ORIGINAL PAPER

Note orders suggest phrase-inserting structure in male Mueller's gibbon songs: a case study

Yoichi Inoue[1] · Waidi Sinun[2] · Shigeto Yosida[3] · Kazuo Okanoya[1,4]

Abstract

Language has a complex and hierarchical structure, which includes context-free grammar and phrase embedding. However, there have been no reports of a phrase-embedding structure in animal vocal communication, although there are several reports of combinational animal sounds that reference different objects and emotions. The songs of male Mueller's gibbons in our study area consist of two notes, "wa" and "oo," and combinations of these are flexible. There are various types of note orders in their song phrases. When the animals sing songs by combining acoustic elements flexibly, a complicated syntax may emerge. When phrase "N" is inserted within another phrase "AB," the generated phrase is shown as "ANB." We named this structure a phrase-inserting structure. If the phrase "N" itself has an inserting structure elsewhere, "ANB" is shown as "AABB." This structure is considered a phrase-embedding structure, as defined by Abe and Watanabe (Nat Neurosci 14:1067–1074, 2011). We hypothesized that there is a phrase-inserting structure in male Mueller's gibbons' songs. We analyzed 70 songs by a single male and found note orders that suggested a phrase-inserting structure. Among them, there were four examples of phrases with structure within phrases, suggesting phrase embedding. Phrase inserting might be considered a sort of precursory level to recursion. Although our sample size was small, our results show that male Mueller's gibbons may be able to produce recursive sequences. This provides an important insight into the study of language evolution.

Keywords Phrase inserting · Phrase embedding · Mueller's gibbon · Song · Recursive

Introduction

The faculty of language is a cognitive system that supports the acquisition and use of certain languages (Pietroski and Crain 2012). Hauser et al. (2002) distinguished between the faculty of language in both the broad sense (FLB) and the narrow sense (FLN). FLB is based on mechanisms shared with non-human animals and includes all of the capacities that support language. It includes the sensory-motor system, the conceptual-intentional system, and the computational

✉ Yoichi Inoue
  gibbonyoichi@hotmail.co.jp

  Waidi Sinun
  wsinun@ysnst.org.my

  Shigeto Yosida
  yoshida.shigeto@gmail.com

  Kazuo Okanoya
  cokanoya@mail.ecc.u-tokyo.ac.jp

[1] Cognition and Behavior Joint Research Laboratory, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

[2] Borneo Refugia Adventure SDN BHD, Level 12, Tun Mustapha Tower Likas Bay Sulaman Road, P. O. Box 11623, 88817 Kota Kinabalu, Sabah, Malaysia

[3] IMS Japan K.K, IQVIA Solutions Japan K.K., Shinagawa Seaside South Tower, 4-12-1 Higashi-shinagawa, Shinagawa-ku, Tokyo 140-0002, Japan

[4] Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan

mechanisms. In contrast, FLN is uniquely human and recursion is the property that distinguishes human language from any other form of animal communication. Recursion is a computational procedure that calls itself or that calls an equivalent kind of procedure (Corballis 2007). In some generative theories of syntax, recursion is usually understood as self-embedding, in the sense of putting an object inside another of the same type (Fitch 2010; Kinsella 2010). Jackendoff and Pinker (2005) disagreed with this hypothesis by Hauser et al. (2002); they claimed that recursion is not the only "human-unique" trait. However, they agreed that true recursion, with hierarchical structures of unlimited depth, does not seem to be present in any other known animal communication system. One characteristic of recursion is a phrase-embedding structure, in which phrases are embedded within phrases. Although there are several reports of combinational animal sounds that reference different objects and emotions (Arnold and Zuberbühler 2006; Beer 1976; Clarke et al. 2006; Crockford and Boesch 2005; Geissmann et al. 2005; Hailman et al. 1985; Inoue et al. 2017; Mitani and Marler 1989; Ouattara et al. 2009; Robinson 1979, 1984; Zuberbühler 2002), there has been no evidence of a phrase-embedding structure in animal vocal communication (Hauser et al. 2014). As to whether animals perceive recursive patterns of sounds, Fitch and Hauser (2004) showed that cotton-top tamarins could not differentiate between finite state grammar and phrase structure grammar. Finite state grammar was defined as the strings of the sounds represented by $(AB)^n$, which generated sequences such as ABAB. In contrast, phrase structure grammar was defined as the embed strings within other strings, represented by A^nB^n, which generated center-embedded, "hierarchical," structures such as A-AB-B. Cotton-top tamarins could easily master the former, but failed to learn the latter. This leads us to question why monkeys cannot learn hierarchical structures. Fitch and Friederici (2012) suggested that recognizing A^nB^n requires both counting and comparing; this computation is difficult for non-human species. Furthermore, Friederici et al. (2006) pointed out that the processing of A^nB^n strings recruits Broca's area, whereas the processing of $(AB)^n$ strings relies on a phylogenetically older cortex, the frontal operculum. Broca's area is uniquely well developed in humans relative to non-human primates (Rilling et al. 2008).

In contrast, Gentner et al. (2006) and Abe and Watanabe (2011) reported that song birds (starlings and Bengalese finches) were able to recognize acoustic patterns defined by a recursive, self-embedding, context-free grammar. However, whether non-human animals can learn to recognize recursive patterns of sounds remains controversial (van Heijningen et al. 2009; Berwick et al. 2011; Beckers et al. 2012; Ten Cate and Okanoya 2012).

Gibbons (Hylobatidae) living in South-East Asia are a small ape, among the closest living relatives of humans. Gibbons are known for their remarkable vocal behavior.

They produce species-specific songs and have several repertoires of notes. Mueller's gibbons (*Hylobates muelleri*) live in the northern region of Borneo (Groves 1972; Marshall and Marshall 1976) and sing sex-specific songs. Males sing solos before or at dawn, and females occasionally sing solos after dawn, during the morning. In addition, males and females sing duets, usually in the morning. The songs of male Mueller's gibbons living in the Danum Valley Conservation Area (DVCA) consist of two note types: "wa" and "oo" (Inoue et al. 2017). The male's songs are long and complex, comprising many multi-note phrases, in which both notes are ordered in various combinations. The males sometimes sing inter- and intragroup antiphonal songs (Inoue et al. 2013). Although the notes of gibbon songs are believed to be genetically determined (Brockelman and Schilling 1984; Geissmann 1984), male Mueller's gibbons' songs may have combinatory rules and chunk structures (Inoue et al. 2017). Terleph et al. (2018) also highlighted the complexity and flexibility of gibbon songs and showed that particular phrase features likely arose from sexual selection pressures and possess similarities to human speech rhythm. We hypothesized that there are phrase-inserting structures, which are considered a sort of precursory level to recursion, in male Mueller's gibbons' songs, and recorded songs by a male in the DVCA for analysis. Recursion or embedding structure is unique to human language. If phrase-embedding structures or phrase-inserting structures are found in male Mueller's gibbons' songs, it may suggest that gibbons' songs have a similarity to human language in their structure.

Methods

Study areas and animals

All observations were conducted in the DVCA located in Sabah, Malaysia (Fig. 1), in the northeast region of Borneo Island. This area consisted of primary rainforest. We studied a gibbon group named "SAPA"; their territory was located around the Borneo Rainforest Lodge (BRL; 5° 01′ N, 117° 44′ E; elevation ~ 190 m), and the group was well habituated to humans.

During the study period from 2001 to 2009, the SAPA group initially consisted of six individuals: an adult male, an adult female, two sub-adults, an adolescent female, and a juvenile male. However, two sub-adults left the SAPA group in 2001–2002, the adult female died in April 2005, the adolescent female left the group in October 2006, and the juvenile male died in April 2008. Therefore, the number of group members was reduced from six to one. Their home range covered approximately 34 ha, the boundaries of which were determined during > 960 h of observations from 2001 to 2009.
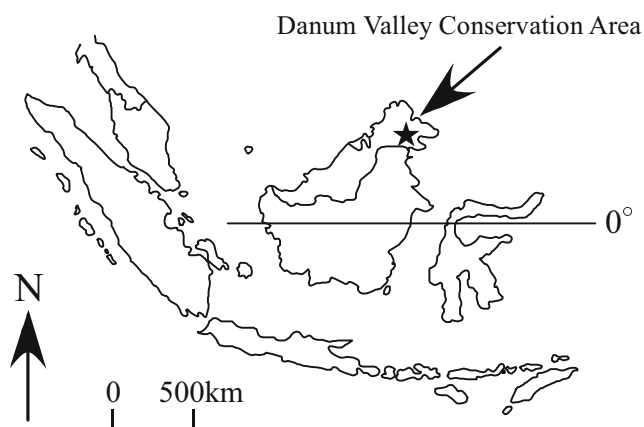
Fig. 1 Location of the Danum Valley Conservation Area (DVCA; arrow), Sabah, Malaysia

Data collection

With the aid of a field assistant, we conducted a following survey over 4–7 successive days, biannually in August and December, from 2001 to 2009. We arrived the sleeping trees where the SAPA group slept on the previous day. We started following at 0500 h and ended about 30 min after the gibbons had arrived at their sleeping trees. We collected behavioral data every 10 min. We started recording gibbon songs, as soon as they started singing. It allowed us to collect almost complete song data. We placed the recording device just under the singing tree and tried to record their songs as clearly as possible.

Recording protocols

We recorded 70 songs in 107 days of observations. We used a digital audio tape recorder (TCD-D100; Sony, Tokyo, Japan) with a microphone (ECM-MS907; Sony, Tokyo, Japan) from 2001 to 2006 and a digital audio recorder (R-09; Roland, Hamamatsu, Japan) with a microphone (ATM57; Audio-Technica, Tokyo, Japan) from 2007 to 2009. We did not find any difference in the quality of the recordings taken before and after 2007. We recorded the gibbon voices under the tree in which the focal male was singing. The recorder was set at a 44.1-kHz sampling rate and 16-bit resolution.

Data analysis

We analyzed 8046 phrases in 70 songs from the SAPA adult male. We converted the recorded sounds to sonograms using the Avisoft-SAS Lab Pro software (Avisoft, Berlin, Germany). First, the sampling frequency was converted to 4000 Hz to allow focusing on the fundamental frequency for the on-screen measurements described below. Next, to remove ambient noise, we processed the sound through a high-pass filter to cut off sound at 500 Hz. Finally, sonograms

were created for on-screen measurements (settings: 256-point fast Fourier transformation and Hamming windows). These configurations yielded final spectral and temporal resolutions of 16 Hz and 16 ms, respectively. We selected a 5% random sample of all notes ($n = 4471$) and measured the onset time, offset time, start frequency, end frequency, middle frequency, maximum frequency, and minimum frequency of each note and inter-note intervals for each song phrase (Fig. 2). Based on the measured data, we calculated the duration, Δ frequency, and Δ frequency/duration of each note.

Acoustic terms and definitions

A song is a series of notes, generally of more than one type, uttered in succession and so related as to form a recognizable sequence or pattern in time (Thorpe 1961). A phrase is a larger, loose collection of several notes preferentially voiced in combination. Intra-phrase intervals are shorter than inter-phrase intervals (Geissmann et al. 2005). Male songs heard around the BRL are composed of many phrases and the interval of the phrases was relatively long. Male gibbons sing phrases with long inserted silences (Appendix Fig. 8). This long interval between phrases might have evolved to facilitate turn-taking among male gibbons, as they may be expecting a neighbor's reply while singing (Inoue et al. 2013). Most of the note intervals in our subject male's songs were less than 2.0 s (90.7%; Fig. 3). Therefore, we identified different phrases within a song as being separated by pauses of > 2 s. Haimoff (1985) reported that adult Mueller's gibbon males produce long and elaborate solos that are characterized by a distinct, progressive elaboration of notes and phrases. He classified male phrases into five note types: "wa," "oo," "oo-wa," "quaver-type notes," and "trill" (Fig. 4b). However, neither our subject male nor the nine neighboring males around BRL sang "oo-wa" or "quaver-type" notes. We confirmed this with 35 male songs from 10 males. We showed a portion of the sonograms produced by six of the males in Appendix Fig. 8. As the trills comprised a short set of "wa" notes, we classified male songs into two note types: "wa" and "oo." The "wa" notes were generally short, with a rapid rise in frequency, and the "oo" notes were relatively monotonal (Table 1 and Figs. 2 and 4a). Two-tailed t tests were used to compare the acoustic characteristics of the two notes (Table 1). Three or more successive "wa" notes were sometimes placed at both the start and end of phrases. Among them, when three or more "wa" notes were produced in rapid succession, we named them trill ("T"). When trills were placed at the start of phrases, the maximum frequency of the successive "wa" notes decreased gradually near the end. In the tables and figures, "w" denotes the "wa" note, and "o" denotes the "oo" note.
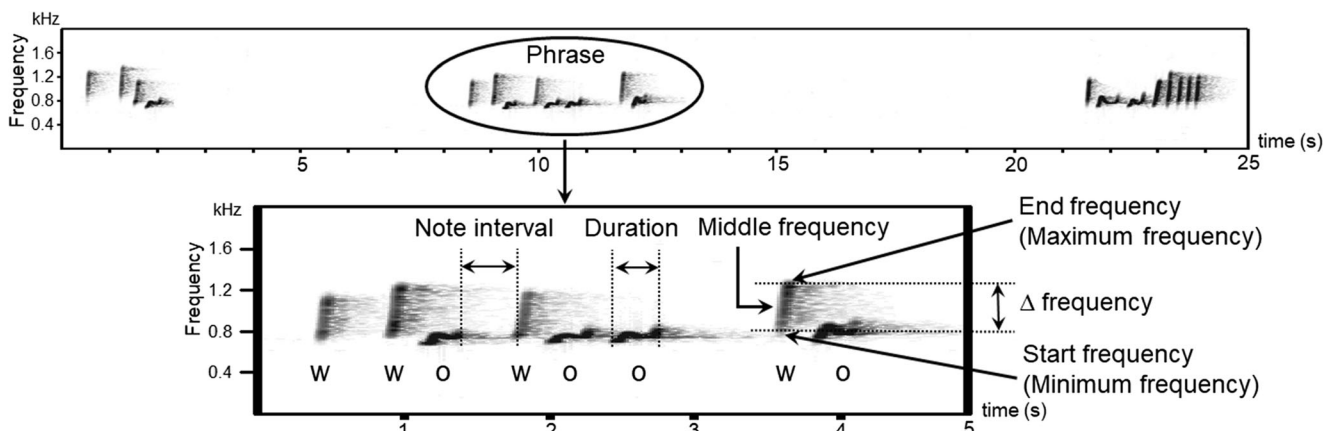
Fig. 2 Sonogram of one phrase composed of "wa" and "oo" notes. Start frequency, end frequency, middle frequency, maximum frequency, minimum frequency, Δ frequency, duration, and note interval are indicated by the arrows. "w" denotes the "wa" note and "o" denotes the "oo" note

Phrase-inserting structure

When phrase "N" is inserted within another phrase "AB," the generated phrase is shown as "A-N-B." We named this structure a phrase-inserting structure. "A" and "B" were parts of a phrase "AB." When trills were placed at both the start and end of AB phrases, we named these AB phrases fixed phrases. Trills were located in front of "A" and at end of "B." In this case, each A and B has a pairwise relationship and always co-occurs in a string. In the middle of a fixed phrase "AB," notes consisting of three or more notes were sometimes included. If the included notes were also sung as a phrase independently, we defined it as "N." There were 78 types of N (Appendix

Table 3) and there were 11 types of A-B (Fig. 5). When we compared A-B with A-N-B, the sonogram pattern of A-B was matched in A-B and A-N-B and the number of "wa" notes in A and B was exactly equal in A-B and A-N-B (examples are shown in Fig. 6). If the phrase "N" itself had an inserting structure elsewhere, "A-N-B" is shown as "A-AB-B." This structure is considered a phrase-embedding structure, as defined by Abe and Watanabe (2011). If phrase A_2B_2 was included within the phrase A_1B_1, the phrase was classified as A_1-A_2B_2-B_1. We defined the structure of the phrase "A_1-A_2B_2-B_1" as phrase-embedding. We analyzed whether or not the corresponding phrases were indistinguishable in terms of the acoustic similarity between those with an inserted
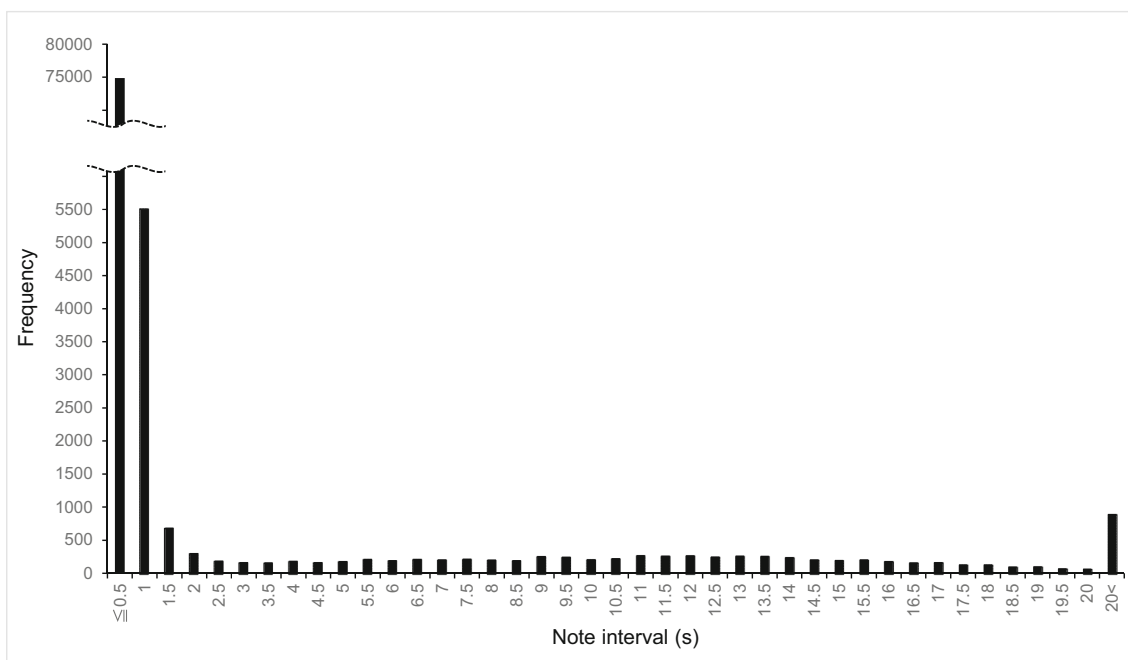
Fig. 3 Histogram of note interval in male songs

Fig. 4 **a** Sonograms of two notes of the SAPA male song. **b** Sonograms of five notes shown by Haimoff (1985). **c** Sonograms of six notes shown by Raemaekers et al. (1984)
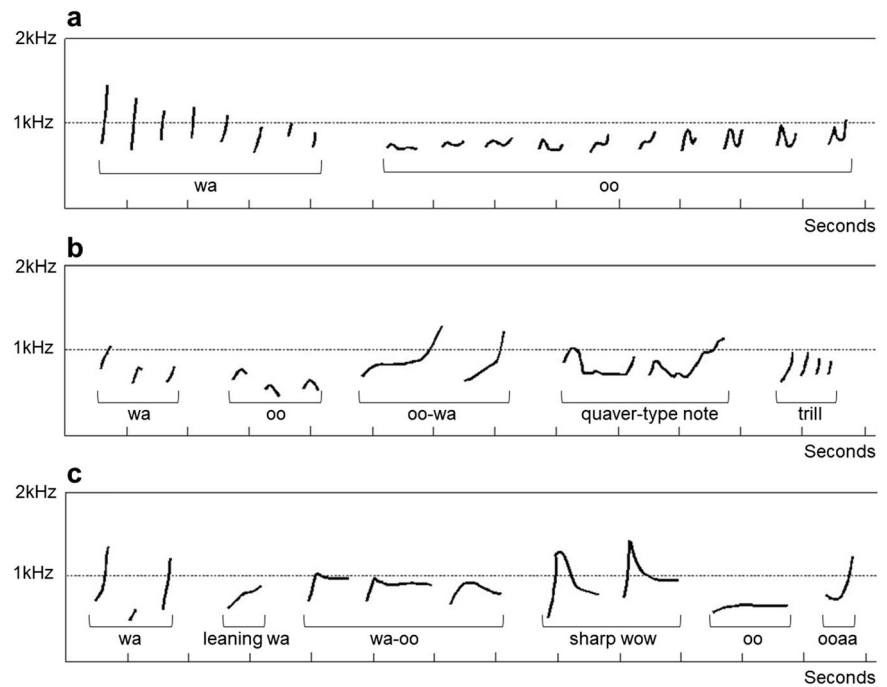

phrase (A-N-B) and those with no inserted phrase (A-B). We selected 20 patterns of phrases which are shown in Figs. 6 and 7 and Appendix Fig. 9, and calculated the acoustical similarity using Avisoft-CORRELATOR (Avisoft, Berlin, Germany) for every pair of same type phrases. For example, for the type I phrases in Fig. 6, we calculated the acoustical similarity between all pairs of "A" notes within the A-N-B phrases (within-class condition), within the A-B phrases (within-class condition), and between the A-N-B and the A-B phrases (between-class condition). The Avisoft-CORRELATOR allowed us to compute cross-correlations between spectrograms by sliding them along with the time axis. The approach of the Avisoft-CORRELATOR can be compared with computing correlations between two grayscale raster images, while

sliding them on the X-axis. The highest correlation coefficients for each pair of sounds were regarded as similarity scores. We examined whether the acoustical similarity was the same regardless of whether the pair of phrases were either from the same class (within-class condition) or from different classes (between-class condition). We performed a linear mixed effects model entering the cross-correlation coefficients as the response variable and the within/between-class condition as the explanatory variable (fixed effects). We entered the phrase types (the abovementioned 20 patterns of phrases) as random intercepts because the average similarity scores were expected to differ among them. We performed model diagnosis by visually assessing normality and homogeneity of residual variance across the random groups and normality of the

Table 1 Acoustic characteristics of "wa" and "oo" notes

	Mean (SD)		t	p
	wa ($n = 3518$)	oo ($n = 953$)		
Start frequency (kHz)	0.797 (0.087)	0.712 (0.056)	45.49	$p < 0.01$
End frequency (kHz)	1.094 (0.167)	0.796 (0.081)	77.50	$p < 0.01$
Middle frequency (kHz)	0.932 (0.111)	0.751 (0.058)	68.17	$p < 0.01$
Δ frequency (kHz)	0.297 (0.146)	0.084 (0.074)	62.08	$p < 0.01$
Minimum frequency (kHz)	0.797 (0.087)	0.694 (0.035)	55.49	$p < 0.01$
Maximum frequency (kHz)	1.094 (0.167)	0.810 (0.079)	74.59	$p < 0.01$
Duration (s)	0.056 (0.027)	0.235 (0.087)	63.01	$p < 0.01$
Δ frequency/duration (kHz/s)	6.211 (3.868)	0.414 (0.467)	86.58	$p < 0.01$
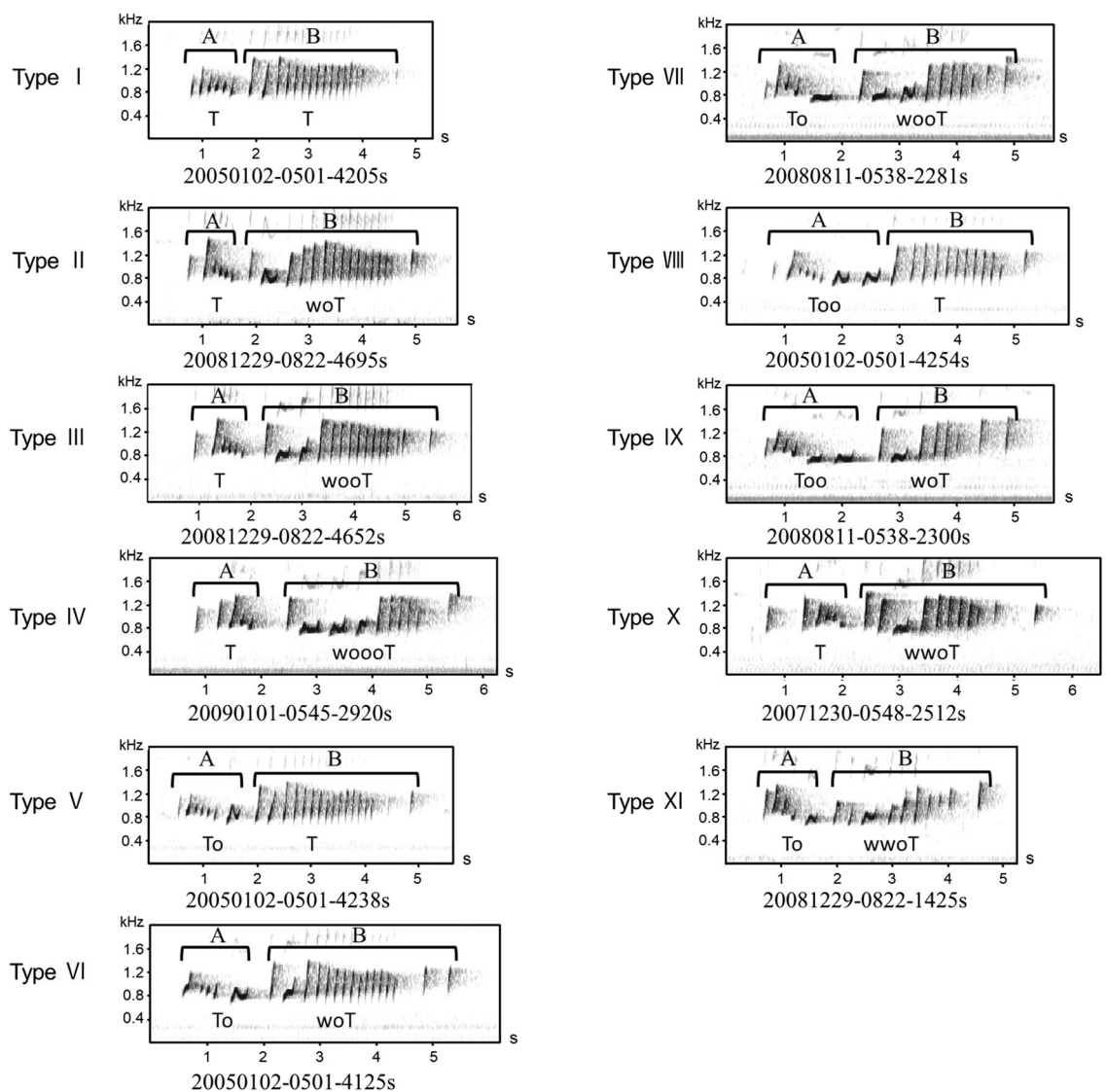
Fig. 5 Examples of 11 types of fixed phrases. A succession of "wa" notes (trill) is placed at both the start and end of the phrase. Fixed phrases were divided into the first half (A) and the second half (B). Date and time are shown under each phrase. For example, 20050102-0501-4205s shows that this phrase was recorded at 4205 s after the song started at 05:01 on 2 January 2005

random intercepts using group-wise boxplots and normal quantile-quantile plots. We reported the mean and standard deviations of the similarity score for the within/between-class condition, and examined the statistical significance using a likelihood ratio test (Pinheiro and Bates 2000). R 3.6.1 (R Core Team 2019) and nlme package (Pinheiro et al. 2019) were used for the analysis.

Results

Note variability

The songs had 2029 types of note order. We classified the notes of the studied gibbon's songs into two notes,

"wa" and "oo," through not only sonogram figures (Fig. 4a), but also acoustic characteristics (Table 1). The mean (SD) values of acoustic parameters calculated for two notes are shown in Table 1. The mean start and end frequencies of "wa" notes were higher than those of "oo" notes. The Δ frequency of "wa" notes was larger than that of "oo" notes. The start frequency and minimum frequency always coincided in "wa" notes. The end frequency and maximum frequency also always coincided in "wa" notes. The mean minimum and maximum frequencies of "oo" notes were 0.694 kHz and 0.810 kHz, respectively. The mean start and end frequencies were slightly out of range. The duration of "wa" notes was shorter than that of "oo" notes. The frequency range of "wa" notes was wider than that of

"oo" notes. "wa" notes were characterized by a fast frequency increase. In contrast, the frequency fluctuation of "oo" notes was relatively small. The mean values of the all examined acoustic parameters were statistically different between the two notes (Table 1). There were no notes in the studied gibbon's songs with acoustic characteristics largely different from "wa" and "oo" notes.

Phrase-inserting structure

In male Mueller's gibbons' songs, a succession of "wa" notes is sometimes placed at both the start and end of phrases. In this study, three or more successive "wa" notes were placed at both the start and end of phrases in 2726 (33.9%) of the 8046 phrases. Among these, we identified 1891 phrases (23.5%) as fixed phrases (Table 2). Phrases consisting of three or more notes were sometimes included within fixed phrases. We found 448 phrases (5.6%) that included phrases consisting of three or more notes (Table 2; examples are shown in Fig. 6). The recordings of type II in Fig. 6 are shown in the Electronic Supplementary Materials. There were a number of A-N-B type II phrases. Among them, we showed 13 examples in Appendix Fig. 9. We identified four phrases as "A_1-A_2B_2-B_1" suggesting phrase embedding (Fig. 7), although A_2B_2 phrases in the examples were not fixed phrases except for example no. 4. Phrase "A_2B_2" in phrase "A_1-A_2B_2-B_1" included the phrase "N" within another phrase (Fig. 7). The number of notes composing a phrase was at times large; there were 785 phrases (9.8%) composed of 20 or more notes.

The overall mean similarities for the within-class and between-class conditions were 0.52 (SD = 0.12, n = 1290) and 0.50 (SD = 0.11, n = 1346), respectively. The diagnostic plots of the linear mixed effects model showed no indication of violating the model assumption (result not shown). The result of the linear mixed effects model showed that the similarity was estimated to be 0.02 (standard error = 0.003) higher in the within-class condition, after controlling the phrase type (X_1 = 33.15, $p < 0.01$).

Discussion

Male Mueller's gibbons living in DVCA have two types of song elements while six note types have been described by Raemaekers et al. (1984). Raemaekers et al. (1984) showed that notes of the white-handed gibbons' songs were classified into six note types; "wa," "leaning wa," "wa-oo," "sharp wow," "oo," and "ooaa" (Fig. 4c). Although we did not find "sharp wow" and "ooaa"

notes in the studied gibbon's songs, the "leaning wa" and "wa-oo" notes were somewhat similar to the "oo" notes of the studied gibbon's songs. However, we included these two notes in the "oo" notes, because the sonograms of the "oo" notes in the studied gibbon's songs resembled each other in shape (Fig. 4a) and their acoustic characteristics were within a certain range (Table 1). We analyzed songs by a single male. The male had various types of note orders in his phrases. In his songs, we identified fixed phrases, which were characterized by three or more successive "wa" notes (trills) placed at both the start and end of phrases; these accounted for 23.5% of all phrases. Phrases consisting of three or more notes were sometimes included within fixed phrases. We found 448 phrases (5.6%) that included phrases consisting of three or more notes. We have concluded that these note orders suggest a phrase-inserting structure. Among them, there were four examples of phrase embedding. A phrase-embedding structure was defined clearly by Abe and Watanabe (2011); when phrase N is inserted in phrase AB, the phrase A-N-B does not always have an embedding structure. If phrase N has a structure such as SV (subject-verb) in human language, the phrase A-N-B has an embedding structure and the phrase A-N-B can be shown as "A-AB-B." In the case of the studied gibbon's songs, we determined whether the phrase N had a structure or not by determining whether the phrase N itself had an inserting structure. When an A_2B_2 phrase had an inserting structure of A_2-N-B_2 elsewhere (it is shown at the lower right of each example in Fig. 7) and was inserted in another A_1-B_1 phrase, we considered that an A_1-A_2B_2-B_1 phrase had a phrase-embedding structure (it is shown at the upper right of each example in Fig. 7). Most of the inserted phrases in a phrase-inserting structure were short and simple, but inserted phrases in a phrase-embedding structure were long and complex. The number of notes composing a phrase of the studied gibbon's songs was sometimes large; phrases composed of 20 or more notes accounted for 9.8% of all phrases. When gibbons sang phrases composed of many notes and the note combinations in phrases were flexible, a phrase-embedding structure might emerge. If the studied gibbon intended to embed phrases within phrases, it suggests that gibbons have the ability to produce hierarchical structures. However, as phrase-embedding occurred only four times in the studied gibbon's songs, further studies are needed to verify this. A phrase-inserting structure occurred when combinations of notes in the male's songs were flexible. Our results lead us to hypothesize that complicated syntax emerges when the animals sing songs by combining acoustic elements flexibly.

A-B　　　　　　　**N**　　　　　　　**A-N-B**

Type I

20050102-0501-2507s　　20060808-0904-1032s　　20081229-0822-8145s

Type II

20050102-0501-3457s　　20090805-0555-983s　　20081229-0822-9720s

Type III

20081229-0822-5162s　　20050803-0533-337s　　20081229-0822-8737s

Type IV

20090805-0555-1871s　　20010804-0605-112s　　20080811-0538-1599s

Type V

20030808-0556-209s　　20050731-0546-226s　　20050102-0501-365s

Type VI

20081229-0822-4411s　　20081229-0822-580s　　20081229-0822-7294s

Type VII

20071226-0821-2377s　　20010804-0605-112s　　20080811-0538-3817s

Type VIII

20070813-0540-786s　　20080811-0538-898s　　20081229-0822-4060s

Type IX

20040809-0810-422s　　20010804-0605-112s　　20080811-0538-1736s

Type X

20050102-0501-3108s　　20031225-0548-596s　　20081229-0822-9044s

Type XI

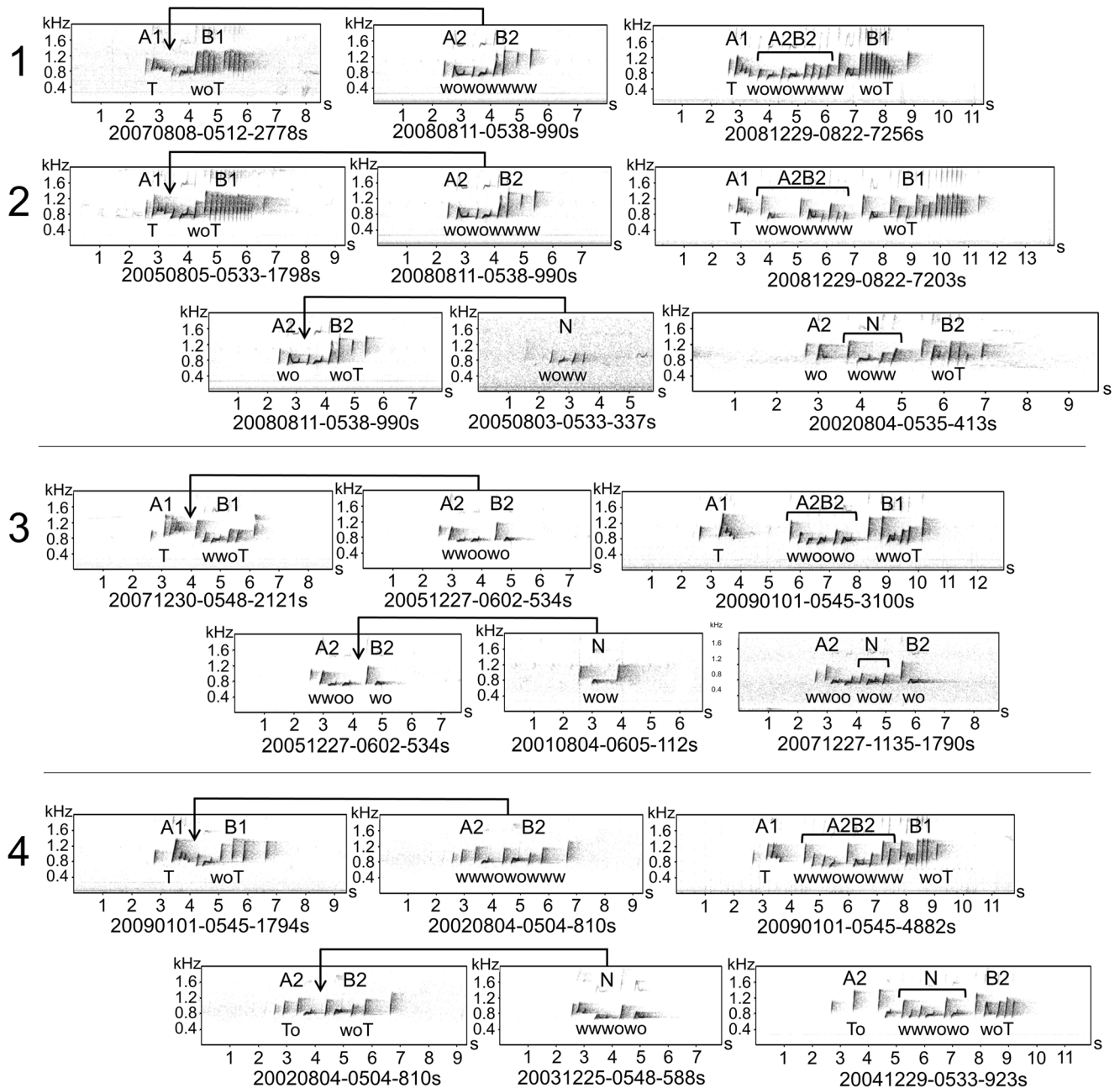20050102-0501-5086s　　20050729-0547-153s　　20070805-0540-1047s

◀ **Fig. 6** Examples of phrase A-N-B included within 11 types of fixed phrases. Arrows indicate where phrase N was included within phrase A-B. Solid line brackets indicate phrase N included within phrase A-B

Corresponding phrases with an inserting structure and no inserting structure were very similar in terms of sonograms, and hardly discriminable by human ear. However, the statistical analysis showed a small but systematic difference in acoustic similarity between them; phrases were more similar among pairs that both had an inserting structure or did not have an inserting structure (within-class condition), than among pairs where one had an inserting structure and the other did not (between-class condition). Further study under experimental conditions will be needed to examine whether gibbons are able to discriminate this small acoustic difference.

Fig. 7 Four examples of phrases shown by A_1-A_2B_2-B_1. Upper figures in examples show that phrase A_2-B_2 is included within phrase A_1-B_1. Lower figures in examples show that phrase N is included within phrase A_2-B_2. Arrows indicate where phrase A_2-B_2 (or N) was included within phrase A_1-B_1 (or A_2-B_2). Solid line brackets indicate phrase A_2-B_2 (or N) included within phrase A_1-B_1 (or A_2-B_2). A_2B_2 phrases were not fixed phrases except for example no. 4

Table 2 Frequency of A-B and A-N-B phrases in male songs. A-B shows fixed phrases. Fixed phrases were classified into 11 types. A-N-B shows that phrase N is included within phrase A-B. "T" is an abbreviation for "trill," which is three or more successive "wa" notes placed at both the start and end of a phrase. "w" denotes the "wa" note and "o" denotes the "oo" note

Phrase type	A-B	Appearance frequency of A-B in male songs	A-N-B	Appearance frequency of A-N-B in male songs	Total
I	T-T	49	T-N-T	52	101
II	T-woT	244	T-N-woT	229	473
III	T-wooT	118	T-N-wooT	47	165
IV	T-woooT	22	T-N-woooT	3	25
V	To-T	293	To-N-T	21	314
VI	To-woT	200	To-N-woT	55	255
VII	To-wooT	43	To-N-wooT	2	45
VIII	Too-T	290	Too-N-T	4	294
IX	Too-woT	53	Too-N-woT	2	55
X	T-wwoT	83	T-N-wwoT	28	111
XI	To-wwoT	48	To-N-wwoT	5	53
	Total (percentage)	1443(17.9%)	Total (percentage)	448 (5.6%)	1891 (23.5%)

One of the main differences between language and non-human animal communication is the grammar used to produce sequences. Human language uses "context-free grammars" that are capable of generating recursive sequences (Chomsky 2002). Among various discussions about the definition of recursion (e.g., Fitch 2010), there is an interpretation that recursion consists of embedding a constituent into a constituent of the same type (e.g., Pinker and Jackendoff 2005; Martins and Fitch 2014). In contrast, animal vocal sequences are usually described as "regular grammars," a simple kind of concatenation system (Berwick et al. 2012). Many researchers have considered that non-human animal vocalizations would belong to regular ones. This paper tested the grammatical structures of primate "songs" with the comparative perspectives of the human language syntax and that of other animals. Our results showed that male gibbon's songs have a phrase-inserting structure which is considered to be a precursory level to recursion. This is the first evidence of a phrase-inserting structure in animal songs. Our data and linguistic perspectives may certainly be of use in future studies to elucidate vocal communication in gibbons and other non-human primates. However, as we collected data from only one male, further studies on many gibbon groups will be necessary to confirm our results. As there are individual differences in male songs, we think it is important to record songs from many individuals. However, neighboring groups at the DVCA were not habituated and it was difficult to record enough songs from neighboring males. Although we recorded 35 songs from 10 neighboring males and found a possibility of phrase insertion in them (for example, the first and third phrases of the D male song in Appendix Fig. 8 seem to have a phrase-inserting structure), the sample size was too small to confirm it in each male song. We are trying to habituate neighboring groups to record more sound data from them

in the future; however, this takes a long time. Whether our study of a single male can be considered representative of the whole species is a difficult question. In the past, there have been linguistic studies where a single individual is primarily studied, and the results are later supplemented by the study of additional individuals. For example, Dr. Matsuzawa, from the Primate Research Institute, Kyoto University, showed that a chimpanzee named "Ai" learned to use Arabic numerals to label sets of real-life objects with the corresponding number (Matsuzawa 1985). Furthermore, "Ai" could represent both the cardinal and the ordinal aspects of number to some extent, using the Arabic numerals 0 through 9 (Biro and Matsuzawa 2001). Later, similar skills were confirmed in other chimpanzees (Inoue and Matsuzawa 2007). We expect that our study will follow the same process as this chimpanzee study. Although it is important to discuss how gibbons perceive syntactic rules, little is known about it. A cognitive test in experimental settings (in this case, artificial song playback is a possible method) is also necessary in the future.
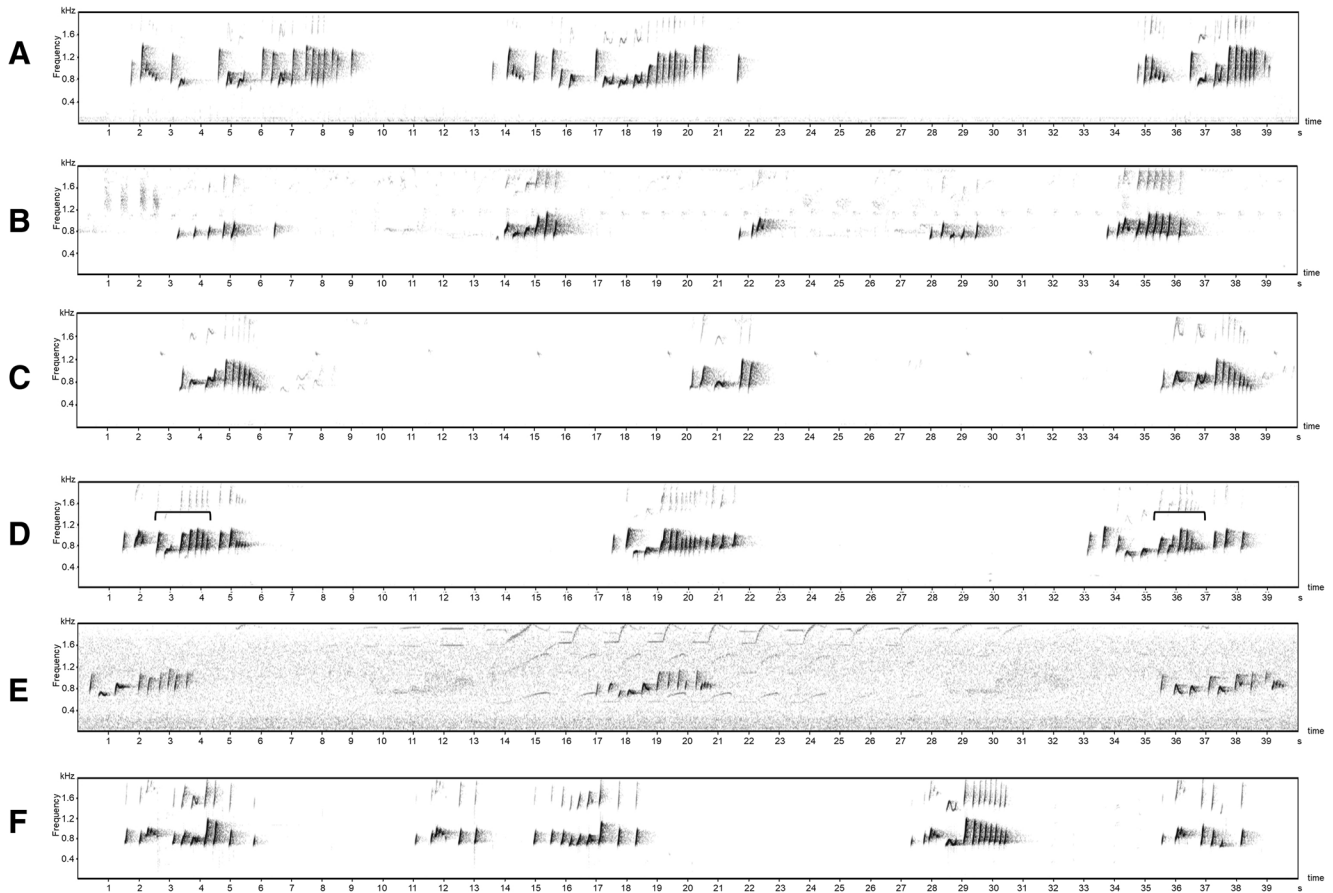
Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

Fig. 8 Portion of the sonograms produced by 6 males around BRL. **A**: SAPA male, **B**: neighbor male located on the west side of the SAPA group, **C**: neighbor male located on the east side of the SAPA group, **D**: neighbor male located on the north side of the SAPA group, **E**: neighbor male located on the south side of the SAPA group. **F**: new group entered in the SAPA group territory after the SAPA group disappeared. In all songs, notes of "oo-wa" and "quaver-type notes" were not found. In **D** male song, the first and third phrases seem to have a phrase-inserting structure (solid line brackets show the inserted phrases)
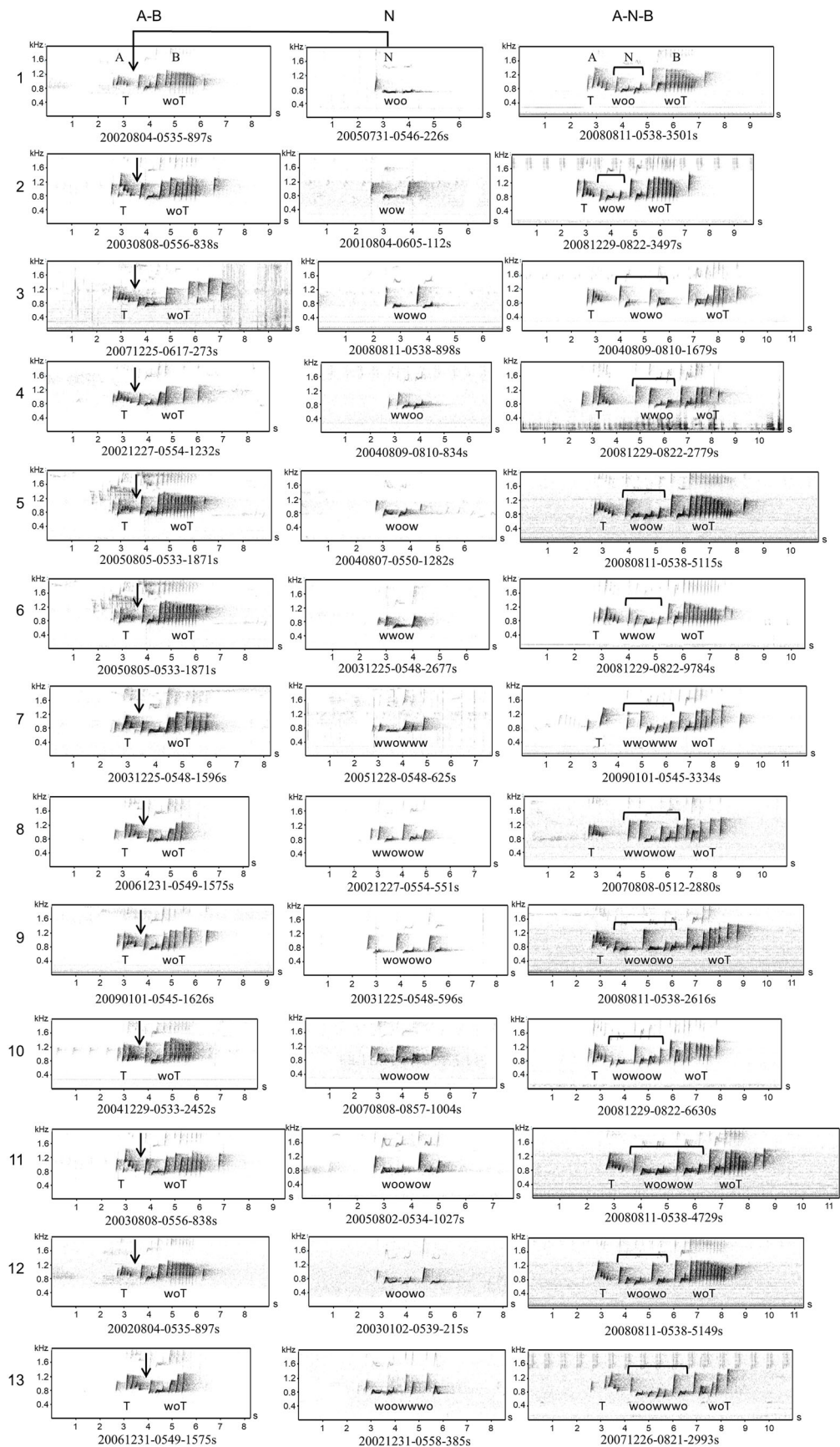
Fig. 9 Examples of phrase A-N-B in Type II phrases

Table 3 Appearance Frequency of N included within A-B phrases and N in male songs

	N	Appearance frequency of N included within A-B phrases	Appearrance frequency of N in male songs
1	wow	80	15
2	woo	42	96
3	wwo	36	36
4	www	27	70
5	wowo	23	21
6	woww	20	8
7	wwow	18	31
8	wwwo	13	36
9	woow	10	40
10	wowow	9	7
11	wowww	9	1
12	wooo	8	115
13	wwww	8	81
14	wwoo	7	97
15	wwwww	7	87
16	oww	6	3
17	wowwo	5	12
18	wwoww	5	25
19	wowowo	5	10
20	ooo	4	11
21	wwowo	4	39
22	wwwow	4	28
23	wwwwo	4	31
24	wowwow	4	2
25	oow	3	2
26	woowo	3	30
27	wowoo	3	6
28	wwwoo	3	46
29	wowwoo	3	2
30	wwowww	3	11
31	wwwoww	3	19
32	wwwwoo	3	12
33	wwwwow	3	18
34	wowwoww	3	1
35	wwwowww	3	14
36	owow	2	1
37	woooo	2	47
38	wowoow	2	2
39	wowoww	2	9
40	wowwww	2	1
41	wwowow	2	29
42	wwowwo	2	16
43	wwwoow	2	44
44	wwwwww	2	37
45	woowwwo	2	7
46	wowowwo	2	1
47	wwowwwo	2	5
48	wwwwoww	2	9
49	wowowwww	2	5
50	oowo	1	2
51	owwoo	1	1
52	owwwo	1	1
53	owwww	1	1
54	wooww	1	23
55	wwooo	1	130
56	wwoow	1	89
57	woooow	1	23
58	woooww	1	16
59	woowow	1	15
60	wwoowo	1	45
61	wwwowo	1	19
62	wooowwo	1	4
63	wooowww	1	11
64	wowwwww	1	4
65	wwowwow	1	7
66	wwwowow	1	26
67	wwwwwoo	1	9
68	woowwwww	1	16
69	wowowwow	1	1
70	wowowwwo	1	2
71	wowwwwww	1	1
72	wwowwwoo	1	2
73	wwwwwoww	1	6
74	wwoowwwow	1	5
75	wwwowowww	1	10
76	wwwwwwoww	1	7
77	wwowwwowww	1	1
78	wwwwwwwwww	1	7
	Total	448	1760

References

Abe K, Watanabe D (2011) Songbirds possess the spontaneous ability to discriminate syntactic rules. Nat Neurosci 14:1067–1074

Arnold K, Zuberbühler K (2006) Language evolution: semantic combinations in primate calls. Nature 441:303

Beckers GJ, Bolhuis JJ, Okanoya K, Berwick RC (2012) Birdsong neurolinguistics: songbird context-free grammar claim is premature. Neuroreport 23(3):139–145

Beer C (1976) Some complexities in the communication behavior of gulls. Ann N Y Acad Sci 280:413–432

Berwick RC, Okanoya K, Beckers GJ, Bolhuis JJ (2011) Songs to syntax: the linguistics of birdsong. Trends Cogn Sci 15:113–121

Berwick RC, Beckers GJ, Okanoya K, Bolhuis JJ (2012) A bird's eye view of human language evolution. Front Evol Neurosci 4:5

Biro D, Matsuzawa T (2001) Use of numerical symbols by the chimpanzee (*Pan troglodytes*): cardinals, ordinals, and the introduction of zero. Anim Cogn 4:193–200

Brockelman WY, Schilling D (1984) Inheritance of stereotyped gibbon calls. Nature 312:634–636

Chomsky N (2002) Syntactic structures, 9th edn. The Hague, The Netherlands: de Gruyter Mouton

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

Corballis MC (2007) Recursion, language, and starlings. Cogn Sci 31(4):697–704

Crockford C, Boesch C (2005) Call combinations in wild chimpanzees. Behaviour 142:397–421

Fitch WT (2010) Three meanings of "recursion": key distinctions for biolinguistics. In: Larson RK, Deprez V, Yamakido H (eds) The evolution of human language: biolinguistic perspectives. Cambridge University Press, Cambridge, pp 73–90

Fitch WT, Friederici AD (2012) Artificial grammar learning meets formal language theory: an overview. Philos. Trans. R. Soc. Lond. B Biol Sci 367:1933–1955

Fitch WT, Hauser MD (2004) Computational constraints on syntactic processing in a nonhuman primate. Science 303:377–380

Friederici AD, Bahlmann J, Heim S, Schubotz RI, Anwander A (2006) The brain differentiates human and non-human grammars: functional localization and structural connectivity. Proc Natl Acad Sci U S A 103(7):2458–2463

Geissmann T (1984) Inheritance of song parameters in the gibbon song, analyzed in 2 hybrid gibbons (*Hylobates pileatus × H. lar*). Folia Primatol 42:216–235

Geissmann T, Bohlen-Eyring S, Heuck A (2005) The male song of the Javan silvery gibbon (*Hylobates moloch*). Contrib Zool 74:1–25

Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC (2006) Recursive syntactic pattern learning by songbirds. Nature 440:1204–1207

Groves CP (1972) Systematics and phylogeny of gibbons. In Rumbaugh DM (ed): Gibbon and siamang, Basel Karger, vol 1, pp 1–89

Hailman JP, Ficken MS, Ficken RW (1985) The "chick-a-dee" calls of *Parus atricapillus*: a recombinant system of animal communication compared with written English. Semiotica 56:191–224

Haimoff EH (1985) The organization of song in Müeller's gibbon (*Hylobates muelleri*). Int J Primatol 6:173–192

Hauser MD, Chomsky N, Fitch WT (2002) The faculty of language: what is it, who has it, and how did it evolve? Science 298:1569–1579

Hauser MD, Yang C, Berwick RC, Tattersall I, Ryan MJ, Watumull J, Chomsky N, Lewontin RC (2014) The mystery of language evolution. Front Psychol 5:401

Inoue S, Matsuzawa T (2007) Working memory of numerals in chimpanzees. Curr Biol 17(23):R1004–R1005

Inoue Y, Sinun W, Yosida S, Okanoya K (2013) Intergroup and intragroup antiphonal songs in wild male Mueller's gibbons (*Hylobates muelleri*). Interact Stud 14:24–43

Inoue Y, Sinun W, Yosida S, Okanoya K (2017) Combinatory rules and chunk structure in male Mueller's gibbon songs. Interact Stud 18:1–25

Jackendoff R, Pinker S (2005) The nature of the language faculty and its implications for evolution of language (reply to Fitch, Hauser, and Chomsky). Cogn 97(2):11–225

Kinsella A (2010) Was recursion the key step in the evolution of the human language faculty? In: Hulst H (ed) Recursion and human language. New York, NY, De Gruyter Mouton, pp 179–191

Marshall JT, Marshall ER (1976) Gibbons and their territorial songs. Science 193:235–237

Martins MD, Fitch WT (2014) Investigating recursion within a domain-specific framework. In: Lowenthal F, Lefebve L (eds) Language and recursion. Springer, New York, pp 15–26

Matsuzawa T (1985) Use of numbers by a chimpanzee. Nature 315:57–59

Mitani JC, Marler P (1989) A phonological analysis of male gibbon singing behavior. Behaviour 109:20–45

Ouattara K, Lemasson A, Zuberbühler K (2009) Campbell's monkeys concatenate vocalizations into context-specific call sequences. Proc Natl Acad Sci U S A 106:22026–22031

Pietroski P, Crain S (2012) The language faculty. In: Margolis E, Samuels R, Stich SP (eds) The Oxford handbook of philosophy and cognitive science. Oxford University Press, New York, pp 361–381

Pinheiro JC, Bates DM (2000) Mixed-effects models in S and S-PLUS. Springer, New York

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core team (2019). nlme: linear and nonlinear mixed effects models. R package version 3.1-143. https://CRAN.R-project.org/package=nlme

Pinker S, Jackendoff R (2005) The faculty of language: what's special about it? Cogn 95:201–236

R Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Raemaekers JJ, Raemaekers PM, Haimoff EH (1984) Loud calls of the gibbon (*Hylobates lar*): repertoire, organization and context. Behaviour 91:146–189

Rilling JK, Glasser MF, Preuss TM, Ma X, Zhao T, Hu X, Behrens TE (2008) The evolution of the arcuate fasciculus revealed with comparative DTI. Nat Neurosci 11:426–428

Robinson JG (1979) An analysis of the organization of vocal communication in the titi monkey Callicebus moloch. Z Tierpsychol 49:381–405

Robinson JG (1984) Syntactic structures in the vocalisations of wedge-capped capuchin monkeys, *Cebus olivaceus*. Behaviour 90:46–78

Ten Cate C, Okanoya K (2012) Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. Philos Trans R Soc Lond Ser B Biol Sci 367:1984–1994

Terleph TA, Malaivijitnond S, Reichard UH (2018) An analysis of white-handed gibbon male song reveals speech-like phrases. Am J Phys Anthropol 166:649–660

Thorpe WH (1961) Bird-song. The biology of vocal communication and expression in birds. Cambridge monographs in experimental biology no. 12. Cambridge: University Press

van Heijningen CA, De Visser J, Zuidema W, Ten Cate C (2009) Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. Proc Natl Acad Sci U S A 106:20538–20543

Zuberbühler K (2002) A syntactic rule in forest monkey communication. Anim Behav 63:293–299