

Combinatory rules and chunk structure in male Mueller's gibbon songs

Yoichi Inoue, Waidi Sinun, Shigeto Yosida and Kazuo Okanoya

The School of Arts and Sciences, The University of Tokyo/Laboratory for Biolinguistics, RIKEN Brain Science Institute / Research and Development Division Yayasan Sabah Group / / Cognition and Behavior Joint Research Laboratory, RIKEN Brain Science Institute/Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo

Understanding whether the long and elaborate songs of male gibbons (*Hylobates muelleri*) have syntax and hierarchical structures (chunks) is an interesting question in the evolution of language, because gibbons are near humans in the phylogenetic tree and a hierarchically organized syntax is considered to be a basic component of human language. We conducted field research at Danum Valley Conservation Area in northern Borneo to test the hypothesis that gibbon songs have syntax and chunks. We followed one Mueller's gibbon group for 1 week in the dry and rainy seasons every year from 2001 to 2009, collecting vocal and behavioral data. Results show that songs emitted by the studied male gibbon were governed by combinatory rules. Some context-dependent songs had different combinatory rules, although they overlapped with the songs whose contexts were uncertain. The male Mueller's songs had characteristics that suggest existence of chunk structure. These results provided an important perspective in the study of language origin.

Keywords: Danum Valley Conservation Area, Mueller's gibbon, male song, combinatory rule, chunk

1. Introduction

In human language, an infinite amount of information is created by combining limited numbers of sound units. In the study of language in nonhuman animals, it is of great interest to determine whether animals combine calls and reference them to specific contexts. There are only limited reports of combinational animal sounds that reference different objects and emotions (Beer, 1976; Robinson, 1979;

Chew, 1981; Cleveland & Snowdon, 1982; Robinson, 1984; Hailman et al., 1985; Mitani & Marler, 1989; Zuberbühler, 2002; Crockford & Boesch, 2005; Geissmann et al., 2005; Templeton et al., 2005; Arnold & Zuberbühler, 2006; Clarke et al., 2006; Ouattara et al., 2009). Moreover, no species other than humans seem to have a comparable capacity to recombine units of sounds into an unlimited variety of larger structures, each differing systematically in meaning. Why are there only limited reports on such vocal behavior in nonhumans? We think that there are two main possibilities. First, the combinational referential vocalizations might genuinely not be common in animals. In most species, their vocal repertoire size may be sufficient and there may be no need for combinational signals in their life history. Second, technical difficulties may hinder studies seeking to examine the existence of combinational referential vocalizations. It is likely that, if a species has several sound elements and there are numerous numbers of possible combinations, we only observe a subset of combinations occurring in a fixed manner. In such situations we are not able to judge whether the vocalization is combinational. Therefore, we might be able to find combinational referential vocalizations in species that have the cognitive ability to reference a variety of objects and emotions, but not a very large vocal repertoire size.

To date, there have been several primate research studies in this field. For example, when Campbell's monkeys (*Cercopithecus campbelli*) combine a 'boom' call with a functionally referential alarm call, they are modifying the information normally transmitted by the alarm call (Zuberbühler, 2002). Ouattara et al. (2009) revealed that adult males of Campbell's monkeys (*Cercopithecus campbelli*) produce loud calls, which they combine into various sequences in highly context-specific ways. Arnold and Zuberbühler (2006) reported that Putty-nosed monkeys (*Cercopithecus nictitans*) combine two predator-specific alarm calls and use them as different meanings of "move away". Chimpanzees (*Pan troglodytes*) also combine calls, and some of them may be linked to specific contexts (Crockford & Boesch, 2005).

Gibbons (Hylobatidae) living in South-East Asia are known for their remarkable vocal behavior. Further, there have also been a very limited number of reports on context-specific combinational signals. Gibbons sing loud and conspicuous songs that transmit over distances of up to 1.1 kilometers through dense forest vegetation. They produce species-specific songs and have a small repertoire of notes. In primates, singing behavior is known in only four genera (*Indri*, *Tarsius*, *Calliebus*, *Hylobates*) and gibbons are the only singing species of ape except humans. Therefore, the study of gibbon songs is thought to be important in verifying the hypothesis that language evolved from music (Darwin, 1859). Haimoff (1985) studied the organization of songs in Mueller's gibbon (*Hylobates muelleri*) and suggested that adult males produce long and elaborate solo phrases, but the songs

are neither sequentially nor interactively organized. Mitani and Marler (1989) examined the phonological structure of wild male agile gibbon (*Hylobates agilis*) songs. They revealed that gibbon songs are produced by reference to a set of rules and the gibbons respond somewhat differently to normal and phonologically rearranged playback songs. Geissmann et al. (2005) investigated the syntax of male Silvery gibbon (*Hylobates moloch*) songs. Based on changes in the proportion of note types, they classified male songs into two phases: introductory phase and main phase. In the main phase, they calculated the first four or five successive note orders in 11 individuals. As a result, male songs appeared to follow certain syntactical rules but it was difficult to recognize recurrent patterns. All study animals appeared to use individually-preferred note sequences. Clarke et al. (2006) compared the acoustic structure of predator-induced songs with regular songs in white-handed gibbons (*Hylobates lar*). They found predator-induced songs were different from normal songs in how the notes were assembled even though the call note repertoires were identical. Moreover, these syntactic differences were meaningful to conspecifics because neighboring groups replied to the predator-induced songs in a different way from normal songs.

Mueller's gibbons (*Hylobates muelleri*) live in the northern and eastern regions of Borneo (Groves, 1972; Marshall & Marshall, 1976). Similar to their congeners, Mueller's gibbons sing sex-specific songs. Males sing solos before or at dawn, and females occasionally sing solos after dawn during the morning hours. In addition, males and females sing duets usually in the morning. Songs of male Mueller's gibbons living in Danum Valley Conservation Area (DVCA) consist of two notes. Male songs are long and complex, comprising many multi-note phrases in which both notes are ordered in various combinations. Males sometimes sing intergroup and intragroup antiphonal songs (Inoue et al., 2013). Although notes of gibbon songs are believed to be genetically determined (Brockelman & Schilling, 1984; Geissmann, 1984), it is unclear whether this is the case for rules governing note order of male songs.

In some species, elaborate vocalizations have hierarchical structures where each vocal unit comprises a group of sounds (chunk) and the chunks form a phrase or bout. Human language and birdsong have chunk structure (ten Cate & Slater, 1991; Williams & Staples, 1992; Berwick et al., 2011) and whale songs also have the kind of sequential organization which is suggestive of concatenated chunks (Suzuki et al., 2006). Whether gibbon songs have chunks is an interesting question to ask. To date, ape vocalizations have been implicitly investigated on the premise of chunk existence. For example, Crockford & Boesch (2005) showed that chimpanzees convey context-specific information by combining different call repertoires, but they didn't examine note orders in each call repertoire. In contrast,

we examined note orders in each song. If gibbon songs have chunks, it would shed light on evolutionary processes shared by gibbon songs and human language.

Based on the features of gibbon songs, we hypothesized: (1) that combinations of notes in songs follow some syntactical rules and (2) that songs have hierarchical structures. By focusing on a single group and collecting song data while successively following them, we obtained a better understanding on the context of the recorded vocal data. The difficulty of following wild gibbons, owing to their sensitivity and wariness of humans, means that few studies have recorded a complete set of vocal data while also collecting non-vocal behavior. However, our subject group has been exposed to many working staff and tourists since the opening of Borneo Rainforest Lodge (BRL) in 1994, and has become habituated to humans. Therefore, we could easily follow the group all day from 5:00 a.m. until their final stop at a sleeping tree, which allowed us to collect a complete set of vocal and behavioral data. The results of our study show that male gibbon songs may have combinatory rules and chunk structures. Although we collected data from only one subject male, our study provides important evidence of a preadaptation of language.

2. Methods

2.1 Study areas and animals

All studies were conducted in the Danum Valley Conservation Area (DVCA) located in Sabah, Malaysia (Figure 1), in the northeast portion of Borneo Island. This area consists of primary rainforest. The Mueller's gibbons we studied were members of SAPA group. The territory of the SAPA group (Figure 1) is located near the Borneo Rainforest Lodge (BRL; 5°01' N, 117°44' E; elevation: approx. 190 m), and the group was habituated to humans.

During our 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 SAPA 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. As neighbor groups were not habituated and we couldn't follow them, it was not clear how much the SAPA's home range overlapped with the neighboring group. Thus, this home range approximation was a maximal value.

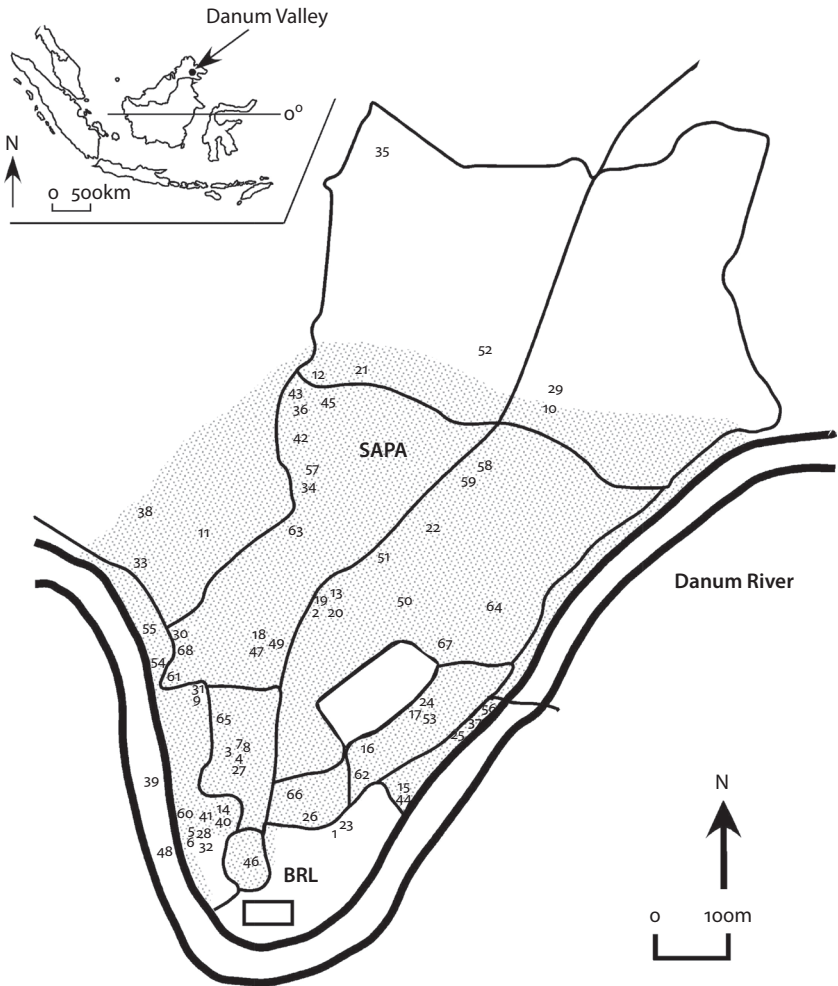


Figure 1. Location of the Borneo Rainforest Lodge (BRL) in the Danum Valley Conservation Area (DVCA; arrow), Sabah, Malaysia and location of the songs. The gray area represents the territory of the SAPA group

Our study was approved by the Research Promotion and Co-ordination Committee, Economic Planning Unit, and Prime Minister's Department, Malaysia.

2.2 Behavioral data collection

We collected behavioral data every 10 minutes while following the gibbon group. We classified gibbon songs according to the following factors: time of day, context (described below), season (dry or rainy), singing place (boundary, non-boundary or another territory), and family composition. The southern part of the home

range surrounded by the river is referred to as a non-boundary area, because neighbor groups can't cross the river. Before dawn, we placed the recording device just under the singing tree and tried to record songs as clearly as possible. We succeeded in recording the male song alone, because the study male sang solo songs in the pre- and/or postdawn hours and other family members didn't sing at the same time except his son. The son produced high pitch voices synchronized with his mother's greatcall until 4-year-old. The son also emitted short voices following his father's songs in rare cases, but it was easy to differentiate the son's voices from the father's songs because they were high pitch voices. When the son was 5.5-year-old, he sang male song for the first time. Then, a song exchange between the study male and the son occurred from 05:46–05:55 (9 minutes). The two subjects were sitting on the branch 30 m above the ground and were 20 m apart from each other. We recorded the songs from under the adult male. We could differentiate the son's song from the study male song because two songs didn't overlap and the son's song was slightly attenuated (Figure 2, wav file was added in supplement). Neighbor males' songs sometimes overlapped, but they were easily excluded from the analysis because they were attenuated. We listed the features of all male songs in Table 1. The location of male singing is shown in Figure 1. We recorded 68 songs in our 107 days' following. The behavioral contexts of songs were classified into five categories: (1) MS (morning song): songs starting before 7 a.m. (2) PB (playback song): songs against the playback of the neighbor male's song. In 2002 and 2003, we conducted playback experiments 4 times inside the non-boundary area of the home range using a radio cassette player (Panasonic RX-ES50; Panasonic, Tokyo, Japan) for sound playback. When the subject male was singing, we played back the neighbor male's song for 10 minutes under the singing tree and analyzed the song sung after sound playback. (3) LMS (late morning song): songs starting after 7 a.m. (4) Enc (encounter song): songs exchanged with neighbor groups at the range boundary or at the river side. (5) FS (family song) alternative songs among family members. For example, the study male sang this song with his 5.5-year-old son alternately (Figure 2) or when he presumably lost sight of another family member. In both cases, songs didn't overlap.

Table 1. List of male songs

No	Date	Start time	End time	Song duration (min)	Context	Season	Place	Family size
1	24-Mar-01	5:50	6:14	24	MS	dry	non-boundary	6
2	4-Aug-01	6:05	6:31	26	MS	dry	non-boundary	6
3	4-Aug-02	5:04	5:55	51	MS	dry	non-boundary	4
4	27-Dec-02	5:54	6:25	31	PB	rainy	non-boundary	4

Table 1. (continued)

No	Date	Start time	End time	Song duration (min)	Context	Season	Place	Family size
5	31-Dec-02	5:58	6:43	45	PB	rainy	non-boundary	4
6	2-Jan-03	5:39	6:12	33	MS	rainy	non-boundary	4
7	8-Aug-03	5:56	6:28	32	PB	dry	non-boundary	4
8	10-Aug-03	5:55	6:18	23	PB	dry	non-boundary	4
9	10-Aug-03	7:15	7:55	40	LMS	dry	non-boundary	4
10	12-Aug-03	6:26	6:40	14	MS	dry	boundary	4
11	25-Dec-03	5:48	6:41	53	MS	rainy	boundary	4
12	3-Aug-04	5:58	6:38	40	Enc	dry	boundary	4
13	5-Aug-04	5:05	6:02	57	MS	dry	boundary	4
14	7-Aug-04	5:50	6:18	28	MS	dry	non-boundary	4
15	9-Aug-04	8:10	8:47	37	LMS	dry	non-boundary	4
16	9-Aug-04	9:04	9:05	1	FS	dry	non-boundary	4
17	26-Dec-04	6:12	6:22	10	MS	rainy	non-boundary	4
18	27-Dec-04	6:16	6:28	12	MS	rainy	non-boundary	4
19	28-Dec-04	6:01	6:11	10	MS	rainy	non-boundary	4
20	28-Dec-04	6:45	7:13	28	MS	rainy	non-boundary	4
21	29-Dec-04	5:33	6:25	52	MS	rainy	boundary	4
22	2-Jan-05	4:49	6:30	101	MS	rainy	non-boundary	4
23	28-Jul-05	5:53	6:11	18	MS	dry	non-boundary	3
24	29-Jul-05	5:47	6:01	14	MS	dry	non-boundary	3
25	30-Jul-05	5:51	6:06	15	MS	dry	non-boundary	3
26	31-Jul-05	5:46	6:10	24	MS	dry	non-boundary	3
27	2-Aug-05	5:34	6:04	30	MS	dry	non-boundary	3
28	3-Aug-05	5:33	5:53	20	MS	dry	non-boundary	3
29	4-Aug-05	5:59	6:10	11	MS	dry	boundary	3
30	5-Aug-05	5:33	6:23	50	MS	dry	boundary	3
31	5-Aug-05	14:11	14:24	13	LMS	dry	non-boundary	3
32	7-Aug-05	5:43	6:04	21	MS	dry	non-boundary	3
33	26-Dec-05	11:09	11:28	19	LMS	rainy	boundary	3
34	27-Dec-05	6:02	6:33	31	MS	rainy	boundary	3
35	28-Dec-05	5:48	6:17	29	MS	rainy	another territory	3
36	29-Dec-05	5:54	6:14	20	MS	rainy	boundary	3

Table 1. (*continued*)

No	Date	Start time	End time	Song duration (min)	Context	Season	Place	Family size
37	30-Dec-05	5:40	6:17	37	MS	rainy	non-boundary	3
38	30-Dec-05	6:14	6:17	3	MS	dry	boundary	3
39	3-Aug-06	14:17	14:30	13	LMS	dry	non-boundary	3
40	4-Aug-06	5:37	6:02	25	FS	dry	non-boundary	3
41	6-Aug-06	6:12	6:32	20	MS	dry	non-boundary	3
42	7-Aug-06	5:46	6:07	21	MS	dry	boundary	3
43	7-Aug-06	6:17	6:21	4	MS	dry	boundary	3
44	8-Aug-06	9:04	9:37	33	LMS	dry	non-boundary	3
45	9-Aug-06	5:43	6:26	43	MS	dry	boundary	3
46	10-Aug-06	5:27	5:33	6	MS	dry	non-boundary	3
47	31-Dec-06	5:49	6:31	42	MS	rainy	non-boundary	2
48	31-Dec-06	8:30	8:50	20	LMS	rainy	non-boundary	2
49	2-Jan-07	5:45	6:12	27	MS	rainy	non-boundary	2
50	4-Jan-07	5:55	6:18	23	MS	rainy	non-boundary	2
51	5-Aug-07	5:40	6:09	29	MS	dry	non-boundary	2
52	5-Aug-07	5:46	5:57	11	MS	dry	boundary	2
53	8-Aug-07	5:12	6:23	71	MS	dry	non-boundary	2
54	8-Aug-07	8:31	9:21	50	Enc	dry	boundary	2
55	8-Aug-07	9:29	9:46	17	LMS	dry	boundary	2
56	10-Aug-07	5:43	6:10	27	MS	dry	non-boundary	2
57	11-Aug-07	6:04	6:18	14	MS	dry	boundary	2
58	11-Aug-07	5:40	6:08	28	MS	dry	boundary	2
59	13-Aug-07	6:17	6:42	25	MS	rainy	boundary	2
60	26-Dec-07	8:21	9:14	53	LMS	rainy	non-boundary	2
61	27-Dec-07	11:35	12:08	33	Enc	rainy	boundary	2
62	29-Dec-07	6:09	6:47	38	MS	rainy	non-boundary	2
63	30-Dec-07	5:48	6:42	54	MS	rainy	boundary	2
64	11-Aug-08	5:38	7:08	90	MS	dry	non-boundary	1
65	29-Dec-08	8:22	11:22	180	LMS	rainy	non-boundary	1
66	1-Jan-09	5:45	7:36	111	MS	rainy	non-boundary	1
67	31-Jul-09	5:44	6:28	44	MS	dry	non-boundary	1
68	5-Aug-09	5:55	6:44	49	MS	dry	boundary	1

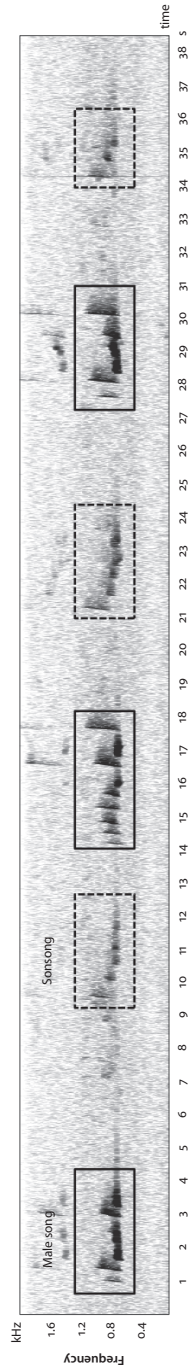


Figure 2. Portion of the sonogram when the study male sang FS (family song) with his 5.5-year-old son alternately. Solid line: study male song

2.3 Recording protocol

We used a digital audio tape recorder (Sony TCD-D100; Sony, Tokyo, Japan) with a microphone (Sony ECM-MS907; Sony, Tokyo, Japan) from 2001–2006 and a digital audio recorder (Roland R-09; Roland, Hamamatsu, Japan) with a microphone (audio-technica-ATM57; Audio-Technica, Tokyo, Japan) from 2007–2009. We didn't find any difference between the quality of recordings taken before 2006 and after 2007. We recorded the gibbon voices under the trees in which the study male was singing. The recorder was set at a 44.1-kHz sampling rate and had 16-bit resolution.

2.4 Data analysis

We converted the recorded sound to sonograms using Avisoft-SAS Lab Pro software (Avisoft, Berlin, Germany). First, the sampling frequency was converted to 4000 Hz. 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 measured onset time, offset time, start frequency and end frequency of each note and inter-notes intervals for each song phrase (Figure 3). For all the statistical analyses, we used GNU R (R Core Team, 2016). Add-on packages which we employed for each analysis were described in corresponding sections.

2.5 Acoustic terms and definitions

A song is a succession of phrases and a phrase is a succession of notes. A note is a minimum unit of song (Haimoff, 1984). When the intervals of two notes are shorter than 1.5 seconds, we defined a series of notes as a phrase. 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". But not only our subject male but also neighboring 9 males around BRL did not sing oo-wa and quaver-type notes. We confirmed it in 34 male songs of 9 males. We showed a portion of the sonograms produced by 5 males in Figure 4. As the trills were comprised of a short set of wa notes, we classified male songs into two note types: wa and oo. The wa notes are generally short in duration, with a rapid rise in frequency, and the oo notes are relatively monotonal sounds (Table 2, Figure 3).

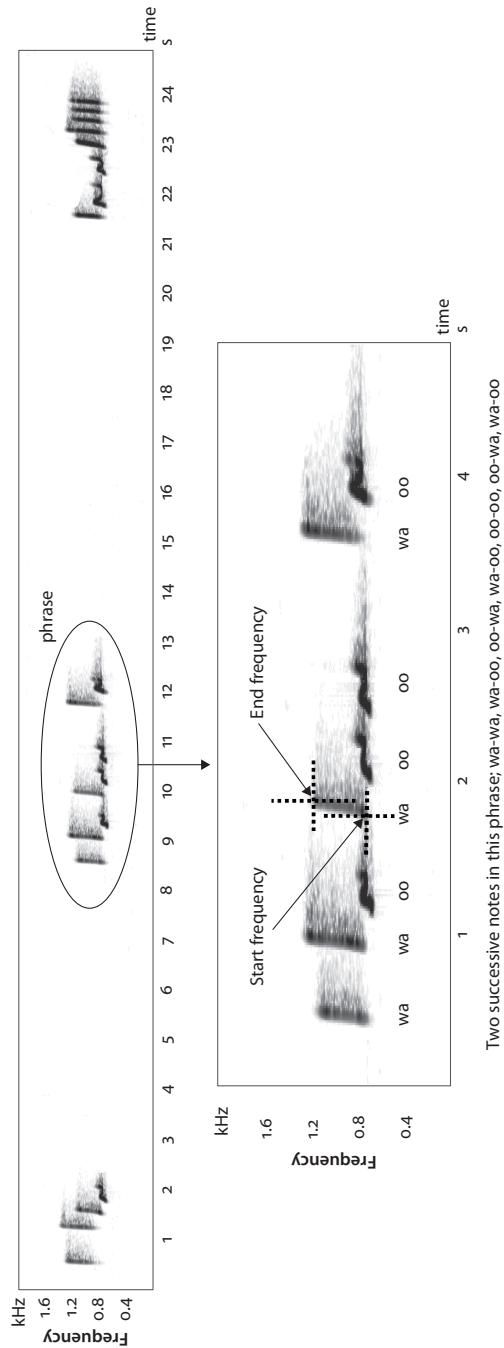


Figure 3. Sonogram of one phrase composed of "wa" and "oo" notes. Start frequency and end frequency are indicated by arrow

Table 2. Acoustic characteristics of wa and oo notes

	Mean (SD)	
	wa	oo
Start frequency (kHz)	0.803 (0.084)	0.710 (0.037)
End frequency (kHz)	1.112 (0.155)	0.790 (0.074)
Duration (second)	0.056 (0.027)	0.211 (0.103)

A: SAPA male, B: neighbor male located on the south west side of the SAPA group, C: neighbor male located on the south east side of the SAPA group, D: neighbor male located on the east side of the SAPA group, E: neighbor male located on the west side of the SAPA group. In all songs, notes of “oo-wa” and “quaver-type notes” were not found

2.6 Analyses of song syntax

We investigated whether gibbons emitted wa and oo notes randomly, or following any syntactical rules. Because compositions of wa and oo notes might vary, we performed analyses separately for songs. We counted number of occurrences of each note type, and those of two successive notes (wa-wa, wa-oo, oo-wa, and oo-oo) in song phrases for each song. Based on the frequencies of wa and oo notes, we computed expected prevalence of the two successive notes. For example, if the relative frequencies of wa and oo notes in a song were 0.8 and 0.2, we computed that the prevalence of wa-wa, oo-oo, and wa-oo / oo-wa notes were 0.64, 0.04 and 0.16, respectively. Then, we examined whether observed number of occurrences of the successive notes deviated from the expected values using goodness-of-fit tests. We obtained p-values by Monte Carlo simulation (Hope, 1968) with 2,000 replicates using the option provided in GNU R’s “chisq.test” function, and performed Bonferroni corrections.

2.7 Syntactical difference among songs

We investigated whether composition ratios of two successive notes vary among songs, and contexts associated with the songs. For this analysis, we focused on MS (morning song) and LMS (late morning song), which we had enough samples for statistical analysis. For each song, we obtained the ratio of two successive notes. We employed a multinomial log-linear model to examine whether the composition ratio of the successive note types varies among contexts, using multinom function in nnet package (Venables & Ripley, 2002). We entered number of each successive note type in songs as response variables, and contexts as an explanatory variable. We performed a likelihood-ratio test against a null hypothesis that the composition ratio of sequence types was identical among contexts.

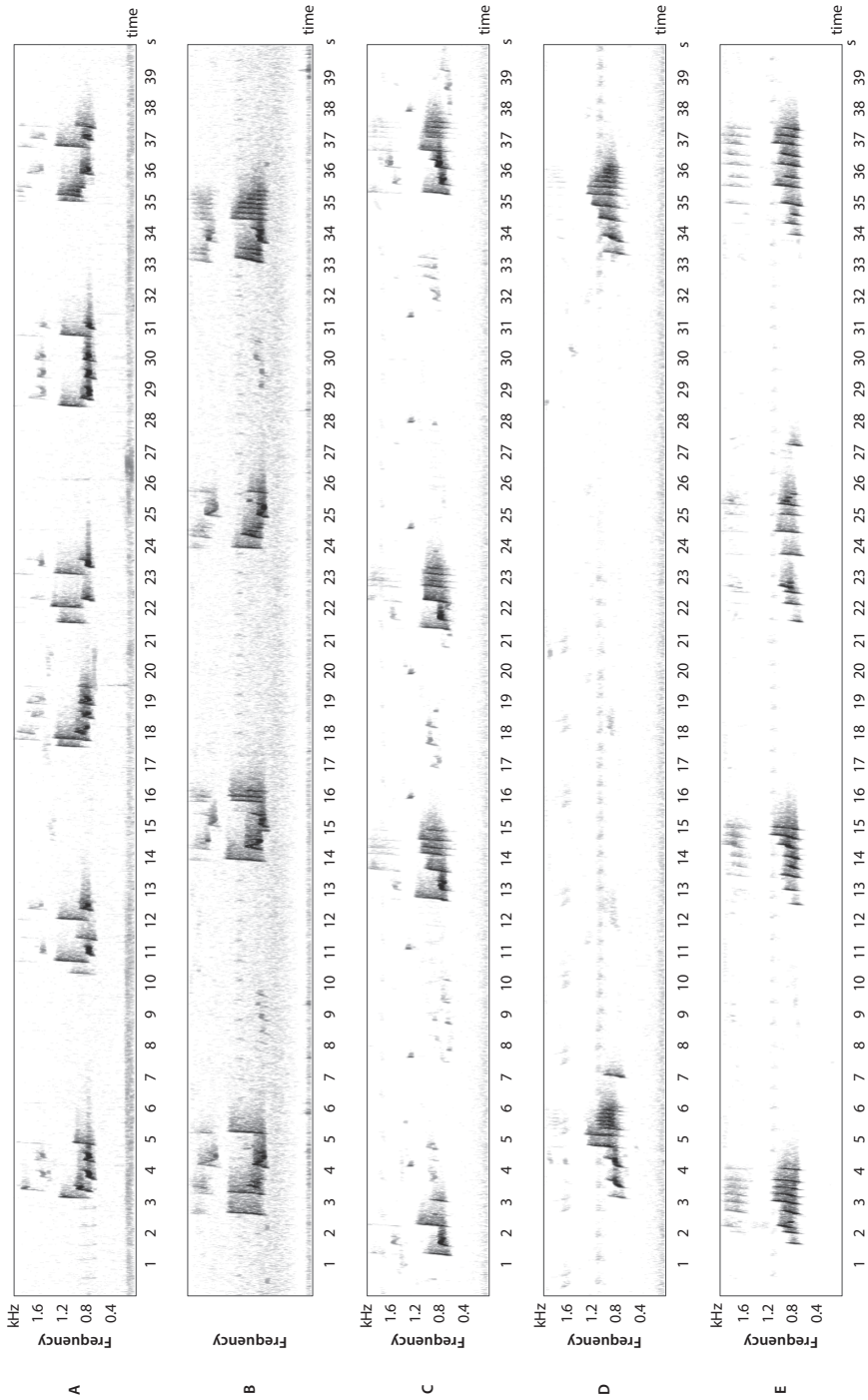


Figure 4. Portion of the sonograms produced by 5 males around BRL

2.8 Analyses of chunk existence

We investigated the possibility of chunk existence in male Mueller's gibbon songs. Because a chunk is defined both from production and perception, studies often involve examination of underlying neuronal mechanisms for chunk generation (production side) and behavioral response of conspecifics (perceptual side). However, neither approach was available for our study, and so we examined the possibility of chunk existence from properties of note sequences. Usually, a chunk is characterized by a fixed sequence of notes separated by relatively long time intervals between previous and succeeding notes (Williams & Staples, 2004). Thus, if we select any sequence of notes, note intervals between onsets of the last note in a sequence and the succeeding notes will be longer in more determined (or fixed) sequences than in less determined ones. In analysis, we extracted all song segments of lengths ranging from 1 note to phrase length - 1. For example, if there was a phrase consists of 5 notes, we could extract 6 patterns of song segments (1st-2nd, 1st-3rd, 1st-4th, ... , 3rd-4th) . Then, we computed onset time intervals between the last notes of the song segment and their succeeding notes and how well the succeeding notes were determined across all songs. For example, suppose we detect a sequence of "wa-wa-oo-wa" that occurred 40 times. We measured intervals between the last wa note and succeeding note (which may be a "wa" or "oo" note) for each occurrence and computed their median value. As a determinacy index, we employed Shannon entropy with Miller-Madow correction (Miller, 1955), using entropy function in *infotheo* library (Meyer, 2014). Shannon entropy takes values larger than zero. If notes following a song segment was highly determined (e.g., always oo notes occurred after wa-wa-oo-wa sequence), then the Shannon entropy value became very close to zero. On the other hand, if notes following a song segment were random (unpredictable), then Shannon entropy became a positive value. We computed Kendall's rank correlations between the median intervals and the determinacy index, expecting that there will be a positive correlation between them.

3. Results

3.1 Analyses of song syntax

In total we recorded 89,299 notes from 68 songs. Among them, we selected 88,772 notes, which are the beginning notes of each phrase and notes observed within 1.5 s from the previous note, and derived 80,855 pairs of successive notes. Acoustic characteristics of wa and oo notes were shown in Table 2. Table 2 shows that duration of wa notes was shorter and frequency range was wider in comparison with oo notes. We also showed compositions of "oo" notes among the two note types

in Figure 5. The mean and SD of “oo” notes ratio were 0.33 and 0.14, respectively.

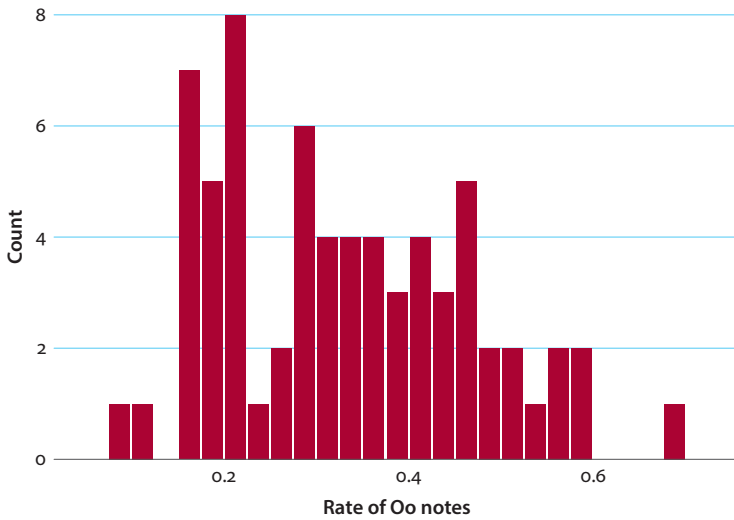


Figure 5. Histogram of oo notes composition ratio. Binwidth = 0.025

Goodness-of-fit tests with Bonferroni corrections revealed that frequencies of the two successive note types differed from expected prevalence in 45 songs (Figure 6).

3.2 Syntactic difference among songs

Composition ratio of the two successive notes in different contexts associated with the songs was shown in Figure 7. The composition ratio was varied but overlapped among contexts. As the result of multinomial log-linear model and likelihood-ratio test which we performed for MS and LMS, there was not significant effect of the context on the composition ratio ($\chi^2_3 = 7.47, p = 0.06$).

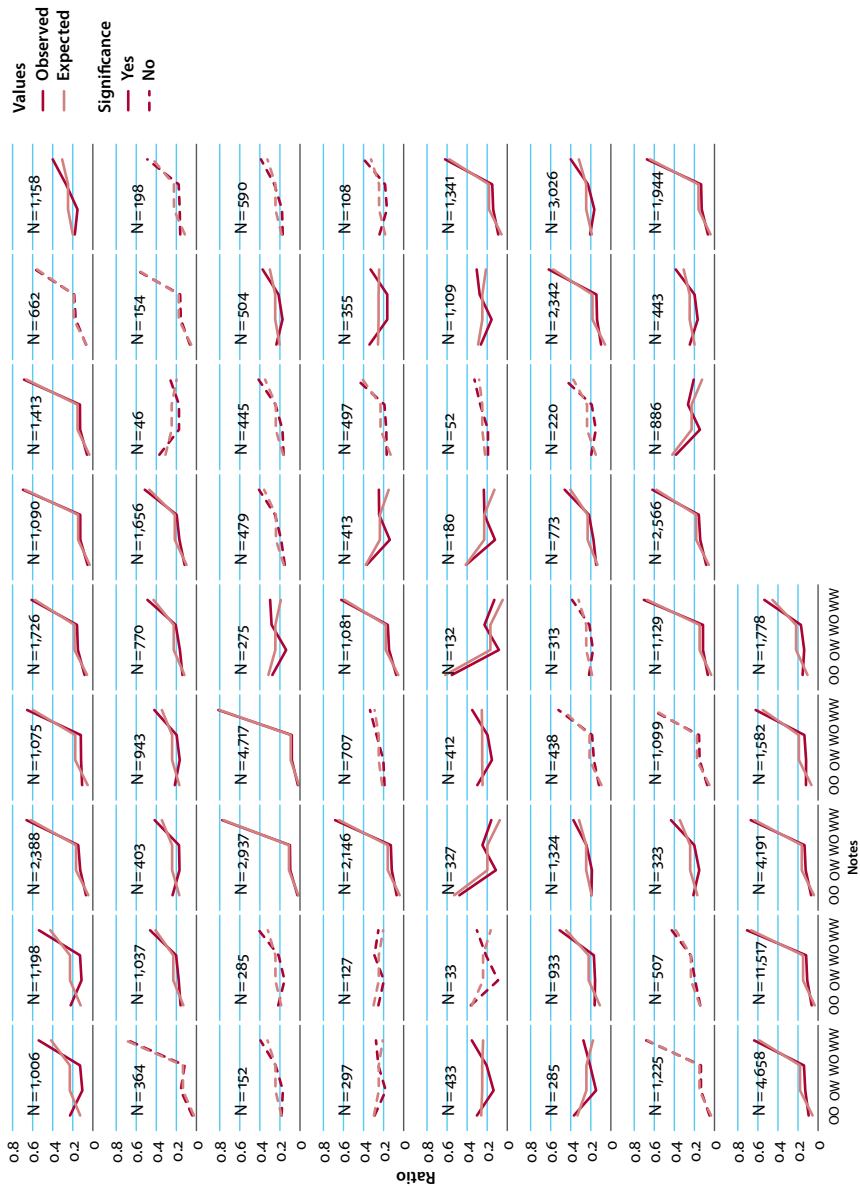


Figure 6. Compositions of observed and expected successive note types for each recording session. From top left to bottom right, songs 1–68 are presented. N indicates number of notes in a song

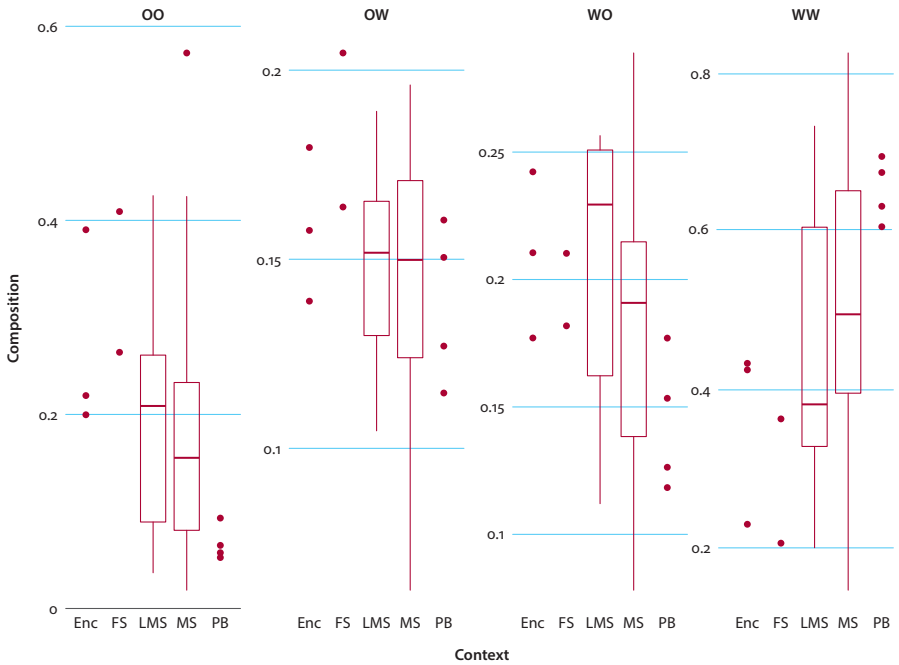


Figure 7. Composition ratio of the two successive notes in different contexts associated with the songs. Enc: encounter song, FS: family song, LMS: late morning song, MS: morning song, PB: playback song. Note that except LMS and MS, all values were plotted since less than five songs were available in each context

3.3 Analyses of chunk existence

From 7,740 phrases, we determined 21,352 types of song segments. Among them, we selected 493,418 segments that occurred more than five times. We calculated entropy and median intervals between succeeding notes of the song segments. The histogram of entropy showed a two-humped pattern; the majority of song segments had entropy greater than 0.1, and subset had entropy smaller than 0.05 (Figure 8). The song segments with small entropy suggest existence of highly deterministic succeeding notes. Figure 9 showed the relationship of the entropy and intervals between succeeding notes of the song segments. In Figure 9, there was a subset of song segments which had small (< 0.05) entropy, corresponding to Figure 8 (histogram), and the subset had relatively shorter (< 0.3 s) intervals. Another subset which had entropy greater than 0.1 showed a trend that the entropy positively correlated with intervals (Figure 9). Kendall's rank correlations between the median intervals and the entropy revealed a positive significant cor-

relation ($\tau = 0.44$, $p < 0.01$), which indicated that less deterministic sequences showed longer note intervals (Figure 9).

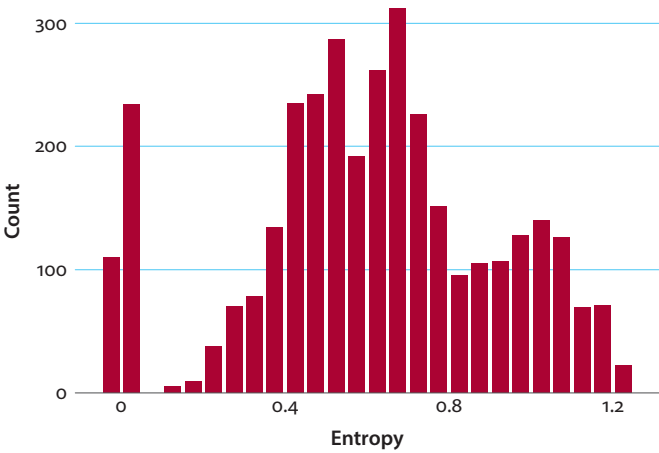


Figure 8. Histogram of Shannon entropy of song segments and a succeeding note (bin width = 0.05)

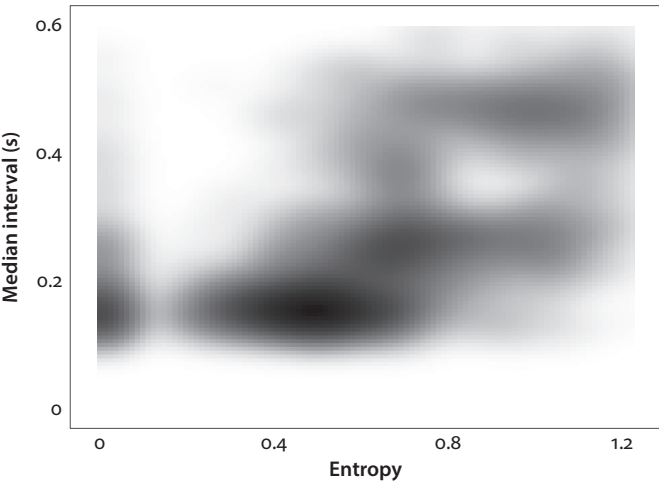


Figure 9. Relationship between median intervals and Shannon entropy. Result of 2D kernel density estimation is shown

4. Discussion

Some animals vocalize by combining distinct acoustic elements in an orderly manner, which often reference external objects and callers’ emotional states.

Such combinational sounds may have evolved when the variety of external objects and emotional states exceeded vocal repertory (Jackendoff, 1999; Arnold & Zuberbühler, 2012). Examples include bird calls and songs (Beer, 1976; Chew, 1981; Hailman et al., 1985; Templeton et al., 2005). For example, in “Chick-a-dee” calls of the black-capped chickadee (*Parus atricapillus*), four notes are variously repeated and combined to produce many signals (Hailman et al., 1985). Furthermore, variation in note compositions could convey a variety of messages to receivers. Four notes are traditionally designated as A, B, C, and D. Templeton et al. (2005) presented live, perched, avian predator models to aviary groups of black-capped chickadees and found a negative correlation between the number of “D” notes and predator size (both wingspan and body length). Several non-human primates also combine pairs of calls into different context-specific call sequences (Zuberbühler, 2002; Ouattara et al. 2009; Arnold & Zuberbühler, 2006). However, no wild animals have semantically compositional syntax (Hurford, 2011).

Gibbons are the only singing ape apart from humans. They live in small groups and also interact with neighboring groups. They sing long, complex songs. Male Mueller's gibbons living in DVCA have two types of song elements, and there are many patterns of note combinations. Their phrase duration, number of notes in a phrase, phrase interval and note interval can be quite variable in the songs of one male gibbon.

We investigated whether two successive notes in male gibbon songs were emitted following any syntactical rules. As a result, frequencies of the two successive note types differed from expected prevalence in 45 songs out of 68 songs. Although we only studied one singing subject, our results indicate that male Mueller's gibbon songs have combinatory rules. Our findings are consistent with previous studies (Mitani & Marler, 1989; Geissmann et al., 2005; Clarke et al., 2006).

We also investigated chunk existence in male Mueller's gibbon songs and obtained a result that matches our prediction that chunk structures occur in songs of males. Chunking involves grouping of elements that form units and is a useful way to learn longer sequences. Studies on humans (Miller, 1956; Simon, 1974) showed that groups of items that can be considered a single unit are learned as quickly as a single item. Chunking is considered to be a basic mechanism in human language acquisition. Although there is no evidence so far that gibbons use chunks as units for song learning, gibbon songs may share a common background with human language.

The possibility that gibbons convey information through their songs and note orders and/or chunks in phrases are related to different kinds of objects, events or emotions, provides an important aspect in our understanding of the origin of language. Composition ratio of the two successive notes in different contexts associated with the songs was shown in Figure 7. We classified gibbon songs into five categories.

The composition ratios of MS and LMS songs were overlapped in Figure 7 and there was not significant effect of the context on the composition ratio among two songs. As MS and LMS songs were defined only by starting time, their behavioral contexts were unclear. So, we discussed about three other songs (PB, Enc and FS) whose contexts were comparatively clear, although sample size was low.

We considered that playback (PB) songs were sung in non-affiliative context and family songs (FS) were sung in affiliative context. In Figure 7, PB songs showed higher composition of ww (wa-wa) than other contexts. In contrast, FS songs showed lower composition of ww than other contexts.

Morton (1977) has argued that among nonhuman animals, lower-frequency, harsh sounds are made when being hostile, and higher-frequency, more pure-tone-like sounds are made when being submissive, appeasing, or friendly. For example, meerkats (*Suricata suricatta*) emit predator-specific alarm calls and each predator-specific call varies with urgency class, with low urgency calls being clearer and more harmonic, and high urgency calls being harsher and noisier (Manser et al., 2002). Thus, meerkat calls contain both continuously varying emotional information and discretely varying referential information. In Mueller's gibbon songs, wa notes were more harsh sounding with wider frequency range than oo notes (Table 2). Therefore, if Mueller's gibbon songs follow the Morton's argument, songs which showed high composition of ww might have been sung in a non-affiliative context. PB songs showed high composition of ww (wa-wa) and low composition of oo (oo-oo) in Figure 7. Our playback experiments were conducted inside the non-boundary area of the home range. Gibbons never intrude into the core area of neighbor group's territory. Songs emanating from the center of the range represent a clear threat to range integrity (Mitani, 1985). So, PB songs were considered to be sung in a non-affiliative context. We provide the following observational information to support this idea. On November 3, 2013, we were recording a male song of another group in DVCA when barking deer came close to us. They were surprised at the sight of us and barked repeatedly for 8 minutes. After the deer started barking, the male gibbon changed his song. The percentage of wa notes in his song increased from 58.0 % to 91.9 %. His song, consisting of many successive wa notes, was thus considered to be sung in a non-affiliative context. Barking deer do not represent a threat to gibbons. When barking deer barked for us, gibbons noticed our existence. As this group was unhabituated and afraid of humans, gibbon's reaction might be addressing not only urgency but also aggressiveness. Gibbons went away in a hurry from us while singing.

By contrast, family songs (FS) showed low composition of ww and high composition of oo in Figure 7. One of FS songs was sung when the study male was separated far from another family member and lost sight of them. The other was sung when the study male sang with his 5.5-year-old son alternately (Figure 2).

Subadult offspring are sometimes forced out of the group when they are from 8 to 10 years old. However, son of this study was not fully matured. Around that time, father and son always moved together and sometimes groomed each other (a total of 7 times within 7 days' observation). Furthermore, play behaviors such as chasing and wrestling were sometimes observed. Therefore, we consider that the function of FS songs is maintenance of family bonds and they are sung in affiliative contexts. Moreover, continuously varying emotion may be represented in discrete note types and their abundance.

Several functions have been attributed to gibbon songs, which emphasize a role in territorial advertisement, mate attraction, and the maintenance of pair and family bonds (Haimoff, 1984; Leighton, 1987; Geissmann, 1999; Geissmann & Orgeldinger, 2000). We considered the function of PB songs as territorial advertisement and threat or warning aiming to prevent the entrance of neighbors into the non-boundary area of the home range. The function of Encounter songs (Enc) were also considered as territorial advertisement and threat or warning, but they had similar composition of ww and oo to FS songs and different composition of ww and oo from PB songs in Figure 7. One of Enc songs was sung on the range boundary. At dawn on 3 August 2004, the study group and the neighbor group were located around the territory boundary and both groups vocalized songs antiphonally from 05:48–06:59. After singing, the neighbor group returned to its territory. The study group followed the neighbor group, intruded 100 m into the neighbor group territory, and then returned to its own territory. No fighting occurred during the encounter. Another two Enc songs were sung across the river. While the study male was singing songs, neighbor group came close at the opposite side of the river and sang antiphonally for about 10 minutes in both cases. The study group and neighbor group faced each other across the river. In all cases of Enc songs, antiphonal songs occurred around the range boundary or across a river. We considered that the functions of Enc songs were not always territorial advertisement but sometimes affiliative. In our 1600 hours' following of four groups in DVCA from 2001 to 2014, we counted 15 encounters within 50 m and observed no physical contacts during encounters. Instead, we observed 11 vocal contacts and only one chasing. Our observations suggest that gibbons are not always offensive during encounters on the range boundary. According to a study in Khao Yai National Park, Thailand, white-handed gibbons (*Hylobates lar*) live in harmony with their neighbors on the range boundary and show few fighting behaviors except chasing (Reichard & Sommer, 1997). Furthermore, Matsudaira et al. (2015) showed the close genetic relationship between neighboring males in Khao Yai gibbons. So, it is not strange that Enc songs are similar to FS songs in composition ratios of the two successive notes.

Although the results shown in Figure 7 do not demonstrate fully that note combinations in gibbon songs were associated with behavioral contexts, male gibbon may change the note order in different situations. Our study is consistent with the previous study in which gibbons assemble a finite number of call units into more complex structures to convey different messages in response to different predator types (Clarke et al., 2006). Future study will be necessary to find out the clear relationship between note orders and contexts.

Our study indicated that the study gibbon songs may have combinatory rules and hierarchical structure. Male gibbon songs are similar to human language in that the vocabulary items are combined syntactically to create phrases and chunks are used for grouping words.

The string-context mutual segmentation hypothesis of language emergence (Okanoya & Merker, 2007) suggests that song strings and behavioral contexts are mutually segmented, while songs having syntax are interacted antiphonally or synchronously among individuals or groups. Our results may support this hypothesis, because gibbon songs have combinatory rules, which may relate to behavioral contexts, and male Mueller's gibbons are known to sing antiphonally (Inoue et al., 2013). Although further research is needed to demonstrate more completely whether specific combinations of notes reflect different meaning, our study of gibbon songs will provide a valuable insight into the study of language origin.

Acknowledgments

We thank the Economic Planning Unit of Sabah State Government, Danum Valley Management Committee and Sabah Biodiversity Centre for permission to conduct this study in Sabah. We are grateful to Etsuko Inoue for her dedicated help with our research. We also thank Donny Sah Itin, Isnadil Mohd and Dennysius Aloysius for gibbon tracking and assistance with data collection. This study was partly supported by Heisei Foundation for Basic Science.

Supplemental data

Wav file of Figure 2. can be found at: <https://doi.org/10.1075/is.18.1.01ino.audio.fig-2>

References

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

- Arnold, K., & Zuberbühler, K. (2012). Call combinations in monkeys: Compositional or idiomatic expressions? *Brain and Language*, 120, 303–309. doi: 10.1016/j.bandl.2011.10.001
- Beer, C. (1976). Some complexities in the communication behavior of gulls. *Annals of the New York Academy of Sciences*, 280, 413–432. doi: 10.1111/j.1749-6632.1976.tb25505.x
- Berwick, R. C., Okanoya, K., Beckers, G.J. L., & Bolhuis, J. J. (2011). Songs to syntax: The linguistics of birdsong. *Trends in Cognitive Sciences*, 16, 113–121. doi: 10.1016/j.tics.2011.01.002
- Brockelman, W. Y., & Schilling, D. (1984). Inheritance of stereotyped gibbon calls. *Nature*, 312, 634–636. doi: 10.1038/312634a0
- Chew, L. (1981). Geographic and individual variation in the morphology and sequential organization of the song of the savannah sparrow (*Passerculus sandwichensis*). *Canadian Journal of Zoology*, 59, 702–713. doi: 10.1139/z81-099
- Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS One*, 1, e73. doi: 10.1371/journal.pone.0000073
- Cleveland, J., & Snowdon, C. T. (1982). The complex vocal repertoire of the adult cotton-top tamarin (*Saguinus oedipus oedipus*). *Zeitschrift für Tierpsychologie*, 58, 231–270. doi: 10.1111/j.1439-0310.1982.tb00320.x
- Crockford, C., & Boesch, C. (2005). Call combinations in wild chimpanzees. *Behaviour*, 142, 397–421. doi: 10.1163/1568539054012047
- Darwin, C. (1859). *On the origin of species*. London: John Murray.
- Geissmann, T. (1984). Inheritance of song parameters in the gibbon song, analyzed in 2 hybrid gibbons (*Hylobates pileatus* × *H. lar*). *Folia Primatologica*, 42, 216–235. doi: 10.1159/000156165
- Geissmann, T. (1999). Duet songs of the siamang, *Hylobates syndactylus*: II, Testing the pair-bonding hypothesis during a partner exchange. *Behaviour*, 136, 1005–1039. doi: 10.1163/156853999501694
- Geissmann, T., & Orgeldinger, M. (2000). The relationship between duet songs and pair bonds in siamangs, *Hylobates syndactylus*. *Animal Behaviour*, 60, 805–809. doi: 10.1006/anbe.2000.1540
- Geissmann, T., Bohlen-Eyring, S., & Heuck, A. (2005). The male song of the Javan silvery gibbon (*Hylobates moloch*). *Contributions to Zoology*, 74, 1–25.
- Groves, C. P. (1972). Systematics and phylogeny of gibbons. In D. M. Rumbaugh (Eds.), *Gibbon and siamang*, vol. 1 (pp. 1–89). Basel: Karger.
- Hailman, J. P., Ficken, M. S., & Ficken, R. W. (1985). The “chick-a-dee” calls of *Parus atricapillus*: a recombinant system of animal communication compared with written English. *Semiotica*, 56, 191–224. doi: 10.1515/semi.1985.56.3-4.191
- Haimoff, E. H. (1984). Acoustic and organizational features of gibbon songs. In: H. Preuschoft, D. J. Chivers, W. Y. Brockelman, & N. Creel (Eds.), *The lesser apes. Evolutionary and behavioural biology* (pp. 333–353). Edinburgh: Edinburgh University Press.
- Haimoff, E. H. (1985). The organization of song in Mueller's gibbon (*Hylobates muelleri*). *International Journal of Primatology*, 6, 173–192. doi: 10.1007/BF02693652
- Hope, A.C. A. (1968). A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society. Series B*, 30, 582–598.
- Hurford, J. R. (2011). *The origins of grammar: language in the light of evolution II* (Vol. 2). Oxford University Press.
- Inoue, Y., Waidi, S., Yosida, S., & Okanoya, K. (2013). Intergroup and intragroup antiphonal songs in wild male Mueller's gibbons (*Hylobates muelleri*). *Interaction Studies*, 14, 24–43. doi: 10.1075/is.14.1.o3ino

- Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences*, 3, 272–279. doi: 10.1016/S1364-6613(99)01333-9
- Leighton, D. R. (1987). Gibbons: Territoriality and monogamy. In B. B. Smuts, D. L. Cheney, R. M. Seyfarth, R. W. Wrangham, & T. T. Struhsaker (Eds.), *Primate societies* (pp. 135–145). Chicago: University of Chicago Press.
- Manser, M. B., Seyfarth, R. M., & Cheney, D. L. (2002). Suricate alarm calls signal predator class and urgency. *Trends in Cognitive Sciences*, 6, 55–57. doi: 10.1016/S1364-6613(00)01840-4
- Marshall, J. T., & Marshall, E. R. (1976). Gibbons and their territorial songs. *Science*, 193, 235–237. doi: 10.1126/science.193.4249.235
- Matsudaira, K., Ishida, T., Reichard, U. H., & Malaivijitnond, S. (2015). Male kin network and dispersal pattern of white-handed gibbons. *The 31st Congress of the Primate Society of Japan*.
- Meyer, P. E. (2014). infotheo: Information-Theoretic Measures. R package version 1.2.0. URL <https://CRAN.R-project.org/package=infotheo>.
- Miller, G. A. (1955). Note on the bias of information estimates. In H. Quastler (Eds.), *Information theory in psychology; problems and methods II-B* (pp. 95–100). Glencoe, IL: Free Press.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits of our capacity for processing information. *Psychology Review*, 63, 81–97. doi: 10.1037/h0043158
- Mitani, J. C. (1985). Location-specific responses of gibbons (*Hylobates muelleri*) to male songs. *Zeitschrift für Tierpsychologie*, 70, 219–224. doi: 10.1111/j.1439-0310.1985.tb00513.x
- Mitani, J. C., & Marler, P. (1989). A phonological analysis of male gibbon singing behavior. *Behaviour*, 109, 20–45. doi: 10.1163/156853989X00141
- Morton, E. S. (1977). Occurrence and significance of motivation structural rules in some bird and mammal sounds. *American Naturalist*, 111, 855–869. doi: 10.1086/283219
- Okanoya, K., & Merker, B. (2007). Neural substrates for string-context mutual segmentation: A path to human language. In C. Lyon, L. Nehaniv, A. Cangelosi & (Eds.), *Emergence of communication and language* (pp. 421–434). Springer-Verlag. doi: 10.1007/978-1-84628-779-4_22
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51), 222026–22031.
- R. Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Reichard, U., & Sommer, V. (1997). Group encounters in wild gibbons (*Hylobates lar*): Agonism, affiliation, and the concept of infanticide. *Behaviour*, 134, 1135. doi: 10.1163/156853997X00106
- Robinson, J. G. (1979). An analysis of the organization of vocal communication in the titi monkey *Callicebus moloch*. *Zeitschrift für Tierpsychologie*, 49, 381–405. doi: 10.1111/j.1439-0310.1979.tb00300.x
- Robinson, J. G. (1984). Syntactic structures in the vocalisations of wedge-capped capuchin monkeys, *Cebus olivaceus*. *Behaviour*, 90, 46–78. doi: 10.1163/156853984X00551
- Simon, H. A. (1974). How big is a chunk? By combining data from several experiments, a basic human memory unit can be identified and measured. *Science*, 183, 482–488. doi: 10.1126/science.183.4124.482
- Suzuki, R., Buck, J. R., & Tyack, P. L. (2006). Information entropy of humpback whale songs. *Journal of the Acoustical Society of America*, 119, 1849–1866. doi: 10.1121/1.2161827
- Templeton, C. N., Green, E., & Davis, K. (2005). Allometry of alarm calls: Black-capped chickadees encode information about predator size. *Science*, 308, 1934–1937. doi: 10.1126/science.1108841

- Ten Cate, C., & Slater, P.J. B. (1991). Song learning in zebra finches: how are elements from two tutors integrated? *Animal Behaviour*, 42, 150–152. doi: 10.1016/S0003-3472(05)80617-7
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Fourth edition. Springer. doi: 10.1007/978-0-387-21706-2
- Williams, H., & Staples, K. (1992). Syllable chunking in zebra finch (*Taeniopygia guttata*) song. *Journal of Comparative Psychology*, 106, 278–286. doi: 10.1037/0735-7036.106.3.278
- Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, 63, 293–299. doi: 10.1006/anbe.2001.1914

Authors' addresses

Yoichi Inoue
The School of Arts and Sciences, The
University of Tokyo
3-8-1 Komaba, Meguro-ku
Tokyo 153-8902
Japan

Shigeto Yosida
Laboratory for Biolinguistics, RIKEN Brain
Science Institute
2-1 Hirosawa, Wako
Saitama 351-0198
Japan

Waidi Sinun
Research and Development Division Yayasan
Sabah Group
Yayasan Sabah Headquarters Complex Level
12, Menara Tun Mustapha
P. O. Box 11622, 88817
Kota Kinabalu, Sabah
Malaysia

Kazuo Okanoya
Laboratory for Biolinguistics, RIKEN Brain
Science Institute
2-1 Hirosawa, Wako, Saitama 351-0198
+81-3-5454-6301
+81-3-5454-6725
Japan

Biographical notes

Yoichi Inoue was a research student of the University of Tokyo. His research interest is interactive behaviors and vocal communication among wild gibbons.

Waidi Sinun is a group manager of Conservation & Environmental Management Division Yayasan Sabah Group and encourages many researchers to carry out research activities in the Conservation Area in Sabah, Malaysia. He got a Ph. D at the University of Manchester, studying hydrologic and geomorphic impacts of forest conversion in the Gunung Kinabalu Highlands Sabah Malaysia.

Shigeto Yosida was a researcher of Laboratory for Biolinguistics, Riken Brain Science Institute. He got a PhD at Chiba University, studying rodent vocal communications.

Kazuo Okanoya is a Professor of Graduate School of Arts and Sciences, the University of Tokyo, and Team Leader of Cognition and Behavior Joint Research Laboratory, Riken. He seeks a computational coding schema for emotional information and for biological/psychological understanding of animal communication including birdsong, human language, and rodent vocalizations.