Animal Behaviour 232 (2026) 123431

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Examining combinatoriality within the pūkeko vocal repertoire

Gabriella E. C. Gall [a, b, c, d, *], Vlad Demartsev [a, c, d], Pranav Minasandra [a, d], Cecilia Baldoni [e], Kristal E. Cain [f], James S. Quinn [g]

[a] Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Konstanz, Germany
[b] Zukunftskolleg, University of Konstanz, Konstanz, Germany
[c] Biology Department, University of Konstanz, Konstanz, Germany
[d] Department for the Ecology of Animal Societies, Max Planck Institute of Animal Behaviour, Konstanz, Germany
[e] Department of Migration, Max Planck Institute of Animal Behaviour, Konstanz, Germany
[f] Te Kura Mātauranga Koiora, School of Biological Sciences, Waipapa Taumata Rau, University of Auckland, Aotearoa, New Zealand
[g] Biology Department, McMaster University, Hamilton, Canada

ARTICLE INFO

Article history:
Received 8 April 2025
Initial acceptance 4 June 2025
Final acceptance 26 September 2025
Available online 9 January 2026
MS. number: 25-00234R

Keywords:
animal communication
Australasian swamphen
call combinations
multilevel combinatoriality
Porphyrio melanotus melanotus
pūkeko
UMAP

Most animals use various vocalizations to communicate with others and coordinate activities. However, animals are limited in the number of sounds they can produce. In humans, language allows for the unrestricted communication of information by generating new meaning from the finite set of sounds. For nonhuman animals, some level of combinatoriality has been observed such that segments of sound can be combined at either one of two levels, within calls or between calls. It is rare to discover evidence that animals combine calls on more than one level. This requires a comprehensive analysis of the combinatorial features that characterize a species' vocal system. Here we studied combinatoriality in the acoustic signals produced by pūkeko, *Porphyrio melanotus melanotus.* We identified 13 sound elements and verified their distinctiveness using uniform manifold approximation and projection. We next assessed the combinatorial abilities of pūkeko using a two-tiered combinatorial system. We first analysed how different sound elements are combined to form calls and found clear structural patterns, where specific sound elements typically serve either as prefixes or suffixes, whereas others serve as connecting (middle) elements. Second, we examined how calls themselves were combined to generate longer call sequences. At the level of call sequences, we specifically focused on yelling-type call sequences, mostly produced in aggressive contexts and found dynamic structural patterns, with calls increasing in duration with the progression of a call sequence. As these sequences unfold, calls undergo systematic transformation through the addition of new terminal elements and modification of existing ones. This hierarchical combinatorial capacity showed how a relatively limited set of acoustic elements can generate an extensive repertoire of calls. Our findings provide empirical evidence for combinatorial processes operating at multiple levels within a single species' communication system and the first evidence of such in a nonvocal learner.

Acoustic communication is widespread across the animal kingdom and serves many functions, such as mate attraction and territorial defence as well as transmitting information about predators, food, or environmental conditions (Bradbury & Vehrencamp, 1998). Vocal repertoires usually consist of a finite and limited number of sounds due to morphological constraints (Bohn et al., 2006; Demery et al., 2021; Montealegre-Z, 2009; Wallschläger, 1980). Rather than increasing the number of sound types produced, many animals show an ability to combine acoustic segments into larger structures, thus increasing the amount of information they can convey (Arnold & Zuberbühler, 2006; Engesser & Townsend, 2019; Suzuki et al., 2018). Human language is known for its 'duality of patterning', whereby a limited number of meaningless sounds (e.g. vowels and consonants) can be combined to create different phonemes (e.g. words) and these phonemes can be further combined into larger structures (e.g. sentences). The meaning of these structures is derived from the specific phonemes they contain and their arrangement (compositionality) (Chomsky, 1957; Hurford, 2012). This duality of patterning has been suggested to be the key to the open-ended combinatorial generativity seen in human languages, such that

* Corresponding author.
E-mail address: gabriella.gall@ab.mpg.de (G. E. C. Gall).

https://doi.org/10.1016/j.anbehav.2025.123431

new words and sentences can be produced from the available elements (Russell & Townsend, 2017; Zuberbühler, 2019).

In nonhuman animals, the acoustic combinatoriality detected so far is simpler. Examples include phonocoding, the combination of meaningless sound elements into sequences or songs; multielement calls, the combination of meaningless sound elements into functionally relevant vocalizations; segmental concatenation, the combination of acoustic segments in a seemingly systematic way; and idiomatic structures, an assemblage of discrete acoustic units into larger sequences, generating information that is unrelated to the meaning of the individual acoustic segments (Engesser & Townsend, 2019). In some animals, combinatoriality is reminiscent of phonology, where meaningless elements are combined into meaningful calls. For instance, the functionally distinct flight call and prompt call of Chestnut-crowned babblers, *Pomatostomus ruficeps*, are multielement calls, composed of the same perceptibly distinct but meaningless elements (Engesser et al., 2019). A simple form of compositionality has been described for pied babblers, *Turdoides bicolor* (Engesser et al., 2016) and Japanese tits, *Parus minor* (Suzuki et al., 2016). Both species combine alert and recruitment calls into a composite mobbing signal, whereby the meaning of the combination is derived from the meaning of the comprising calls (Engesser et al., 2016; Suzuki et al., 2016). Although this form of combinatoriality shows some similarity to linguistic structures in human language (Engesser & Townsend, 2019), it seems to lack the open-ended generativity of human language.

One of the challenges in understanding animal combinatoriality is that, in the past, many studies have delineated the acoustic units using silence intervals as delimiters, with the most basic unit being uninterrupted sound, disregarding any spectral changes within (Kershenbaum et al., 2016; Mann et al., 2021). This approach likely overlooked subtle but potentially meaningful combinations of sound elements and underestimated the combinatorial abilities of some species. Current research is adopting a more nuanced approach, annotating sound elements based on abrupt spectral shifts within continuous sounds (Leroux et al., 2021; Walsh et al., 2019, 2023). For example, Australian magpies, *Gymnorhina tibicen dorsalis*, combine four distinct sound elements to create a larger number of calls within their nonsong vocal repertoire (Walsh et al., 2023). These calls are further combined into larger sequences, displaying a potential for a multilevel combinatoriality (Walsh et al., 2023). This approach has been used especially for animal song (Berwick et al., 2011) and is becoming more frequent in the study of animal communication more broadly. In contrast to calls, songs are thought of as being louder, longer, and more complex and are assumed to be learnt to some extent (Rose et al., 2022), though the line between songs and calls in some birds is not as distinct as once thought (Rose et al., 2022). The current knowledge of multilevel combinatoriality in birds extends mostly to vocal learners (e.g. songbirds), with less being known about how nonlearners might combine calls and whether the function of the call sequences is derived from the order and the structure of the composing calls. The above approach allows for a more comprehensive investigation of the rules governing acoustic combinations, the potential for syntax-like structures in animal communication and their communicative significance across species.

Building on the principles of this refined approach to studying combinatoriality in animal calls, we investigated the potential for multilevel combinatoriality in the vocalizations of the Australasian swamphen, or pūkeko, *Porphyrio melanotus melanotus*. Pūkeko are large rails (~ 1 kg) occurring in New Zealand, where they live in social groups of 3–20 individuals (Dey et al., 2012; Jamieson, 1997). In pūkeko groups, all adults may breed, and if multiple females do breed, they lay eggs within the same nest (joint-laying). Although only breeding adults take turns incubating a clutch, all group members take care of chicks (Craig, 1980). Pūkeko use a range of displays (Craig, 1977) and highly variable vocalizations (Clapperton & Jenkins, 1984) to communicate during social interactions. A previous study identified 25 call types in nine broad categories, including contact calls, for example, the individually distinct n'yip (Clapperton & Jenkins, 1987) and hiccup (Clapperton & Jenkins, 1984), as well as territorial calls, such as crowing calls and three distinct types of yelling calls (Clapperton & Jenkins, 1984). Crowing calls were shown to be sexually dimorphic and, in the case of males, individually distinct (Clapperton, 1983; Clapperton & Jenkins, 1987), whereas yelling calls are used during aggressive displays between neighbours. Many of the calls of pūkeko are described as being composed of several distinct sound elements (for example, n'yip, n'yick, hiccup, crowing, yelling, flight call, crow-like calls, defence calls and food call), with some elements shared between call types (Clapperton & Jenkins, 1984).

This study aimed to evaluate whether and how pūkeko combine different sound elements into calls and whether and how calls are organized into structured sequences. To do so, we used recordings collected from individual pūkeko nests, which are located towards the centre of the small group territories (from 0.7 to 3 ha) (Craig, 1976). Thus, we were likely to record most group members with minimal amounts of stress to the birds. From these data, we evaluated the distinctiveness of the sound elements comprising the adult pūkeko repertoire. We then investigated how the birds combine sound elements into sound combinations, or 'calls', and explored the sequential arrangement of these calls into larger sequences. Throughout the text, we use the following definitions: 'sound element', similar spectral structure; 'call', a sequence of sounds separated by silence (made up of one or more sound elements); and 'call sequence' (or call bout), intracall intervals are shorter than intercall intervals.

METHODS

Data for this study were collected at Tāwharanui Regional Park (−36.371520°, 174.829835°) on the North Island of New Zealand, between October 5 and November 17, 2022. At the time of data collection, some individuals in focal groups and many nonfocal groups were unbanded. Thus, determining individual and potentially group identity would have been difficult when collecting focal audio recordings. In addition, such recordings would have had to be collected from within a blind and at great distances from the vocalizing birds. Because we were also especially interested in the vocalizations of chicks at the nest for a separate project, we decided to record the calls of adults and chicks (though chicks were not analysed for this study) from within a pūkeko group's nest. We found nests by watching for nesting behaviour and by searching for suitable habitats, close to water and vegetative cover (Craig, 1980; Craig & Jamieson, 1990). We placed a miniature audio recorder (Soroka 18E, TS Market) wrapped in parafilm about 2 cm underneath the eggs within the nest bowl. We programmed the audio units for continuous 24 h recording at either 16 kHz, 32-bit or 24 kHz, 32-bit uncompressed WAV files. Pūkeko are accustomed to frequent nest checks at this site and quickly resumed normal behaviours after this disturbance. Sampled nests were visited on consecutive days to retrieve the recorder and replace it with a new and fully charged one. Overall, we recorded the soundscape at 19 different nests, with clutches of varying sizes, composition and incubation stages. Each nest was recorded between 1 and 7 days continuously (median = 2 days), with one nest being recorded only for ~3 h and another only for 7 h, due to recorder malfunction (see Table S1 for details). Nests with eggs about to hatch or with new hatchlings were recorded for longer periods of time for a

different project to obtain more detailed information on this critical moment. We did not know individual identity and did not account for nest identity. Similarly, we did not know the actual context for any of the calls due to the recording method. However, we used a previously published repertoire to distinguish between adult, juvenile and chick vocalizations and were able to infer some context due to the type of vocalization emitted (for example, territorial calls; Clapperton & Jenkins, 1984). Here, we only focused on the vocalizations of adult callers at or near the nest.

Ethical Note

This research was approved and facilitated by the Ngāti Manuhiri and the Department of Conservation, New Zealand (Wildlife Act Authority No. 102034-FAU). This research was further approved by the McMaster University Ethics Committee (AUP# 21-12-38) and was in line with the ASAB Ethical Committee/ABS Animal Care Committee Guidelines. Banders were qualified under the New Zealand National Bird Banding Scheme. We recorded at 19 nests as we tried to cover three different stages of incubation: early incubation/laying, mid-incubation and pipping/hatching. We dropped microphones into a nest or retrieved them from nests during regular nest checks, which were part of a different project. For this, nests were visited once per day (during laying and hatching) to check for new eggs or newly hatched chicks and every 3 days for nests that were in the middle of incubation. All preparations for the recordings (turning microphones on, wrapping in parafilm, etc.) were conducted before approaching the nest to minimize the disturbance time at the nest.

Sound Annotations

We viewed and manually annotated the recorded audio files in Adobe Audition 2021 (Adobe Inc., San Jose, CA, U.S.A.), with spectrograms generated using a 1024-point FFT size, Blackman-Harris window. Following studies by Hedwig and Kohlberg (2024); Kershenbaum et al. (2016); and Walsh et al. (2024), we annotated 'sound elements', characterized as a section of pūkeko sound either separated by a 'silent gap' (maximum 0.08 s) from another sound or by an abrupt spectral shift; for example, the sound may become noisy. We identified and categorized sound elements manually by visual–auditory inspection and annotated them using the Adobe Audition 'markers' function. Sound annotations were done by GECG and two trained assistants. The annotations of the two assistants were verified by GECG and then assessed using uniform manifold approximation and projection for dimension reduction (UMAP) (see below) to ensure sound distinctiveness. We annotated the recordings from 17 of the 19 sampled nests using two different annotation schemes: (1) in six nests (see Table S1), we annotated ~13 h of continuous recording (N = 11 034 sound elements, Table S2), and (2) in 11 nests, we selectively annotated rarer sounds to increase the sample size in the data set for sound distinctiveness analysis. For the spectral analysis, we visually examined all annotated sounds and selected 5465 sound element exemplars with a high signal-to-noise ratio from 13 different sound types across 17 nests (Table S3). The last two nests were not used for call supplementing, as they included hatchling chicks, making it more difficult to find noise-free adult calls.

Distinctiveness of Sound Elements

To verify the manual annotation and to confirm that the sound element types are spectrally distinct, we used a spectrogram-based UMAP dimensionality reduction approach following the methods outlined in Thomas et al. (2022) with custom modifications to account for acoustic parameters specific to pūkeko. (1) We filtered all recordings using a Butterworth bandpass filter set at a 0.5 to 7 kHz range to encompass most of the spectral content across all sound elements. (2) We generated spectrograms using the absolute value taken from a one-sided short-time Fourier transformation. To logarithmically scale the spectrogram, we used a Mel filter set at the same frequency range as used in the Butterworth bandpass filter. (3) We used the UMAP to project spectrograms into a three-dimensional latent space where each time-frequency bin was treated as an independent dimension, using the Python package umap-learn (McInnes et al., 2018) and the code provided by Thomas et al. (2022). As an alternative validation, we also verified these results by training a Random Forest Classifier using Mel-frequency-cepstral-coefficients (see Supplementary Materials Section C).

To evaluate to what degree similarly annotated sound elements are grouped in latent space, we assessed the manual label of the 40 nearest neighbours of each datapoint in the latent space. We chose 40 nearest neighbours for our comparisons, as it was approximately a 50% sample of our rarest sound element ('screach', N = 75; Table 2). We created an evaluation matrix, where each cell represents the probability for data points of label A to have a

Table 1
Frequency distribution of the number of elements contained in the various calls

No. of elements in call	Total no. of calls
1	3507
2	1128
3	961
4	323
5	119
6	20
7	5
8	8
9	4
10	2
11	5
12	4
13	2
16	2
19	2
21	1

Table 2
Number of annotated sound elements used for sound combination and sequence analysis

Sound element	Total no. of sound elements	*N* emitted as unigram	*N* emitted as unigram odds ratio
a.squeek	829	5	−3.97
brr	939	853	1.05
brumm	700	43	−1.64
bup	2081	1136	0.54
buzz	984	665	0.75
di	1154	477	0.26
drr	317	28	−1.28
screach	47	18	−4.23
scream	431	2	−4.23
snort	403	27	−1.56
squak	1392	205	−0.77
squeek	1001	41	−2.05
tut	756	7	−3.54

Positive odds ratios indicate sound elements that are more likely to be emitted as unigrams, zero indicates that sounds are emitted with the same frequency as unigrams, as expected by chance and negative odds ratio values indicate that sounds occur less frequently as unigrams than expected by chance.

neighbour with label B in latent space, given the observed frequencies of different labels among the 40 nearest neighbours (Thomas et al., 2022). To account for unbalanced sample sizes among the different sound elements, we normalized the probability of encountering same-class data points among the 40 nearest neighbours to the overall probability of encountering same-class data points. This score can be interpreted as the amount of increase in the likelihood of same-class neighbours over the random chance expectation. We then applied a log2 transformation, such that the score is symmetric around zero and can be interpreted as the amount of increase or decrease (positive or negative values, respectively) in likelihood over the random chance expectation: for example, 2 is twice as much as the random chance expectation, and −2 is half as much as the random chance expectation.

To visualize the degree to which different sound elements are found in close vicinity of each other within latent space, we constructed a UMAP-based neighbourhood graph in which signal types are presented as nodes connected with edges. The length of an edge represents the relative probability of the connected signal types being among their 40 nearest neighbours in the UMAP embedding. Such an approach can help identify sound types that constitute a gradient and sound types that are clearly distinct from each other. To compare our labels with UMAP-projected sounds, we constructed silhouette plots considering all manual labels on the resulting UMAP (Fig. S1a). Furthermore, we used hierarchical density-based spatial clustering of applications with noise (HDBSCAN) for all sound elements (Fig. S1b) to analyse how sounds were distributed within UMAP space.

From Sound Elements to Calls

We examined how sound elements of each type were used to build sound combinations (calls). The following analyses were conducted in R (version 4.3.2) using the annotated continuous data.

First, to test whether sound elements were more likely to be emitted as unigrams (that is, by themselves, not as parts of sound combinations), we calculated the probability for each sound type to occur as a unigram. We then computed the expected number of instances for sounds of each sound type to occur as unigrams by multiplying the probability of a sound being emitted by the total number of unigram sounds. Finally, for each sound, we reported the log-odds ratio by dividing the real frequency of unigram occurrence by the expected frequency of unigram occurrence. In these and other log-odds ratios described below, positive values indicate sounds occurring as unigrams more than expected by chance, zero indicates that sounds are emitted with the same frequency by themselves as expected by chance, and negative values indicate that sounds occur less frequently by themselves than expected by chance.

Different sound elements were considered to be within a sound combination (referred to as a 'call' in the further text) if two or more sound elements followed each other within 0.04 s. We chose this arbitrary time, as it was high enough to aggregate identified calls while accounting for differences in annotation precision and small enough not to clump together multiple different calls into one. To probe sound sequentially within calls, for each sound element type, we computed the probabilities that a sound of each other sound element type followed it. To compute these transition probabilities, we divided the sound combination into a sequence of overlapping bigrams (for example, for sound combination ABBCA, transitions would be A-B, B−B, B−C and C-A). To account for the different frequencies of each sound element in our data set, we calculated the probability of transitioning from sound element type i to sound element type j under the Markov assumption (that is, assuming that the decision of which sound element type to emit is contingent only on the sound type of the previous sound) as:

$$Pr(s_2 = j \mid s_1 = i) = \frac{Pr(s_1 = i \,\&\, s_2 = j)}{Pr(s_1 = i)}$$

with $Pr(s_1 = i \,\&\, s_2 = j) = \frac{n(ij)}{N_t}$ and $Pr(s_1 = i) = \frac{n(i)}{N}$. Here, $n(ij)$ is the number of transitions from sound i to sound j; N_t, the total number of transitions between sounds; $n(i)$, the total number of occurrences of sound i; and N, the total number of sounds in the data set. Calculating these transitions for each possible pair of sound types i and j, completed a matrix of transition probabilities, interpreted here as a directed network where each sound type is represented as a node, and the transition probabilities between two sound types are represented as the edges connecting them. As interpretable characteristics of these networks, we calculated betweenness centrality for each node, which measures the number of shortest paths connecting different nodes passing through a specific node. This corresponds to measuring the extent to which, for a given sound element, sound combinations or calls contain a sound of that element. We further calculated the ratio between nodes' in- and out-degrees, which indicates whether a specific node or sound element is more likely to be at the start (prefix) or end (suffix) of a combination of sound elements.

To establish more directly whether sounds from each sound type are more likely to serve as prefixes (i.e. beginnings) or suffixes (i.e. endings) of sound combinations, we calculated the odds ratio for sound elements of each sound type to be at the start, in the middle, or at the end of sound combinations. We first calculated the number of times sounds of each sound type occur either at the start, the middle, or the end of a sound combination. We included in our analyses sound combinations composed of up to nine elements. We calculated the expected number of sounds occurring in each position within the sound combination under the assumption that sounds are emitted at random, that is, a sound element would have an equal probability to occur at any location within the sound combination. Finally, we used the real and expected number of occurrences at each position to obtain the log-odds ratio. In addition, to assess whether these sequential preferences are stable independent of the length of a sound combination, we separately calculated these log-odds ratios for sequences of length two to five (for longer sequences, we did not have enough examples to calculate reasonable odds ratios).

From Calls to Call Sequences

Pūkeko emit some calls within larger sequences, where each call is emitted within a certain time lag from the previous call. We considered calls to be part of a call sequence if they were emitted within 5 s of each other. We decided on a 5 s interval, as follows: a visual inspection of our recordings. This time interval led to the reliable identification of individual call bouts without mixing separate bouts. Call sequences differ by the types of calls, as distinguished by their sound-element composition, emitted within them. We used these differences to categorize call sequences into six types, generally following the categories laid out by Clapperton and Jenkins (1984). However, we focused predominantly on the sound elements included within these sequences and ended up with two types of yelling sequences (including mostly 'tut', 'a.squeek', 'brumm' or 'scream'), hiccup sequences (including snorts), two types of food call sequences (including only 'bup' or mostly 'bup and di' elements) and 'n'yip' sequences (including mostly squeeks and squaks). In the main text, we explored yelling sequences that included tut-a.squeek-brumm

calls, as these were some of the least variable and for which we had the highest sample size (N= 181; see also Table S5). In addition, these are the sequences for which the context is the clearest for interpretation, given our recording setup. Details for the other sequences can be found in the supplementary material (Section D, Figs. S5–8).

For the identified call sequences we investigated (1) the correlation between the number of sound elements in a call and the relative call position within the sequence (earlier-later); (2) the correlation between the number of sound elements within a call and call duration; and (3) the intercall interval (measured from the start of each call to the start of the consecutive call) across the sequence. For each of these analyses, we used Kendall's rank correlation, as the residuals of the data were not normally distributed, and we could not fit generalized linear models. In the next step, we determined the probability of where specific sound elements appear within the call sequence, that is, in the first call, the middle calls or the last call of a sequence.

RESULTS

We identified 13 different sound elements for adult pūkeko in our recordings (Fig. 1). The identified sound elements did not separate into distinct clusters in the UMAP space but formed overlapping clusters (Fig. 2a, Fig. S1), suggesting a partial gradient of the sound element acoustic structures. Such a partial gradient is also supported by silhouette scores (Fig. S1a), which show that certain sound element types had significant overlap with their nearest sound element types. In addition, HDBSCAN (Fig. S1b) could not assign a large number of sounds to detected clusters. Nevertheless, a comparison of the 40 nearest neighbours of each sound element within latent space indicates that each sound element type is more likely to be surrounded by other sound elements of the same type than by other sound elements (Fig. S1b). Our separate supervised learning approach (a random forest model), also based on spectrograms, similarly struggled to distinguish between these edge cases (Fig. S2). This suggests that information needed to cleanly distinguish these sound-element types is not always present in the spectrograms.

From Sound Elements to Calls

In the analysed continuous recordings from six nests, 31.8% of sounds were emitted as a unigram; the rest were found within a combination of sound elements of up to 21 elements (Table 1). Specifically, we found brr, bup, buzz, di and screach to be overrepresented as unigrams, whereas a.squeek, brumm, drr, scream, snort, squak, squeek and tut were more likely to be emitted together with other elements than expected by chance (Table 2).

Considering paired sounds (bigrams), while some sound elements were very likely to be repeated (that is, 'bup', 'di', 'brr' and 'buzz'), most sound elements co-occurred with specific other sound elements and in a certain order. For example, tut-a.squeek appeared as a stable pair and never in the reversed order. The tut-a.squeek construct was most likely followed by 'squeek' or 'squak' and unlikely to be directly followed by any other sound element (Fig. 3a). The structure of the sound sequences is further highlighted by the betweenness centrality (Table 3), with six of the 13 sound elements (screach, drr, scream, tut, brumm and buzz) having a betweenness centrality of 0, indicating that no shortest path to other elements passes through these sound elements and that they are mostly at the beginning and end of the sequence or as repeated elements of the same element type. The other seven sound elements have values ranging from 51 (squak) to 11 (snort), indicating their importance in appearing in the middle part of a sequence, connecting other sound elements. When calculating the ratio between a node's in- and out-degree (Table 3), we found clear indications for some sound elements to typically appear as an prefix (tut, screach), whereas others seem to appear as a suffix (squeek, brumm, bup and drr).

To further confirm whether different sound elements are more likely to appear as a prefix or suffix, we calculated the odds ratio for specific sound elements to be at the start (first element), in the middle, or at the end (last element) of a call. Our results here agree with our bigram analysis, showing that overall, there are stable prefixes (tut, squak, brr, buzz and di) and suffixes (bup, screach and brumm). Other elements appear more flexible within calls and can appear both in the middle and at the end (snort, squeek and drr), or only in the middle (scream and a.squeek) (Fig. 4). Teasing this apart by the number of elements included in each call reveals

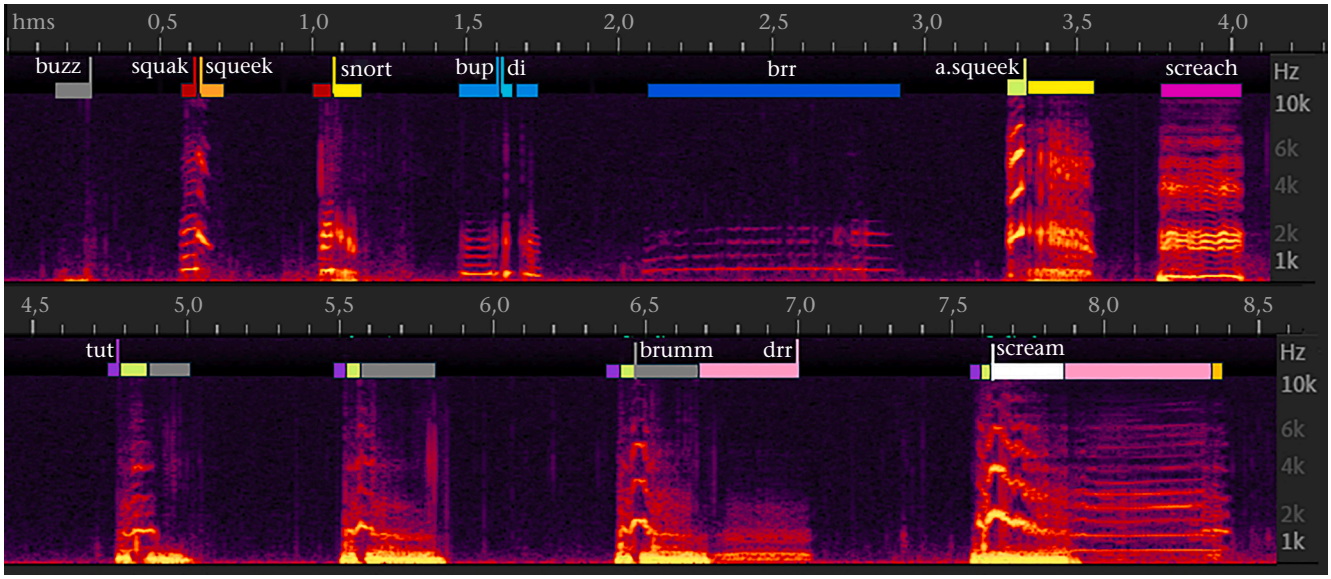

Figure 1. Examples of pūkeko calls with different sound elements. Sound elements are defined by smaller silent periods or sudden spectral shifts. In the bottom row is an example of a yelling sequence (with reduced intercall interval to improve visualization in this plot; the real mean intercall interval within yelling sequences is 2.94 s).

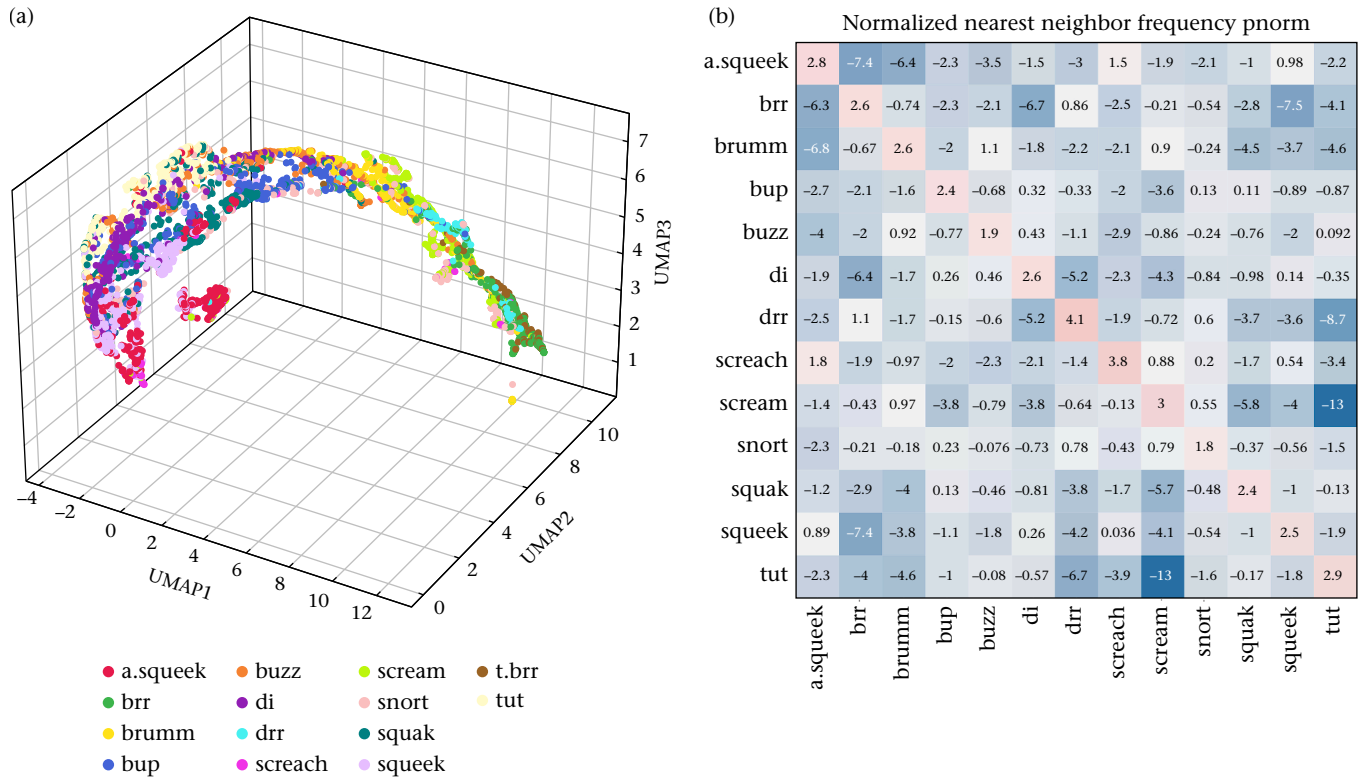

Figure 2. (a) Latent space representation of the different sound elements. (b) Log2-normalized nearest neighbour frequency of sound elements within latent space. Positive numbers (highlighted in red) indicate that two sounds are more likely to be each other's nearest neighbour than expected by chance, and negative values (blue) indicate that they are very unlikely to be found close to each other.

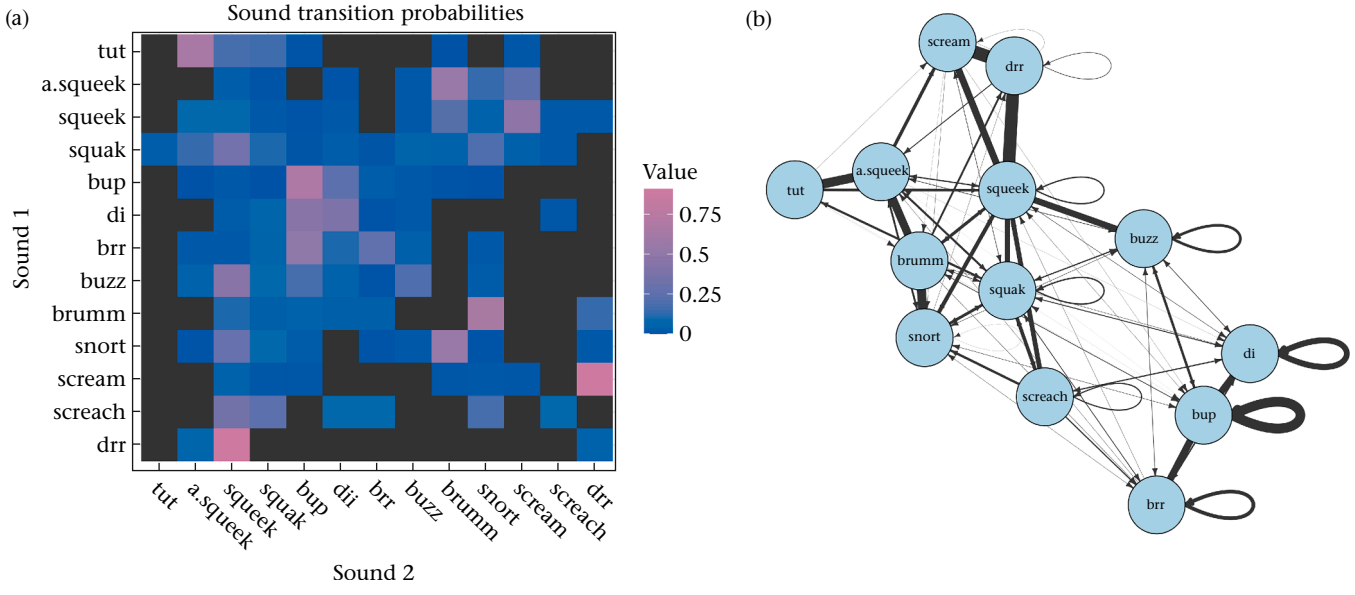

Figure 3. (a) Transition matrix between sound elements within sound combinations. Blue values indicate very unlikely transitions, whereas pink indicates very high transition probabilities. Grey values indicate that the probability of the transition occurring is 0. (b) A network calculated from the sound transition data, where each sound element is represented as a node and the transition probabilities between sounds as directed weighted edges.

Table 3
Network metrics for the bigram transitions for sound combinations

Sound element	Betweenness centrality	In-degree/out-degree
bup	47	1.19
buzz	0	0.36
squak	51	0.84
squeek	12	2.84
di	13	0.96
brr	30	0.41
snort	11	1.25
a.squeek	42	1.01
brumm	0	1.40
tut	0	0.03
scream	0	0.77
drr	0	1.11
screach	0	0.11

Shown are betweenness centrality scores and the ratio between a node's in- and out-degree. Betweenness centrality measures the number of shortest paths connecting different nodes and passing through a specific node. The ratio between a node's in- and out-degree indicates whether a specific node or sound element is more likely to be an prefix or a suffix, for example, at the start or end of a combination of sound elements; that is, low ratios indicate that a sound is an prefix, whereas high ratios indicate that a sound is a suffix.

more detailed structural rules. Some sounds are very stable, whereas others shift position within a sound sequence or appear/disappear, depending on the total number of elements (Fig. S3). For instance, all sounds can appear in two-element calls, but some elements (screach) never appear in calls of length 3 or 4. The position (start, middle and end) of elements becomes much more fixed within longer calls; for example, tut and squak are very stable prefixes across different sequence lengths, a.squeek appears within the middle of a call independent of sequence length, whereas brumm, scream and drr shift from the end to end towards the middle when occurring within longer calls. Some sounds (e.g. bup, di and buzz) are distributed somewhat randomly within the sound sequences that they appear within (see Fig. S3).

From Calls to Call Sequences

We explored the structure of yelling call sequences that include tut-a.squeek-brumm calls. We found a positive correlation between the number of sound elements within a call and the normalized position of a call within a call sequence, such that calls at the beginning of a sequence had fewer elements than calls at the end of a yelling sequence (Kendall's rank correlation: $z = 13.296$, tau $= 0.45626$, $P < 0.001$; Fig. 5a); this led to calls in longer sequences being on average longer than calls in shorter sequences (Fig. S4). Additionally, we found that calls with more elements were longer in duration (Kendall's rank correlation: $z = 20.792$, tau $= 0.671$, $P < 0.001$, Fig. 5b). Calls within a yelling sequence were emitted on average every 2.94 s. However, the intercall interval was not constant; instead, the interval decreased such that calls towards the end of a yelling sequence were emitted at shorter intervals than the calls at the beginning of a sequence (Kendall's rank correlation: $z = -3.250$, tau $= -0.102$, $P = 0.001$; Fig. 5c). Within yelling sequences, we only found 9 of the 13 identified sound elements (Fig. 5d). Several sound elements tended to appear at a certain phase of the call sequence while being completely absent from others (Fig. 5d). Overall, we found that with the progression of the yelling sequences, the calls are produced at a faster rate, becoming longer and changing in terms of their structure (more sound elements and more variable sound element composition).

DISCUSSION

Defining a basic communicative unit is conceptually challenging, and the annotation of animal vocalizations is easier and faster when they are delimited by silence. However, using silence to define units can lead to bias and overlook fine syntactic structures within uninterrupted sound (calls) that potentially indicate combinatorial complexity (Kershenbaum et al., 2016;

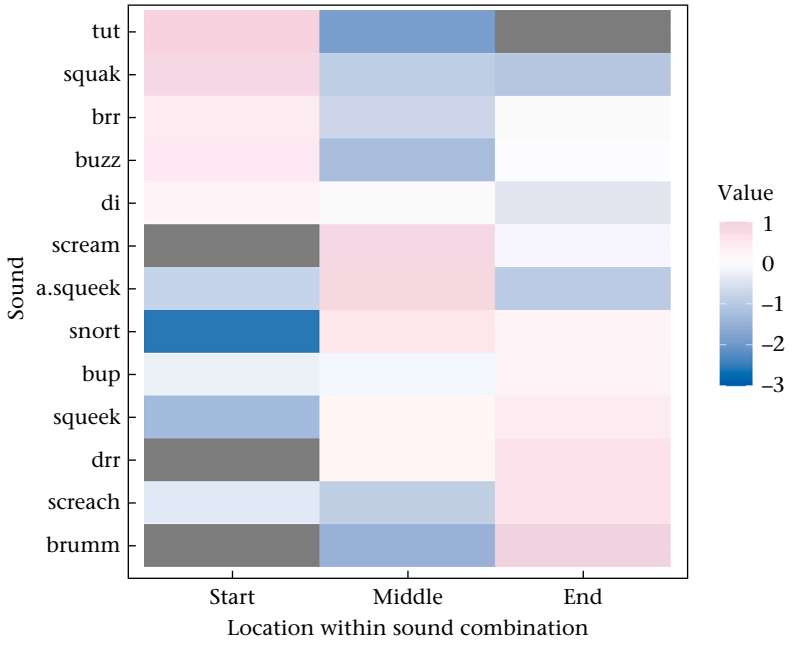

Figure 4. Location of sound elements within sound combinations for all sound combinations combined. Presented are the log-odds ratios of each sound being at the start, middle, and end of a call, respectively. Negative values (in blue) indicate sounds being underrepresented at a specific location, 0 (white) indicates that sound elements appear at a location similar to the chance expectation, whereas positive values (in red) indicate that sound elements are more likely to be found at that location than expected by chance. Grey cells indicate that the sound element never appeared at that location.

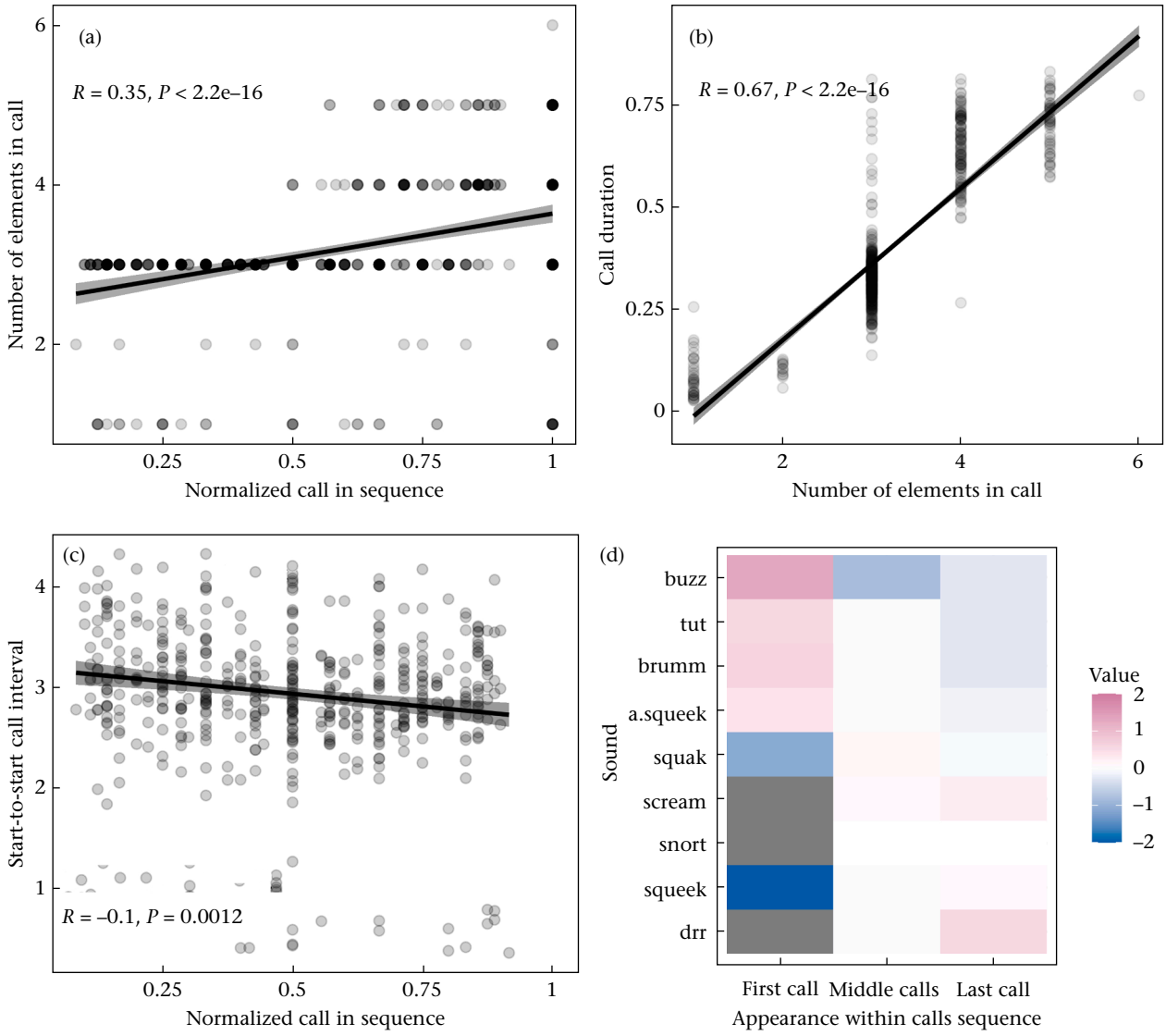

Figure 5. (a) Correlation between number of elements in a call and the position of the call within a sequence; (b) correlation between the duration of calls and the number of elements they are made up of; (c) correlation between intercall interval and position of calls within a sequence; (d) appearance of specific sound elements within the calls within yelling call sequences. Negative values (in blue) indicate sounds being underrepresented within a specific location, 0 (white) indicates that sound elements appear at a location near the chance expectation, whereas positive values (in red) indicate that sound elements are more likely to be found at that location than expected by chance. Grey cells indicate that the sound element never appeared within that location.

Mann et al., 2021). Our study reinforces this view by demonstrating that the calls of a nonvocal learning species, the pūkeko, are composed from 13 basic sound elements that form the building blocks of its acoustic repertoire. We took a systematic and hierarchical approach like the one used in human language and defined sound elements, calls and call sequences while attempting to also resolve the organizational rules at each of the levels. This approach showed that stereotypic organizational patterns appear at all levels.

Certain sounds were mostly produced as unigrams, suggesting that they bear independent communicative functions. Other sounds were mostly produced as a part of multicomponent calls, acting as prefixes, suffixes or connecting elements. Physical constraints of the vocal apparatus (Demery et al., 2021) may affect whether a specific sound element serves as a prefix, suffix or connecting element, such that some sounds might be easier to produce at the start or end of a call. None the less, sound elements that rarely/never appear as unigrams are likely not to have independent functions but rather to serve as modifiers to the information encoded in the core sound elements. What the specific function and modifications are for pūkeko sound elements is currently unknown and will have to be tested with specifically designed experiments. Here, the distinction of sound elements and their position within common vocalizations can aid in deciphering the meaning of signals. Specifically, the information may serve as a baseline for signal manipulation experiments, in which specific elements could be switched or removed before playing them back to birds in order to determine their function (Engesser et al., 2019).

Higher-level structural combinatoriality (call sequences) showed stereotypic progression of signalling events. For instance, in yelling bouts, later calls in the bout were longer, produced faster, contained more sound elements and were more variable in sound element composition than earlier calls within yelling bouts. Such patterns are typically shown and studied in the context of songs

(for example, birds, gibbons, whales, hyraxes and humans) (Berwick et al., 2011; Cannon, 2023; Clarke et al., 2006; Demartsev et al., 2017). For instance, singing is driven by motor constraints (Rajan & Doupe, 2013), whereby birds may 'warm up' to improve performance over time. Similarly, pūkeko may warm up during calling sequences, starting with simpler calls and increasing complexity with progressive calling. Additionally, in both human and animal songs, such dynamic progression songs have been suggested to prevent habituation and to maintain the attentiveness of receivers (Demartsev et al., 2023; Hughes, 2011; Osborn, 2013; Rothenberg et al., 2014). Although call sequences are not equivalent to songs, similar processes might apply, especially in long-range territorial or advertisement call sequences, such as the yelling call sequences analysed here.

The shift in syntactic structure along the call sequences (inclusion/omission of sounds depending on the position of the call) can imply dynamic informational content of a call sequence, such that information encoded in later calls relies on the context set by the earlier ones. Changes in call structure could also indicate changes in caller arousal states, such that longer and more variable calls indicate increased arousal. In primates, high-arousal calls are longer (Liao et al., 2018; Ordóñez-Gómez et al., 2023; Rendall, 2003; Slocombe & Zuberbü, 2007) and have a higher fundamental frequency (Hz) than low-arousal calls (Fichtel et al., 2001; Schrader & Todt, 1993; Schwartz et al., 2022; & Zuberbühler, 2007), though see (Ordóñez-Gómez et al., 2023). We found pūkeko yelling calls to increase in duration and to include more high-pitched sound elements (e.g. scream instead of brumm) in later calls within calling bouts. Thus, we speculate that pūkeko might become more aroused as the bout proceeds. As pūkeko yelling calls are part of territorial displays, callers might increase their call rate and the complexity of their calls when the threat of 'invading' neighbours is higher. Certain sounds were most likely to appear at the start of calls (prefix) and generally at the beginning of call sequences. Such a position of the sound elements could indicate an alerting function recruiting receivers' attentional energy before the main messaging components of the signalling sequence. The alerting function is often achieved through the evocative or degradation-resistant acoustic structure of the vocal units (Demartsev et al., 2025; Ord & Stamps, 2008).

The increase in duration of calls along calling sequences stands in contrast to Menzerath's Law, which suggests that longer call sequences are composed of shorter elements (Menzerath, 1954). Menzareth law has been observed in multiple avian and mammalian systems with songs ranging from shorter to longer than an average pūkeko yelling sequence; for example, gibbons (Huang et al., 2020), penguins (Favaro et al., 2020), zebra finches (James et al., 2021) and, most recently, 11 whale species (Youngblood, 2025). Such communicative efficiency is hypothesized to be driven by minimizing articulation effort and vocal production costs (Gustison et al., 2016). Thus, the contrasting results found in the pūkeko sequences are somewhat unexpected. However, animal communication does not always adhere to the compression (Zipf, 2016) and the physical efficiency drivers originally defined for human languages (Gibson et al., 2019). The effort-determining factor in vocal production can differ across species (Demartsev et al., 2019), and selection can drive signals away from theoretical efficiency due to unique ecological or functional constraints (Clink et al., 2020). The aggressive context of the yelling sequences would suggest that conveying honest information about the individual traits and reducing signal ambiguity might take functional precedence over communicative compression and economy. Aggressive and territorial displays are often exaggerated to facilitate detection (Smith & Evans, 2011) or as a handicap (Zahavi, 1977), with both processes potentially acting in opposition to Menzareth' law and minimizing the signalling costs.

We identified 13 sound elements and how they are organized on two hierarchical levels, suggesting a sophisticated and structured communication system in adult pūkeko, reminiscent of the one described previously for Australian magpies (Walsh et al., 2023). However, we currently lack detailed knowledge of the meaning and function of sound elements and calls and are unable to specify the type of combinatoriality pūkeko actually show (Engesser & Townsend, 2019). For instance, we do not know whether all sound elements are meaningless or whether sound elements appearing predominantly as unigrams contain separate meaning, though our results suggest that this is likely the case. Accordingly, we do not know whether pūkeko show phonocoding, multielement calls or neither (Engesser & Townsend, 2019). Similarly, we do not know whether variation in call composition and structure along yelling call sequences or other call sequences has any functional significance. It remains to be shown if pūkeko call sequences could resemble the call combinations of chimpanzees who combine pant hoots and food calls in a structured way within highly specific behavioural contexts (LeRoux et al., 2021) or pied babblers and Japanese tits who combine alert and recruitment calls into mobbing sequences (Engesser et al., 2016; Suzuki et al., 2016). Much of the focus of studies on combinatoriality within vocal systems in birds has focused on species that are considered vocal learners, for example, songbirds. Our study highlights the fact that a bird species that is not known to learn (parts of) its repertoire may also show hierarchically structured communication systems similar to primates.

Although our recordings were restricted to the breeding season and limited to the immediate vicinity of the nest, our recording methodology generated a more extensive and uniform sample of vocalizations than would have been feasible using manual audio recordings. Although we likely missed less frequent calls and sound elements that occur in other contexts or seasons (e.g. mating calls), our findings establish a fundamental framework for understanding the hierarchical and structural complexity of pūkeko vocalizations. This presents opportunities for future research to expand upon our foundational acoustic framework. We note that the sound elements emitted by pūkeko are on a gradient (Fig. 2a). Both UMAP and random forest analysis independently demonstrate broad agreement with our manual classification; however, some sound element types are consistently misidentified. Human annotators take into account the visual spectrographic representations of the sounds but also the audible sound itself. Additionally, humans may perceive changes in sound due to the distance to the recording device, as well as coarticulation effects resulting from transitions between different sound types. Consequently, humans may distinguish finer details between certain sound element types or, conversely, group variable sounds into the same type. In such instances, one way to confirm whether such confounding pairs of sound element types are indeed separate classes is through playback experiments (Engesser et al., 2019).

As we exclusively focus on adult vocalizations, an exciting avenue for a follow-up study would be examining how chick vocalizations, which comprise their own diverse acoustic repertoire, eventually transform into adult forms, potentially involving similar or entirely different sound elements and combinatorial rules. The ontogenetic trajectory of these vocalizations could provide critical insights into both learning and innate components of avian vocalizations (Loo & Cain, 2021). Although we were unable to identify individual vocalizers, future work could focus on individual differences in acoustic production, accounting for factors such as sex, status and age. For instance, certain vocalizations

may be produced primarily by males during territory defense or by dominant females to defend breeding status, or the complexity of combinations may increase with age as juveniles gradually master the full adult repertoire.

The hierarchical combinatorial structure documented in our study represents a fundamental contribution to understanding how complexity emerges in animal communication systems. The observation that most sound elements rarely appear in isolation but instead participate in various combinations across different calls suggests that selection may have favored flexibility in information encoding rather than an expanded inventory of acoustically distinct signals. The structural parallels between this system and those found in other species, from primates to songbirds, provide an opportunity to examine convergent evolutionary processes in communication. Looking forward, this acoustic framework we have established opens numerous research directions, including playback experiments to test the function and perception of specific sound elements and their combinations, comparative studies across populations and related species to trace evolutionary trajectories and detailed investigations into how social complexity may drive or constrain communicative complexity.

Author Contributions

Gabriella E.C. Gall: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Vlad Demartsev:** Writing – review & editing, Visualization, Methodology, Formal analysis, Conceptualization. **Pranav Minasandra:** Writing – review & editing, Methodology. **Cecilia Baldoni:** Writing – review & editing, Validation. **Kristal E. Cain:** Writing – review & editing, Conceptualization. **James S. Quinn:** Writing – review & editing, Project administration, Investigation.

Data Availability

Data and code are available as supplementary material.

Declaration of Interest

We declare that we have no competing interests.

Acknowledgments

We thank the Tāwharanui Open Sanctuary Society and Auckland Council staff, especially Larissa Bickers Cherrie and Matt Maitland. We further thank Quinlan Mann for assistance with data collection, Felix Greisinger and Ersila Xhakollari for their help annotating sound files, Ariana Strandburg-Peshkin for helpful discussion of the analysis, Carel van Schaik for helpful and insightful discussion of the topic more generally and two anonymous reviewers for helpful comments. This study was funded through a Postdoctoral fellowship to GECG from the Zukunftskolleg and the Centre for the Advanced Study of Collective Behaviour at the University of Konstanz, which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2117–422037984, as well as a small grant from the Centre for the Advanced Study of Collective Behaviour, University of Konstanz to GECG.

Supplementary Material

Supplementary material associated with this article is available at https://doi.org/10.1016/j.anbehav.2025.123431.

References

Arnold, K., & Zuberbühler, K. (2006). Language evolution: Semantic combinations in primate calls. *Nature, 441*(7091), 303. https://doi.org/10.1038/441303a

Berwick, R. C., Okanoya, K., Beckers, G. J. L., & Bolhuis, J. J. (2011). Songs to syntax: The linguistics of birdsong. *Trends in Cognitive Sciences, 15*(3), 113–121. https://doi.org/10.1016/j.tics.2011.01.002

Bohn, K. M., Moss, C. F., & Wilkinson, G. S. (2006). Correlated evolution between hearing sensitivity and social calls in bats. *Biology Letters, 2*(4), 561–564. https://doi.org/10.1098/rsbl.2006.0501

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of Animal Communication* (Sinauer).

Cannon, C. (2023). A theoretical account of whale song syntax: A new perspective for understanding human language structure. *Proceedings of the Linguistic Society of America, 8*(1), 5571. https://doi.org/10.3765/plsa.v8i1.5571

Chomsky, N. (1957). *Syntactic structures*. Mouton. https://doi.org/10.1515/9783112316009

Clapperton, B. K. (1983). Sexual differences in pukeko calls. *Notornis, 30*(1), 69–70. https://doi.org/10.63172/908431megtre

Clapperton, B. K., & Jenkins, P. F. (1984). Vocal repertoire of the pukeko (Aves: Rallidae). *New Zealand Journal of Zoology, 11*(1), 71–84. https://doi.org/10.1080/03014223.1984.10428230

Clapperton, B. K., & Jenkins, P. F. (1987). Individuality in contact calls of the pukeko (Aves: Rallidae). *New Zealand Journal of Zoology, 14*(1), 19–28. https://doi.org/10.1080/03014223.1987.10422678

Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS One, 1*(1), Article e73. https://doi.org/10.1371/journal.pone.0000073

Clink, D. J., Ahmad, A. H., & Klinck, H. (2020). Brevity is not a universal in animal communication: Evidence for compression depends on the unit of analysis in small ape vocalizations. *Royal Society Open Science, 7*(4), Article 200151. https://doi.org/10.1098/rsos.200151

Craig, J. L. (1976). An interterritorial hierarchy: An advantage for a subordinate in a communal territory. *Zeitschrift für Tierpsychologie, 42*, 200–205.

Craig, J. L. (1977). The behaviour of the pukeko, porphyrio porphyrio melanotus. *New Zealand Journal of Zoology, 4*(4), 413–433. https://doi.org/10.1080/03014223.1977.9517966

Craig, J. L. (1980). Pair and group breeding behaviour of a communal gallinule, the pukeko, Porphyrio p. melanotus. *Animal Behaviour, 28*(2), 593–603.

Craig, J. L., & Jamieson, I. G. (1990). Pukeko: Different approaches and some different answers. In P. B. Stacey, & W. D. Koenig (Eds.), *Cooperative Breeding in Birds: Longterm Studies of Ecology and Behavior* (pp. 385–412). Cambridge University Press. https://doi.org/10.1017/CBO9780511752452.014.

Demartsev, V., Geva, Y., Alba-González, P., Koren, L., Ilany, A., & Geffen, E. (2025). Alerting components in animal vocalization. *Animal Behaviour, 230*, Article 123373. https://doi.org/10.1016/j.anbehav.2025.123373

Demartsev, V., Gordon, N., Barocas, A., Bar-Ziv, E., Ilany, T., Goll, Y., Ilany, A., & Geffen, E. (2019). The "Law of Brevity" in animal communication: Sex-specific signaling optimization is determined by call amplitude rather than duration. *Evolution Letters, 3*(6), 623–634. https://doi.org/10.1002/evl3.147

Demartsev, V., Haddas-Sasson, M., Ilany, A., Koren, L., & Geffen, E. (2023). Male rock hyraxes that maintain an isochronous song rhythm achieve higher reproductive success. *Journal of Animal Ecology, 92*(8), 1520–1531. https://doi.org/10.1111/1365-2656.13801

Demartsev, V., Ilany, A., Kershenbaum, A., Geva, Y., Margalit, O., Schnitzer, I., Barocas, A., Bar-Ziv, E., Koren, L., & Geffen, E. (2017). The progression pattern of male hyrax songs and the role of climactic ending. *Scientific Reports, 7*, 2794. https://doi.org/10.1038/s41598-017-03035-x

Demery, A. J. C., Burns, K. J., & Mason, N. A. (2021). Bill size, bill shape, and body size constrain bird song evolution on a macroevolutionary scale. *Ornithology, 138*(2). https://doi.org/10.1093/ornithology/ukab011

Dey, C. J., Jamieson, I. G., & Quinn, J. S. (2012). Reproductive skew and female trait elaboration in a cooperatively breeding rail. *Ibis, 154*(3), 452–460. https://doi.org/10.1111/j.1474-919X.2012.01223.x

Engesser, S., Holub, J. L., O'Neill, L. G., Russell, A. F., & Townsend, S. W. (2019). Chestnut-crowned babbler calls are composed of meaningless shared building blocks. *Proceedings of the National Academy of Sciences of the United States of America, 116*(39), 19579–19584. https://doi.org/10.1073/pnas.1819513116

Engesser, S., Ridley, A. R., & Townsend, S. W. (2016). Meaningful call combinations and compositional processing in the southern pied babbler. *Proceedings of the National Academy of Sciences, 113*(21), 5976–5981. https://doi.org/10.1073/pnas.1600970113

Engesser, S., & Townsend, S. W. (2019). Combinatoriality in the vocal systems of nonhuman animals. *Wiley Interdisciplinary Reviews: Cognitive Science, 10*(4), Article e1493. https://doi.org/10.1002/wcs.1493

Gibson, E., Futrell, R., Piandadosi, S. T., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive*

Sciences, 23(5), 389–407. Elsevier Ltd. https://doi.org/10.1016/j.tics.2019.02.003.

Favaro, L., Gamba, M., Cresta, E., Fumagalli, E., Bandoli, F., Pilenga, C., Isaja, V., Mathevon, N., & Reby, D. (2020). Do penguins' vocal sequences conform to linguistic laws? *Biology Letters, 16*(2), Article 20190589. https://doi.org/10.1098/rsbl.2019.0589

Fichtel, C., Hammerschmidt, K., & Jürgens, U. (2001). On the vocal expression of emotion. A multi-parametric analysis of different states of aversion in the squirrel monkey. *Behaviour, 138*(1), 97–116. https://doi.org/10.1163/156853901510670 94Gibson

Gustison, M. L., Semple, S., Ferrer-i-Cancho, R., & Bergman, T. J. (2016). Gelada vocal sequences follow Menzerath's linguistic law. *Proceedings of the National Academy of Sciences, 113*(19), E2750–E2758. https://doi.org/10.1073/pnas.1522072113

Hedwig, D., & Kohlberg, A. (2024). Call combination in African forest elephants *Loxodonta cyclotis*. *PLoS One, 19*(3), Article e0299656. https://doi.org/10.1371/journal.pone.0299656

Huang, M., Ma, H., Ma, C., Garber, P. A., & Fan, P. (2020). Male gibbon loud morning calls conform to Zipf's law of brevity and Menzerath's law: Insights into the origin of human language. *Animal Behaviour, 160*, 145–155. https://doi.org/10.1016/j.anbehav.2019.11.017

Hughes, A. L. (2011). Stereotyped and non-stereotyped features of the temporal patterning of singing sessions in the ovenbird Seiurus auricapillus. *Behavioural Processes, 87*, 165–170. https://doi.org/10.1016/j.beproc.2011.02.010

Hurford, J. (2012). *The origins of grammar*. Oxford University Press.

James, L. S., Mori, C., Wada, K., & Sakata, J. T. (2021). Phylogeny and mechanisms of shared hierarchical patterns in birdsong. *Current Biology, 31*(13), 2796–2808.e9. https://doi.org/10.1016/j.cub.2021.04.015

Jamieson, I. G. (1997). Testing reproductive skew models in a communally breeding bird, the pukeko, *Porhyrio porphyrio*. *Proceedings of the Royal Society B: Biological Sciences, 264*(1380), 335–340. https://doi.org/10.1098/rspb.1997.0048

Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., Bohn, K., Cao, Y., Carter, G., Cäsar, C., Coen, M., Deruiter, S. L., Doyle, L., Edelman, S., Ferrer-i-Cancho, R., Freeberg, T. M., Garland, E. C., Gustison, M., Harley, H. E., … Zamora-Gutierrez, V. (2016). Acoustic sequences in non-human animals: A tutorial review and prospectus. *Biological Reviews, 91*(1), 13–52. https://doi.org/10.1111/brv.12160

Leroux, M., Bosshard, A. B., Chandia, B., Manser, A., Zuberbühler, K., & Townsend, S. W. (2021). Chimpanzees combine pant hoots with food calls into larger structures. *Animal Behaviour, 179*, 41–50. https://doi.org/10.1016/j.anbehav.2021.06.026

Liao, D. A., Zhang, Y. S., Cai, L. X., & Ghazanfar, A. A. (2018). Internal states and extrinsic factors both determine monkey vocal production. *Proceedings of the National Academy of Sciences of the United States of America, 115*(15), 3978–3983. https://doi.org/10.1073/pnas.1722426115

Loo, Y. Y., & Cain, K. E. (2021). A call to expand avian vocal development research. *Frontiers in Ecology and Evolution, 9*, Article 757972. https://doi.org/10.3389/fevo.2021.757972

Mann, D. C., Fitch, W. T., Tu, H. W., & Hoeschele, M. (2021). Universal principles underlying segmental structures in parrot song and human speech. *Scientific Reports, 11*(1), 776. https://doi.org/10.1038/s41598-020-80340-y

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*. http://arxiv.org/abs/1802.03426.

Menzerath, P. (1954). *Die Architektonik des deutschen Wortschatzes. Dümmler, F.*

Montealegre-Z, F. (2009). Scale effects and constraints for sound production in katydids (Orthoptera: Tettigoniidae): Correlated evolution between morphology and signal parameters. *Journal of Evolutionary Biology, 22*(2), 355–366. https://doi.org/10.1111/j.1420-9101.2008.01652.x

Ordóñez-Gómez, J. D., Schamberg, I., & Hammerschmidt, K. (2023). The acoustic structure of spider monkey (Ateles geoffroyi) calls is related both to caller goal and arousal. *American Journal of Primatology, 85*(8), Article e23508. https://doi.org/10.1002/ajp.23508

Ord, T. J., & Stamps, J. A. (2008). Alert signals enhance animal communication in "noisy" environments. *Proceedings of the National Academy of Sciences, 105*(48), 18830–18835. www.pnas.orgcgidoi10.1073pnas.0807657105.

Osborn, B. (2013). Subverting the verse-chorus paradigm: Terminally climactic forms in recent rock music. *Music Theory Spectrum, 35*, 23–47. https://doi.org/10.1525/mts.2013.35.1.23

Rajan, R., & Doupe, A. J. (2013). Behavioral and neural signatures of readiness to initiate a learned motor sequence. *Current Biology, 23*(1), 87–93. https://doi.org/10.1016/j.cub.2012.11.040

Rendall, D. (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *Journal of the Acoustical Society of America, 113*(6), 3390–3402. https://doi.org/10.1121/1.1568942

Rose, E. M., Prior, N. H., & Ball, G. F. (2022). The singing question: Re-conceptualizing birdsong. *Biological Reviews, 97*(1), 326–342. https://doi.org/10.1111/brv.12800

Rothenberg, D., Roeske, T. C., Voss, H. U., Naguib, M., & Tchernichovski, O. (2014). Investigation of musicality in birdsong. *Hearing Research, 308*, 71–83. https://doi.org/10.1016/j.heares.2013.08.016

Russell, A. F., & Townsend, S. W. (2017). Communication: Animal steps on the road to syntax? *Current Biology. Cell Press, 27*(15), R753–R755. https://doi.org/10.1016/j.cub.2017.06.066

Schrader, L., & Todt, D. (1993). Contact call parameters covary with social context in common marmosets, Callithrix j. jacchus. *Animal Behaviour, 46*(5), 1026–1028.

Schwartz, J. W., Sanchez, M. M., & Gouzoules, H. (2022). Vocal expression of emotional arousal across two call types in young rhesus macaques. *Animal Behaviour, 190*, 125–138.

Slocombe, K. E., & Zuberbü, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences, 104*(43), 17228–172233. https://doi.org/10.1073/pnas.0706741104

Smith, C. L., & Evans, C. S. (2011). Exaggeration of display characteristics enhances detection of visual signals. *Behaviour, 148*(3), 287–305. https://doi.org/10.1163/000579511X556592

Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nature Communications, 7*(1), Article 10986. https://doi.org/10.1038/ncomms10986

Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2018). Call combinations in birds and the evolution of compositional syntax. *PLoS Biology, 16*(8), Article e2006532. https://doi.org/10.1371/journal.pbio.2006532

Thomas, M., Jensen, F. H., Averly, B., Demartsev, V., Manser, M. B., Sainburg, T., Roch, M. A., & Strandburg-Peshkin, A. (2022). A practical guide for generating unsupervised, spectrogram-based latent space representations of animal vocalizations. *Journal of Animal Ecology, 91*(8), 1567–1581. https://doi.org/10.1111/1365-2656.13754

Wallschläger, D. (1980). Correlation of song frequency and body weight in passerine birds. *Experientia, 36*(4), 412. https://doi.org/10.1007/BF01975119

Walsh, S. L., Engesser, S., Townsend, S. W., & Ridley, A. R. (2023). Multi-level combinatoriality in magpie non-song vocalizations. *Journal of the Royal Society Interface, 20*(199), Article 20220679. https://doi.org/10.1098/rsif.2022.0679

Walsh, S. L., Townsend, S. W., Engesser, S., & Ridley, A. R. (2024). Call combination production is linked to the social environment in Western Australian magpies (*Gymnorhina tibicen dorsalis*). *Philosophical Transactions of the Royal Society B: Biological Sciences, 379*, Article 20230198. https://doi.org/10.1098/rstb.2023.0198, 1905.

Walsh, S. L., Townsend, S. W., Morgan, K., & Ridley, A. R. (2019). Investigating the potential for call combinations in a lifelong vocal learner. *Ethology, 125*(6), 362–368. https://doi.org/10.1111/eth.12860

Youngblood, M. (2025). Language-like efficiency in whale communication. *Science Advances, 11*(6), eads6014. https://doi.org/10.1126/sciadv.ads6014

Zahavi, A. (1977). The cost of honesty (further remarks on the handicap principle). *Journal of Theoretical Biology, 67*(3), 603–605. https://doi.org/10.1016/0022-5193(77)90061-3

Zipf, G. K. (2016). Human behavior and the principle of least effort: An introduction to human ecology. *Ravenio books*.

Zuberbühler, K. (2019). Evolutionary roads to syntax. *Animal Behaviour Academic Press, 151*, 259–265. https://doi.org/10.1016/j.anbehav.2019.03.006