dimension — spectral properties of clicks — are potentially meaningful in the communication system of sperm whales. We argue that recurrent spectral patterns can be observed across individual whales. We describe these patterns, suggesting that sperm whales actively control spectral properties which have the potential to carry meaning.

For this purpose, we invoke the source-filter theory from human speech production (Fant, 1971; Stevens, 1998). Sperm whales vocalize by sending air through phonic lips which vibrate and result in clicks (Huggenberger et al., 2016; Madsen et al., 2002b, 2023). Under our proposed view, whale clicks are equivalent to the pulses of vocal folds in human speech production. In other words, we treat clicks as the source and the sperm whales' resonant body (the nasal complex, including the spermaceti organ and distal air sack) as the filter that modulates resonant frequencies. We show that codas feature two discrete patterns in their resonant frequencies that match formants in human speech. We show that these patterns are recurrent across different whales.

Under this approach, the *coda types* correspond to human vowel duration (the number of pulses/clicks) and pitch (F0) in human speech. Recurrent spectral properties observed in whale codas (that we call *coda vowels*) correspond to formant frequencies, i.e. vowel identity in human speech (Table 1). Pitch (F0) in humans is to a large degree orthogonal to vowel quality. For example, in tonal languages such as Mandarin Chinese, syllables with vowels [a] or [i] can feature all four tones (Duanmu, 2007). We argue that spectral patterns in sperm whales (called *coda vowels*) are similarly orthogonal to the coda type: all described vocalic and diphthongal patterns are possible on different coda types.

We observe at least two distinct spectral patterns in sperm whales codas across studied whales: (i) codas with a single pronounced spectral peak below 10kHz (at approximately 5800 Hz) and (ii) codas with two spectral peaks below 10kHz (at approximately 3700 Hz and 6200 Hz). We term the first pattern the "*a*-vowel" and the latter the "*i*-vowel". Figure 1 illustrates the distinction

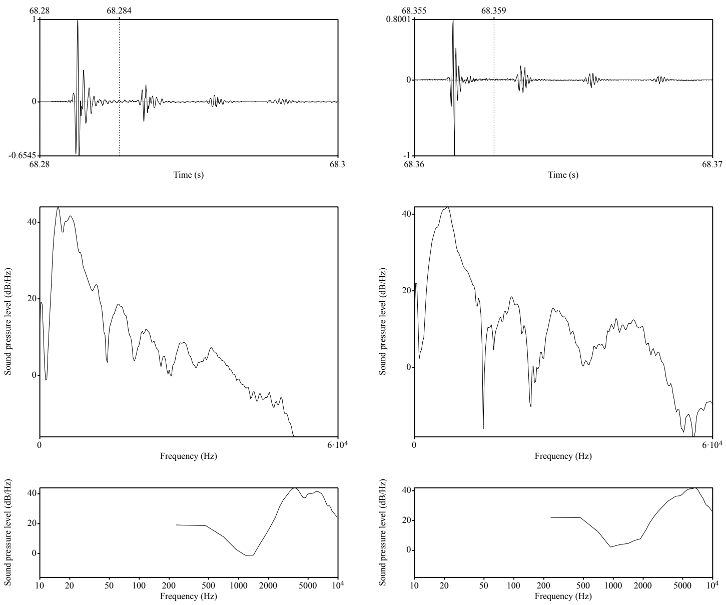

Figure 1: Waveforms and spectra (4ms window) of the first pulse of the first clicks of two codas, (**left**) one with the characteristic *i*-vowel pattern (coda 6911) and (**right**) a coda (coda 6912) with the characteristic *a*-vowel pattern. Both codas were produced consecutively by the same whale, 'Pinchy' (whale #5560), during the same bout. The bottom two figures show two log spectra in the 10-10,000 Hz range.

between two vowels with two single-pulse spectra. These patterns are discrete: a coda is either of type *a* or type *i*, which means that all clicks in a coda are of the same type. This coda-level pattern is a crucial new proposal. The two types are easily distinguishable by a simple spectral analysis. While it is possible that further more fine-grained distinction exist, all codas that we observed are either of the *a*-type of the *i*-type.

While there are substantial similarities in human vowels and sperm whale codas, there is one important difference between the two: human vowels are phonemic, which means they distinguish meaning. No referential meaning relationship has yet been established for sperm whale codas. While it is possible or even likely that codas do distinguish or carry meaning, this has not yet been established. In this respect, the coda vowels are an observed pattern that does not have an established function yet. Despite this difference, conceptualizing sperm whale codas as vowels is justified by the common mechanisms in both the production and the acoustics of the two systems as well as highly practical for representational purposes (e.g. for transcribing dialogues in Tables 2 and 4).

Acoustic properties with spectral analyses of sperm whale clicks have been studied before (Moore et al., 1993; Thode et al., 2002; Whitehead, 2003), but no recurrent coda-level patterns that might be meaningful had been observed in the spectral domain. That some clicks can have multiple peaks which shift with depth has been observed in Thode et al. (2002)(Goold and Jones 1995; Lin et al. 2017 also report one or two peaks, Huggenberger et al. 2016 reports one). Thode et al. (2002) conclude that these shifting peaks might be at least partly controlled by the whales, but the multiple peaks that shift with depth were analyzed primarily as a feature of movement (how deep the whales dived), which is contrary to what we claim here. Thode et al. (2002) only reports acoustic properties of echolocation clicks which serve a different purpose from codas. None of the prior works report any patterns at the coda level that would not be movement-related and recurrent at the coda-level (rather than at the click-level) across whales.

Other odontocetes seem to primarily modulate the fundamental frequency (F0), i.e. the frequency of their phonic lip vibration when communicating. This is primarily obvious in whistle-like vocalizations of dolphins, orcas, or beluga whales (Filatova et al., 2015; Panova et al., 2016, 2019). Even pulsed call vocalizations seemed to primarily be modulated in the fundamental frequency (F0) (Sportelli, 2019; Wellard et al., 2020). The same is true for vocalizations that Panova et al. (2016) call "vowel"-like based on an acoustic impression. The work on orcas, dolphins, or beluga whales does not describe spectral patterns that would be independent of the fundamental frequency (=formants that originate from manipulating the filter part in the articulatory process) as is the case in sperm whales. Vocalizations of odontocetes are not usually analyzed in terms of source filter theory or in terms of equivalents between human vowels and whale vocalizations except when studying acoustic correlates of whale size (Samarra and Miller, 2006) or based on impressionistic similarity (Panova et al., 2016). To our knowledge, no reports exist that odontocetes modulate resonant frequencies which would result in simple discrete patterns that are discoverable without dimensionality reduction techniques and orthogonal in terms of the *source* features (such as F0) and *filter* features (such as formant modulation).

Humpback whale (*Megaptera novaeangliae*) vocalizations have been compared to human vowels, but the described patterns require clustering techniques (Pines, 2019). Additionally, the primary function of humpback song is likely mating or male collaboration (Herman, 2017; Darling et al., 2006; Whitehead and Rendell, 2015), which stands in contrast to sperm whale and other odontocetes where vocalizations likely have non-mating social communicative function (Whitehead and Weilgart, 1991). To be sure, other animals are able to modulate formant frequencies both in their vocalizations in the wild (Fitch, 2000; Fitch and Hauser, 2003; Stansbury and Janik, 2019; Stoeger et al., 2012) and when imitating human vocalizations (such as gray seals, parrots, or elephants; Stansbury and Janik 2019; Patterson and Pepperberg 1998; Klatt and Stefanski 2005; Stoeger et al. 2012). Compared to the observed pattern in sperm whales, formant modulations in other species seem less discrete and orthogonal (between the source and filter features).

Establishing the sperm whale coda vowel patterns from underwater hydrophones that are not placed directly on the whale through a tag would pose a substantial challenge, as underwater acoustics can distort the signal substantially and the apparent patterns might be attributed to spectral disturbances. For this reason, we only analyze data from hydrophones placed directly on the vocalizing whale. We have analyzed one of the largest datasets of tagged sperm whales recorded in Dominica from 2014 to 2018 by The Dominica Sperm Whale Project (Gero et al., 2014).

That spectral properties might be meaningful in sperm whale communication system has been recently proposed in Beguš et al. (2023). A fiwGAN model (Beguš, 2021) was trained to imitate sperm whale codas and learn to embed information into the learned vocalizations. Building on Beguš (2020), an introspection technique was developed in Beguš et al. (2023) to test for what meaningful properties a model learns from unknown data and applied to sperm whale communication. We show that the model learned properties previously considered meaningful: the number of clicks and their inter-click intervals. Additionally, however, we uncovered that the network learned acoustic properties as meaningful: spectral mean and acoustic regularity. These two properties have not been previously considered as meaningful.

While the interpretability technique in Beguš et al. (2023) pointed to candidate properties learned as meaningful by the models, here we explicitly uncover and describe the spectral patterns. In other words, AI models have been shown to be useful as hypothesis space reduction techniques (Andreas et al., 2022; Jumper et al., 2021; Stokes et al., 2020; Davies et al., 2021) that can offer clues to researchers, but do not yet provide the explicit mechanisms. In our case, the interpretability

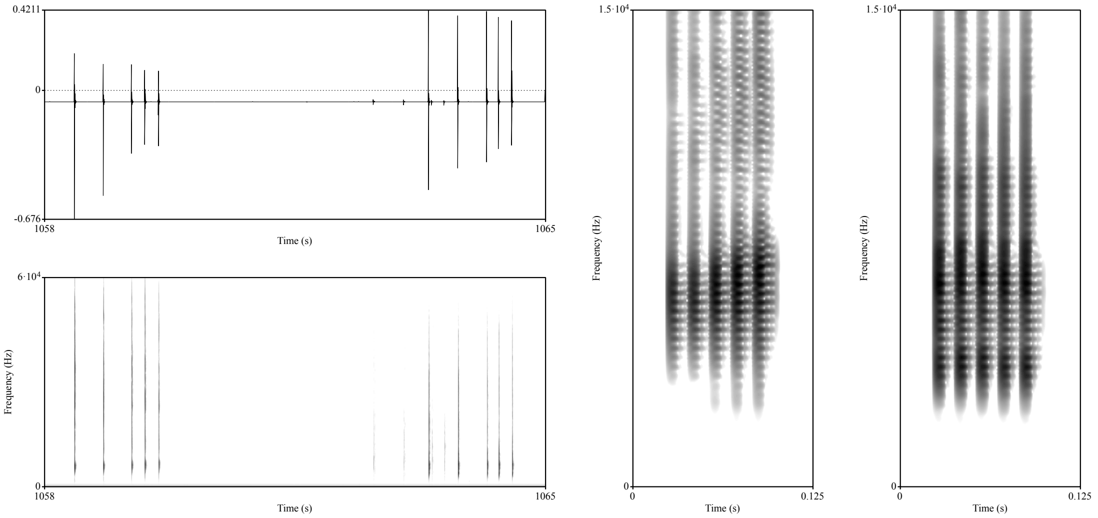

Figure 2: (**left**) A waveform and a spectrogram (0-60 kHz) of two consecutive focal codas by Pinchy (6931 and 6932) when timing is not removed. (**right**) Spectrograms (0-15 kHz) of the same two codas using our visualization approach. This visualization uncovers a clear pattern: the first coda features a single formant (is of the *a*-vowel type), the second has two formants (is of the *i*-vowel type). Similar click-level visualizations have been made in Thode et al. (2002); Lin et al. (2017), but the differences were ascribed to depth in Thode et al. (2002).

technique uncovered that an AI agent considered spectral properties to be meaningful. This present work is thus a post-hoc explicit analysis of a clue provided by an AI agent and an interpretability technique. We explicitly describe several acoustic patterns that we observe during the acoustic analysis and argue that they might be meaningful in sperm whale vocalizations. Findings of this paper thus suggest that the approach proposed in Beguš et al. (2023) where a fiwGAN is trained on the unknown data and the *causal disentanglement with extreme values* (CDEV; Beguš et al. 2023) technique is applied on learned representations can be successful for uncovering meaningful properties in unknown communication systems.

2 Data and methods

The data stems from following and recording social units of female and immature sperm whales along the western coast of the Island of Dominica between 2005 and 2018 (details in the Appendix). This resulted in a dataset of 3948 codas that were annotated as above maintaining temporal ordering and associating speaker identities. To explore the hypothesis that spectral properties carry meaning in sperm whales, the analysis needs to be as controlled as possible for potential underwater acoustic effects or sperm whale movement. All analyzed data is from a single channel (either right or left) from hydrophones on tags limited to only vocalizations of the focal, tagged, whales (except when explicitly stated otherwise with regards to dialogue data where we also analyze the non-focal whale who is engaging in a dialogue with the focal whale). This means that the hydrophone is equidistant from the whale throughout the recording period. Most of our analysis focuses on the band between 0 and 15 kHz. We removed the DC offset and peak-normalized each click to value 1 or -1.

We propose a new visualization technique for the codas which removes the inter-click timing and only represents spectral trends throughout the codas. To achieve this, we extract 15ms from the beginning of each annotated click and concatenate all clicks from the same coda produced by a given whale. This visualization technique allows us to remove the effects of ICI and visually observe patterns in spectral properties that would be obscured if timing was left in the visualization. For example, Figure 2 illustrates the effect of our visualization technique on the ability to find patterns in spectral structure of coda. Two codas (6931 and 6932) by tagged whale, Pinchy, are visualized in waveforms and spectrograms with timing between clicks preserved on the left. On the right of the figure are two spectrograms where timing has been removed and the spectrograms are limited to the 0-15kHz interval.

All spectrograms in figures have a window length of 10ms and are produced in the Praat software (Boersma and Weenink, 2015). Our acoustically analyzed corpus includes 7,022 coda clicks from 1,344 codas.

We checked for the effect of the tag/hydrophone placement on the spectral patterns by examining a subset of codas (n=17) that were recorded by two tags, one on the focal whale and one on non-focal whale (see Section 3.3). We further examined the effect of depth on the spectal patterns by using the Dtag's pressure sensor and contrasted this with the formant frequencies in the codas.

3 Results

3.1 Vowels

The visualization technique described in Section 2 allows us to more effectively observe spectral patterns in coda vocalizations. Spectral analysis of the entire corpus reveal an easily observable and recurrent pattern: sperm whale interchange codas with one and two formants below the 10kHz range. Figure 3 illustrates a slice of the entire corpus with clearly distinguishable one- and two-formant codas.

To estimate mean values of peaks, we analyzed 55 focal *a*-vowel clicks and 35 focal *i*-vowel clicks as transcribed in Table 2. The spectrum of each click was analyzed by *scipy*'s *find_peaks* function (Virtanen et al., 2020), which found as many as four peaks in the range of 1000-12000 Hz. The frequencies of the two tallest spectral peaks were selected and labelled in order as F1 (the lower frequency peak) and F2 (the higher frequency peak). If only one peak was found its frequency was recorded as F1 and F2 was recorded as missing. In 5 out of 55 clicks the *a*-vowel was mistakenly identified as having two peaks. All *i*-vowel codas were correctly labeled. The mean F1 peak in the *a*-vowel coda is 5787 Hz (SD = 793 Hz). The mean F1 peak of the *i*-vowel is 3683 Hz (SD = 166 Hz); the mean F2 is 6174 Hz (SD = 311 Hz).

We term the two observed patterns where the coda's spectral properties have a single or two formants as *coda vowels*. This is by analogy to human vowels which differ in their formant frequencies. Codas with a single formant will be referred as the *a*-vowel coda, codas with two formants as the *i*-vowel codas.

Both coda vowels appear on various coda types according to the traditional classification, although some coda types are rarer in our data. For example, Figure 3 shows an interchange between *a*- and *i*-vowel codas on the 1+1+3 coda type. Figure 4 shows three cases of 5R2 coda types with the *i*-vowel pattern by Pinchy and the same coda type with the *a*-vowel pattern by Fork. The exchange between the two vowels can happen within the same bout (a bout is a set of codas where timing between two codas is not greater than 10s). In other words, nearly all coda types in these figures are of the 1+1+3 type, yet they show a lot of variability between the *a* and the *i* codas.

That coda vowels are likely actively controlled by whales is also suggested by the fact that we have not observed mixing between one and the other spectral pattern within codas. In other words, we observe that if a coda is of the *a*-type, all clicks will feature a single formant and vice

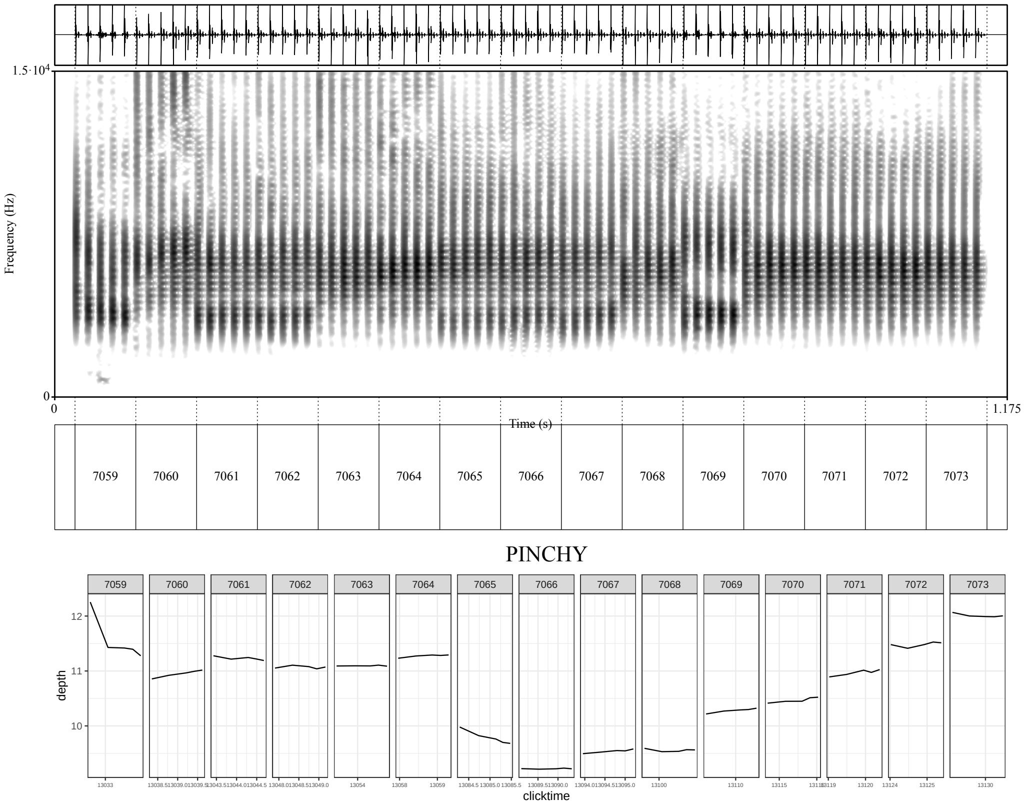

Figure 3: (**top**) Waveforms and spectrograms 0–15,000 (single right channel) of 15 codas from two consecutive bouts by focal Pinchy with timing removed and all clicks peak-normalized. The second bout starts with coda 7065. All codas are of the 1+1+3 type. **(bottom)** Depth values (in m) for each coda.

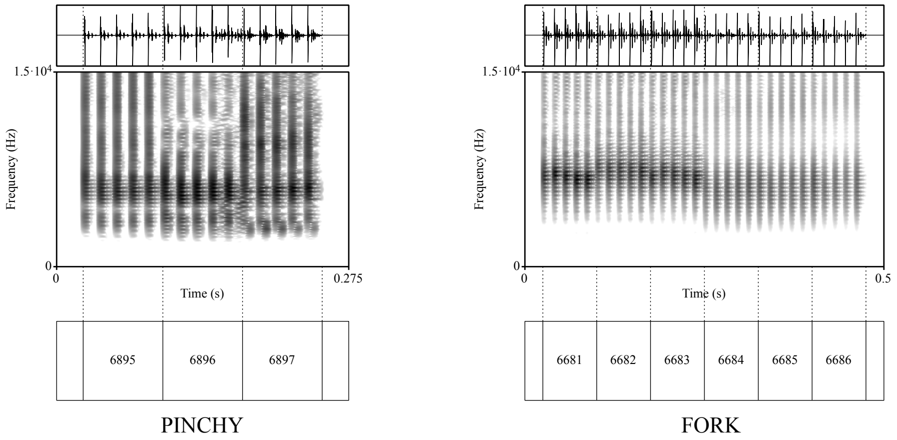

Figure 4: Waveforms and spectrograms 0–15,000 Hz (based on single, right channel) displaying the *i*-vowel pattern (left) by Pinchy and the *a*-vowel pattern (right) by 'Fork' (whale #5151) of 5R2 codas recorded on each of their tags respectively.

versa: if a coda is of the *i*-type, all clicks have two formants. If the distinction was random or an automatic consequence of some external factor, we would expect mixing between *a*-like and *i*-like clicks. Figures 3, 11, 12, and 13 that the two coda vowels completely coincide with coda boundaries.

The observed pattern is also not an idiosyncratic property of a single whale, but a recurrent feature across whales. Figures 3, 11, 12, and 13 show the *a*-/*i*- interchange pattern on four different whales within the same bout and the same 1+1+3 coda type. Additionally, Sam (whale #5726) shows the *a*/*i* pattern on the 5R1 coda (Figure 14) and Jocasta (whale #5987) features some *i*-coda vowels in addition to *a*-coda vowels. The *a*/*i* pattern is thus present in at least six whales. There are total 14 focal whales in our data (Sally, whale #6052, is excluded because she was recorded together with TBB, whale #5759), but some whales have only a few focal codas recorded. Lady Oracle (whale #5712) has only 7, Nalgene (whale #5133) only 13 codas, Soursop (whale #5719) a total of 34. In these whales, the relatively low number of recorded codas makes it difficult to observe the coda vowel pattern. Atwood (whale #5586), Fork, Pinchy, and TBB represent over 2/3 of all analyzed clicks in the corpus; the *a*-/*i*-coda vowels are found in all these well-represented whales. In some whales the *i*-vowel is missing, but it is difficult to establish whether these whales lack the *i*-vowel pattern completely or is this just an accidental gap due to data sparsity/noisy recordings in some of these whales.

3.2 Diphthongs

In addition to the two patterns (*a*- vs. *i*-coda vowel), we observe that some codas feature upwards, downwards or other types of trajectories in formant frequencies. Diphthongal patterns are rarer in our data, compared to level codas but they are useful for understanding how whale movement or hydrophone placement might affect formant trajectories of codas. We argue that diphthongs present additional evidence that the observed spectral patterns are controlled by whales and are not artifacts as we show that formant trajectories do not automatically follow from whale movement or depth. While clear diphthongal pattern exist, it is currently difficult to quantify what counts as a diphthong. This is parallel to human diphthongs that can vary in the amount of formant trajectories substantially (Lindau et al., 1990).

Figure 5 shows three diphthongal and two level codas by Fork. The first four codas are consecutive which means that Fork first vocalizes two codas with a clear upward trajectory, the next coda with a clear downward trajectory and a level coda right after that. The first four codas (three diphthongal and two level) are of the *a*-vowel type. The last coda is not consecutive, but added here to represent Fork's level *i*-vowel. The formant trajectories below the 10kHz range on spectrograms in Figure 5 illustrate these diphthongal patterns. The figure also illustrates that formant trajectory can be both upward or downward on the same coda type, and is thus likely not influenced by the number of clicks or their timing. The movement and depth values show no clear differences in movemement or depth between level and diphthongal codas.

Figure 6 even more strongly illustrates that depth, head, pitch, and roll do not play a crucial role in formation of diphthongs. Figure 6 features three codas from another whale, TBB: one coda with a level trajectory and two codas with substantial falling or rising trajectories in formant frequencies. The figure on the left shows a spectrogram for an *a*-vowel coda 8624 with a constant formant frequency across the coda. The second and third spectrograms show a substantial trajectory in formant frequencies: falling and rising. The movement data, however, point in the opposite direction: the level coda features more head, pitch, and roll movement than the diphthong codas. The change in depth appears comparable across the three codas. It appears that the formant frequency trajectories are independent of movement of whales. It appears that these diphthongs (falling or rising trajectories in formant frequencies) are actively controlled by sperm whale's articulators and

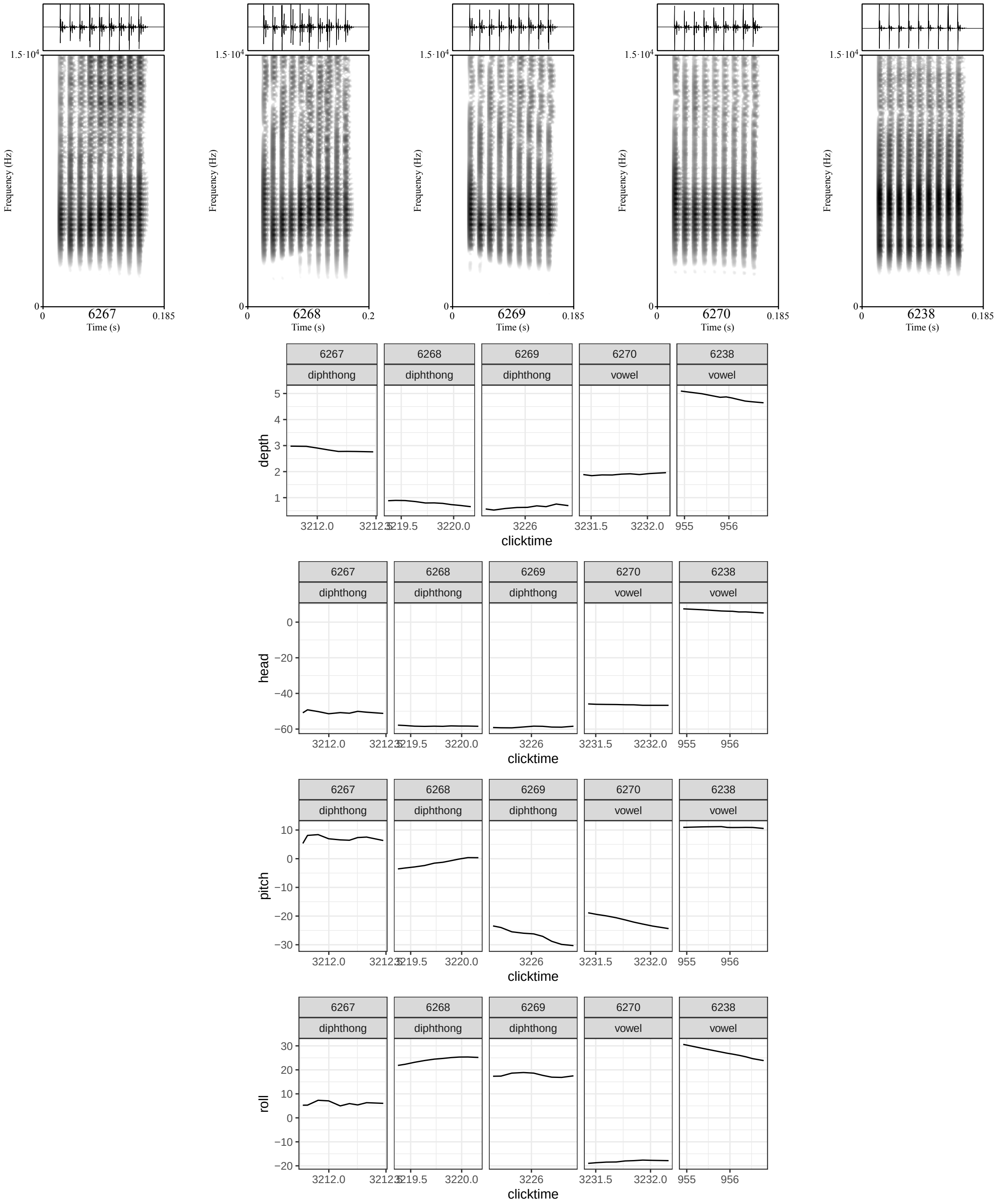

Figure 5: (top) Waveforms and spectrograms 0–15,000 Hz (single, right channel) from four consecutive codas and an additional coda (rightmost) recorded from a tag on Fork. The first threee figures from the left illustrate dipthongal patterns; the two figures on the right illustrate level formants. The codas are of the 9i, 10i, 9i, 9i, and 9i type. All codas are from the same bout. **(bottom)** Corresponding position data for the observed codas: depth (in m), head, pitch, and roll (all in degrees).

are not an automatic consequence of movement.

The described diphthongal patterns where the formant is in a rising or a falling trajectory tend to be featured on codas with more clicks. This is parallel to human diphthongs where many diphthongal patterns have two targets and are thus generally longer (i.e. vowels with more glottal pulses; Lindau et al. 1990). Figure 7 features several diphthongs on long codas. In addition to the rising and falling pattern, we also observe a rising-falling trajectory in Tweak's (whale #6070) diphthongs.

Occasionally, diphthongal patterns occur on shorter codas too (Figure 8). It appears that sperm whales can control both the distinction between *a*-vowel and *i*-vowel codas and the distinction between flat and diphthongal formant patterns. The majority of diphthongal patterns are, however, observed on the *a*-type coda vowels.

Another type of diphthong that our analysis uncovers across several whales are *a*-vowel codas in which the first click has a substantially higher formant frequency than the rest of the coda. These types of codas are given in Figure 10. This pattern is independently present in Atwood, Fork, Fruit Salad, and Jocasta (see also Figures 15 and 7 for more examples of this type). This pattern is so clear that it is recoverable in one instance even from a non-focal whale.

3.3 Controlling for depth, movement, and hydrophone placement

In principle, the observed spectral properties of sperm whale clicks could be attributed to hydrophone placement or whale movement. Here, we present evidence suggesting that the patterns we describe are not artifacts, but are actively controlled by sperm whales. Hydrophone placement and whale movement (towards or away from the hydrophone) have been shown to modulate spectral information in odontocetes (Miller, 2002; Branstetter et al., 2012). All our observations are made on focal whales only, i.e. the hydrophone is largely equidistant from the source throughout the recording for each whale. We show that head, pitch, and roll have no effect on spectral peaks. Depth has a weak correlation with spectral peak, but we show that all described patterns are possible at different depths/depth trajectories and provide several pieces of evidence suggesting that vowel and diphthong patterns are not crucially affected by depth. Additionally, a simultaneous recording of a set of clicks from two whales suggest that hydrophone placement does not crucially alter our described patterns.

To test whether diphthongal patterns or spectral patterns in general are an automatic consequence of whale movement (i.e. correlate with whale movements) or are controlled by whales irrespective of their movement patterns, we perform a correlation test between depth, head, pitch, and roll recording from tags and spectral peaks in their vocalizations. Welch spectra were extracted with the *scipy.signal.welch* function and correlated to movement data using the Pearson correlation test. For all correlation analysis, we removed Sally's and Jocasta's codas because Sally's recording includes TBB's codas and because Jocasta's codas were partly misaligned. Head, pitch, and roll have no correlations with spectral peaks: ($r = 0.004$ for head, $r = -0.002$ for pitch, $r = 0.006$ for roll). Depth has a weak correlation ($r = 0.236$) with spectral peak, but as will be shown throughout this paper, both diphthongal patterns and the *a*-/*i*-coda vowels are made at various depths or depth changes. For example, diphthongs like the ones produced by Fork in Figure 5 or codas with substantially higher first click (Figures 10 and 15) can happen close to the water surface with very little depth movement of the whale.

Because depth is the only parameter that shows some correlation with spectral peak, we primarily focus on showing that all described patterns can be produced at different depths. We also show other parameters (head, pitch, roll) do not affect the patterns (Figure 5 and 6), but to a lesser degree than depth because the correlation tests show no correlations with these other parameters.

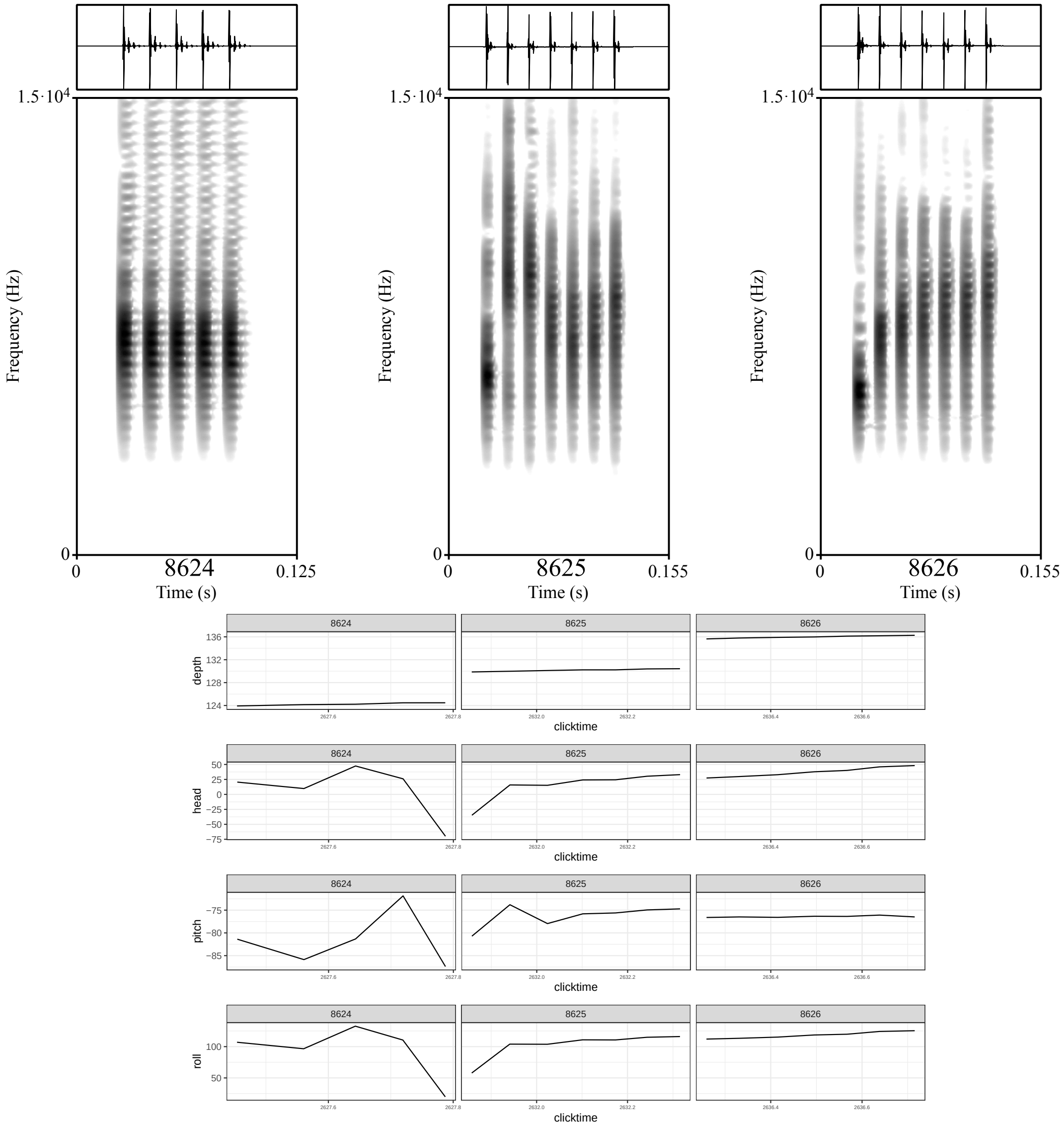

Figure 6: (top) Waveforms and spectrograms 0–15,000 Hz (single right channel) from three consecutive codas recorded from a tag on TBB. The codas are of the 5R1, 7i, and 7i type. All codas are from the same bout. **(bottom)** Corresponding position data from the for the observed codas: depth (in m), head, pitch, and roll (all in degrees).

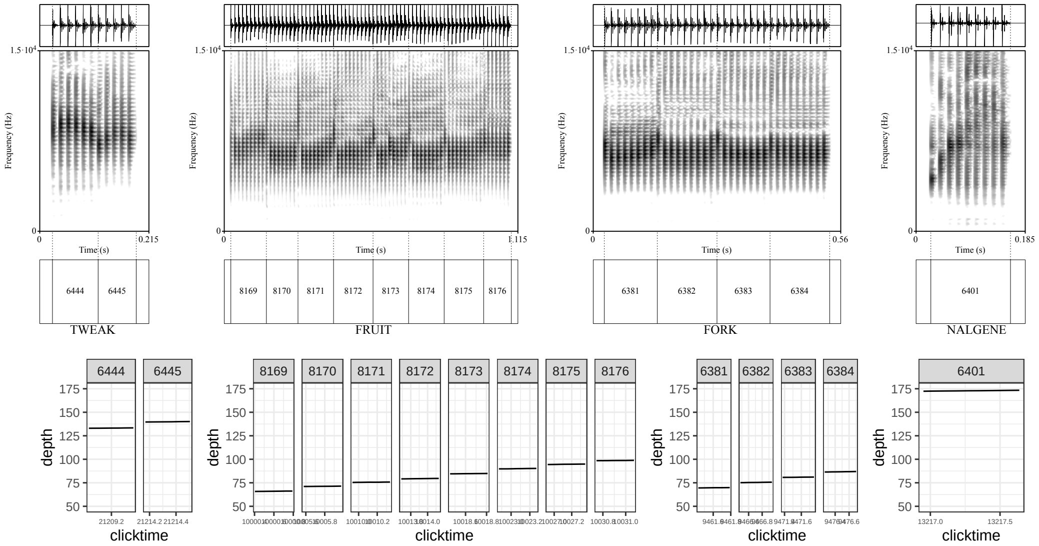

Figure 7: (**top**) Waveforms and spectrograms (0–15,000 Hz; single right channel) of diphthongs with substantial formant trajectories from four different whales. (**bottom**) Depth values (in m) for each corresponding coda.

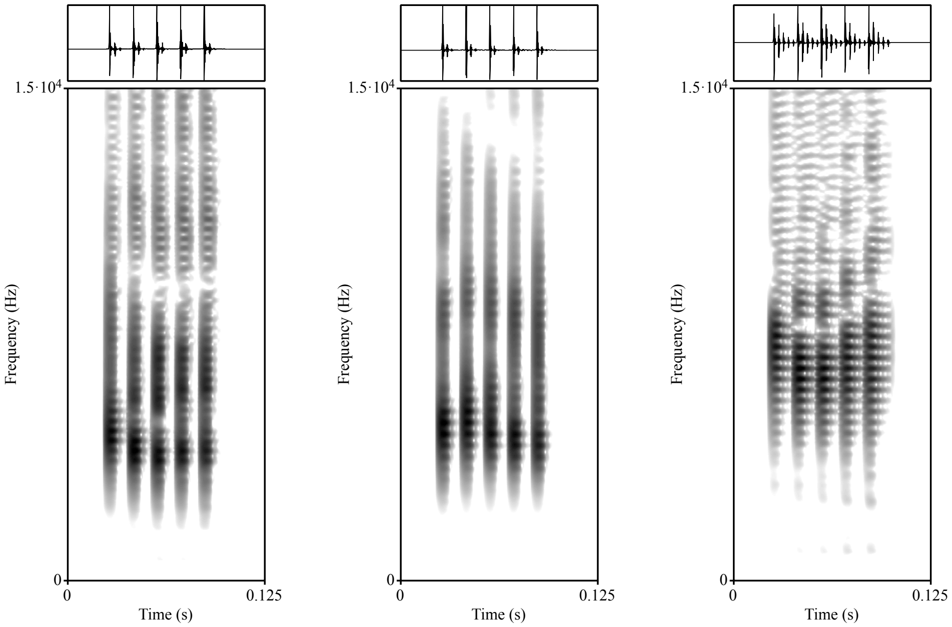

Figure 8: Waveforms and spectrograms (0–15,000 Hz; single, right channel) of (**left**) coda 6638 of the 1+1+3 type by Fork with the *i*-vowel pattern; (**center**) coda 8570 of the 1+1+3 type by TBB with the *i*-vowel pattern; and (**right**) coda 6156 of the 5R1 type by Atwood with the *a*-vowel pattern.

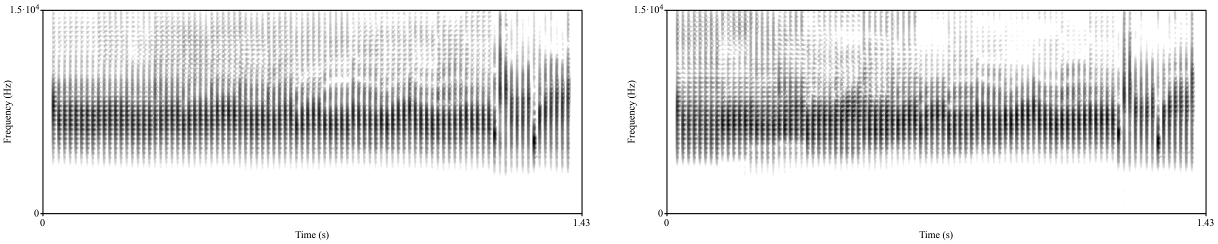

Figure 9: Spectrograms (0–15,000 Hz; single right channel) of 17 codas uttered by TBB (left) and recorded on Sally's tag (right). The left channel captures the same patterns as well, meaning that the relative position of the whale likely does not crucially affect spectral patterns when whales are in close proximity either.

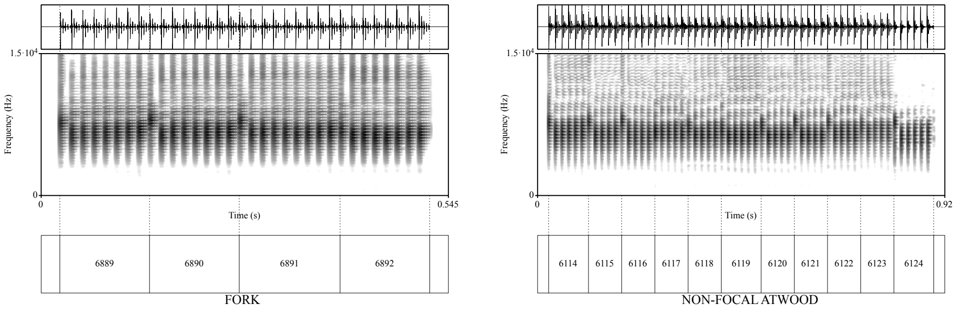

Figure 10: Waveforms and spectrograms (0–15,000, single right channel) of codas from Fork and non-focal Atwood (single left channel) where the first click has a substantially higher peak formant frequency.

Additionally, if depth crucially affected formant frequencies, we would expect to see unidirectional diphthongs when depth increases or decreases. The diphthongal patterns in Tweak (the rising-falling pattern in Figure 7) and especially the pattern observed in TBB point to the contrary. For example, Figure 6 shows codas of TBB when she was diving. The penult coda has a substantial falling pattern, while the ultimate coda has a substantial rising pattern (the antepenult coda has the flat pattern), despite TBB moderately diving throughout these codas.

We have evidence suggesting that microphone placement on the whale, and therefore relative orientation of their nasal complex to the hydrophone, does not critically affect the described spectral properties either. The best piece of evidence suggesting this comes from a recording session when two whales, Sally and TBB, were both tagged. Figure 9 illustrates a recording of a set of codas produced by TBB, recorded both on TBB's and Sally's tag. Judging by amplitude of the recording, TBB and Sally were not as close in the first 4 codas in Figure 9, but became very close in the second half of the spectrogram in Figure 9. Sally's tag recording of TBB's vocalization illustrate practically the exact same patterns as the TBB's actual tagged data when TBB was uttering these codas.

3.4 Dialogue

Sperm whales are known to engage in dialogue with each other, in which individuals exchange codas, of varying types, in sequence (Schulz et al., 2008). This often results in overlapping codas, where two whales produce codas at the same time, frequently by making codas of the same type within milliseconds (the so-called echo-codas; Weilgart 1990; Whitehead 2003). The overlapping codas have been studied in detail, primarily from the perspective of timing and coda types: it appears that whales start converging on coda types and timing during dialogues. To our knowledge, no spectral analyses of dialogues have been conducted so far.

Our new approach is also useful for analyzing dialogues. If the non-focal whale is in close proximity to the focal whale, we can observe the interchange between *a*- and *i*-codas in non-focal whales as well. Figures 16 and 17 feature codas from two bouts of Pinchy as well as a non-focal whale when Pinchy was tagged and the whales were engaging in coda exchange. A clear pattern of exchange between *a*- and *i*-vowel codas are observed in both the focal Pinchy and the corresponding non-focal whale.

Table 2 analyzes these two bouts of dialogue between Pinchy and her interlocutor in terms of the *a*- and *i*-vowel. We annotate them for the two proposed type (annotations can be verified by the spectrograms in Figures 16 and 17). We observe a distinct pattern where the whales engage in an interchange of *a*- and *i*-codas. Another similar dialogue is observed on two non-focal whales (Table 4 and Figure 18).

If only traditional coda types were analyzed, this dialogue would appear as a simple repetition of coda types. With the analysis of vowels on codas, it appears that the whales are exchanging the two different elements (*a* and *i*). When two whales are vocalizing simultaneously (echo codas), they can use the same coda vowel (*a*-*a* or *i*-*i*) or different coda vowels (*a*-*i* or *i*-*a*; see Table 2).

The fact that *a*- and *i*-codas are visible from non-tagged whales also means that the whales can not only produce but also hear (and likely perceive) this difference. In other words, hydrophones capture most of the observed spectral properties on both focal and non-focal whales (Figure 9). It is likely that these properties get distorted at distance, but the effects of distance are currently difficult to estimate. Even if underwater acoustics distorts the signal at distance substantially, it is likely that the difference between the *a*- and *i*-coda vowels have multiple spectral cues beyond the spectral peaks that we describe, which would facilitate perception of the two codas. The hearing ability of sperm whales is strong in the frequency range where we observe the patterns, and it can exceed 30 kHz (Schmidt et al., 2018). The *a*-/*i*- difference is perceptible even to human listeners.

Table 2: A dialogue between Pinchy and a non-focal whale from two bouts (divided by a horizontal line). If two codas are less than 1 s apart, they transcribed in the same line. In the second bout, two codas (unmarked here) are probably from a third whale.

Pinchy #	Non-focal #	Pinchy Type	Non-focal Type	Pinchy Vowel	Non-focal Vowel
6893		9i		i	
6894		9i		i	
6895	6915	5R2	5R2	i	a
6896	6916	5R2	5R2	i	i
6897		5R2		i	
6898		1+1+3		i	
	6917		1+1+3		a
	6918		1+1+3		i
6899	6919	1+1+3	1+1+3	i	i
	6920		1+1+3		i
6900	6921	1+1+3	1+1+3	a	i
6901	6922	1+1+3	1+1+3	a	i
6902	6923	1+1+3	1+1+3	a	i
6903	6924	1+1+3	1+1+3	a	a
6904		1+1+3		a	
	6925		1+1+3	i	
	6926		1+1+3	i	
6905		1+1+3		i	
6906		1+1+3		i	
6907		1+1+3		i	
6908		1+1+3		i	
6909		1+1+3		i	
6910	6927	1+1+3	1+1+3	i	i
6911	6928	1+1+3	1+1+3	i	i
6912		1+1+3		a	
6913	6929	1+1+3	1+1+3	a	i
6914		1+1+3		i	
6930	6938	1+1+3	1+1+3	a	a
6931		1+1+3		a	
6932		1+1+3		i	
6933		1+1+3		i	
6934		1+1+3		i	
6935		1+1+3		i	
6936	6939	1+1+3	1+1+3	i	i
6937		1+1+3		a	
6940	6959	1+1+3	1+1+3	a	i
6941	6960	1+1+3	1+1+3	a	i
6942	6961	1+1+3	1+1+3	a	a
6943	6962	1+1+3	1+1+3	a	a
6944	6963	1+1+3	1+1+3	a	i
6945	6964	1+1+3	1+1+3	i	i
	6965		1+1+3		i?
	6966		1+1+3		i
	6967		1+1+3		i
6946		1+1+3		a	
6947		1+1+3		i	
6948		1+1+3		i	
6949		1+1+3		i	
6950		1+1+3		i	
6951		1+1+3		i	
6952		1+1+3		a	
6953		1+1+3		a	
6954		1+1+3		a	
6955		1+1+3		a	
6956		1+1+3		i	
6957		1+1+3		i	
6958		7i		i	

4 Discussion

4.1 Articulatory control

Our proposal suggests that spectral patterns (vocalic and diphthongal) require substantial articulatory control in sperm whales. While there are many aspects of sperm whale articulation that are not yet fully understood, recent work has suggested that sperm whales and other odontocetes can control articulators to a larger degree than previously thought (Madsen et al., 2023). Weir et al. (2007) argues that a vocalization of sperm whales that is different from the coda vocalizations (clicks) — squeals — might be controlled by the whales which results in spectral modulations of squeals. Sperm whales have also been shown to be able to produce other types of vocalizations, such as trumpets (Pace et al., 2021), which additionally points to at least some level of active articulatory control. This line of work, however, focuses on registers that produce different kinds of vocalizations and not on differences within codas. Nevertheless, the work in Weir et al. (2007), Pace et al. (2021), and Madsen et al. (2023) suggest that active modulation of vocalizations might be possible. We also know that echolocation clicks are acoustically distinct from coda clicks, which is perhaps achieved by distal air sac shape (Madsen et al., 2002b). Additionally, it has been suggested that sperm whales may be able to create conformational changes to their nasal complex that could change the distance between reflective air sacs by 10 percent through the contraction of longitudinal muscles which could pull soft parts of the nose back towards the skull (Bøttcher et al., 2018). It is possible that similar changes in the nasal complex would lead to the described changes in spectral properties of coda vowels, but articulatory predictions are difficult to currently test in sperm whales. Given that the spermaceti organ is surrounded by muscles on the sides and air sacks on the ends of the organ (Huggenberger et al., 2016), it is not impossible to assume that the whales can control changes in the resonant body that are substantial enough to result in our observed patterns.

It has long been established that the source of coda vocalizations are the phonic lips. While this is unconfirmed at this point, we speculate that the the distal air sac acts as a filter resulting in the observed spectral properties. Phonic lips appear at the entrance of the distal air sac (Cranford, 1999; Huggenberger et al., 2016), which is parallel to human articulators where vocal folds appear at the entrance of the vocal tract. If we model the distal air sac as a simple tube closed at both ends, the first resonant frequency will be at 5800 Hz (F1 of the *a*-coda vowel) if the length of the tube is just under 3 cm (assuming the speed of sound in air at 343 m/s). For the resonant frequency of 3700 Hz (F1 of the *i*-coda vowel), the tube length would be just over 4.5 cm. The observed spectral trajectories can thus be achieved by controlling the shape of the distal air sac by only a few centimeters or less. The exact articulatory mechanisms behind the observed patterns as well as which tube model is the most appropriate for it is left for future work.

4.2 Orthogonality

The traditional sperm whale coda types appear to be independent of the proposed coda vowels, which means that the source features (number of pulses and timing) is highly orthogonal to filter features (spectral properties). The *a*-vowel and the *i*-vowel can appear on 1+1+3 or 5R2 codas, for example, as do the diphthong patterns.

We argue that sperm whale codas are composed of several independent features. The first two features in Table 3 have already been established. Additionally, Sharma et al. (2023) recently argue that timing/click number properties are highly combinatorial: their established rhythm, tempo, rubato, and ornamentation have the potential to make the traditionally observed properties even more complex. Here, we propose that in addition to the number of clicks and their timing, spectral properties constitute a new set of features that can independently combine with already established features. In sum, a list of properties that are potentially meaningful now also includes formant patterns and formant trajectories (Table 3).

Table 3: A list of potentially meaningful properties.

source features:	Number of clicks Timing
filter features:	Formant patterns: the *a*-vowel and the *i*-vowel Formant trajectories: level, rising, falling dipthongs

Many of these patterns are fully orthogonal. We have shown that the *a*-/*i*-vowel distinction is possible on different coda types (1+1+3, 5R1, or 5R2) and that diphthongal patterns can also surface on different coda types, including codas with 5 clicks. We also observe that diphthongs are more common on codas with more clicks. It is possible that further such distributional tendencies exist or that they differ across clans. Establishing such distributional patterns is left for future work.

5 Conclusion

This paper uncovers a new pattern in sperm whale coda vocalizations and suggests that a new dimension — spectral properties of clicks in codas — might be a meaningful feature in the sperm whale communication system. Traditionally, sperm whale codas have been primarily analyzed in terms of the number of clicks and the timing between clicks. These two parameters have been used to classify codas into several traditional coda types (Weilgart and Whitehead, 1993) or recently more fine grained combinatorial timing/click features (Sharma et al., 2023).

Human spoken language employs acoustic properties to convey meaning. Vowels in human speech can differ in length (or the number of vocal pulses), timing between pulses (or the fundamental frequency F0), formant frequencies (or the quality of vowels such as *a* vs. *i*) and trajectory of vowels (monophthongs like *i* and diphthongs like *ai*).

It appears that sperm whale codas feature equivalents to all these characteristics. The number of clicks and their inter-click timing can be broadly understood as the duration of the coda and the fundamental frequency (F0). We do not doubt that these two properties are meaningful in sperm whale vocalizations, as has been previously proposed. In human speech, the fundamental frequency (F0) and duration of vowels carry meaning-distinguishing information. For example, the four Mandarin tones can change the meaning of segmentally the same syllable: *mā* 'mother', *má* 'hemp', *mǎ* 'horse', and *mà* 'scold' (Duanmu, 2007, 225).

This paper suggests that, in addition to the number of clicks and ICI, resonant frequencies (formants) in sperm whale codas and their trajectories are potentially meaningful as well. We uncover at least two clear patterns in spectral properties of codas: the *a*-type and the *i*-type coda vowels. We also show that the two coda vowels can be actively exchanged in sperm whale dialogues. Finally, we argue that individual codas can also have rising and falling trajectories (or a combination of the two), a pattern that we call *coda diphthongs.* These patterns are likely not artifacts resulting from whales' movement or depth position. We argue that the patterns are recurrent across whales, controlled by whales, perceivable, and discrete (in the sense that a coda is either of one type or the other with no mixing of types within coda).

Exploration of the acoustic properties in codas was prompted by a deep neural network architecture called fiwGAN (Beguš, 2021). The network was trained to imitate sperm whale codas and embed information into these vocalizations. An interpretability technique called CDEV (Beguš et al., 2023) suggested that several spectral properties might be meaningful in this communication system. By uncovering recurrent patterns in spectral properties that repeat across whales and appear to be actively controlled by the whales, we make explicit the possibility that spectral properties are meaningful, as suggested by the fiwGAN model (Beguš, 2021) and the CDEV method in Beguš et al. (2023).

These findings have the potential to add to communicative complexity of sperm whale vocalizations and open up several new possibilities for research. Our paper suggests that the sperm whale communcation system is not a Morse code-like system, but that spectral properties of codas are acoustically differentiated. How the spectral properties of codas are realized in other clans and how they relate to referential meaning is the logical next step of this research.

References

M. Amano, A. Kourogi, K. Aoki, M. Yoshioka, and K. Mori. Differences in sperm whale codas between two waters off Japan: possible geographic separation of vocal clans. *Journal of Mammalogy*, 95(1):169–175, 02 2014. ISSN 0022-2372. doi: 10.1644/13-MAMM-A-172. URL https://doi.org/10.1644/13-MAMM-A-172.

T. O. S. Amorim, L. Rendell, J. Di Tullio, E. R. Secchi, F. R. Castro, and A. Andriolo. Coda repertoire and vocal clans of sperm whales in the western atlantic ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, 160:103254, 2020. ISSN 0967-0637. doi: https://doi.org/10.1016/j.dsr.2020.103254. URL https://www.sciencedirect.com/science/article/pii/S096706372030042X.

J. Andreas, G. Beguš, M. M. Bronstein, R. Diamant, D. Delaney, S. Gero, S. Goldwasser, D. F. Gruber, S. de Haas, P. Malkin, N. Pavlov, R. Payne, G. Petri, D. Rus, P. Sharma, D. Tchernov, P. Tønnesen, A. Torralba, D. Vogt, and R. J. Wood. Toward understanding the communication in sperm whales. *iScience*, 25(6):104393, 2022. ISSN 2589-0042. doi: https://doi.org/10.1016/j.isci.2022.104393. URL https://www.sciencedirect.com/science/article/pii/S2589004222006642.

T. Arnbom. Individual identification of sperm whales. *Report of the International Whaling Commission*, 37(20):1–204, 1987.

G. Beguš. Generative adversarial phonology: Modeling unsupervised phonetic and phonological learning with neural networks. *Frontiers in Artificial Intelligence*, 3:44, 2020. ISSN 2624-8212. doi: 10.3389/frai.2020.00044. URL https://www.frontiersin.org/article/10.3389/frai.2020.00044.

G. Beguš. CiwGAN and fiwGAN: Encoding information in acoustic data to model lexical learning with Generative Adversarial Networks. *Neural Networks*, 139:305–325, 2021. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2021.03.017. URL https://www.sciencedirect.com/science/article/pii/S0893608021001052.

G. Beguš, A. Leban, and S. Gero. Approaching an unknown communication system by latent space exploration and causal inference. *arXiv*, 2303.10931:1–25, 2023.

P. Boersma and D. Weenink. Praat: doing phonetics by computer [computer program]. version 5.4.06. Retrieved 21 February 2015 from http://www.praat.org/, 2015.

B. K. Branstetter, P. W. Moore, J. J. Finneran, M. N. Tormey, and H. Aihara. Directional properties of bottlenose dolphin (Tursiops truncatus) clicks, burst-pulse, and whistle sounds. *The Journal of the Acoustical Society of America*, 131(2):1613–1621, 02 2012. ISSN 0001-4966. doi: 10.1121/1.3676694. URL https://doi.org/10.1121/1.3676694.

A. Bøttcher, S. Gero, K. Beedholm, H. Whitehead, and P. T. Madsen. Variability of the inter-pulse interval in sperm whale clicks with implications for size estimation and individual identification. *The Journal of the Acoustical Society of America*, 144(1):365–374", 2018.

T. W. Cranford. The sperm whale's nose: Sexual selection on a grand scale?1. *Marine Mammal Science*, 15(4):1133–1157, 1999. doi: https://doi.org/10.1111/j.1748-7692.1999.tb00882.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1748-7692.1999.tb00882.x.

J. D. Darling, M. E. Jones, and C. P. Nicklin. Humpback whale songs: Do they organize males during the breeding season? *Behaviour*, 143(9):1051 – 1101, 2006. doi: https://doi.org/10.1163/156853906778607381. URL https://brill.com/view/journals/beh/143/9/article-p1051_1.xml.

A. Davies, P. Veličković, L. Buesing, S. Blackwell, D. Zheng, N. Tomašev, R. Tanburn, P. Battaglia, C. Blundell, A. Juhász, M. Lackenby, G. Williamson, D. Hassabis, and P. Kohli. Advancing mathematics by guiding human intuition with AI. *Nature*, 600(7887):70–74, 2021. doi: 10.1038/s41586-021-04086-x. URL https://doi.org/10.1038/s41586-021-04086-x.

S. Duanmu. *The phonology of standard Chinese*. The phonology of the world's languages series. Oxford University Press, Oxford ;, 2nd ed. edition, 2007. ISBN 9780199215799.

G. Fant. *With Calculations based on X-Ray Studies of Russian Articulations*. De Gruyter Mouton, Berlin, Boston, 1971. ISBN 9783110873429. doi: doi:10.1515/9783110873429. URL https://doi.org/10.1515/9783110873429.

O. A. Filatova, F. I. Samarra, V. B. Deecke, J. K. Ford, P. J. Miller, and H. Yurk. Cultural evolution of killer whale calls: background, mechanisms and consequences. *Behaviour*, 152(15): 2001 – 2038, 2015. doi: https://doi.org/10.1163/1568539X-00003317. URL https://brill.com/view/journals/beh/152/15/article-p2001_1.xml.

W. T. Fitch. The phonetic potential of nonhuman vocal tracts: Comparative cineradiographic observations of vocalizing animals. *Phonetica*, 57(2-4):205–218, 2000. doi: doi:10.1159/000028474. URL https://doi.org/10.1159/000028474.

W. T. Fitch and M. D. Hauser. *Unpacking "Honesty": Vertebrate Vocal Production and the Evolution of Acoustic Signals*, pages 65–137. Springer New York, New York, NY, 2003. ISBN 978-0-387-22762-7. doi: 10.1007/0-387-22762-8_3. URL https://doi.org/10.1007/0-387-22762-8_3.

S. Gero, M. Milligan, C. Rinaldi, P. Francis, J. Gordon, C. Carlson, A. Steffen, P. Tyack, P. Evans, , and H. Whitehead. Behavior and social structure of the sperm whales of dominica, west indies. *Marine Mammal Science*, 30:905–922,, 2014.

S. Gero, A. Bøttcher, H. Whitehead, and P. T. Madsen. Socially segregated, sympatric sperm whale clans in the atlantic ocean. *Royal Society Open Science*, 3(6):160061, 2016a. doi: 10.1098/rsos.160061. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsos.160061.

S. Gero, H. Whitehead, and L. Rendell. Individual, unit and vocal clan level identity cues in sperm whale codas. *Royal Society Open Science*, 3(1):150372, 2016b. doi: 10.1098/rsos.150372. URL `https://royalsocietypublishing.org/doi/abs/10.1098/rsos.150372`.

J. C. Goold and S. E. Jones. Time and frequency domain characteristics of sperm whale clicks. *The Journal of the Acoustical Society of America*, 98(3):1279–1291, 09 1995. ISSN 0001-4966. doi: 10.1121/1.413465. URL `https://doi.org/10.1121/1.413465`.

L. M. Herman. The multiple functions of male song within the humpback whale (megaptera novaeangliae) mating system: review, evaluation, and synthesis. *Biological Reviews*, 92(3):1795–1818, 2017. doi: https://doi.org/10.1111/brv.12309. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/brv.12309`.

S. Huggenberger, M. André, and H. H. A. Oelschläger. The nose of the sperm whale: overviews of functional design, structural homologies and evolution. *Journal of the Marine Biological Association of the United Kingdom*, 96(4):783–806, 2016. doi: 10.1017/S0025315414001118.

L. A. E. Huijser, V. Estrade, I. Webster, L. Mouysset, A. Cadinouche, and V. Dulau-Drouot. Vocal repertoires and insights into social structure of sperm whales (physeter macrocephalus) in mauritius, southwestern indian ocean. *Marine Mammal Science*, 36(2):638–657, 2020. doi: https://doi.org/10.1111/mms.12673. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/mms.12673`.

M. Johnson and P. L. Tyack. A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE Journal of Oceanic Engineering*, 28(1):3–12, 2003.

J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021. doi: 10.1038/s41586-021-03819-2. URL `https://doi.org/10.1038/s41586-021-03819-2`.

D. H. Klatt and R. A. Stefanski. How does a mynah bird imitate human speech? *The Journal of the Acoustical Society of America*, 55(4):822–832, 08 2005. ISSN 0001-4966. doi: 10.1121/1.1914607. URL `https://doi.org/10.1121/1.1914607`.

C.-F. Lin, Y.-C. Chung, J.-D. Zhu, S.-H. Chang, C.-C. Wen, I. A. Parinov, and S. N. Shevtsov. The energy based characteristics of sperm whale clicks using the Hilbert Huang transform analysis methoda). *The Journal of the Acoustical Society of America*, 142(2):504–511, 08 2017. ISSN 0001-4966. doi: 10.1121/1.4996106. URL `https://doi.org/10.1121/1.4996106`.

M. Lindau, K. Norlin, and J.-O. Svantesson. Some cross-linguistic differences in diphthongs. *Journal of the International Phonetic Association*, 20(1):10–14, 1990. doi: 10.1017/S0025100300003996.

P. Madsen, M. Wahlberg, and B. Møhl. Male sperm whale (physeter macrocephalus) acoustics in a high-latitude habitat: implications for echolocation and communication. *Behavioral Ecology and Sociobiology*, 53(1):31–41, 2002a. doi: 10.1007/s00265-002-0548-1. URL `https://doi.org/10.1007/s00265-002-0548-1`.

P. T. Madsen, R. Payne, N. U. Kristiansen, M. Wahlberg, I. Kerr, and B. Møhl. Sperm whale sound production studied with ultrasound time/depth-recording tags. *Journal of Experimental Biology*, 205(13):1899–1906, 2002b. ISSN 0022-0949. doi: 10.1242/jeb.205.13.1899. URL https://doi.org/10.1242/jeb.205.13.1899.

P. T. Madsen, U. Siebert, and C. P. H. Elemans. Toothed whales use distinct vocal registers for echolocation and communication. *Science*, 379(6635):928–933, 2023. doi: 10.1126/science.adc9570. URL https://www.science.org/doi/abs/10.1126/science.adc9570.

P. J. Miller. Mixed-directionality of killer whale stereotyped calls: a direction of movement cue? *Behavioral Ecology and Sociobiology*, 52(3):262–270, 2002. doi: 10.1007/s00265-002-0508-9. URL https://doi.org/10.1007/s00265-002-0508-9.

B. Møhl, M. Wahlberg, P. T. Madsen, A. Heerfordt, and A. Lund. The monopulsed nature of sperm whale clicks. *The Journal of the Acoustical Society of America*, 114(2):1143–1154, 07 2003. ISSN 0001-4966. doi: 10.1121/1.1586258. URL https://doi.org/10.1121/1.1586258.

K. E. Moore, W. A. Watkins, and P. L. Tyack. Pattern similarity in shared codas from sperm whales (physeter catodon). *Marine Mammal Science*, 9(1):1–9, 1993. doi: https://doi.org/10.1111/j.1748-7692.1993.tb00421.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1748-7692.1993.tb00421.x.

D. S. Pace, C. Lanfredi, S. Airoldi, G. Giacomini, M. Silvestri, G. Pavan, and D. Ardizzone. Trumpet sounds emitted by male sperm whales in the mediterranean sea. *Scientific Reports*, 11(1):5867, 2021. doi: 10.1038/s41598-021-84126-8. URL https://doi.org/10.1038/s41598-021-84126-8.

E. Panova, A. Agafonov, R. Belikov, and F. Melnikova. Characteristics and microgeographic variation of whistles from the vocal repertoire of beluga whales (Delphinapterus leucas) from the White Sea. *The Journal of the Acoustical Society of America*, 146(1):681–692, 07 2019. ISSN 0001-4966. doi: 10.1121/1.5119249. URL https://doi.org/10.1121/1.5119249.

E. M. Panova, R. A. Belikov, A. V. Agafonov, O. I. Kirillova, A. D. Chernetsky, and V. M. Bel'kovich. Intraspecific variability in the "vowel"-like sounds of beluga whales (delphinapterus leucas): Intra- and interpopulation comparisons. *Marine Mammal Science*, 32(2):452–465, 2016. doi: https://doi.org/10.1111/mms.12266. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mms.12266.

D. K. Patterson and I. M. Pepperberg. Acoustic and articulatory correlates of stop consonants in a parrot and a human subject. *The Journal of the Acoustical Society of America*, 103(4):2197–2215, 04 1998. ISSN 0001-4966. doi: 10.1121/1.421365. URL https://doi.org/10.1121/1.421365.

H. Pines. Mapping the phonetic structure of humpback whale song units: extraction, classification, and Shannon-Zipf confirmation of sixty sub-units. *Proceedings of Meetings on Acoustics*, 35(1):010003, 01 2019. ISSN 1939-800X. doi: 10.1121/2.0000957. URL https://doi.org/10.1121/2.0000957.

L. Rendell, S. L. Mesnick, M. L. Dalebout, J. Burtenshaw, and H. Whitehead. Can genetic differences explain vocal dialect variation in sperm whales, physetermacrocephalus? *Behavior Genetics*, 42(2):332–343, 2012. doi: 10.1007/s10519-011-9513-y. URL https://doi.org/10.1007/s10519-011-9513-y.

L. E. Rendell and H. Whitehead. Vocal clans in sperm whales (¡i¿physeter macrocephalus¡/i¿). *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1512):225–231, 2003. doi: 10.1098/rspb.2002.2239. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2002.2239.

F. Samarra and P. Miller. The role of the spectral characteristics of discrete calls in killer whale (Orcinus orca) communication. *The Journal of the Acoustical Society of America*, 119 (5 Supplement):3373–3373, 05 2006. ISSN 0001-4966. doi: 10.1121/1.4786575. URL https://doi.org/10.1121/1.4786575.

F. N. Schmidt, M. M. Delsmann, K. Mletzko, T. A. Yorgan, M. Hahn, U. Siebert, B. Busse, R. Oheim, M. Amling, and T. Rolvien. Ultra-high matrix mineralization of sperm whale auditory ossicles facilitates high sound pressure and high-frequency underwater hearing. *Proceedings of the Royal Society B: Biological Sciences*, 285(1893):20181820, 2018. doi: 10.1098/rspb.2018.1820. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2018.1820.

T. M. Schulz, H. Whitehead, S. Gero, and L. Rendell. Overlapping and matching of codas in vocal interactions between sperm whales: insights into communication function. *Animal Behaviour*, 76 (6):1977–1988, 2008. ISSN 0003-3472. doi: https://doi.org/10.1016/j.anbehav.2008.07.032. URL https://www.sciencedirect.com/science/article/pii/S0003347208004120.

P. Sharma, S. Gero, R. Payne, D. F. Gruber, D. Rus, A. Torralba, and J. Andreas. Contextual and combinatorial structure in sperm whale vocalisations. *bioRxiv*, 2023. doi: 10.1101/2023.12.06.570484. URL https://www.biorxiv.org/content/early/2023/12/08/2023.12.06.570484.

J. Sportelli. Killer whale (Orcinus orca) pulsed calls in the eastern canadian arctic. Master's thesis, UC San Diego, 2019. URL https://escholarship.org/uc/item/9750f0m6.

A. L. Stansbury and V. M. Janik. Formant modification through vocal production learning in gray seals. *Current Biology*, 29(13):2244–2249.e4, 2019. ISSN 0960-9822. doi: https://doi.org/10.1016/j.cub.2019.05.071. URL https://www.sciencedirect.com/science/article/pii/S0960982219306852.

K. N. Stevens. *Acoustic Phonetics.* MIT Press, 1998.

A. S. Stoeger, D. Mietchen, S. Oh, S. de Silva, C. T. Herbst, S. Kwon, and W. T. Fitch. An asian elephant imitates human speech. *Current Biology*, 22(22):2144–2148, 2012. ISSN 0960-9822. doi: https://doi.org/10.1016/j.cub.2012.09.022. URL https://www.sciencedirect.com/science/article/pii/S096098221201086X.

J. M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N. M. Donghia, C. R. MacNair, S. French, L. A. Carfrae, Z. Bloom-Ackermann, V. M. Tran, A. Chiappino-Pepe, A. H. Badran, I. W. Andrews, E. J. Chory, G. M. Church, E. D. Brown, T. S. Jaakkola, R. Barzilay, and J. J. Collins. A deep learning approach to antibiotic discovery. *Cell*, 180(4):688–702.e13, 2020. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2020.01.021. URL https://www.sciencedirect.com/science/article/pii/S0092867420301021.

A. Thode, D. K. Mellinger, S. Stienessen, A. Martinez, and K. Mullin. Depth-dependent acoustic features of diving sperm whales (Physeter macrocephalus) in the Gulf of Mexico. *The Journal of the Acoustical Society of America*, 112(1):308–321, 07 2002. ISSN 0001-4966. doi: 10.1121/1.1482077. URL https://doi.org/10.1121/1.1482077.

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

W. A. Watkins and W. E. Schevill. Sperm whale codas. *The Journal of the Acoustical Society of America*, 62(6):1485–1490, 1977. ISSN 0001-4966. doi: 10.1121/1.381678. URL `https://doi.org/10.1121/1.381678`.

L. Weilgart and H. Whitehead. Coda communication by sperm whales (physeter macrocephalus) off the galápagos islands. *Canadian Journal of Zoology*, 71(4):744–752, 1993. doi: 10.1139/z93-098. URL `https://doi.org/10.1139/z93-098`.

L. S. Weilgart. *Vocalizations of the sperm whale (Physeter macrocephalus) off the Galapagos Islands as related to behavioral and circumstantial variables.* PhD thesis, Dalhousie University, 1990.

C. R. Weir, A. Frantzis, P. Alexiadou, and J. C. Goold. The burst-pulse nature of 'squeal' sounds emitted by sperm whales (physeter macrocephalus). *Journal of the Marine Biological Association of the United Kingdom*, 87(1):39–46, 2007. doi: 10.1017/S0025315407054549.

R. Wellard, R. L. Pitman, J. Durban, and C. Erbe. Cold call: the acoustic repertoire of ross sea killer whales (orcinus orca, type c) in McMurdo Sound, Antarctica. *Royal Society Open Science*, 7(2):191228, 2020. doi: 10.1098/rsos.191228. URL `https://royalsocietypublishing.org/doi/abs/10.1098/rsos.191228`.

H. Whitehead. *Sperm whales : social evolution in the ocean.* University of Chicago Press, Chicago, 2003. ISBN 0226895173.

H. Whitehead and L. Rendell. *The cultural lives of whales and dolphins.* The University of Chicago Press, Chicago, 2015. ISBN 9780226895314.

H. Whitehead and L. Weilgart. Patterns of visually observable behaviour and vocalizations in groups of female sperm whales. *Behaviour*, 118(3-4):275 – 296, 1991. doi: https://doi.org/10.1163/156853991X00328.

L. V. Worthington and W. E. Schevill. Underwater sounds heard from sperm whales. *Nature*, 180 (4580):291–291, 1957. doi: 10.1038/180291a0. URL `https://doi.org/10.1038/180291a0`.

Author Contributions

The pattern of coda vowels and diphthongs was discovered and established by G.B. G.B.: Conceptualization, Methodology, Formal analysis, Validation, Investigation, Software, Visualization, Writing - Original Draft, Supervision. R.S.: Software, Investigation. A.L.: Software. M.S.: Investigation (exploratory analysis of click-level acoustics). S.G.: Resources, Fieldwork, Data Collection, Data Curation, Software.

Acknowledgements

This study was funded by Project CETI via grants from Dalio Philanthropies and Ocean X; Sea Grape Foundation; Rosamund Zander/Hansjorg Wyss, Chris Anderson/Jacqueline Novogratz through The Audacious Project: a collaborative funding initiative housed at TED. We thank Pratyusha Sharma, Antonio Torralba, Jacob Andreas, Daniela Rus, Roee Diamant, and the rest of the Project CETI team for their comments.

Fieldwork for The Dominica Sperm Whale Project during tagging years in 2014-2018 was supported by through a FNU fellowship for the Danish Council for Independent Research supplemented by a Sapere Aude Research Talent Award, a Carlsberg Foundation expedition grant, a grant from Focused on Nature, two Explorer Grants from the National Geographic Society, and supplementary grants from the Arizona Center for Nature Conservation, Quarters For Conservation, the Dansk Akustisks Selskab, Oticon Foundation, and the Dansk Tennis Fond all to SG. Further funding was provided by a Discovery and Equipment grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) to Hal Whitehead of Dalhousie University and a FNU large frame grant and a Villum Foundation Grant to Peter Madsen of Aarhus University. We are grateful to Kristian Beedholm for CodaSorter, as well as Mark Johnson and Peter L. Tyack for their in-kind contribution of Dtags and associated code during the research. We thank the Chief Fisheries Officers and the Dominica Fisheries Division officers for research permits and their collaboration in data collection; all the crews of R/V Balaena and The DSWP team for data collection, curation, and annotation; as well as Dive Dominica, Al Dive, and W.E.T. Dominica for logistical support while in Dominica.

A Appendix

A.1 Methods

The following paragraphs in this section are not original to our paper, but taken from several other papers authored by S.G. with only minimal paraphrasing and a few additions. These paragraphs provide only facts about data collection that are important for analyzing our work but not in any way crucial to our original argument. Because the paragraphs primarily describe facts about data collection, we believe reproducing the text with this acknowledgment is more appropriate than paraphrasing the facts with different frames and presenting the work as original.

> Well-known social units sperm whales were tracked along the western coast of the Island of Dominica (N15.30 W61.40) between 2014 and 2018. Codas were recorded through the deployment of animal-borne sound and movement tags (DTag generation 3, Johnson and Tyack 2003). Tagging was accomplished on an 11-meter rigid-hulled inflatable boat suing a hand pole. DTags record two-channel audio at 120 kHz or 125 kHz with a 16-bit resolution, providing a flat (±2 dB) frequency response between 0.4 and 45kHz. Pressure and acceleration were sampled at a rate of 500 Hz with a 16-bit resolution, and were decimated to 25 Hz for analysis.
>
> Whales, including the tagged whales, were photographically identified (Arnbom, 1987). Only tag deployments from one of the two sympatric clans (EC1, the Eastern Caribbean Clan) were included in the analysis to control for any differences in repertoires between vocal clans (Gero et al., 2016a).
>
> To define codas, absolute inter-click intervals were measured as in Gero et al. (2016b), using Coda Sorter, a custom-written tool (K. Beedholm, Marine Bioacoustics Lab, Aarhus University) in LabView (National Instruments, TX, USA). Determining if codas were produced by the tagged whales or non-focal animals was accomplished in CodaSorter using estimates for each click for the angle of arrival, channel delay, centroid frequency, and inter-pulse interval (IPI, the time between the onset of the first pulse and the onset of the next pulse in the multi-pulse structure of sperm whales clicks, Møhl et al. 2003). It is possible that some clicks are misclassified as focal or non-focal, we believe the error rate is minor enough such that final conclusions of this paper are not affected. Photo-identification supported this process by identifying which whales were present and associated with the tagged whales at each surfacing. During annotation, rare, long codas were excluded from analysis (greater than 10 clicks, less than 5% of all codas recorded).

A.2 Figures

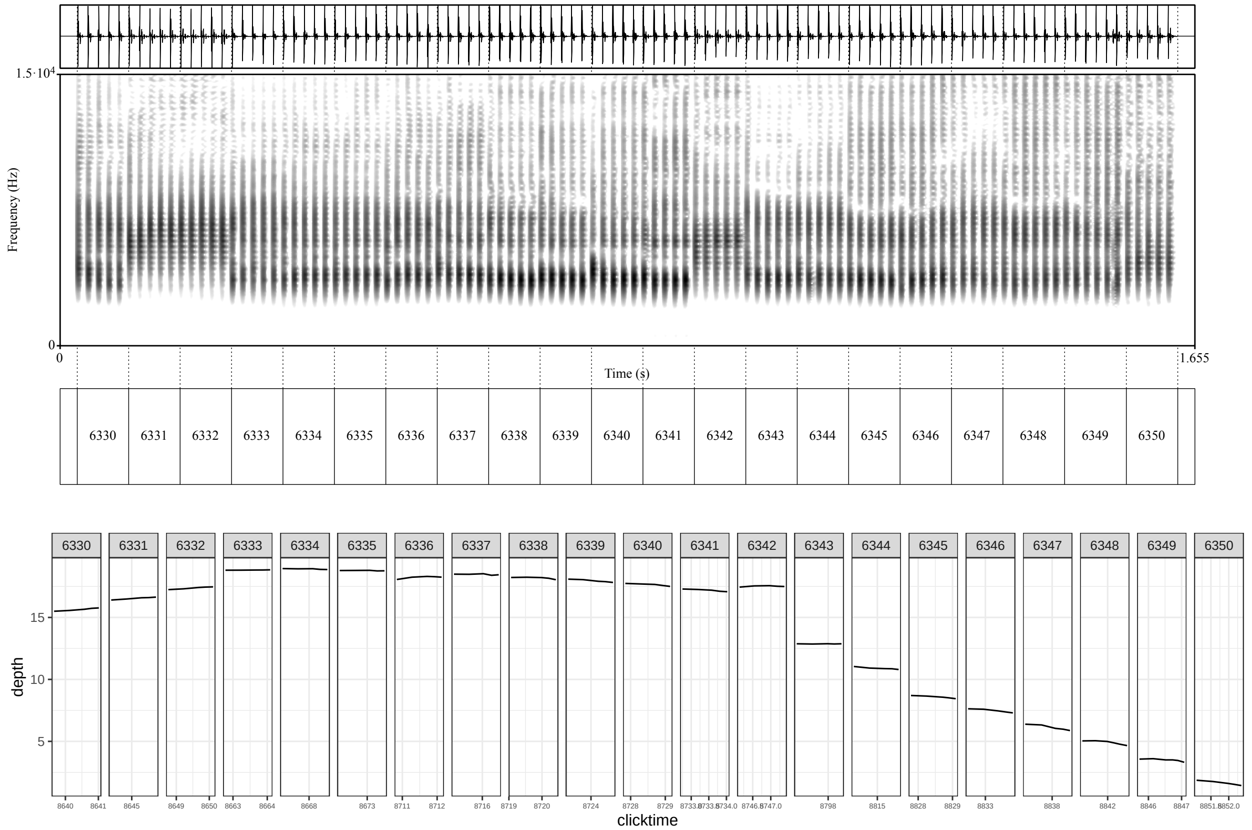

Figure 11: (**top**) Waveforms and spectrograms 0–15,000 (single right channel) of 21 codas from a single bout by focal Fork with timing removed and all clicks are peak-normalized. All codas are of the 1+1+3 type except for codas 6348 and 6349 (both 6 clicks). (**bottom**) Depth values (in m) for each coda.

Table 4: A dialogue between two non-focal whales when Fork was wearing a tag. If two codas are less than 1 s apart, they are marked as concurrent.

Non-focal 1 #	Non-focal 2 #	Non-focal 1 Type	Non-focal 2 Type	Non-focal 1 Vowel	Non-focal2 Vowel
7255	7268	1+1+3	1+1+3	a	i
7256	7269	1+1+3	1+1+3	i	i
7257	7270	1+1+3	7-noise	i	i
7258	7271	1+1+3	1+1+3	i	i
7259	7272	1+1+3	1+1+3	i	i
7260	7273	1+1+3	1+1+3	a	i
7261	7274	1+1+3	1+1+3	a	i
7262	7275	1+1+3	1+1+3	a	i
7263	7276	1+1+3	1+1+3	i	i
	7277		1+1+3		a
	7278		1+1+3		a
	7279		1+1+3		i
7264	7280	1+1+3	1+1+3	a	i
	7281		1+1+3		a
7265	7282	1+1+3	1+1+3	i	a
7266	7283	1+1+3	1+1+3	a	a
7267	7283	1+1+3		a	

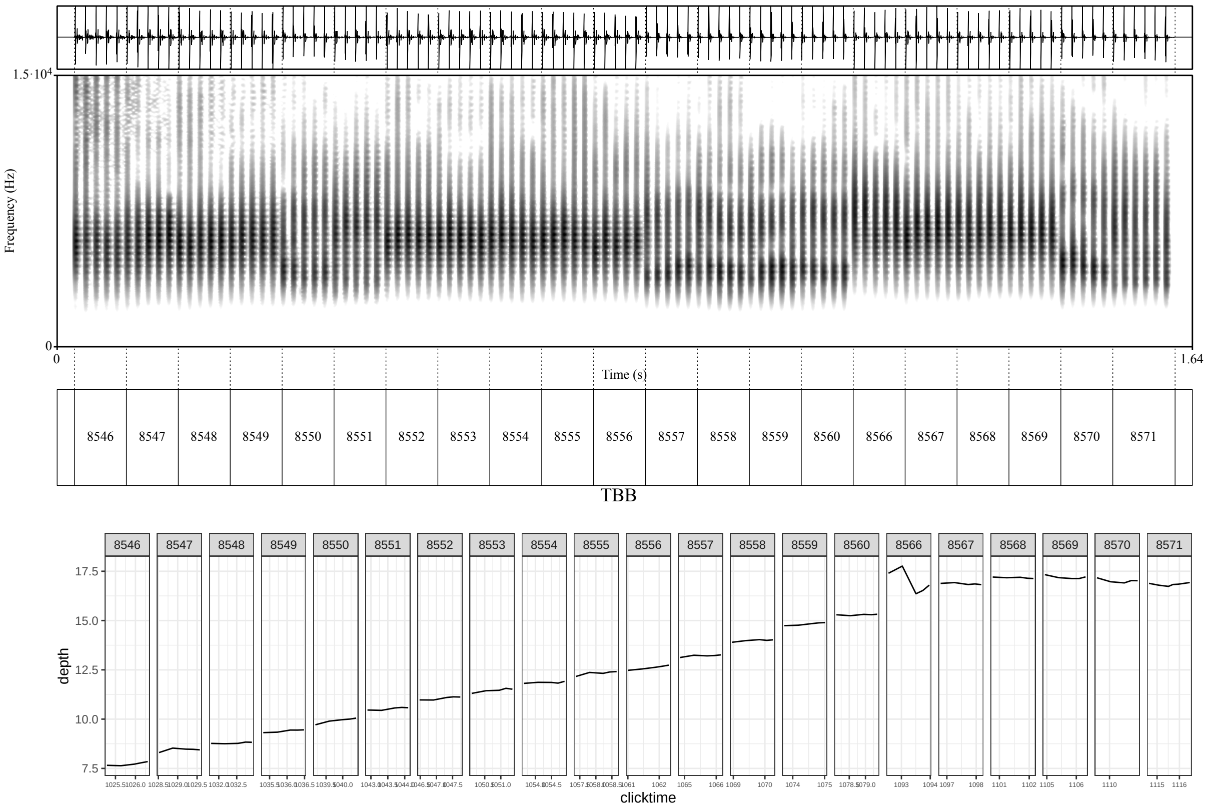

Figure 12: (**top**) Waveforms and spectrograms 0–15,000 (single right channel) of 21 codas from two bouts by focal TBB with timing removed and all clicks are peak-normalized. The new bouts begin with coda 8566. All codas are of the 1+1+3 type except coda 8671 (6 clicks). (**bottom**) Depth values (in m) for each coda.

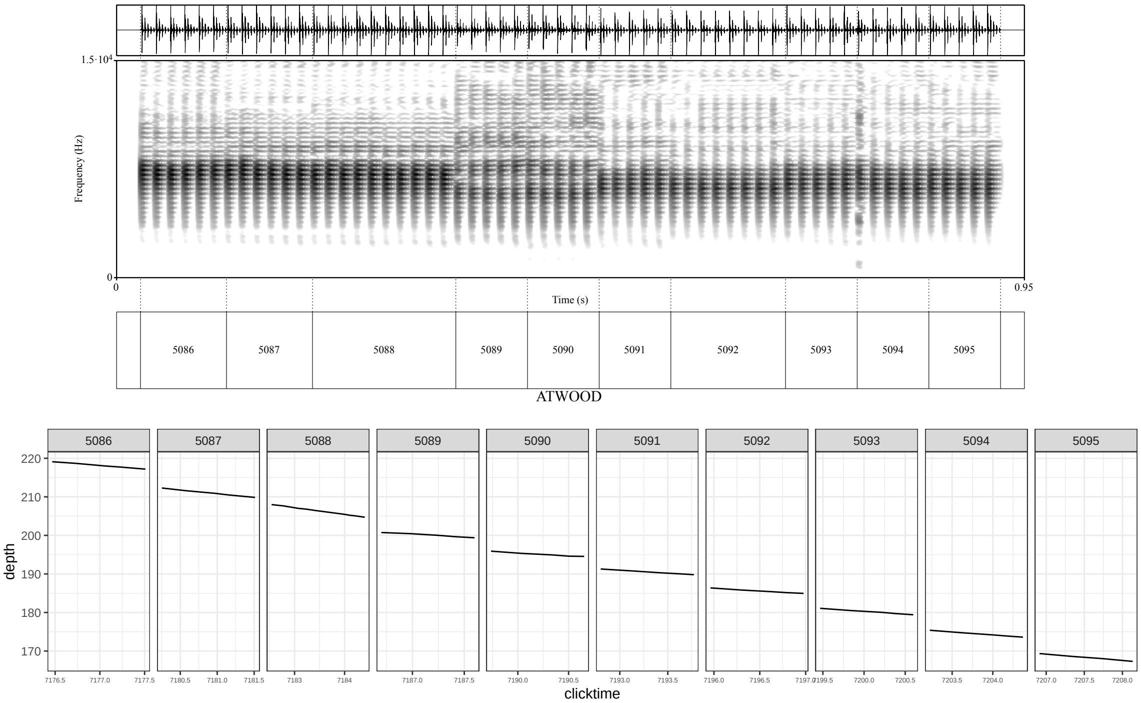

Figure 13: (**top**) Waveforms and spectrograms 0–15,000 (single right channel) of 10 codas from one bout by focal Atwood with timing removed and all clicks are peak-normalized. Coda types are: 6-NOISE 6-NOISE, 10R, 1+1+3, 1+1+3, 1+1+3, 8-NOISE, 1+1+3, 1+1+3, 1+1+3. (**bottom**) Depth values (in m) for each coda.

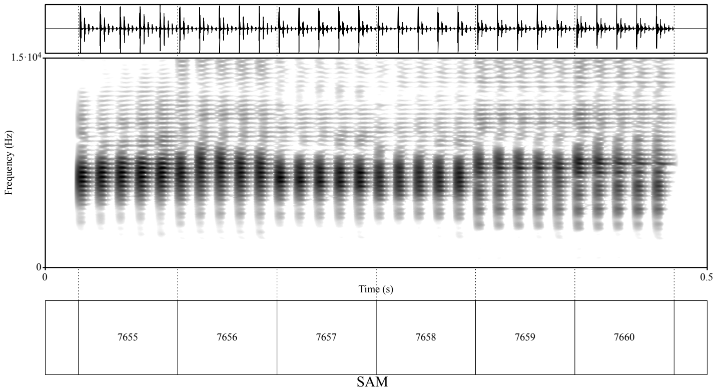

Figure 14: Waveforms and spectrograms 0–15,000 (single right channel) of 6 codas from one bout by focal Sam with timing removed and all clicks are peak-normalized. All coda types are 5R1.

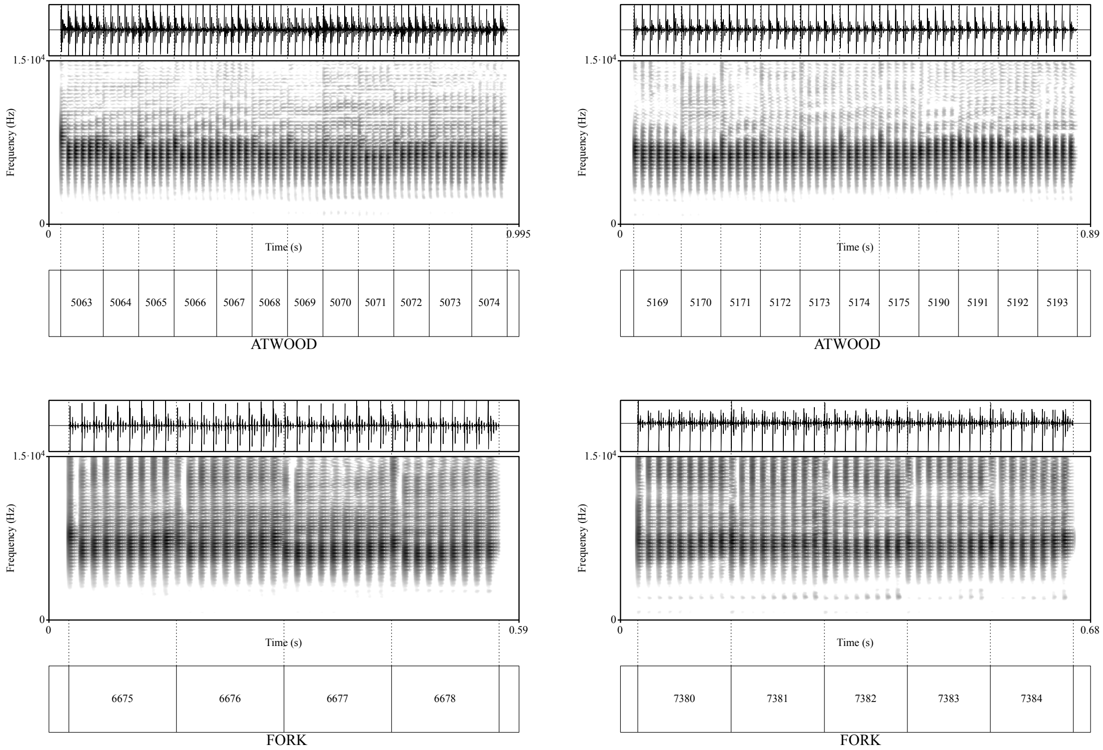

Figure 15: Waveforms and spectrograms (0–15,000, single right channel) of codas from Fork and Atwood, where the first click has a substantially higher formant frequency.

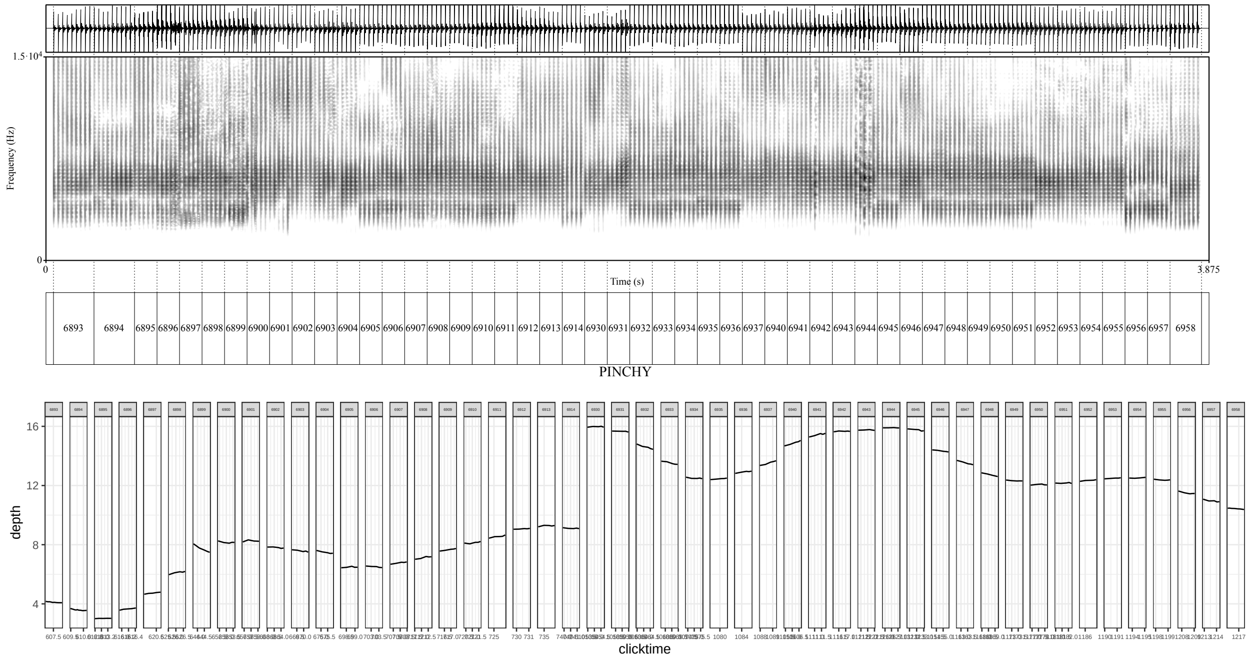

Figure 16: (**top**) Waveforms and spectrograms 0–15,000 (single right channel) of Pinchy for codas analyzed in Section 3.4. (**bottom**) Depth values (in m) for each coda.

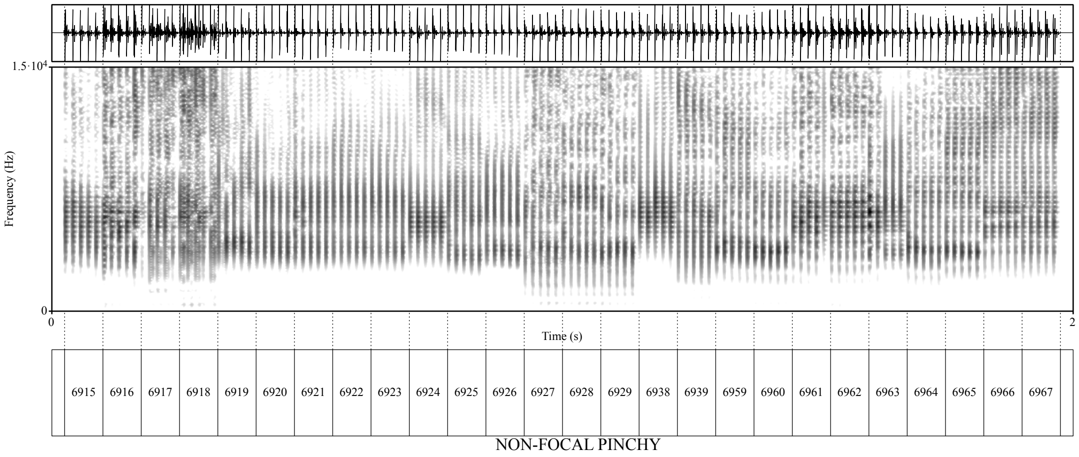

Figure 17: Waveforms and spectrograms 0–15,000 (single left channel) of the non-focal whale for codas analyzed in Section 3.4.

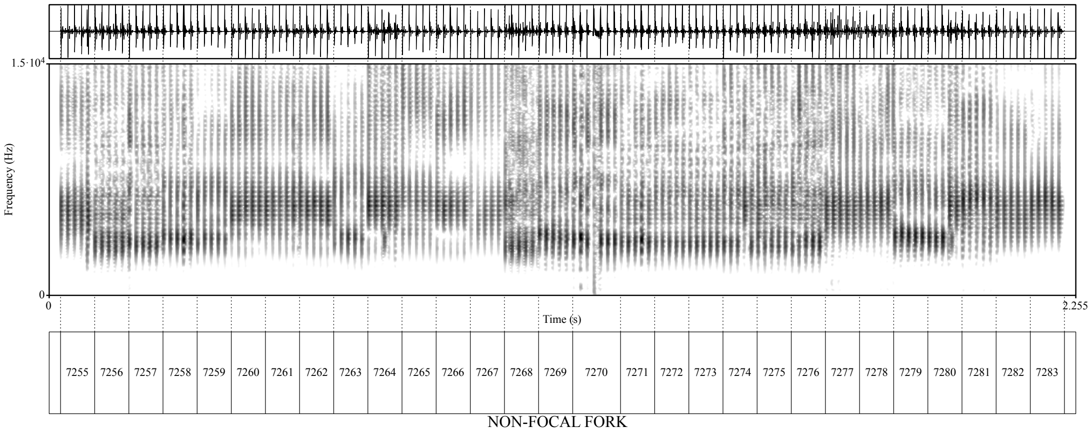

Figure 18: Waveforms and spectrograms 0–15,000 (single left channel) of a dialogue in two non-focal whales analyzed in Table 4.